CRISPR'S GAMBIT

Novel tactics for genetic engineering of human and bacterial cells

by Despoina Trasanidou

Propositions

1. The base-editing technology is the most promising development in the current genetic engineering field. *(this thesis)*

2. Unconventional use of anti-CRISPR proteins enhances the versatility of the CRISPR-Cas toolbox. *(this thesis)*

3. Gene therapy is an emerging approach of personalised medicine.

4. Intermittent fasting has tremendous potential for prevention of diseases.

5. Artificial Intelligence severely impacts academic integrity.

6. Chess is a simulation of life.

Propositions belonging to the thesis, entitled

CRISPR's Gambit Novel tactics for genetic engineering of human and bacterial cells

Despoina Trasanidou
Wageningen, 16 May 2023

CRISPR'S GAMBIT

Novel tactics for genetic engineering of human and bacterial cells

Despoina Trasanidou

Thesis committee

Promotor
Prof. Dr John van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Co-promotor
Dr Raymond H.J. Staals
Assistant Professor at the Laboratory of Microbiology
Wageningen University & Research

Other members
Prof. Dr Niels Geijsen, Leiden University, Leiden
Prof. Dr Michiel Kleerebezem, Wageningen University & Research
Dr Daan Swarts, Wageningen University & Research
Dr Elleke Bosma, Freya Biosciences ApS, Copenhagen, Denmark

This research was conducted under the auspices of VLAG Graduate School (Biobased, Biomolecular, Chemical, Food, and Nutrition sciences)

CRISPR'S GAMBIT

Novel tactics for genetic engineering of human and bacterial cells

Despoina Trasanidou

Thesis

Submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
In the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 16 May 2023
at 11 a.m. in the Omnia Auditorium.

Despoina Trasanidou
CRISPR'S GAMBIT Novel tactics for genetic engineering of human and bacterial cells, 547 pages

PhD thesis, Wageningen University, Wageningen, The Netherlands (2023)
With references, with summary in English

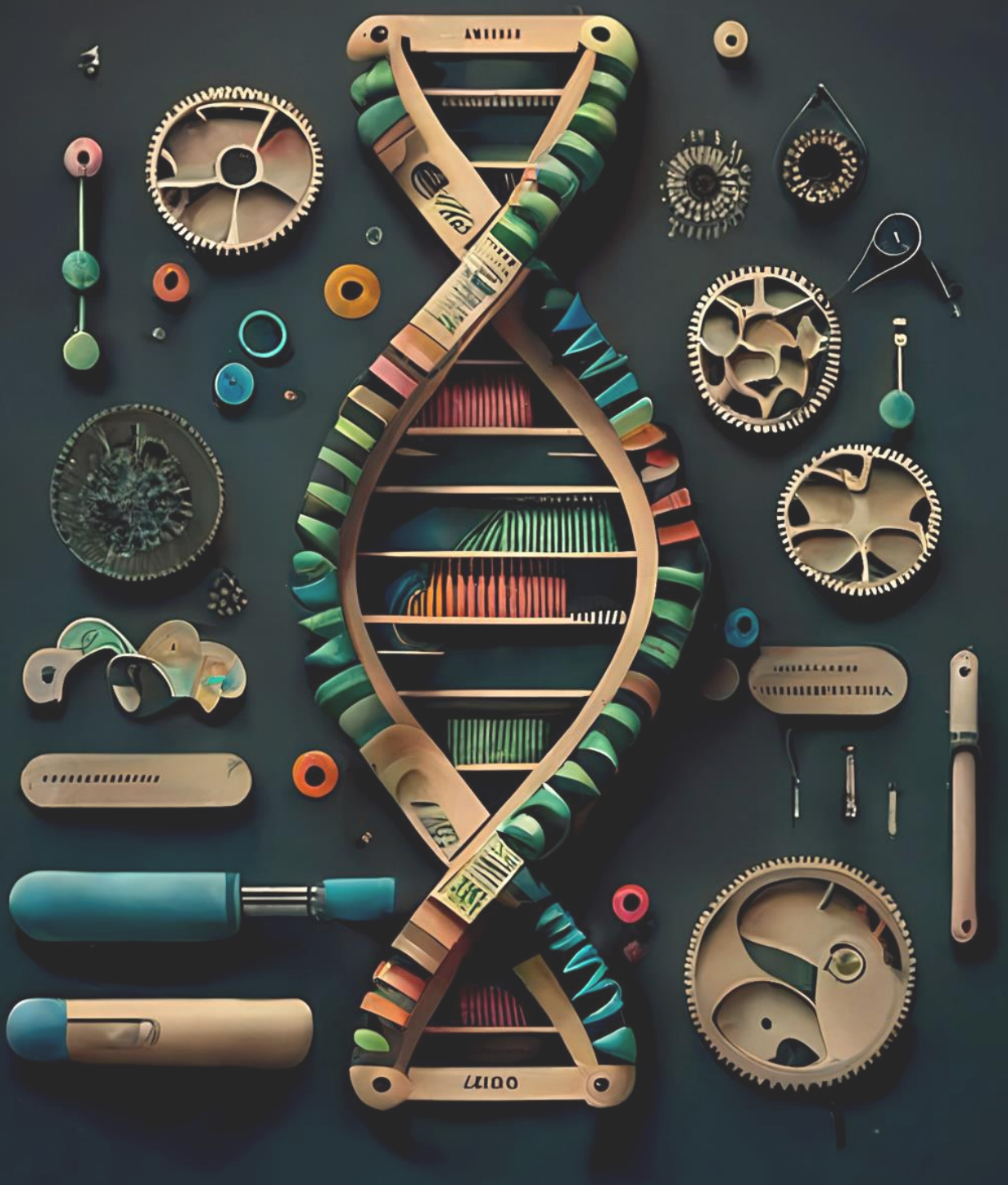
ISBN: 978-94-6447-662-0
DOI: https://doi.org/10.18174/628837

Table of contents

Chapter 1	General introduction & Thesis outline	9
Chapter 2	Base-editing: from biology to technology	51
Chapter 3	Characterizing a type III CRISPR-Cas-associated adenosine deaminase	137
Chapter 4	Keeping CRISPR in check: diverse mechanisms of phage–encoded anti–CRISPRs	203
Chapter 5	*In vivo* characterization of the AcrIIC1 anti–CRISPR protein for Cas9–based genome engineering	233
Chapter 6	Efficient genome and base editing in human cells using ThermoCas9	317
Chapter 7	Rational and random engineering of ThermoCas9 for optimal activity at moderate temperatures	383
Chapter 8	Summary & General Discussion	451

Altering the DNA is like …

… a chess game!



Chapter 1

General introduction & Thesis outline

Despoina Trasanidou

History of genetic engineering

Artificial selection and molecular biology

Exploring the infinite complexity of nature, humankind has always been seeking for new means to satisfy primary needs, such as food and health (Fig. 1). In prehistoric years, hunting and gathering of wild animals and plants were the predominant modes of subsistence, being steadily replaced by domestication and selective breeding in early societies[1,2]. The term 'selective breeding' was first coined by Robert Bakewell during the British Agricultural Revolution (1783)[3], spurring the conceptualization of natural selection (by Charles Darwin, 1859)[4], genetic inheritance (by Gregor Mendel, 1865)[5], (pan)gene as the unit of hereditary information (by Hugo de Vries, 1889)[6], and genetics (by William Bateson, 1905)[7]. Although the study and the development of artificial selection techniques facilitated screening of animals and plants for desired traits, direct intervention in the genetic variation was unattainable. In the late 1920s, Hermann Joseph Muller subjected flies to high temperature or X−ray radiation to deliberately introduce random genetic mutations, which were inherited to the next generation[8-10]. In the same period, Frederick Griffith's experiment showed that genetic information can be naturally transmitted among bacteria via a process called transformation[11]. This genetic information was later identified as deoxyribonucleic acid (DNA) (by Oswald Avery, Colin MacLeod and Maclyn McCarty)[12] and its double−helix structure was resolved in 1953 (by James Watson, Francis Crick, Maurice Wilkins, and Rosalind Franklin)[13]. In 1958, the Central Dogma of molecular biology revealed that genetic information passes from DNA (in the form of a gene) to RNA and then to protein, allowing for the genetic information stored in genes to be expressed as proteins with a wide spectrum of catalytic and structural functions[14]. In the period 1950−1970, a series of seminal discoveries (plasmid DNA, DNA ligases, restriction enzymes) and the invention of artificial (chemical) transformation led to creation of the first recombinant DNA (Berg, 1972)[15-19].

Recombinant DNA technology and genome editing

The generation of the first genetically modified (GM) bacteria, animals, and plants symbolized the dawn of the genetic engineering era[20-22] (Fig. 1). In 1977, a human protein (somatostatin) produced by GM *Escherichia coli* was commercialized for the first time. Shortly after, pivotal technological advances (Sanger sequencing, Polymerase Chain

Reaction, Electroporation, Gene gun) broadened and sped up the possibilities to use recombinant DNA technology[23-26], setting the stage for commercialization of several GM organisms[27-32]. In 1987, Oliver Smithies and Mario Capecchi pioneered site−directed genomic modification (called gene targeting) in mice, exploiting the cellular homologous recombination machinery[33,34]. In the following years, the bacteriophage−derived Cre recombinase, the yeast−derived Flp recombinase as well as the bacteriophage λ Red and the bacterial Rac prophage RecET systems were used as alternative approaches for site-specific recombination/recombineering[35-38]. However, the poor efficiency, the requirement for selectable markers, and the genomic scars left after editing limited their application. After the discovery that double−strand DNA breaks (DSDBs) largely enhance the genomic integration of foreign DNA in eukaryotic cells[39], meganucleases progressed the mammalian genome editing field even further, despite their low cleavage efficiency and fixed recognition sites. Alternatively, DNA binding modules of zinc fingers and of transcription activator–like effector (TALE) proteins have been fused to the restriction enzyme FokI, resulting in customizable nucleases called Zinc Finger Nucleases (ZFNs) and TALE Nucleases (TALENs), respectively[40,41]. Despite their programmability, these endonucleases require laborious, expensive, and time−consuming protein engineering to enable targeting of different sites[42,43].

The CRISPR–Cas era

In 1987, an array of Clustered Regularly Interspaced Short Palindromic Repeats was initially observed in the genome of *E. coli*[44], and was later named CRISPR[45] (Fig. 1). In 2000, aided by the rapid developments in the field of whole genome sequencing[46-50], Mojica *et al.* noticed that CRISPR arrays are widespread in prokaryotes[51]. CRISPR arrays were found to associate with a typical set of genes, encoding for CRISPR−associated (Cas) proteins[45]. The Cas genes were predicted to encode for helicases, nucleases and polymerases, implying their possible role in DNA repair[52]. In 2005, three independent studies suggested that CRISPR−Cas are adaptive immune systems against mobile genetic elements (MGEs), such as viruses and plasmids[53-55]. This hypothesis was experimentally proven two years later by Barrangou *et al*[56], and the first insights into the molecular mechanism of CRISPR−Cas immunity were provided by Brouns *et al*[57]. A plethora of follow−up studies that are still ongoing today focused on the characterization of the impressive diversity and functionality of CRISPR−Cas systems[58-72]. The ability of CRISPR−Cas systems to recognize and cleave specific DNA sequences generating double-strand breaks (DSBs) led, in 2013,

to their repurpose as next generation tools for genetic engineering in bacteria, plants, and human cells[73-78]. The need for control of these tools resulted in the discovery of natural inhibitors of CRISPR−Cas systems, known as anti−CRISPR proteins (Acrs)[79]. Acrs are small proteins encoded from prokaryotic viruses or prophage regions, and reflect the evolutionary complexity of CRISPR−Cas systems[79,80]. To date, Acrs have been extensively used for temporal and spatial control of CRISPR−mediated genome editing and transcriptional control[80]. A major next innovation was to utilize Cas proteins as a vehicle for delivery of desired proteins/catalytic function to a locus of choice, by creating (partially) catalytically inactive Cas fusion proteins. Of particular interest are the fusions with deaminases (base-editors) and reverse transcriptases (prime-editors) that allow for editing of DNA without DSBs, thus limiting the production of undesired by-products[81-87]. These so-called base−editing and prime−editing technologies, together with the CRISPR−Cas and anti−CRISPR technologies, have revolutionized gene therapy, crop improvement, and strain development[80,88,89]. Notably, the exploitation of all these technologies had paved the way for several, currently ongoing CRISPR clinical trials for the treatment of Acquired Immunodeficiency Syndrome (AIDs), cancer, blindness, β−thalassemia, hereditary angioedema, diabetes, and urinary tract infection[90-92].

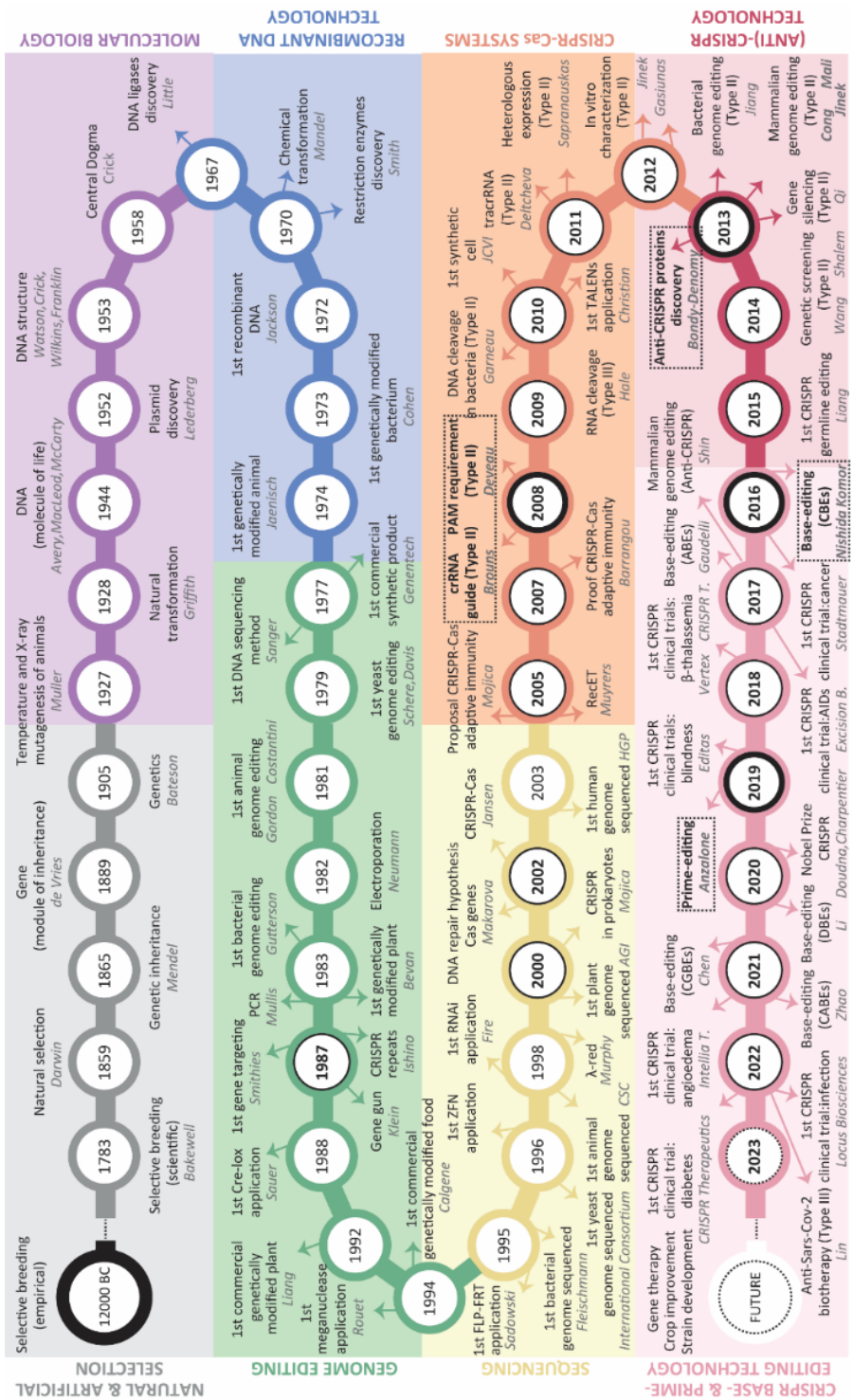


Fig. 1 History of genetic engineering. Modern genetic engineering is steadily shaped through eight different periods of development. These include the periods of natura and artificial selection (grey), molecular biology (purple), genome editing (green), recombinant DNA technology (blue), sequencing (yellow), CRISPR-Cas technology (orange), anti-CRISPR technology (dark pink), CRISPR base- and prime-editing technologies (light pink).

CRISPR-Cas systems

Architecture

CRISPR−Cas systems are adaptive immune mechanisms present in most archaea and many bacteria against invading MGEs, such as plasmids and viruses[93-98]. The CRISPR−Cas locus typically consists of a CRISPR array, a leader sequence, and a Cas−gene cluster. The CRISPR array is composed of short (21−47 bp), partially palindromic, repetitive segments (repeats) interspaced by short (20−72 bp), variable DNA sequences (spacers) derived from previous exposures to foreign genetic material[96,99]. The leader sequence contains a promoter region and other regulator elements, and is commonly adjacent to the CRISPR array to allow for acquisition of new spacers and for transcription of the CRISPR array[100]. The Cas−gene cluster encodes numerous and highly diverse Cas proteins that are used for the classification of the CRISPR-Cas systems (see below) and, together with the CRISPR array and the leader sequence, play a major role in CRISPR−Cas immunity[72].

Mechanism

CRISPR–Cas immunity develops in three stages: adaptation, expression, and interference, as initially proposed by Van der Oost *et al.* (2009)[93] (Fig. 2a, b). At the adaptation stage, a Cas protein complex (consisting of at least the Cas1 and Cas2 proteins) processes a short fragment of the foreign DNA (protospacer) and incorporates it as a new spacer in the CRISPR array, after duplication of the leader-proximal CRISPR repeat[101]. In many CRISPR-Cas systems, acquisition is limited to protospacers that are flanked by a variable short sequence, called protospacer adjacent motif (PAM)[102]. Spacer acquisition triggered by a partially or fully complementary pre−existing spacer (primed adaptation) is more efficient than spacer acquisition from an unknown source (naïve adaptation)[103-105]. Adaptation may sometimes require the activity of non−Cas/ host exonucleases (RecBCD, AddAB, DnaQ, ExoT), polymerase(s), and/or ligase(s)[106-114]. Newly acquired spacers form immunological memory for a sequence−specific defense against future invasions of complementary MGEs[102]. At the expression stage, the CRISPR array is typically transcribed into a long, precursor CRISPR RNA (pre−crRNA) that is further processed by Cas ribonucleases or by non-Cas host RNases into smaller, mature crRNA molecules. Each mature crRNA contains a single spacer and parts of the flanking repeat sequences. The mature crRNA molecules remain bound or associate with other Cas protein(s), forming effector CRISPR

ribonucleoprotein complexes (crRNPs)[57,61]. At the interference stage, crRNPs bind and cleave invading DNA or RNA, through partial or full base−pair complementarity between the spacer and the target protospacer sequences. Similar to adaptation, recognition of foreign genetic material usually requires the presence of a PAM. Lethal self−targeting (auto−immunity) is avoided by the absence of PAM in the CRISPR array of the host chromosome[58,60,93,115].

Classification

Since their discovery, an impressive plethora of different CRISPR−Cas systems has been identified. The current classification scheme groups CRISPR−Cas systems into two classes, six types, and at least thirty-three subtypes[72]. Class 1 systems contain a multi−subunit effector complex (with the exception of type III-E), while Class 2 systems make use of a single multidomain effector protein (Fig. 2c). Class 1 systems are categorized into type I, type III, and type IV systems, while Class 2 systems are classified into type II, type V, and type VI systems, based on their signature Cas genes.

1. Class 1 systems

In type I systems (Fig. 2a), like in most types, the adaptation module is composed of a Cas1-Cas2 complex, formed by four Cas1 monomers bridged by two Cas1 monomers[116]. Cas1 shows endonuclease activity on ssDNA, dsDNA, and ssRNA and is essential for spacer acquisition[117-120]. Although Cas2 also presents endonuclease activity on dsDNA and ssRNA, its main role is to bridge the Cas1 monomers, assist their binding to the CRISPR locus, and stabilize the pre-spacer DNA[116,121-123]. In some cases, DnaQ (DNA exonuclease) and Cas4 (DNA nuclease) trim the 3' end of the PAM-specific pre-spacer to optimize the efficiency and fidelity of pre-spacer integration into the CRISPR array[124-126]. Type I systems are grouped into seven subtypes (I-A to I-G). Processing of pre-crRNA is generally performed by Cas6, and rarely (subtype I-C) by Cas5[72,127]. The multi−subunit effector complex in type I systems is a Cascade complex that recruits the Cas3 nuclease to the target dsDNA for PAM-dependent cleavage[72] (Fig. 3).

Type III systems are grouped into six subtypes (III-A to III-F)[72]. They do not always have an adaptation module, suggesting they work in conjunction with the adaptation module of another resident CRISPR-Cas system. Since type III systems target RNA, they acquire

spacers from transcriptionally active regions of DNA and make use of reverse-transcriptase fused to Cas1. Adaptation in type III systems is PAM-independent and does not depend on Cas6 or interference-related Cas proteins[111-114,128]. In contrast, maturation of pre-crRNA is mediated by Cas6, which is usually provided by other CRISPR-cas loci. Target binding induces sequence-unspecific DNA cleavage by the HD domain of Cas10 and cyclic oligoadenylate (cOA) production by the Palm domain (although some lack DNase activity)[69,72,127-129]. These signaling cOA molecules trigger diverse CARF and SAVED proteins thought to provoke cell death (abortive infection)[130,131].

Type IV systems are the most enigmatic CRISPR systems to date, as they are mainly expressed from plasmids, they sometimes lack adaptation genes and they have no apparent nuclease activity[132-134]. Type IV systems are grouped into five subtypes (IV-A to IV-E) and usually co-exist with type I systems, possibly taking advantage of their adaptation machineries[72,134-138]. Binding of the Cascade:crRNA to the target DNA may trigger the recruitment of DinG, which in turn unwinds target dsDNA and causes gene downregulation or probably recruits host nucleases[139,140,141-144]. Alternatively, Cascade binds small heterogeneous RNAs, probably recruiting CysH that stabilizes AMPylation of specific substrates for immune response, gene regulation, or generation of secondary messengers[145,146]. Cascade may also sponge up and inactivate small guide RNAs to protect the plasmid from RNA-guided defense mechanisms and provide selective advantage[132-134,147,148].

2. Class 2 systems

Type II systems (Fig. 2b) are grouped into three subtypes (II-A, II-B, and II-C)[72]. Adaptation involves the Cas1-Cas2 complex, which usually interacts with Cas9 via Csn2 or with Cas4[149-155]. Maturation of the pre-crRNA requires binding of a transactivating CRISPR RNA (tracrRNA) molecule and of the Cas9 protein. The tracrRNA is composed of three stem loops and an anti−repeat sequence that base-pairs with the crRNA repeat. Cas9 stabilizes the tracrRNA:crRNA duplex and recruits the host ribonuclease RNase III to process the duplexed repeat[57,61,66,67]. During interference, Cas9 binds to target dsDNA via PAM recognition and crRNA-DNA base-pairing, creating a blunt DSDB (Fig. 3). Exceptionally, some type II-A and type II-C Cas9 proteins are able to additionally cleave ssRNA, in a PAM-independent manner[156-158].

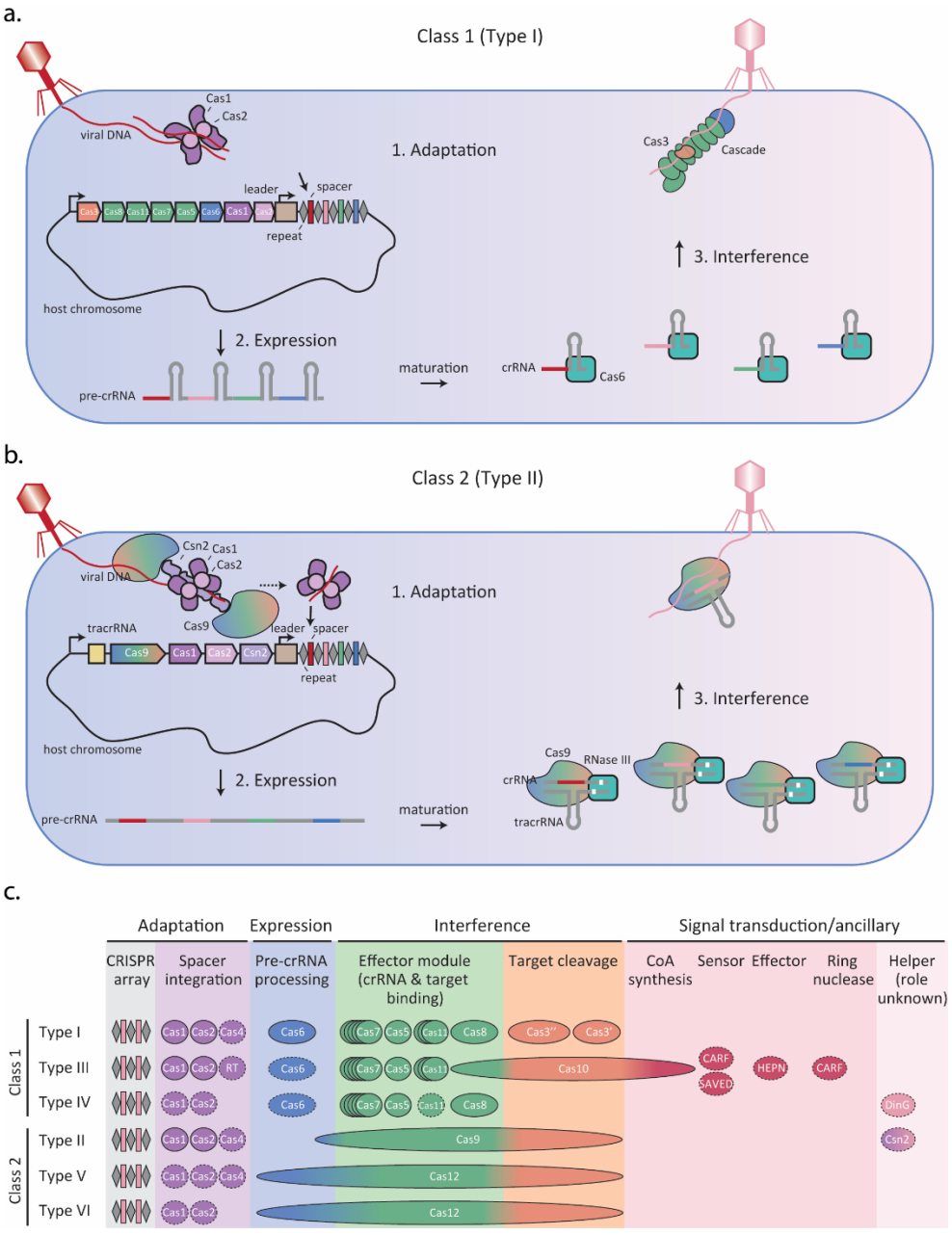


Fig. 2 General mechanism and classification of CRISPR-Cas systems. Schematic representation of the mechanism of immunity in (**a**) Class 1 and (**b**) Class 2 systems. As examples, the mechanisms of (**a**) type I and (**b**) type II systems are depicted. (**c**) Current classification of CRISPR-Cas systems (adapted from Makarova *et al.*[72]).

Type V systems are grouped into thirteen subtypes (V-A to V-K as well as V-M and V-U), either lacking an adaptation module or expressing Cas1 usually with Cas2, Cas4, and/or Cas12[72]. Although processing of pre-crRNA is usually performed solely by the effector complex, sometimes tracrRNA and/or RNase III activity are required[61,159-168]. During interference, most Cas12 effector proteins bind to target DNA via PAM recognition and crRNA-DNA base-pairing, creating staggered DSDB[160,164,168-175] (Fig. 3). Some of them, additionally exhibit non-specific ssDNase activity in a PAM-independent manner[162,164,176]. Alternatively, they may nick the non-target DNA strand of the dsDNA leading to collateral ssDNA cleavage, they may target ssRNA and then collaterally degrade both ssDNA and ssRNA, or they may direct insertion of Tn7-like transposons downstream of the protospacer on the target dsDNA via RNA-guided Tn7 transposition[161,171,177,178]. In few cases, they may bind dsDNA in a PAM-dependent manner without exhibiting cleavage activity[165,179].

Type VI systems are divided into six subtypes (VI-A to VI-D as well as VI-X and VI-Y)[72,180]. Type VI systems either encode the Cas1-Cas2 complex or lack adaptation module using Cas1 from type II-C systems *in trans*. Alternatively, they employ a RT-Cas1 protein for spacer acquisition from RNA[72,181,182]. Processing of pre-crRNA is often performed by the effector protein. All characterized Cas13 effector proteins mediate crRNA-guided ssRNA-cleavage in a PAM-independent manner, followed by promiscuous RNA degradation[72,183-186] (Fig. 3). However, some Cas13 proteins require a double-sided protospacer flanking site (PFS) for target recognition[184,187,188]. Their activity may be repressed and enhanced by the accessory proteins Csx27 and Csx28/ WYL1, respectively[184,186].

In general, multiple CRISPR-Cas loci may co-exist in the same host, often exhibiting cross-talk at any of the three stages of immunity. Moreover, CRISPR-Cas systems sometimes present functions other than adaptive immunity, including role in pathogenicity, biofilm formation, and group behaviour[189].

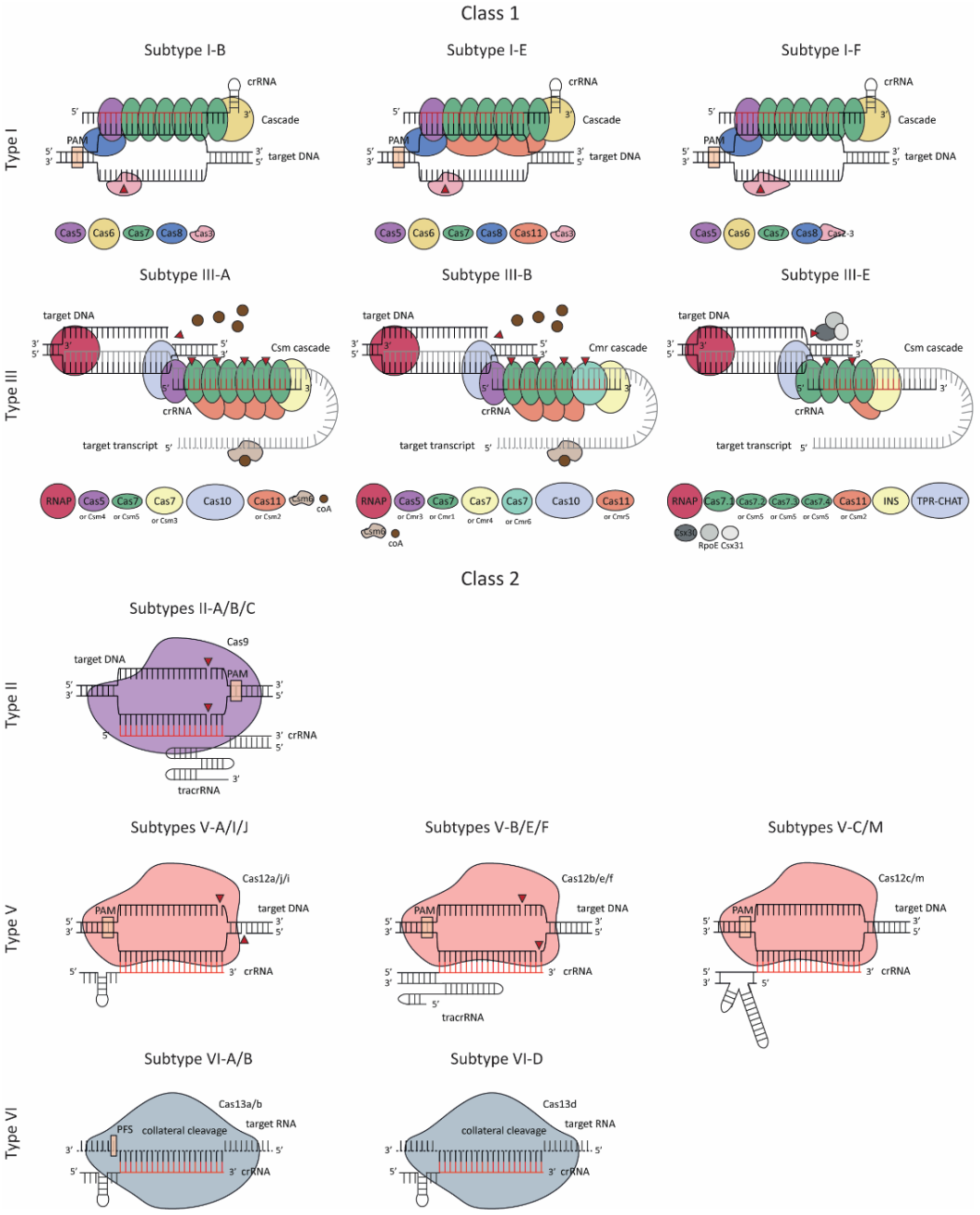


Fig 3 CRISPR-Cas effector complexes that have been used for genome editing and/or transcriptional regulation. Class 1 systems express a multi−subunit effector complex (Cascade), while Class 2 systems express a single multidomain effector complex. Red arrows indicate cleavage sites.

CRISPR-Cas technology

In nature, prokaryotes use the RNA-guided Cas endonucleases of the variable CRISPR–Cas systems to bind and cleave foreign DNA or RNA mainly to confer adaptive immunity (Fig. 3). The first proof that Cas endonucleases can be programmed for sequence-specific genome editing and transcriptional control in bacterial, plant, and human cells constituted the outset of the running rampant field of CRISPR–Cas discovery, engineering and application[73-78]. Here, we discuss the current developments of CRISPR-Cas technology and approaches to control excessive engineering.

Genome editing

DNA endonucleases of Class2 CRISPR-Cas systems (types II and V) have taken CRISPR tool development by storm, due to their minimal, single-protein effector domains. Few applications have also embraced the multisubunit effector of type I CRISPR-Cas systems. Recent genome editing techniques make use of Cas nucleases alone or combined with other enzymes, forming base-editors, prime-editors, Cas-related transposases, and Cas-related recombinases (Fig. 4).

1. DNA nucleases

Class 2 nucleases scan the genome for presence of PAM, locally unwind PAM-flanking dsDNA, and, after complementary spacer-protospacer base-pairing, they introduce DSDB within the protospacer region. Cas9 nucleases (type II) guided by the crRNA:tracrRNA chimera (or the synthetic single-guide RNA fusion; sgRNA) typically recognize a G/C-rich PAM and introduce a blunt-end DSDB at the region upstream of the PAM[66,67]. In contrast, most crRNA-guided Cas12 nucleases (type V) recognize an A/T-rich PAM and create a staggered DSDB at the region downstream of the PAM[160,164,168-175] (Fig. 3). In both cases, the DSDB is repaired via end-joining (EJ) mechanisms or homology-directed repair (HDR), depending on the cell type, cell state and the nature of DSDBs[190,191] (Fig. 4). EJ mechanisms are efficient in most eukaryotes, while they are scarce or inactive in most prokaryotes[192]. EJ, mainly represented by non-homologous end-joining (NHEJ), seals the DSDB through uncontrolled insertion and/or deletion of nucleotides (indels) that result in disruption of the target sequence. Hence, nucleases are usually programmed to target open reading

frames, cis-regulatory elements within promoters or enhancers as well as non-coding DNA regions[193-199]. Moreover, multiple guide RNAs can be used for introduction of Cas-mediated DSDBs at different sites simultaneously[75,76,200]. Pairs of guide RNAs can also be programmed to target neighboring genomic sites for deletion of the intervening DNA region[75,76,200]. Alternatively, the class 1 Cas3 nuclease (type I) is recruited by the crRNA-guided Cascade to the target dsDNA, after recognition of a PAM, and processively nicks the non-target strand, subsequently mediating collateral ssDNA damage for deletion of large DNA fragments[64,201-203]. Homology-independent insertion of large DNA fragments via EJ has also been reported, resulting in disruption of the target sequence[204-206]. Interestingly, HDR of Cas-mediated DSDBs enables precise editing with various types of modifications, from point mutations to gene deletion, insertion or replacement[206]. Both class 1 (type I and III) and class 2 (type II and V) systems have been combined with HDR, either triggering the DSDB repair or counter-selecting the non-edited cells[207-213]. However, HDR requires donor ssDNA or dsDNA template, is active only in dividing cells (exclusively during cell replication), and typically exhibits low efficiencies[74-76,214-216]. The biggest drawback of both genome editing applications based on EJ and HDR is the requirement of Cas-mediated DSDBs, which often trigger undesired chromosomal translocations, rearrangements, and activation of p53 responses[217-220].

2. DNA base-editors

CRISPR base-editors introduce precise, single-nucleotide transitions or transversions on target dsDNA, without the requirement of DSDBs, donor DNA templates, and HDR activity (Fig. 4). Current DNA base-editors are composed of a catalytically inactive Cas9 (nickase, nCas9; or 'dead'; dCas9) or Cas12 (dead; dCas12a) protein fused to at least one ssDNA deaminase enzyme and often proteins that influence DNA repair mechanisms. Five main types of base-editors have been developed to date, including (a) cytosine base-editors (CBEs) that convert C•G to T•A; (b) adenine base-editors (ABEs) that convert A•T to G•C; (c) cytosine-to-guanine base-editors (CGBEs) that convert C•G to G•C; (d) cytosine-to-adenine base-editors (CABEs) that convert C•G to A•T; and (e) dual base-editors (DBEs) that simultaneously convert C•G to T•A and A•T to G•C[81-85]. Base-editing technology is used not only for correction of single-nucleotide polymorphisms (SNPs) in dividing and non-dividing cells but also for gene (in)activation, editing of cis-regulatory elements, and site-directed mutagenesis[221-224]. Moreover, base-editing has been applied on mitochondrial and cytoplasmic DNA[224-229]. Major bottlenecks of base-editing technology

are the undesired by-stander and off-target edits, the suboptimal product purity, and the large size of the base-editing fusion proteins.

3. Prime-editors

Similar to base-editors, CRISPR prime-editors introduce point mutations on target dsDNA without DSDBs, donor DNA, and HDR activity (Fig. 4). However, primer-editors additionally enable insertion, deletion, replacement, integration and inversion of small or large DNA fragments. Current prime-editors consist of a partially inactive Cas9 (nCas9) protein fused to a reverse transcriptase (RT) that creates the desired modification using a small extension of an engineered guide RNA (pegRNA) as primer binding site (PBS) and reverse-transcriptase template (RTT). Four main types of prime-editors have been generated to date, including (a) prime-editor 1 (PE1) that contains the wild-type Moloney murine leukemia virus (M-MLV) RT; (b) prime-editor 2 (PE2) that contains an engineered penta-mutant of M-MLV RT; (c) prime-editor 3 (PE3) that combines PE2 with a conventional guide RNA in addition to pegRNA; (d) twin prime-editor (twinPE) that uses two pegRNAs and sometimes is co-expressed with a site-specific serine recombinase. Prime-editors present similar efficiency compared to Cas9-mediated EJ, avoiding the uncontrolled nature of DSB repair. Moreover, prime editing exhibits lower bystander and off-target editing compared to Cas9 base-editors but also lower on-target activity and product purity. Nevertheless, enhanced on-target efficiency is observed with engineered pegRNAs (epegRNAs). Like base-editors, the large size of prime-editors still remains a major bottleneck[87,230,231].

4. Transposases

Natural CRISPR-associated transposases and engineered dCas9-transposase fusions enable RNA-guided integration of large DNA fragments, independent of the stage of cell cycle (Fig. 4). Two natural CRISPR-associated transposases have been exploited to date, the type I-F and the type V-k systems. The type I-F system (Tn6677) is composed of Cas proteins (Cas6, Cas7, and Cas8-Cas5), the transposase operon (TnsA, TnsB and TnsC), the guide RNA, and the donor LE–cargo–RE transposase DNA substrate. The type V-K system (CAST) consists of Cas12k, the transposase operon (TnsB, TnsC, and TniQ), the guide RNA, and the donor LE–cargo–RE transposase DNA substrate. Both systems perform DNA integration approximately 60 base pairs downstream of the PAM but CAST shows higher off-targeting. The engineered Himar transposase–dCas9 fusion utilizes paired sgRNAs for

large cargo integration into genomic TA motifs. However, this system has high off-targeting editing and is restricted to applications in bacteria[232-235].

5. Recombinases

The engineered dCas9-Ginβ recombinase fusion (recCas9) uses paired sgRNAs to mediate targeted deletion of large DNA fragments (Fig. 4). Site-specific recombinases catalyze cleavage, strand exchange, and re-ligation of the dsDNA strands. Contrary to DNA nucleases, recombinase-mediated cleavage does not provoke indel formation (higher product purity), does not depend on cellular repair mechanisms, and is active also in non-dividing cells. However, despite the low off-target editing, these systems also present low on-target editing and stringent sequence requirements, limiting their applicability[236].

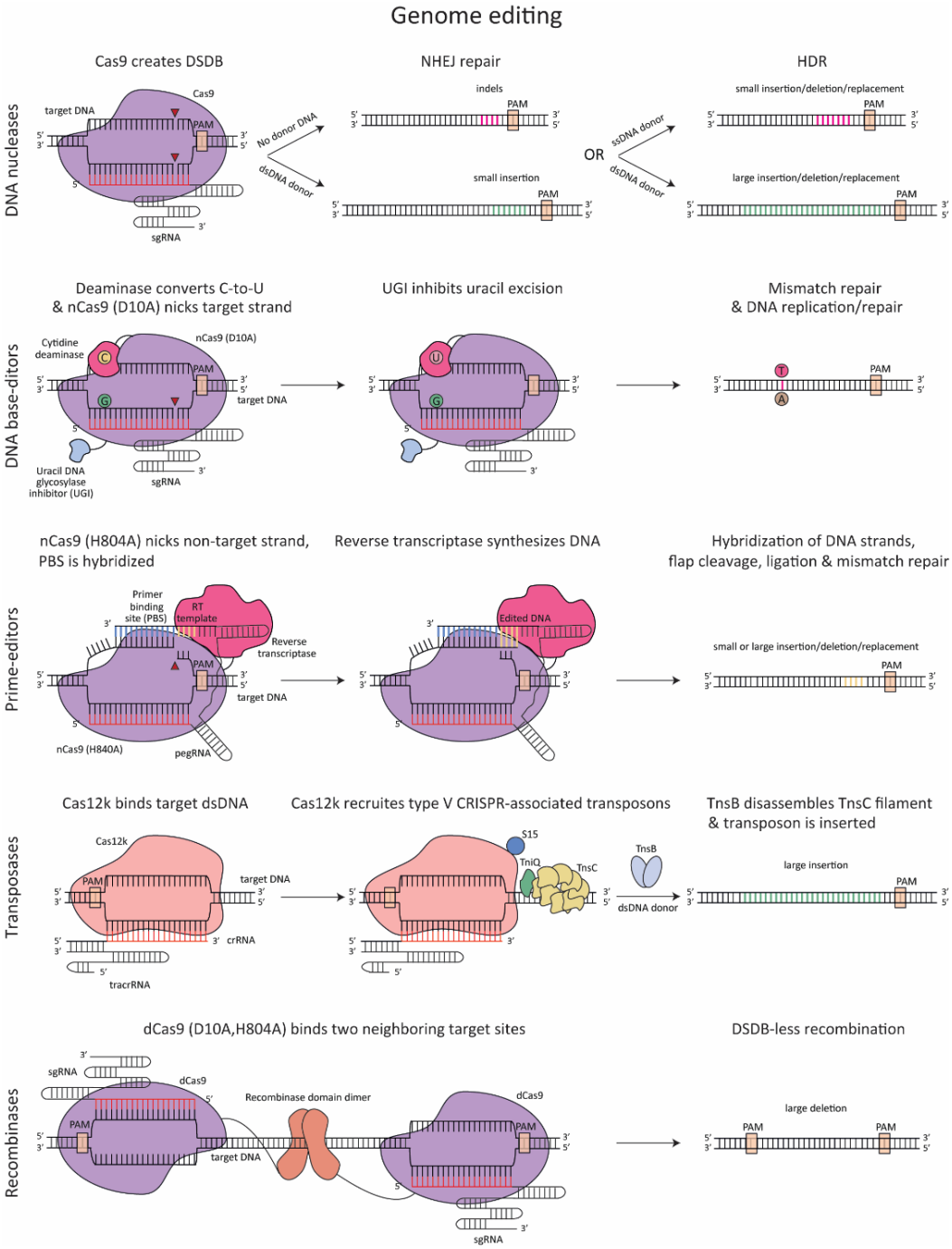


Fig 4 Overview of genome editing approaches. Five main strategies of genome editing have been applied to date. These include DNA nucleases, DNA base-editors, prime-editors, transposases, and recombinases. Red arrows indicate cleavage sites.

Transcriptional regulation

Recent transcriptional regulation techniques make use of inactive DNA nucleases (dCas9, dCas12), wild-type RNA nucleases as well as inactive RNA nucleases fused to deaminase enzymes (Fig. 5).

1. Inactive DNA nucleases fused to transcriptional effector domains

Catalytically inactive dCas9 and dCas12 (dCas12a, dCas12b) proteins maintain their dsDNA binding ability. Therefore, they are usually programmed to bind promoter regions or the beginning of coding regions in order to sterically block RNAP-mediated transcription initiation or elongation, respectively. This results in repression of gene expression, a process named CRISPR interference (CRISPRi)[74,78] (Fig. 5). Although very efficient in bacteria, CRISPRi is usually inefficient in eukaryotes, probably due to the difficulty in interrupting RNA polymerase II activity[78]. To improve CRISPRi efficiency, dCas9 was fused to a repressive effector domain, such as the KRAB domain of Kox1, the CS domain of HP1α, the WRPW domain of Hes1, the SID4X domain, the Max-interacting protein 1 (Mxi1), or the SRDX repression domain. Of them, the dCas9-KRAB protein was the most efficient[237-240]. Similarly, dCas9 was fused to an activating effector domain, such as the Herpes simplex VP16 or the nuclear factor kappa B (p65) for activation of gene expression, a process named CRISPR activation (CRISPRa) (Fig. 5). Multiple copies of VP16 (VP64 or VP160) have also been used for enhanced activation but the efficiency remained low[238]. Implementation of multiple sgRNAs targeting the promoter region as well as more structured dCas9 fusions (dCas9–SunTag, dCas9–VPR) further improved CRIPSRa efficiency[241-244]. The dCas9–SunTag system is composed of dCas9 fused to a multimeric array of peptides binding to single-chain variable fragments (scFv) that are fused to VP64 (or sometimes p65AD-HSF1), while the dCas9–VPR system consists of dCas9 fused to VP64, p65D, and Epstein–Barr virus R transactivator (Rta)[243-245]. Replacement of VP64 with p65AD-HSF1 in the dCas9–SunTag system (SPH system), a combination of dCas9–SunTag with dCas9–SAM were also applied[245]. Alternatively, RNA aptamers were fused to the guide RNA to recruit activating effector domains, forming the dCas9–SAM and scRNA systems. The dCas9–SAM system contains a dCas9–VP64 fusion protein and a modified guide RNA bearing two MS2 hairpins that interact with MS2 binding proteins (MCPs), which are fused to the trans-activator domain of p65 and the human heat shock factor 1 (HSF1)[246]. In the scRNA system, the guide RNA is fused to protein-binding RNA aptamers

that recruit cognate RNA-binding proteins (RBPs) to form RNA aptamer–RBP pairs (e.g., MS2–MCP, com–Com, PP7–PCP). Interestingly, the scRNA system enables simultaneous repression and activation of different genes[247]. Similarly, when dCas12a was fused to the transcriptional factor RemA and guided to bind a coding region it downregulated the target gene, while when guided to bind the promoter region it upregulated the target gene[248]. The dCas12a protein has also been used for multiplex silencing, targeting different genes or the same gene at different locations simultaneously[249,250]. Moreover, truncated guides have been combined with active Cas12a (or Cas9), allowing target binding but not cleavage. Providing multiple guides of different lengths, simultaneous genome editing and CRISPRi was achieved[78,251]. Other Cas12 variants that have been successfully applied for CRISPRi and/or CRISPRa are the engineered dCas12b, dCas12e, dCas12f (dCasMINI), and dCas12j as well as the naturally deoxyribonuclease-inactive Cas12c and Cas12m[165,179,252-256]. Except from Class 2-based gene regulation, the Class 1 type I-B, I-E, I-F, and III-A CRISPR-Cas system have been harnessed for CRISPRi and CRISPRa purposes, usually through fusion of the Cascade to (multiple copies of) repression and/or activation domains[213,257-262]. Nevertheless, the large size of the CRISPR transcriptional regulator systems still limits their applications.

2. RNA nucleases

Except from deactivated Cas DNA nucleases that stably bind dsDNA to block transcription initiation or elongation, type-VI RNA nucleases (Cas13a, Cas13b, and Cas13d) have been used to cleave mRNA transcripts and thus repress gene expression[160,184,186,187,263-265] (Fig. 5). In addition, although engineered and certain naturally occurring type II-A and type II-C Cas9 variants with both dsDNA and ssRNA nuclease activity have been applied, they come with the risk of DNA off-target editing[156-158,265-273].

3. RNA base-editors

In addition to transcriptional regulation and RNA cleavage, RNA editing can also limit the number of functional transcripts, through generation of stop codons or mRNA mutagenesis. Current CRISPR RNA base-editors are composed of dCas13b or truncated dCas13X fused to at least one RNA deaminase enzyme (Fig. 5). Three main types of CRISPR RNA base-editors have been developed to date, including (a) CBEs that convert C-to-U

(RESCUEr16 and RESCUE-S systems); (b) ABEs that convert A-to-I (REPAIRv1, REPAIRv2, REPAIRx, SNAP-ADAR1, SNAP-ADAR2, and HALO-ADAR1 systems); and (c) DBEs that simultaneously convert C-to-U and A-to-I (RESCUEr16 and RESCUE-S systems with tailored guide RNAs, and HALO-ADAR1-APOBEC1 system)[180,274-277]. However, the undesired by-stander and off-target edits, the suboptimal product purity, and the large size of these editors are still challenging.

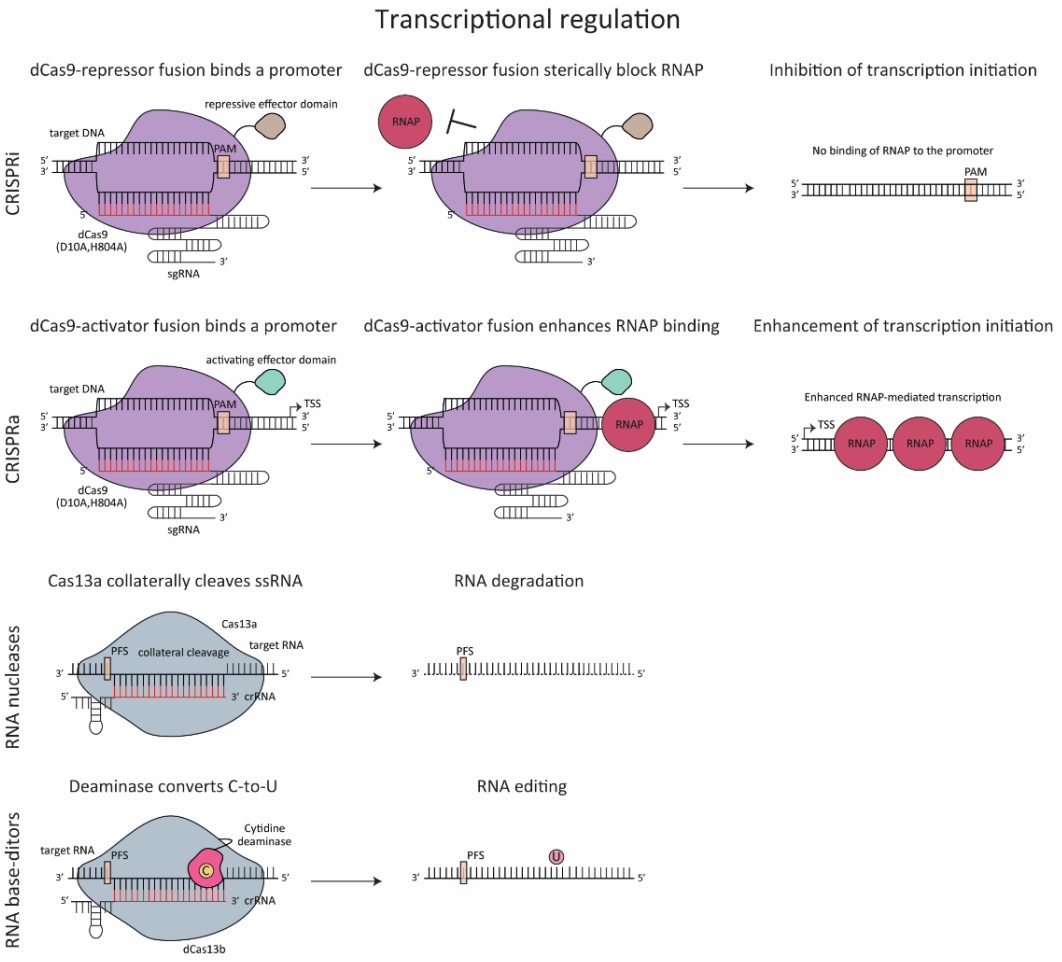


Fig 5 Overview of translational regulation approaches. Four main strategies of transcriptional regulation have been applied to date. These include CRISPR interference (CRISPRi), CRISPR activation (CRISPRa), RNA nucleases, and RNA base-editors.

Control of CRISPR-Cas technology

Excessive genome editing and transcriptional regulation can be controlled using chemicals, light, splitting, small RNAs and anti-CRISPR proteins (Fig. 6).

1. Ligand-inducible control

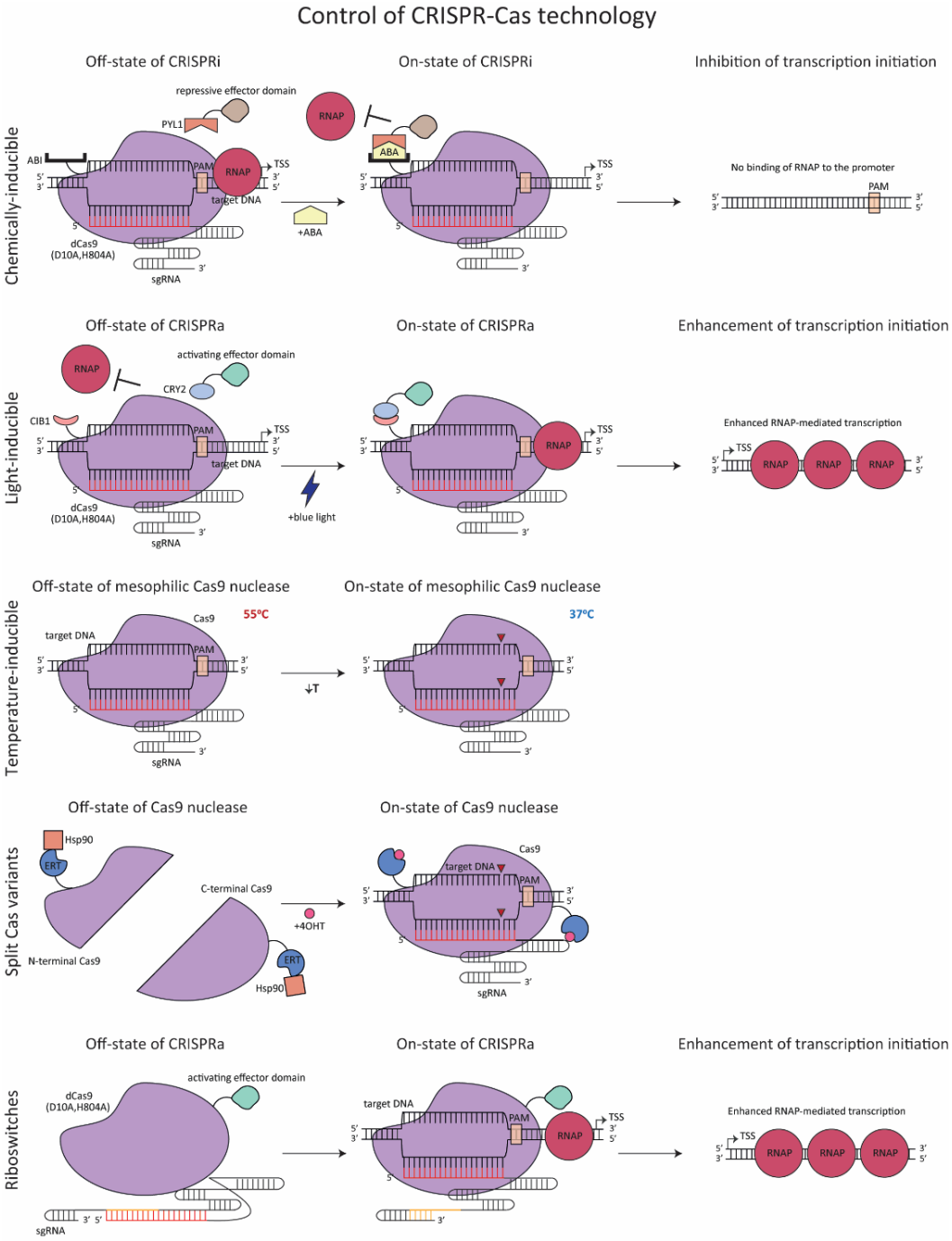
The dCas9 or dCas12a protein and its transcriptional effector domain have been separately fused to several chemically-induced dimerizing domains (CIDs), including abscisic acid (ABA)-inducible ABI–PYL1, gibber-ellin (GA)-inducible GID1–GAI24, and rapamycin-inducible FKBP–FRB[278-282] (Fig. 6). Similarly, dCas9 or Cas13a and its transcriptional effector domain have been separately coupled to optogenetically-inducible dimerizing domains (OIDs), such as blue light-inducible CRY2-CIB1 or VVD-GAVPO, Magnet pMag–nMag, and phytochrome-based red light-inducible PhyB–PIF[283-288] (Fig. 6). Presence of ligands (chemical or light) triggers dimerization of CIDs or OIDs, positioning the effector domain to the dCas9 or Cas13a binding site[257-259,278-288]. Alternatively, riboswitches embedded either in the guide RNA or the *cas9* coding sequence have been applied[289-296] (Fig. 6). Moreover, a temperature switch can temporarily affect the activity of Cas9 or Cas12a[297-299] (Fig. 6). Combination of optogenic and temperature control of Cas9 has also been developed[300]. To additionally confer spatiotemporal control of gene expression, split dCas9 variants or (split) dCas9 and dCas12a fusions to cellular receptors have been implemented[288,301-303] (Fig. 6). Characteristic examples are the split dCas9-Tango GPCR system, the ChaCha system, the receptor tyrosine kinase (RTK) system, and the modular extracellular sensor architecture (MESA) system. Split dCas9-Tango GPCR system contains split dCas9 fused to the intracellular domain of a G protein-coupled receptor (GPCR) and the adaptor protein Beta Arrestin (ARRB2) fused to a specific protease (tobacco etch virus, TEV). The ChaCha system contains a TEV protease as well as dCas9 fused to ARRB2 through a TEV site. The RTK system consists of a split dCas9 fused to RTK, while MESA is composed of dCas9 fused to a dimerizing target chain via a cleavable linker and of TEV protease fused to the other protease chain[304-309]. Nevertheless, the high ligand-independent background, the low efficiency and signal sensitivity as well as the large size of these fusion proteins significantly limit their applications.

2. Complementary oligonucleotides

Oligonucleotide-based inhibitors can also be used to inhibit the activity of Cas9 and Cas12a (Fig. 6). Characteristic examples are guide-complementary DNA oligonucleotides, DNA oligonucleotides with phosphorothioate linkages, RNA-DNA hybrids, truncated guide RNAs that allow Cas9 binding but not target dsDNA cleavage, and chemically-modified DNA oligonucleotides that are complementary to the PAM-interacting domain of Cas9 and/or to the repeat sequence of the guide RNA[310-315].

3. Anti-CRISPR proteins

Anti-CRISPRs, encoded by diverse MGEs, are small proteins that naturally inhibit CRISPR-Cas systems at different stages of immunity. To date, at least 89 families of anti-CRISPR proteins have been discovered, inhibiting type I, II, III, V, and VI CRISPR-Cas systems. Despite their recent characterization, several anti-CRISPR proteins have already been harnessed as off-switches in genome editing and transcriptional regulation applications in order to reduce off-target effects, off-cell activities, and cell toxicity (Fig. 6). The most commonly used anti-CRISPR proteins are those blocking type II (AcrIIA2, AcrIIA4, AcrIIA5, AcrIIC1, and AcrIIC3), and type VI (AcrVIA2, AcrVIA5) CRISPR-Cas effector proteins[316-323].


Control of CRISPR-Cas technology
Chemically-inducible
Off-state of CRISPRi
On-state of CRISPRi
Inhibition of transcription initiation
repressive effector domain
PYL1
ABI
RNAP
PAM
TSS
target DNA
dCas9
(D10A,H804A)
sgRNA
ABA
+ABA
No binding of RNAP to the promoter
Light-inducible
Off-state of CRISPRa
On-state of CRISPRa
Enhancement of transcription initiation
activating effector domain
CRY2
CIB1
+blue light
Enhanced RNAP-mediated transcription
Temperature-inducible
Off-state of mesophilic Cas9 nuclease
On-state of mesophilic Cas9 nuclease
Cas9
55°C
37°C
↓T
Split Cas variants
Off-state of Cas9 nuclease
On-state of Cas9 nuclease
Hsp90
ERT
C-terminal Cas9
N-terminal Cas9
+4OHT
Riboswitches
Off-state of CRISPRa
On-state of CRISPRa
Enhancement of transcription initiation
activating effector domain
dCas9
(D10A,H804A)
sgRNA
target DNA
Enhanced RNAP-mediated transcription

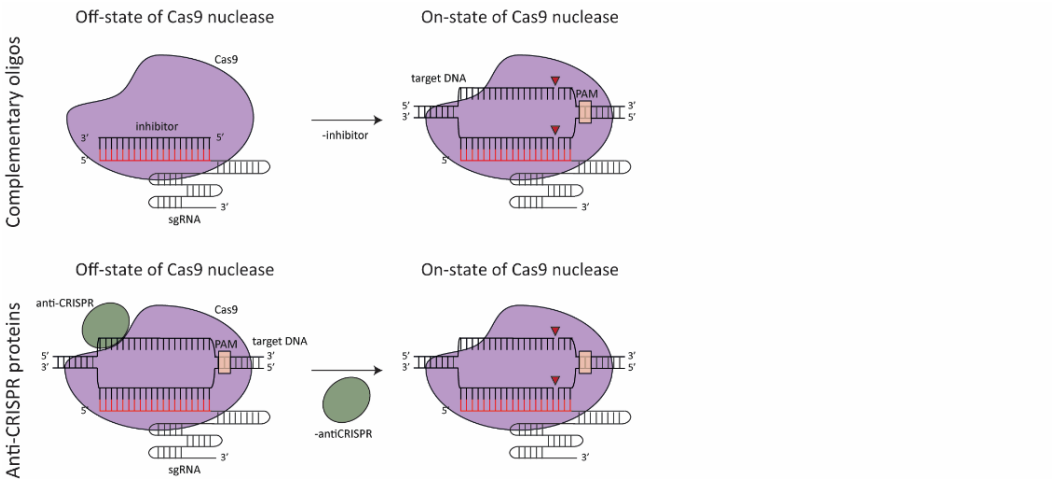


Fig 6 Overview of CRISPR-Cas control approaches. Seven main strategies of CRISPR-Cas control have been applied to date. These include chemically-inducible, light-inducible and temperature-inducible systems, split Cas variants, riboswitches, complementary oligos, and anti-CRISPR proteins. Red arrows indicate cleavage sites.

The evolving CRISPR-Cas technology has sparked the imagination of scientists around the world to develop cutting-edge genetic engineering approaches. The dream of browsing a catalog and simply pick a hyper-accurate, highly efficient, and unconditionally safe tool for any desired application is getting closer with each new innovation. We are just starting to witness how one of the most fascinating discoveries in science will shape our future. Hence, it is imperative to ensure that every step towards the new *status quo* is made with utmost care.

Thesis outline

Chapter 1 | General Introduction

The first chapter chronicles the ground-breaking events in the field of genetic engineering, starting from artificial selection and molecular biology moving towards recombinant DNA technology, genome editing, and the CRISPR−Cas era. The CRISPR revolution has brought forward a set of seminal discoveries revolving around the biology of CRISPR-Cas systems and their exploitation as next-generation genetic engineering tools. Here, we describe the architecture, mechanism of action, and current classification of CRISPR-Cas systems, shedding light on their impressive modular and functional complexity. Moreover, we summarize the current developments of CRISPR-Cas technology in genome editing and transcriptional regulation, providing strategies for their controlled and safe application.

Chapter 2 | Base-editing: from biology to technology

The second chapter reviews the biological characteristics of deaminase enzymes and their exploitation for precise genome and transcriptome editing. Deaminase enzymes are widespread in all kingdoms of life, playing a pivotal role in health and survival. We classify them based on their substrate specificity, and we scrutinize their mechanism of action, biological role, and association with diseases. Further on, we describe their development into CRISPR-guided, programmable base-editors, and we summarize the currently available base-editing tools. Finally, we discuss applications, bottlenecks, and future perspectives of CRISPR base-editing technology.

Chapter 3 | Characterizing a type III CRISPR–Cas–associated adenosine deaminase

The third chapter focuses on the characterization of an adenosine deaminase enzyme fused to a CARF domain from *Methylacidiphilum fumariolicum*, named mfCARF-AD. During type III CRISPR-Cas interference, cyclic oligoadenylate (cOA) signaling molecules are generated and bind CARF proteins to allosterically activate their effector domains for enhanced immunity. We show that mfCARF-AD, mainly existing in an oligomeric state, is a putative cOA-degrading ring-nuclease and cOA-activated adenosine deaminase with no

activity on free adenosine molecules. We postulate that mfCARF-AD interacts with purine salvage and carbon metabolism proteins, implying its role in crucial biological processes. In parallel, mfCARF-AD genetically associates with a type III-B CRISPR-Cas system, possibly connecting its biological role with bacterial immunity. However, the exact mechanism of action of this enzyme remains elusive.

Chapter 4 | Keeping CRISPR in check: diverse mechanisms of phage–encoded anti–CRISPRs

The fourth chapter reviews the biological traits of anti-CRISPR proteins and their utilization for spatiotemporal control of CRISPR-Cas technology. Anti-CRISPR proteins are small, natural inhibitors of CRISPR-Cas systems, evolved during the eternal arms-race between MGEs and prokaryotes. We describe strategies for anti-CRISPR protein discovery, and we categorize the identified anti-CRISPR proteins based on the class and type of CRISPR-Cas systems that they inhibit. Moreover, we analyze their highly diverse molecular mechanisms against CRISPR immunity, and we present their biological roles. Last, we summarize the latest advances of the anti-CRISPR technology and we provide future directions.

Chapter 5 | *In vivo* characterization of the AcrIIC1 anti–CRISPR protein for Cas9–based genome engineering

The fifth chapter is dedicated to the *in vivo* characterization of an anti-CRISPR protein from *Neisseria meningitidis*, called AcrIIC1. Although anti-CRISPR proteins are predominantly used to block the cleavage activity of Cas proteins, they can also be harnessed for several other CRISPR-based applications. We investigate the genome editing capabilities of the thermoactive type II-C Cas9 variants from *Geobacillus thermodenitrificans* T12 (ThermoCas9) and *Geobacillus stearothermophilus* (GeoCas9) in *Escherichia coli*. We show that AcrIIC1 strongly suppresses their DNA cleavage activity, but not their DNA binding ability. Harnessing this unique trait, we use the AcrIIC1:Cas9 complexes to develop novel gene silencing and the first anti-CRISPR base-editing tools. In this way, we provide novel approaches of the anti-CRISPR technology, and we propose their utilization in other mesophilic and thermophilic bacteria.

Chapter 6 | Efficient genome and base editing in human cells using ThermoCas9

The sixth chapter applies the knowledge gained from chapter 4 on ThermoCas9 activity at mesophilic temperatures to develop genome and base-editing tools for human cells. The type II-A CRISPR-Cas9 nuclease from *Streptococcus pyogenes* (SpyCas9) is used in the majority of genetic engineering applications that have been documented so far, limiting the genome-targeting range. We prove that the small, naturally specific and thermostable ThermoCas9 variant is active in human cells and that it can be used for genome editing as an alternative to SpyCas9. Furthermore, we establish a ThermoCas9-mediated base-editor (ThermoBE4) for C-to-T conversions in human genomes. ThermoBE4 has a three times wider window of activity than the equivalent SpyCas9 base-editor (BE4), which could be helpful for applications involving gene mutagenesis. ThermoCas9 thus offers an alternative platform that broadens the targeting range for genome- and base-editing in human cells.

Chapter 7 | Rational and random engineering of ThermoCas9 for optimal activity at moderate temperatures

The seventh chapter describes our efforts towards the optimization of ThermoCas9 cleavage activity at temperatures between 20-37°C. Due to its thermophilic nature, ThermoCas9 is active at low temperatures only under optimal conditions (full-length guides, perfect PAMs), as we observed using a previously developed toxin-targeting system in *E. coli*. Therefore, we generated a ThermoCas9 library via *in vitro* random mutagenesis to enable *in vivo* selection of catalytically enhanced ThermoCas9 variants under suboptimal conditions. Moreover, we detected key residues that may play a role in high protein stability and that are possible rational engineering targets for higher structural flexibility. Our initial findings provide the basis for the random and/or rational generation of ThermoCas9 variants with optimal activity at low temperatures, and thus with potential in biotechnological and medical applications.

Chapter 8 | Thesis Summary & General Discussion

The last chapter summarizes the work performed in this PhD thesis. It further discusses the most outstanding outcomes of this work, their novelty, and their overall contribution to the modern genetic engineering portfolio. Finally, novel ideas for further optimization of the current CRISPR-Cas technology are provided, leaving an open door for surprising advances in the field.

References

1. N Skoglund, P., Ersmark, E., Palkopoulou, E., & Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Current Biology, 25(11), 1515-1519.
2. Colledge, S., & Conolly, J. (Eds.). (2007). The origins and spread of domestic plants in Southwest Asia and Europe (pp. 1-424). Walnut Creek: Left Coast Press.
3. Joubert, D. M. (1977). Animal breeding--from Bakewell to Hammond: 1725-1964. Journal of the South African Veterinary Association, 48(4), 233-236.
4. Darwin, C. (1859). The Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life. Books, Incorporated, Pub..
5. Fisher, R. A. (1936). Has Mendel's work been rediscovered?. Annals of science, 1(2), 115-137.
6. De Vries, H. (1889). Intracellular Pangenesis, pp. 1-212.
7. Bateson, W. (1906). The progress of genetic research. In Report of the Third International Conference on Genetics (pp. 90-97). London: Spottiswoode & Co.
8. Muller, H. J. (1928). The measurement of gene mutation rate in Drosophila, its high variability, and its dependence upon temperature. *Genetics*, *13*(4), 279.
9. Muller, H. J. (1927). Artificial transmutation of the gene. *Science*, *66*(1699), 84-87.
10. Muller, H. J. (1928). The problem of genic modification. Verh. 5. internat. *Kongr. Vererbgsw., Berlin*, *1*.
11. Griffith, F. (1928). The significance of pneumococcal types. *Epidemiology & Infection*, *27*(2), 113-159.
12. Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *Die Entdeckung der Doppelhelix*, 97.
13. Watson, J. D., & Crick, F. H. (1953). A structure for deoxyribose nucleic acid.
14. Crick, F. H. (1958, January). On protein synthesis. In *Symp Soc Exp Biol* (Vol. 12, No. 138-63, p. 8).
15. Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiological reviews*, *32*(4), 403-430.
16. Little, J. W., Zimmerman, S. B., Oshinsky, C. K., & Gellert, M. (1967). Enzymatic joining of DNA strands, II. An enzyme-adenylate intermediate in the dpn-dependent DNA ligase reaction. *Proceedings of the National Academy of Sciences*, *58*(5), 2004-2011.
17. Smith, H. O., & Welcox, K. W. (1970). A restriction enzyme from Hemophilus influenzae: I. Purification and general properties. *Journal of molecular biology*, *51*(2), 379-391.
18. Mandel, M., & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *Journal of molecular biology*, *53*(1), 159-162.
19. Jackson, D. A., Symons, R. H., & Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proceedings of the National Academy of Sciences*, *69*(10), 2904-2909.
20. Cohen, S. N., & Chang, A. C. (1973). Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants. *Proceedings of the National Academy of Sciences*, *70*(5), 1293-1297.
21. Jaenisch, R., & Mintz, B. (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proceedings of the national academy of sciences*, *71*(4), 1250-1254.
22. Bevan, M. W., Flavell, R. B., & Chilton, M. D. (1983). A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, *304*(5922), 184-187.
23. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.
24. Mullis, K. B. (1994). *The polymerase chain reaction* (Vol. 41, No. 5). Springer science & business media.
25. Neumann, E., Schaefer-Ridder, M., Wang, Y., & Hofschneider, P. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. *The EMBO journal*, *1*(7), 841-845.
26. Klein, T. M., Wolf, E. D., Wu, R., & Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, *327*(6117), 70-73.
27. Struhl, K., Stinchcomb, D. T., Scherer, S., & Davis, R. W. (1979). High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proceedings of the National Academy of Sciences*, *76*(3), 1035-1039.
28. Gordon, J. W., & Ruddle, F. H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science*, *214*(4526), 1244-1246.
29. Gutterson, N. I., & Koshland Jr, D. E. (1983). Replacement and amplification of bacterial genes with sequences altered in vitro. *Proceedings of the National Academy of Sciences*, *80*(16), 4894-4898.

30. Bevan, M. W., Flavell, R. B., & Chilton, M. D. (1983). A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, *304*(5922), 184-187.
31. James, C. (1997). Global status of transgenic crops in 1997. *ISAAA briefs*, *5*, 1-31.
32. Kramer, M. G., & Redenbaugh, K. (1994). Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR™ tomato story. *Euphytica*, *79*(3), 293-297.
33. Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., & Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, *330*(6148), 576-578.
34. Thomas, K. R., & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, *51*(3), 503-512.
35. Sauer, B., & Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proceedings of the National Academy of Sciences*, *85*(14), 5166-5170.
36. Sadowski, P. D. (1995). The Flp Recombinase of th 2-µm Plasmid of Saccharomyces cerevisiae. *Progress in nucleic acid research and molecular biology*, *51*, 53-91.
37. Murphy, K. C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. *Journal of bacteriology*, *180*(8), 2063-2071.
38. Muyrers, J. P., Zhang, Y., Benes, V., Testa, G., Rientjes, J. M., & Stewart, A. F. (2004). ET Recombination. In *Bacterial Artificial Chromosomes* (pp. 107-121). Humana Press.
39. Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and cellular biology*, *14*(12), 8096-8106.
40. Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, *93*(3), 1156-1160.
41. Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., ... & Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, *186*(2), 757-761.
42. Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Stoddard, B. L., & Baker, D. (2006). Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, *441*(7093), 656-659.
43. Ménoret, S., Fontanière, S., Jantz, D., Tesson, L., Thinard, R., Rémy, S., ... & Anegon, I. (2013). Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. *The FASEB Journal*, *27*(2), 703-711.
44. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, *169*(12), 5429-5433.
45. Jansen, R., Embden, J. D. V., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, *43*(6), 1565-1575.
46. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... & Venter, J. C. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, *269*(5223), 496-512.
47. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., ... & Oliver, S. G. (1996). Life with 6000 genes. *Science*, *274*(5287), 546-567.
48. C. elegans Sequencing Consortium*. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*, *282*(5396), 2012-2018.
49. Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M., Rooney, T., Tallon, L. J., ... & Somerville, M. C. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796-815.
50. Guttmacher, A. E., & Collins, F. S. (2003). Welcome to the genomic era. New England Journal of Medicine, 349(10), 996-998.
51. Mojica, F. J., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, *36*(1), 244-246.
52. Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B., & Koonin, E. V. (2002). A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic acids research*, *30*(2), 482-496.
53. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*(2), 174-182.
54. Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551-2561.
55. Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653-663.
56. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712.

57. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., ... & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960-964.
58. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., ... & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology*, *190*(4), 1390-1400.
59. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., ... & Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, *139*(5), 945-956.
60. Garneau, J. E., Dupuis, M. È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., ... & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67-71.
61. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., ... & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602-607.
62. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research*, *39*(21), 9275-9282.
63. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO journal*, *30*(7), 1335-1342.
64. Westra, E. R., van Erp, P. B., Künne, T., Wong, S. P., Staals, R. H., Seegers, C. L., ... & van der Oost, J. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell*, *46*(5), 595-605.
65. Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *science*, *322*(5909), 1843-1845.
66. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science*, *337*(6096), 816-821.
67. Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39), E2579-E2586.
68. Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., ... & Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467-477.
69. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., ... & Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722-736.
70. Makarova, K. S., & Koonin, E. V. (2015). Annotation and classification of CRISPR-Cas systems. *CRISPR*, 47-75.
71. Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology*, *37*, 67-78.
72. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., ... & Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67-83.
73. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, *31*(3), 233-239.
74. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *elife*, *2*, e00471.
75. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.
76. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823-826.
77. Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D. L., Wei, P., ... & Zhu, J. K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell research*, *23*(10), 1229-1232.
78. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173-1183.
79. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, *493*(7432), 429-432.
80. Marino, N. D., Pinilla-Redondo, R., Csörgő, B., & Bondy-Denomy, J. (2020). Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature methods*, *17*(5), 471-479.
81. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424.
82. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., ... & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, *353*(6305), aaf8729.
83. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464-471.

84. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., ... & Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nature biotechnology*, *38*(7), 875-882.
85. Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., ... & Zhang, X. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nature biotechnology*, *39*(1), 35-40.
86. Chen, L., Park, J. E., Paa, P., Rajakumar, P. D., Prekop, H. T., Chew, Y. T., ... & Chew, W. L. (2021). Programmable C: G to G: C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nature communications*, *12*(1), 1-7.
87. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., ... & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149-157.
88. Anzalone, A. V., Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature biotechnology*, *38*(7), 824-844.
89. Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., ... & Huang, J. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & cell*, *6*(5), 363-372.
90. Henderson, H. (2022). CRISPR Clinical Trials: A 2022 Update. *Innovative Genomics Institute*.
91. Reardon, S. (2016). First CRISPR clinical trial gets green light from US panel. *Nat News*.
92. Lin, P., Shen, G., Guo, K., Qin, S., Pu, Q., Wang, Z., ... & Wu, M. (2022). Type III CRISPR-based RNA editing for programmable control of SARS-CoV-2 and human coronaviruses. *Nucleic acids research*, *50*(8), e47-e47.
93. Van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., & Brouns, S. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in biochemical sciences*, *34*(8), 401-407.
94. Deveau, H., Garneau, J. E., & Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annual review of microbiology*, *64*, 475-493.
95. Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, *327*(5962), 167-170.
96. Karginov, F. V., & Hannon, G. J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. *Molecular cell*, *37*(1), 7-19.
97. Koonin, E. V., & Makarova, K. S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 biology reports*, *1*.
98. Sorek, R., Kunin, V., & Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, *6*(3), 181-186.
99. Barrangou, R., & Van Der Oost, J. (2013). CRISPR-Cas systems. *RNA-Mediated Adaptative Immunity in Bacteria and Archaea. 1010079783642346576th ed. Heidelberg SVB, editor*.
100. Alkhnbashi, O. S., Shah, S. A., Garrett, R. A., Saunders, S. J., Costa, F., & Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics*, *32*(17), i576-i585.
101. Amitai, G., & Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. *Nature Reviews Microbiology*, *14*(2), 67-76.
102. Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*, *21*(6), 528-534.
103. Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications*, *3*(1), 1-7.
104. Swarts, D. C., Mosterd, C., Van Passel, M. W., & Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PloS one*, *7*(4), e35888.
105. Fineran, P. C., Gerritzen, M. J., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S. A., ... & Brouns, S. J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences*, *111*(16), E1629-E1638.
106. Ivančić-Baće, I., Cass, S. D., Wearne, S. J., & Bolt, E. L. (2015). Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic acids research*, *43*(22), 10821-10830.
107. Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., ... & Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, *520*(7548), 505-510.
108. Radovčić, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E. L., & Ivančić-Baće, I. (2018). CRISPR–Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5′ ssDNA exonucleases. *Nucleic acids research*, *46*(19), 10173-10183.
109. Ramachandran, A., Summerville, L., Learn, B. A., DeBell, L., & Bailey, S. (2020). Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. *Journal of Biological Chemistry*, *295*(11), 3403-3414.
110. Aviram, N., Thornal, A. N., Zeevi, D., & Marraffini, L. A. (2022). Different modes of spacer acquisition by the Staphylococcus epidermidis type III-A CRISPR-Cas system. *Nucleic acids research*, *50*(3), 1661-1672.

111. Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., ... & Fire, A. Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. *Science*, *351*(6276), aad4234.
112. González-Delgado, A., Mestre, M. R., Martínez-Abarca, F., & Toro, N. (2019). Spacer acquisition from RNA mediated by a natural reverse transcriptase-Cas1 fusion protein associated with a type III-D CRISPR–Cas system in Vibrio vulnificus. *Nucleic acids research*, *47*(19), 10202-10211.
113. Zhang, X., Garrett, S., Graveley, B. R., & Terns, M. P. (2022). Unique properties of spacer acquisition by the type III-A CRISPR-Cas system. *Nucleic acids research*, *50*(3), 1562-1582.
114. Aviram, N., Thornal, A. N., Zeevi, D., & Marraffini, L. A. (2022). Different modes of spacer acquisition by the Staphylococcus epidermidis type III-A CRISPR-Cas system. *Nucleic acids research*, *50*(3), 1661-1672.
115. Gasiunas, G., Sinkunas, T., & Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and molecular life sciences*, *71*(3), 449-465.
116. Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*, *21*(6), 528-534.
117. Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., ... & Yakunin, A. F. (2011). A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair. *Molecular microbiology*, *79*(2), 484-502.
118. Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, *17*(6), 904-912.
119. Kim, T. Y., Shin, M., Yen, L. H. T., & Kim, J. S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. *Biochemical and biophysical research communications*, *441*(4), 720-725.
120. Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research*, *40*(12), 5569-5576.
121. Beloglazova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., ... & Yakunin, A. F. (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *Journal of Biological Chemistry*, *283*(29), 20361-20371.
122. Nam, K. H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M. P., & Ke, A. (2012). Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *Journal of Biological Chemistry*, *287*(43), 35943-35952.
123. Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell*, *163*(4), 840-853.
124. Drabavicius, G., Sinkunas, T., Silanskas, A., Gasiunas, G., Venclovas, Č., & Siksnys, V. (2018). DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. *EMBO reports*, *19*(7), e45543.
125. Kim, S., Loeff, L., Colombo, S., Jergic, S., Brouns, S. J., & Joo, C. (2020). Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature*, *579*(7797), 141-145.
126. Almendros, C., Nobrega, F. L., McKenzie, R. E., & Brouns, S. J. J. (2019). Cas4–Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. *Nucleic acids research*, *47*(10), 5223-5230.
127. Charpentier, E., Richter, H., van der Oost, J., & White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS microbiology reviews*, *39*(3), 428-441.
128. Schmidt, F., Cherepkova, M. Y., & Platt, R. J. (2018). Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature*, *562*(7727), 380-385.
129. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., & Siksnys, V. (2016). Spatiotemporal control of type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. *Molecular cell*, *62*(2), 295-306.
130. Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č., & Koonin, E. V. (2020). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic acids research*, *48*(16), 8828-8847.
131. Steens, J. A., Salazar, C. R. P., & Staals, R. H. (2022). The diverse arsenal of type III CRISPR–Cas-associated CARF and SAVED effectors. *Biochemical Society Transactions*, *50*(5), 1353-1364.
132. Koonin, E. V., & Makarova, K. S. (2017). Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome biology and evolution*, *9*(10), 2812-2825.
133. Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B*, *374*(1772), 20180087.
134. Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2020). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic acids research*, *48*(4), 2000-2012.
135. Makarova, K. S., Aravind, L., Wolf, Y. I., & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology direct*, *6*(1), 1-27.

136. Staals, R. H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., Van Duijn, E., ... & Shinkai, A. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. *Molecular cell*, *52*(1), 135-145.
137. Elmore, J., Deighan, T., Westpheling, J., Terns, R. M., & Terns, M. P. (2015). DNA targeting by the type IG and type IA CRISPR–Cas systems of Pyrococcus furiosus. *Nucleic acids research*, *43*(21), 10353-10363.
138. Bernheim, A., Bikard, D., Touchon, M., & Rocha, E. P. (2020). Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements. *Nucleic acids research*, *48*(2), 748-760.
139. Grodick, M. A., Segal, H. M., Zwang, T. J., & Barton, J. K. (2014). DNA-mediated signaling by proteins with 4Fe–4S clusters is necessary for genomic integrity. *Journal of the American Chemical Society*, *136*(17), 6470-6478.
140. Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A type IV-A CRISPR-Cas system in Pseudomonas aeruginosa mediates RNA-guided plasmid interference in vivo. *The CRISPR journal*, *2*(6), 434-440.
141. Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., ... & Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature microbiology*, *4*(1), 89-96.
142. Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA biology*, *16*(10), 1438-1447.
143. Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2020). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic acids research*, *48*(4), 2000-2012.
144. Guo, X., Sanchez-Londono, M., Gomes-Filho, J. V., Hernandez-Tamayo, R., Rust, S., Immelmann, L. M., ... & Randau, L. (2022). Characterization of the self-targeting Type IV CRISPR interference system in Pseudomonas oleovorans. *Nature Microbiology*, *7*(11), 1870-1878.
145. Zhou, Y., Bravo, J. P., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *Iscience*, *24*(3), 102201.
146. Taylor, H. N., Laderman, E., Armbrust, M., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. *Frontiers in microbiology*, 1236.
147. Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., ... & Koonin, E. V. (2019). CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nature Reviews Microbiology*, *17*(8), 513-525.
148. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proceedings of the National Academy of Sciences*, *115*(23), E5307-E5316.
149. Ka, D., Jang, D. M., Han, B. W., & Bae, E. (2018). Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. *Nucleic acids research*, *46*(18), 9805-9815.
150. Wei, Y., Terns, R. M., & Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes & development*, *29*(4), 356-361.
151. Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., & Marraffini, L. A. (2015). Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature*, *519*(7542), 199-202.
152. Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., & Wigley, D. B. (2019). Structure of the DNA-bound spacer capture complex of a type II CRISPR-Cas system. *Molecular cell*, *75*(1), 90-101.
153. Arslan, Z., Wurm, R., Brener, O., Ellinger, P., Nagel-Steger, L., Oesterhelt, F., ... & Pul, Ü. (2013). Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic acids research*, *41*(12), 6347-6359.
154. Mir, A., Edraki, A., Lee, J., & Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS chemical biology*, *13*(2), 357-365.
155. Hooton, S. P., & Connerton, I. F. (2015). Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Frontiers in microbiology*, *5*, 744.
156. Dugar, G., Leenay, R. T., Eisenbart, S. K., Bischler, T., Aul, B. U., Beisel, C. L., & Sharma, C. M. (2018). CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the Campylobacter jejuni Cas9. *Molecular cell*, *69*(5), 893-905.
157. Rousseau, B. A., Hou, Z., Gramelspacher, M. J., & Zhang, Y. (2018). Programmable RNA cleavage and recognition by a natural CRISPR-Cas9 system from Neisseria meningitidis. *Molecular cell*, *69*(5), 906-914.
158. Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A., & Doudna, J. A. (2018). RNA-dependent RNA targeting by CRISPR-Cas9. *elife*, *7*, e32724.
159. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, *532*(7600), 517-521.
160. Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., ... & Koonin, E. V. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*, *60*(3), 385-397.

161. Zhang, H., Li, Z., Xiao, R., & Chang, L. (2020). Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nature structural & molecular biology*, *27*(11), 1069-1076.
162. Huang, X., Sun, W., Cheng, Z., Chen, M., Li, X., Wang, J., ... & Wang, Y. (2020). Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nature communications*, *11*(1), 1-14.
163. Zhang, B., Luo, D., Li, Y., Perčulija, V., Chen, J., Lin, J., ... & Ouyang, S. (2021). Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12i1. *Nature Communications*, *12*(1), 1-13.
164. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C. A., Li, Z., Cress, B. F., ... & Doudna, J. A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science*, *369*(6501), 333-337.
165. Wu, W., Mohanraju, P., Liao, C., Adiego-Pérez, B., Creutzburg, S. C., Makarova, K. S., Keessen, K., ... & van der Oost, J. The Miniature CRISPR-Cas12m Effector Binds DNA To Block Transcription.
166. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., ... & Ji, Q. (2021). Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nature Chemical Biology*, *17*(11), 1132-1138.
167. Kurihara, N., Nakagawa, R., Hirano, H., Okazaki, S., Tomita, A., Kobayashi, K., ... & Nureki, O. (2022). Structure of the type VC CRISPR-Cas effector enzyme. *Molecular Cell*, *82*(10), 1865-1877.
168. Harrington, L. B., Ma, E., Chen, J. S., Witte, I. P., Gertz, D., Paez-Espino, D., ... & Doudna, J. A. (2020). A scoutRNA is required for some type V CRISPR-Cas systems. *Molecular cell*, *79*(3), 416-424.
169. Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. *Cell*, *167*(7), 1814-1828.
170. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., ... & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759-771.
171. Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., ... & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science*, *363*(6422), 88-91.
172. Liu, J. J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L., ... & Doudna, J. A. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*, *566*(7743), 218-223.
173. Karvelis, T., Bigelyte, G., Young, J. K., Hou, Z., Zedaveinyte, R., Budre, K., ... & Siksnys, V. (2020). PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic acids research*, *48*(9), 5016-5023.
174. Wang, Y., Wang, Y., Pan, D., Yu, H., Zhang, Y., Chen, W., ... & Ji, Q. (2022). Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease. *Cell Reports*, *40*(13), 111418.
175. Chen, L. X., Al-Shayeb, B., Méheust, R., Li, W. J., Doudna, J. A., & Banfield, J. F. (2019). Candidate phyla radiation roizmanbacteria from hot springs have novel and unexpectedly abundant CRISPR-Cas systems. *Frontiers in microbiology*, *10*, 928.
176. Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., ... & Doudna, J. A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*, *362*(6416), 839-842.
177. Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, *360*(6387), 436-439.
178. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, *365*(6448), 48-53.
179. Huang, C. J., Adler, B. A., & Doudna, J. A. (2022). A naturally DNase-free CRISPR-Cas12c enzyme silences gene expression. *Molecular Cell*, *82*(11), 2148-2160.
180. Xu, C., Zhou, Y., Xiao, Q., He, B., Geng, G., Wang, Z., ... & Yang, H. (2021). Programmable RNA editing with compact CRISPR–Cas13 systems from uncultivated microbes. *Nature methods*, *18*(5), 499-506.
181. Toro, N., Mestre, M. R., Martínez-Abarca, F., & González-Delgado, A. (2019). Recruitment of reverse transcriptase-Cas1 fusion proteins by type VI-A CRISPR-Cas systems. *Frontiers in microbiology*, *10*, 2160.
182. Hoikkala, V., Ravantti, J., Díez-Villaseñor, C., Tiirola, M., Conrad, R. A., McBride, M. J., ... & Sundberg, L. R. (2021). Cooperation between different CRISPR-Cas types enables adaptation in an RNA-targeting system. *MBio*, *12*(2), e03338-20.
183. Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., ... & Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299), aaf5573.
184. Smargon, A. A., Cox, D. B., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., ... & Zhang, F. (2017). Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Molecular cell*, *65*(4), 618-630.
185. Meeske, A. J., Nakandakari-Higa, S., & Marraffini, L. A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*, *570*(7760), 241-245.

186. Yan, W. X., Chong, S., Zhang, H., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2018). Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Molecular cell*, *70*(2), 327-339.
187. Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., ... & Zhang, F. (2017). RNA targeting with CRISPR–Cas13. *Nature*, *550*(7675), 280-284.
188. East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, *538*(7624), 270-273.
189. Mohanraju, P., Saha, C., van Baarlen, P., Louwen, R., Staals, R. H., & van der Oost, J. (2022). Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology*, *20*(6), 351-364.
190. Chapman, J. R., Taylor, M. R., & Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell*, *47*(4), 497-510.
191. Yeh, C. D., Richardson, C. D., & Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nature cell biology*, *21*(12), 1468-1478.
192. Finger-Bou, M., Orsi, E., van der Oost, J., & Staals, R. H. (2020). CRISPR with a happy ending: Non-templated DNA repair for prokaryotic genome engineering. *Biotechnology Journal*, *15*(7), 1900404.
193. Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, *343*(6166), 80-84.
194. Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., ... & Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, *343*(6166), 84-87.
195. Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., ... & Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature biotechnology*, *32*(12), 1262-1267.
196. Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., ... & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, *34*(2), 184-191.
197. Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., ... & Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nature biotechnology*, *34*(2), 192-198.
198. Zhu, S., Li, W., Liu, J., Chen, C. H., Liao, Q., Xu, P., ... & Wei, W. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nature biotechnology*, *34*(12), 1279-1286.
199. Montalbano, A., Canver, M. C., & Sanjana, N. E. (2017). High-throughput approaches to pinpoint function within the noncoding genome. *Molecular cell*, *68*(1), 44-59.
200. Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *cell*, *153*(4), 910-918.
201. Dolan, A. E., Hou, Z., Xiao, Y., Gramelspacher, M. J., Heo, J., Howden, S. E., ... & Zhang, Y. (2019). Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-Cas. *Molecular cell*, *74*(5), 936-950.
202. Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E., ... & Mashimo, T. (2019). CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nature communications*, *10*(1), 1-13.
203. Cameron, P., Coons, M. M., Klompe, S. E., Lied, A. M., Smith, S. C., Vidal, B., ... & Sternberg, S. H. (2019). Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nature biotechnology*, *37*(12), 1471-1477.
204. Auer, T. O., Duroure, K., De Cian, A., Concordet, J. P., & Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome research*, *24*(1), 142-153.
205. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E. J., ... & Belmonte, J. C. I. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, *540*(7631), 144-149.
206. Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. *Annual review of biochemistry*, *79*, 181.
207. Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology*, *5*(11), a012740.
208. Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., ... & She, Q. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research*, *44*(4), e34-e34.
209. Cheng, F., Gong, L., Zhao, D., Yang, H., Zhou, J., Li, M., & Xiang, H. (2017). Harnessing the native type IB CRISPR-Cas for genome editing in a polyploid archaeon. *Journal of Genetics and Genomics*, *44*(11), 541-548.
210. Zhang, J., Zong, W., Hong, W., Zhang, Z. T., & Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metabolic engineering*, *47*, 49-59.

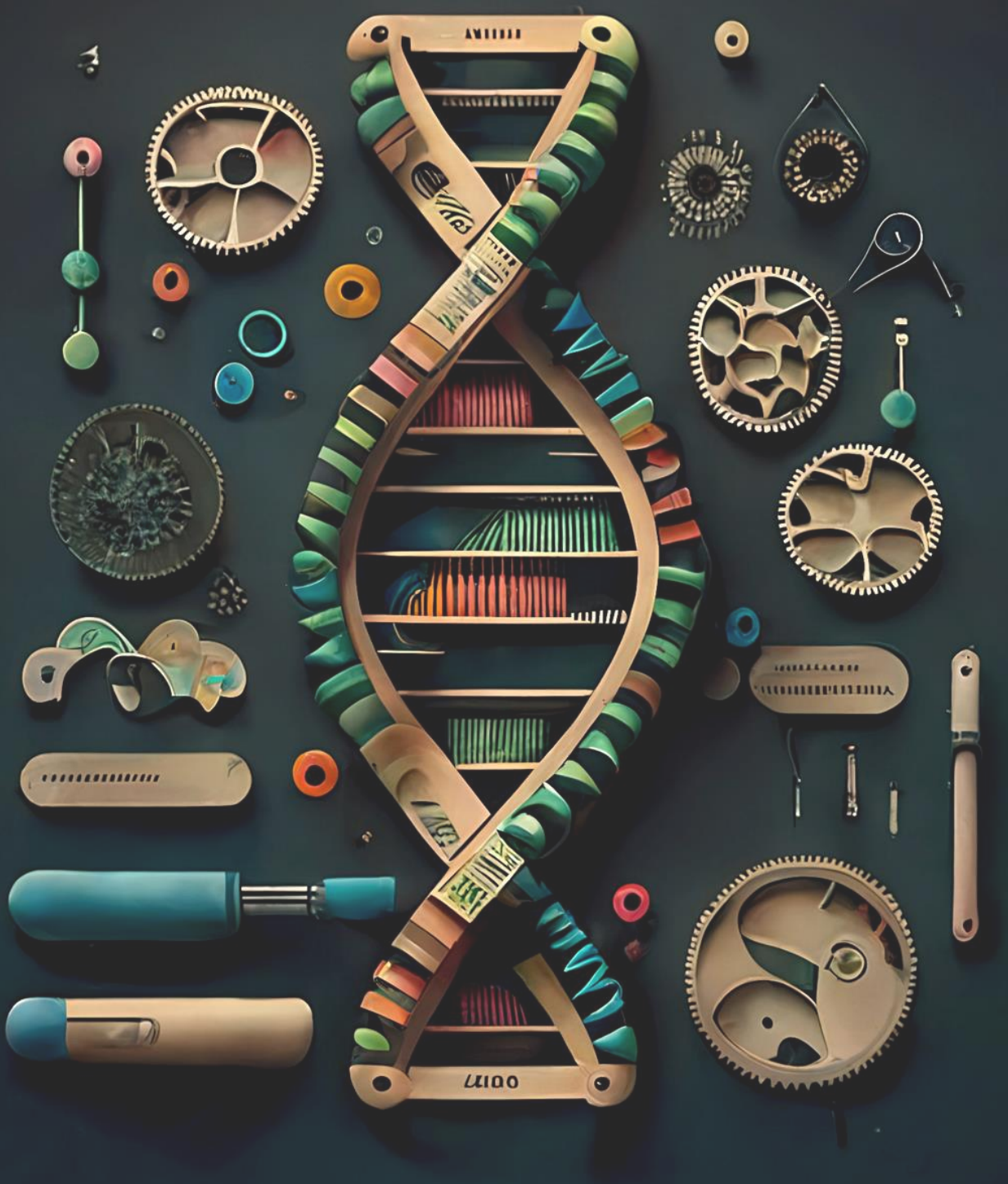
211. Hidalgo-Cantabrana, C., Goh, Y. J., Pan, M., Sanozky-Dawes, R., & Barrangou, R. (2019). Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. *Proceedings of the National Academy of Sciences*, 116(32), 15774-15783.
212. Ichikawa, H. T., Cooper, J. C., Lo, L., Potter, J., Terns, R. M., & Terns, M. P. (2017). Programmable Type III-A CRISPR-Cas DNA Targeting Modules. *PloS one*, *12*(4), e0176221.
213. Guan, J., Wang, W., & Sun, B. (2017). Chromosomal targeting by the type III-A CRISPR-Cas system can reshape genomes in Staphylococcus aureus. *Msphere*, *2*(6), e00403-17.
214. Heyer, W. D., Ehmsen, K. T., & Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annual review of genetics*, *44*, 113.
215. Moynahan, M. E., & Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature reviews Molecular cell biology*, *11*(3), 196-207.
216. Lin, S., Staahl, B. T., Alla, R. K., & Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *elife*, *3*, e04766.
217. Kosicki, M., Tomberg, K., & Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature biotechnology*, *36*(8), 765-771.
218. Cullot, G., Boutin, J., Toutain, J., Prat, F., Pennamen, P., Rooryck, C., ... & Bedel, A. (2019). CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature communications*, *10*(1), 1-14.
219. Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., ... & Kaykas, A. (2018). p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature medicine*, *24*(7), 939-946.
220. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature medicine*, *24*(7), 927-930.
221. Yeh, W. H., Chiang, H., Rees, H. A., Edge, A. S., & Liu, D. R. (2018). In vivo base editing of post-mitotic sensory cells. *Nature communications*, *9*(1), 1-10.
222. Wang, Y., Liu, Y., Zheng, P., Sun, J., & Wang, M. (2021). Microbial base editing: a powerful emerging technology for microbial genome engineering. *Trends in Biotechnology*, *39*(2), 165-180.
223. Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., & Gao, C. (2020). Fine-tuning sugar content in strawberry. *Genome biology*, *21*(1), 1-14.
224. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., ... & Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nature biotechnology*, *38*(7), 875-882.
225. Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., ... & Liu, D. R. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature, 583(7817), 631-637.
226. Kang, B. C., Bae, S. J., Lee, S., Lee, J. S., Kim, A., Lee, H., ... & Kim, J. S. (2021). Chloroplast and mitochondrial DNA editing in plants. Nature Plants, 7(7), 899-905.
227. Nakazato, I., Okuno, M., Yamamoto, H., Tamura, Y., Itoh, T., Shikanai, T., ... & Arimura, S. I. (2021). Targeted base editing in the plastid genome of Arabidopsis thaliana. Nature Plants, 7(7), 906-913.
228. Li, R., Char, S. N., Liu, B., Liu, H., Li, X., & Yang, B. (2021). High-efficiency plastome base editing in rice with TAL cytosine deaminase. Molecular Plant, 14(9), 1412-1414.
229. Mok, B. Y., Kotrys, A. V., Raguram, A., Huang, T. P., Mootha, V. K., & Liu, D. R. (2022). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. Nature Biotechnology, 1-10.
230. Anzalone, A. V., Gao, X. D., Podracky, C. J., Nelson, A. T., Koblan, L. W., Raguram, A., ... & Liu, D. R. (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nature Biotechnology*, *40*(5), 731-740.
231. Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., ... & Liu, D. R. (2022). Engineered pegRNAs improve prime editing efficiency. *Nature biotechnology*, *40*(3), 402-410.
232. Klompe, S. E., Vo, P. L., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219-225.
233. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, *365*(6448), 48-53.
234. Chen, S. P., & Wang, H. H. (2019). An engineered Cas-transposon system for programmable and site-directed DNA transpositions. *The CRISPR journal*, *2*(6), 376-394.
235. May, E. W., & Craig, N. L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. *Science*, *272*(5260), 401-404.
236. Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., & Liu, D. R. (2016). A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic acids research*, *44*(20), 9758-9770.
237. Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., ... & Weissman, J. S. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*, *159*(3), 647-661.

238. Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., ... & Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*(2), 442-451.
239. Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Cong, L., ... & Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, *500*(7463), 472-476.
240. Lawhorn, I. E., Ferreira, J. P., & Wang, C. L. (2014). Evaluation of sgRNA target sites for CRISPR-mediated repression of TP53. *PLoS One*, *9*(11), e113232.
241. Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., ... & Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research*, *23*(10), 1163-1171.
242. Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., & Joung, J. K. (2013). CRISPR RNA–guided activation of endogenous human genes. *Nature methods*, *10*(10), 977-979.
243. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S., & Vale, R. D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, *159*(3), 635-646.
244. Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., PR Iyer, E., ... & Church, G. M. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nature methods*, *12*(4), 326-328.
245. Zhou, H., Liu, J., Zhou, C., Gao, N., Rao, Z., Li, H., ... & Yang, H. (2018). In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR–dCas9-activator transgenic mice. *Nature neuroscience*, *21*(3), 440-446.
246. Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., ... & Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, *517*(7536), 583-588.
247. Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., ... & Lim, W. A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, *160*(1-2), 339-350.
248. Wu, Y., Liu, Y., Lv, X., Li, J., Du, G., & Liu, L. (2020). CAMERS-B: CRISPR/Cpf1 assisted multiple-genes editing and regulation system for Bacillus subtilis. *Biotechnology and Bioengineering*, *117*(6), 1817-1825.
249. Zhang, X., Wang, J., Cheng, Q., Zheng, X., Zhao, G., & Wang, J. (2017). Multiplex gene regulation by CRISPR-ddCpf1. *Cell discovery*, *3*(1), 1-9.
250. Zhang, J. L., Peng, Y. Z., Liu, D., Liu, H., Cao, Y. X., Li, B. Z., ... & Yuan, Y. J. (2018). Gene repression via multiplex gRNA strategy in Y. lipolytica. *Microbial cell factories*, *17*(1), 1-13.
251. Ramesh, A., Ong, T., Garcia, J. A., Adams, J., & Wheeldon, I. (2020). Guide RNA engineering enables dual purpose CRISPR-Cpf1 for simultaneous gene editing and gene regulation in Yarrowia lipolytica. *ACS Synthetic Biology*, *9*(4), 967-971.
252. Ming, M., Ren, Q., Pan, C., He, Y., Zhang, Y., Liu, S., ... & Qi, Y. (2020). CRISPR–Cas12b enables efficient plant genome engineering. *Nature plants*, *6*(3), 202-208.
253. Pan, C., Wu, X., Markel, K., Malzahn, A. A., Kundagrami, N., Sretenovic, S., ... & Qi, Y. (2021). CRISPR–Act3. 0 for highly efficient multiplexed gene activation in plants. *Nature Plants*, *7*(7), 942-953.
254. Escobar, M., Li, J., Patel, A., Liu, S., Xu, Q., & Hilton, I. B. (2022). Quantification of Genome Editing and Transcriptional Control Capabilities Reveals Hierarchies among Diverse CRISPR/Cas Systems in Human Cells. *ACS Synthetic Biology*, *11*(10), 3239-3250.
255. Mills, C., Riching, A., Keller, A., Stombaugh, J., Haupt, A., Maksimova, E., ... & Strezoska, Z. (2022). A Novel CRISPR Interference Effector Enabling Functional Gene Characterization with Synthetic Guide RNAs. *The CRISPR Journal*.
256. Xu, X., Chemparathy, A., Zeng, L., Kempton, H. R., Shang, S., Nakamura, M., & Qi, L. S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Molecular Cell*, *81*(20), 4333-4345.
257. Pickar-Oliver, A., Black, J. B., Lewis, M. M., Mutchnick, K. J., Klann, T. S., Gilcrest, K. A., ... & Gersbach, C. A. (2019). Targeted transcriptional modulation with type I CRISPR–Cas systems in human cells. *Nature biotechnology*, *37*(12), 1493-1501.
258. Young, J. K., Gasior, S. L., Jones, S., Wang, L., Navarro, P., Vickroy, B., & Barrangou, R. (2019). The repurposing of type IE CRISPR-Cascade for gene activation in plants. *Communications biology*, *2*(1), 1-7.
259. Stachler, A. E., & Marchfelder, A. (2016). Gene repression in haloarchaea using the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas IB system. *Journal of Biological Chemistry*, *291*(29), 15226-15242.
260. Chang, Y., Su, T., Qi, Q., & Liang, Q. (2016). Easy regulation of metabolic flux in Escherichia coli using an endogenous type IE CRISPR-Cas system. *Microbial cell factories*, *15*(1), 1-9.
261. Tarasava, K., Liu, R., Garst, A., & Gill, R. T. (2018). Combinatorial pathway engineering using type I-E CRISPR interference. *Biotechnology and Bioengineering*, *115*(7), 1878-1883.
262. Chen, Y., Cheng, M., Song, H., & Cao, Y. (2022). Type IF CRISPR-PAIR platform for multi-mode regulation to boost extracellular electron transfer in Shewanella oneidensis. *Iscience*, *25*(6), 104491.
263. Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., & Hsu, P. D. (2018). Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell*, *173*(3), 665-676.

264. Zhao, X., Liu, L., Lang, J., Cheng, K., Wang, Y., Li, X., ... & Nie, G. (2018). A CRISPR-Cas13a system for efficient and specific therapeutic targeting of mutant KRAS for pancreatic cancer treatment. *Cancer letters*, *431*, 171-181.
265. Nelles, D. A., Fang, M. Y., Aigner, S., & Yeo, G. W. (2015). Applications of Cas9 as an RNA-programmed RNA-binding protein. *BioEssays*, *37*(7), 732-739.
266. Engreitz, J., Abudayyeh, O., Gootenberg, J., & Zhang, F. (2019). CRISPR tools for systematic studies of RNA regulation. *Cold Spring Harbor perspectives in biology*, *11*(8), a035386.
267. O'Connell, M. R. (2019). Molecular mechanisms of RNA targeting by Cas13-containing type VI CRISPR–Cas systems. *Journal of molecular biology*, *431*(1), 66-87.
268. Batra, R., Nelles, D. A., Pirie, E., Blue, S. M., Marina, R. J., Wang, H., ... & Yeo, G. W. (2017). Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. *Cell*, *170*(5), 899-912.
269. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L., & Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, *497*(7448), 254-257.
270. Price, A. A., Sampson, T. R., Ratner, H. K., Grakoui, A., & Weiss, D. S. (2015). Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proceedings of the National Academy of Sciences*, *112*(19), 6164-6169.
271. Liu, Y., Chen, Z., He, A., Zhan, Y., Li, J., Liu, L., ... & Huang, W. (2016). Targeting cellular mRNAs translation by CRISPR-Cas9. *Scientific reports*, *6*(1), 1-9.
272. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62-67.
273. O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M., & Doudna, J. A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*, *516*(7530), 263-266.
274. Abudayyeh, O. O., Gootenberg, J. S., Franklin, B., Koob, J., Kellner, M. J., Ladha, A., ... & Zhang, F. (2019). A cytosine deaminase for programmable single-base RNA editing. *Science*, *365*(6451), 382-386.
275. Li, R., Char, S. N., Liu, B., Liu, H., Li, X., & Yang, B. (2021). High-efficiency plastome base editing in rice with TAL cytosine deaminase. *Molecular Plant*, *14*(9), 1412-1414.
276. Cox, D. B., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., & Zhang, F. (2017). RNA editing with CRISPR-Cas13. *Science*, *358*(6366), 1019-1027.
277. Liu, Y., Mao, S., Huang, S., Li, Y., Chen, Y., Di, M., ... & Chi, T. (2020). REPAIR x, a specific yet highly efficient programmable A> I RNA base editor. *The EMBO Journal*, *39*(22), e104748.
278. Gao, Y., Xiong, X., Wong, S., Charles, E. J., Lim, W. A., & Qi, L. S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nature methods*, *13*(12), 1043-1049.
279. Bao, Z., Jain, S., Jaroenpuntaruk, V., & Zhao, H. (2017). Orthogonal genetic regulation in human cells using chemically induced CRISPR/Cas9 activators. *ACS synthetic biology*, *6*(4), 686-693.
280. Chen, T., Gao, D., Zhang, R., Zeng, G., Yan, H., Lim, E., & Liang, F. S. (2017). Chemically controlled epigenome editing through an inducible dCas9 system. *Journal of the American Chemical Society*, *139*(33), 11337-11340.
281. Zetsche, B., Volz, S. E., & Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature biotechnology*, *33*(2), 139-142.
282. Tak, Y. E., Kleinstiver, B. P., Nuñez, J. K., Hsu, J. Y., Horng, J. E., Gong, J., ... & Joung, J. K. (2017). Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. *Nature methods*, *14*(12), 1163-1166.
283. Nihongaki, Y., Furuhata, Y., Otabe, T., Hasegawa, S., Yoshimoto, K., & Sato, M. (2017). CRISPR–Cas9-based photoactivatable transcription systems to induce neuronal differentiation. *Nature methods*, *14*(10), 963-966.
284. Nihongaki, Y., Kawano, F., Nakajima, T., & Sato, M. (2015). Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature biotechnology*, *33*(7), 755-760.
285. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., & Sato, M. (2015). CRISPR-Cas9-based photoactivatable transcription system. *Chemistry & biology*, *22*(2), 169-174.
286. Zetsche, B., Volz, S. E., & Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature biotechnology*, *33*(2), 139-142.
287. Levskaya, A., Weiner, O. D., Lim, W. A., & Voigt, C. A. (2009). Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature*, *461*(7266), 997-1001.
288. Qi, F., Tan, B., Ma, F., Zhu, B., Zhang, L., Liu, X., ... & Cheng, B. (2019). A synthetic light-switchable system based on CRISPR Cas13a regulates the expression of LncRNA MALAT1 and affects the malignant phenotype of bladder cancer cells. *International journal of biological sciences*, *15*(8), 1630.
289. Liu, Y., Zhan, Y., Chen, Z., He, A., Li, J., Wu, H., ... & Cai, Z. (2016). Directing cellular information flow via CRISPR signal conductors. *Nature methods*, *13*(11), 938-944.
290. Patinios, C., Creutzburg, S. C., Arifah, A. Q., Adiego-Pérez, B., Gyimah, E. A., Ingham, C. J., ... & Staals, R. H. (2021). Streamlined CRISPR genome engineering in wild-type bacteria using SIBR-Cas. *Nucleic acids research*, *49*(19), 11392-11404.

291. Zhao, J., Inomata, R., Kato, Y., & Miyagishi, M. (2021). Development of aptamer-based inhibitors for CRISPR/Cas system. *Nucleic acids research*, 49(3), 1330-1344.
292. Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., & Minton, N. P. (2019). RiboCas: a universal CRISPR-based editing tool for Clostridium. *ACS synthetic biology*, *8*(6), 1379-1390.
293. Tang, W., Hu, J. H., & Liu, D. R. (2017). Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nature communications*, *8*(1), 1-8.
294. Siu, K. H., & Chen, W. (2019). Riboregulated toehold-gated gRNA for programmable CRISPR–Cas9 function. *Nature chemical biology*, *15*(3), 217-220.
295. Kundert, K., Lucas, J. E., Watters, K. E., Fellmann, C., Ng, A. H., Heineike, B. M., ... & Kortemme, T. (2019). Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nature communications*, *10*(1), 1-11.
296. Park, S. V., Yang, J. S., Jo, H., Kang, B., Oh, S. S., & Jung, G. Y. (2019). Catalytic RNA, ribozyme, and its applications in synthetic biology. *Biotechnology advances*, *37*(8), 107452.
297. Moreno-Mateos, M. A., Fernandez, J. P., Rouet, R., Vejnar, C. E., Lane, M. A., Mis, E., ... & Giraldez, A. J. (2017). CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature communications*, *8*(1), 1-9.
298. Mougiakos, I., Bosma, E. F., Weenink, K., Vossen, E., Goijvaerts, K., van der Oost, J., & van Kranenburg, R. (2017). Efficient genome editing of a facultative thermophile using mesophilic spCas9. *ACS synthetic biology*, *6*(5), 849-861.
299. Mohanraju, P., Mougiakos, I., Albers, J., Mabuchi, M., Fuchs, R. T., Curcuru, J. L., ... & van der Oost, J. (2021). Development of a Cas12a-based genome editing tool for moderate thermophiles. *The CRISPR journal*, *4*(1), 82-91.
300. Richter, F., Fonfara, I., Bouazza, B., Schumacher, C. H., Bratovič, M., Charpentier, E., & Möglich, A. (2013). Engineering of temperature-and light-switchable Cas9 variants. *Nucleic acids research*, *44*(20), gkw930.
301. Nguyen, D. P., Miyaoka, Y., Gilbert, L. A., Mayerl, S. J., Lee, B. H., Weissman, J. S., ... & Wells, J. A. (2016). Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nature communications*, *7*(1), 1-10.
302. Hill, Z. B., Martinko, A. J., Nguyen, D. P., & Wells, J. A. (2018). Human antibody-based chemically induced dimerizers for cell therapeutic applications. *Nature chemical biology*, *14*(2), 112-117.
303. Liu, Y., Han, J., Chen, Z., Wu, H., Dong, H., & Nie, G. (2017). Engineering cell signaling using tunable CRISPR–Cpf1-based transcription factors. *Nature communications*, *8*(1), 1-8.
304. Nihongaki, Y., Otabe, T., Ueda, Y., & Sato, M. (2019). A split CRISPR–Cpf1 platform for inducible genome editing and gene activation. *Nature chemical biology*, *15*(9), 882-888.
305. Schwarz, K. A., Daringer, N. M., Dolberg, T. B., & Leonard, J. N. (2017). Rewiring human cellular input–output using modular extracellular sensors. *Nature chemical biology*, *13*(2), 202-209.
306. Baeumler, T. A., Ahmed, A. A., & Fulga, T. A. (2017). Engineering synthetic signaling pathways with programmable dCas9-based chimeric receptors. *Cell reports*, *20*(11), 2639-2653.
307. Daringer, N. M., Dudek, R. M., Schwarz, K. A., & Leonard, J. N. (2014). Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS synthetic biology*, *3*(12), 892-902.
308. Barnea, G., Strapps, W., Herrada, G., Berman, Y., Ong, J., Kloss, B., ... & Lee, K. J. (2008). The genetic design of signaling cascades to record receptor activation. *Proceedings of the National Academy of Sciences*, *105*(1), 64-69.
309. Kipniss, N. H., Dingal, P. C., Abbott, T. R., Gao, Y., Wang, H., Dominguez, A. A., ... & Qi, L. S. (2017). Engineering cell sensing and responses using a GPCR-coupled CRISPR-Cas system. *Nature communications*, *8*(1), 1-10.
310. Li, B., Zeng, C., Li, W., Zhang, X., Luo, X., Zhao, W., ... & Dong, Y. (2018). Synthetic oligonucleotides inhibit CRISPR-Cpf1-mediated genome editing. *Cell reports*, *25*(12), 3262-3272.
311. Barkau, C. L., O'Reilly, D., Rohilla, K. J., Damha, M. J., & Gagnon, K. T. (2019). Rationally designed anti-CRISPR nucleic acid inhibitors of CRISPR-Cas9. *nucleic acid therapeutics*, *29*(3), 136-147.
312. Rose, J. C., Popp, N. A., Richardson, C. D., Stephany, J. J., Mathieu, J., Wei, C. T., ... & Fowler, D. M. (2020). Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. *Nature communications*, *11*(1), 1-11.
313. Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., ... & Church, G. (2015). Cas9 gRNA engineering for genome editing, activation and repression. *Nature methods*, *12*(11), 1051-1054.
314. Dahlman, J. E., Abudayyeh, O. O., Joung, J., Gootenberg, J. S., Zhang, F., & Konermann, S. (2015). Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nature biotechnology*, *33*(11), 1159-1161.
315. Swartjes, T., Shang, P., Van Den Berg, D. T., Künne, T., Geijsen, N., Brouns, S. J., ... & Notebaart, R. A. (2022). Modulating CRISPR-Cas genome editing using guide-complementary DNA oligonucleotides. *The CRISPR journal*, *5*(4), 571-585.

316. Li, C., Psatha, N., Gil, S., Wang, H., Papayannopoulou, T., & Lieber, A. (2018). HDAd5/35++ adenovirus vector expressing anti-CRISPR peptides decreases CRISPR/Cas9 toxicity in human hematopoietic stem cells. *Molecular Therapy-Methods & Clinical Development*, *9*, 390-401.
317. Shin, J., Jiang, F., Liu, J. J., Bray, N. L., Rauch, B. J., Baik, S. H., ... & Doudna, J. A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. *Science advances*, *3*(7), e1701620.
318. Lin, P., Qin, S., Pu, Q., Wang, Z., Wu, Q., Gao, P., ... & Wu, M. (2020). CRISPR-Cas13 inhibitors block RNA editing in bacteria and mammalian cells. *Molecular cell*, *78*(5), 850-861.
319. Aschenbrenner, S., Kallenberger, S. M., Hoffmann, M. D., Huck, A., Eils, R., & Niopek, D. (2020). Coupling Cas9 to artificial inhibitory domains enhances CRISPR-Cas9 target specificity. *Science advances*, *6*(6), eaay0187.
320. Liang, M., Sui, T., Liu, Z., Chen, M., Liu, H., Shan, H., ... & Li, Z. (2020). AcrIIA5 suppresses base editors and reduces their off-target effects. *Cells*, *9*(8), 1786.
321. Hoffmann, M. D., Aschenbrenner, S., Grosse, S., Rapti, K., Domenger, C., Fakhiri, J., ... & Niopek, D. (2019). Cell-specific CRISPR–Cas9 activation by microRNA-dependent expression of anti-CRISPR proteins. *Nucleic acids research*, *47*(13), e75-e75.
322. Hirosawa, M., Fujita, Y., & Saito, H. (2019). Cell-type-specific CRISPR activation with microRNA-responsive AcrllA4 switch. *ACS Synthetic Biology*, *8*(7), 1575-1582.
323. Lee, J., Mou, H., Ibraheim, R., Liang, S. Q., Liu, P., Xue, W., & Sontheimer, E. J. (2019). Tissue-restricted genome editing in vivo specified by microRNA-repressible anti-CRISPR proteins. *Rna*, *25*(11), 1421-1431.



Chapter 2

Base-editing: from biology to technology

Despoina Trasanidou[1], John van der Oost[1*], Raymond H. J. Staals[1*]

[1]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

*Corresponding authors

Manuscript in preparation

Abstract

The deaminase superfamily is a group of ubiquitous enzymes that catalyse reactions essential for health and survival. They are categorised mainly in cytidine deaminases (CDs) and adenosine deaminases (ADs), based on their substrate specificity. CDs and ADs have similar phylogenetic origins, and are widespread in both prokaryotes and eukaryotes. Here, we summarise the mechanism of action and the biological role of these enzymes, used for their classification in several functional subgroups. Based on their functionality, we further address the link between dysregulation of these enzymes with mammalian diseases. In addition, we describe the development of CDs and ADs into programmable base-editors, and we provide an overview of the characteristics of all currently available base-editing tools. Moreover, we discuss applications, challenges, and future directions of base-editing technology.

Keywords

Cytidine deaminase, adenosine deaminase, base-editing, classification

Introduction

The 1990-2010s was a dynamic period in deaminase research, during which it was demonstrated that the deaminase superfamily shares a unique fold of α-helices and β-strands (called 'a+b fold' or 'deaminase-like fold') with four other superfamilies[1-5]. Despite the presence of the deaminase-like fold in all superfamilies, only the deaminase superfamily is able to catalyse deamination of nucleobases, due to its characteristic active site residues (C/H)xExnPCxxC (where x is any residue and n is any number of residues) that capture a zinc ion[1,6]. The zinc ion activates a water molecule that forms a tetrahedral intermediate with the carbon atom that is linked to the amine group, leading to deamination of the nucleobase (Fig. 1a)[7]. The substrate binding and catalytic residues of the deaminase superfamily proteins can vary among the different nucleobase substrate specificities[1]. As such, the deaminase superfamily is mainly composed of cytidine deaminases (CDs) and adenosine deaminases (ADs). CDs (EC 3.5.4.5) are enzymes responsible for the deamination of (deoxy)cytidine ((d)C) to (deoxy)uridine ((d)U), whereas ADs (EC 3.5.4.4) convert (deoxy)adenosine ((d)A) to (deoxy)inosine ((d)I)[8,9].

Ancestral members of the deaminase superfamily, found in bacteria and archaea, act solely on free bases, nucleosides, or nucleotides. During evolution, these enzymes acquired more complex biological activities, allowing them to act also on the bases of nucleic acid polymers. Eventually, nucleotide and nucleic acid deaminases evolved in all domains life (Fig. 1b). Specifically, ancestral cytidine deaminases (CDAs) acting on free nucleotides evolved into the tRNA adenosine deaminase A (TadA) early during bacterial evolution, as reflected by their ubiquitous distribution[10,11]. After the evolution of the eukaryotic cell through a fusion of an archaeon and a bacterium, the bacterial *tadA* gene ended up in the ancestral eukaryotic genome. After divergence of the eukaryotic cells, TadA evolved into the essential pan-eukaryotic adenosine deaminase 2 acting on tRNA (ADAT2)[12]. ADAT2 evolved into ADAT3, through gene duplication, and into ADAT1, by acquiring an additional deaminase motif that shifted the site-specificity to another A-base of tRNA[12]. After animal/fungal divergence, ADAT1 acquired a double-stranded RNA (dsRNA) binding domain in metazoans, resulting in adenosine deaminases that predominantly act on pre-mRNA secondary structures (ADARs)[10,12,13]. Inactivating mutations in the Zn-binding site of the ADAR enzymes, resulted in the so called adenosine deaminase domain-containing proteins (ADADs). Notably, given the common ancestry and despite the adenosine deaminase specificity of TadA, ADATs, ADARs and ADADs, the

catalytic domain (active site residues) of these proteins is more similar to CDAs and to cytidine deaminases acting on RNA (CDARs) than to the adenosine deaminases acting on free nucleotides (ADAs)[10,12]. CDARs originate from bacterial toxin cytidine deaminases (BTCDAs), and mainly represent a vertebrate-specific expansion that evolved later and probably independently of the ADATs[14-16]. Specifically, the activation-induced deaminase (AID) and apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) family emerged at the origin of vertebrates[14,15]. First, AID and APOBEC2 (A2) appeared in jawless and cartilaginous fish[15,17]. During tetrapod evolution, duplication of the *aid* gene resulted in APOBEC1 (A1). APOBEC4 (A4) probably emerged in early metazoans, and was later transmitted to photosynthetic eukaryotes through the algal-cnidarian symbiosis, while APOBEC5 (A5) evolved in amphibians[16,18-21]. Last, in placental mammals, specific host adaptation to viruses and retrotransposons has triggered duplications and rearrangements of a primordial AID/A2-like *apobec3* (A3) gene, resulting in an A3 locus diversified among different lineages[15,22-26]. In addition, secreted novel AID/APOBEC-like deaminases (SNADs) seem to have emerged through positive selection in early metazoan evolution, being present in molds and vertebrates[16].

Here, we review the current insights in the classification, the mechanisms, and the biological roles of CDs and ADs. In addition, we describe their exploitation as programmable single-base editors and we summarize all the currently available base-editing tools, discussing their applications and future challenges.

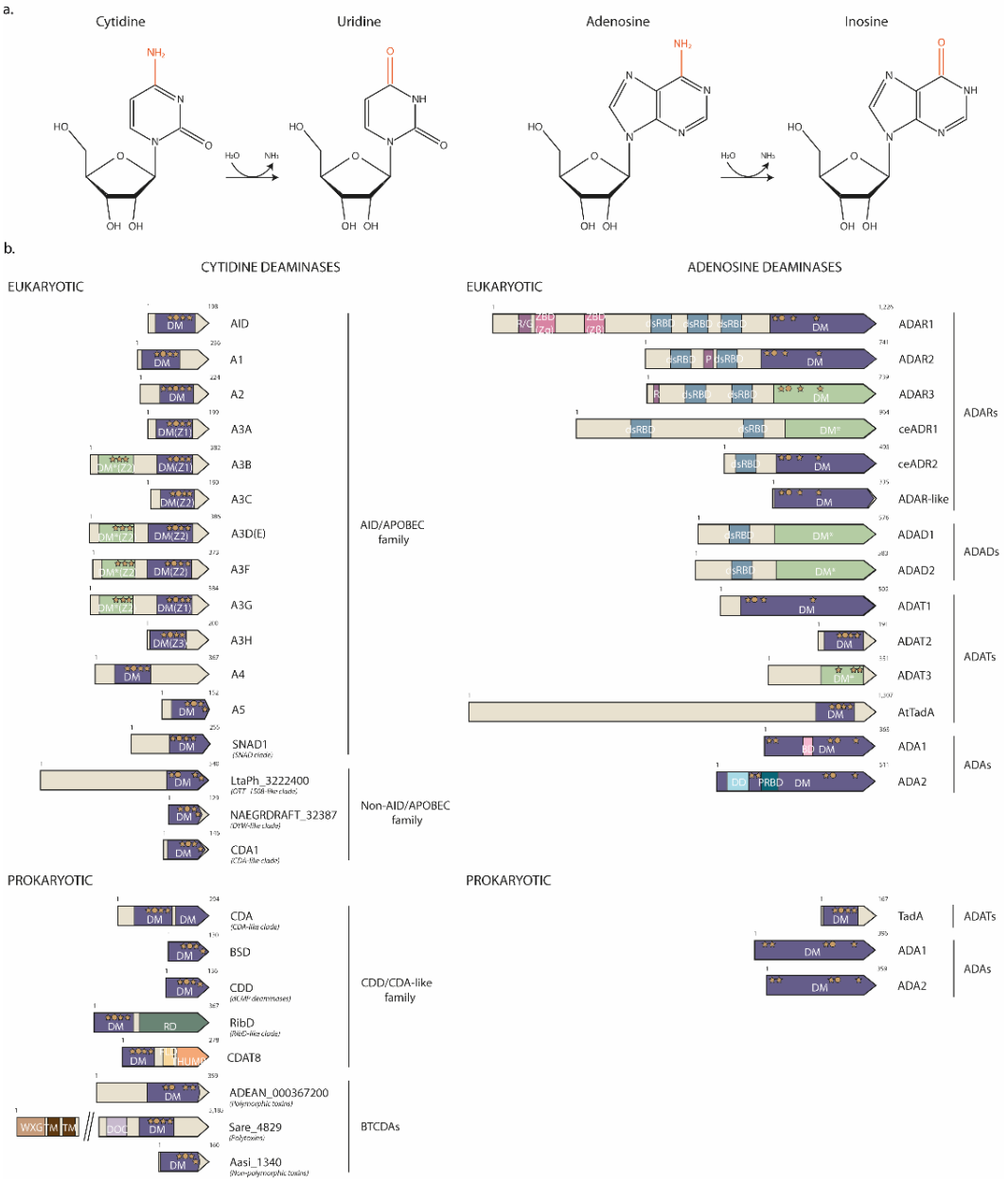


Fig. 1 Enzymatic activity and architecture of CDs and ADs. **a.** Enzymatic activity of CDs and ADs. An activated, zinc-bound water molecule attacks the electrophilic carbon (C6) of the pyrimidine (left) or the purine (right) ring, adding a nucleophilic hydroxyl group and generating a tetrahedral intermediate in the active site of CDs and ADs. The amino group is then protonated and eliminated to form the uracil (left) or inosine (right) product. **b.** Architecture of CDs and ADs. All deaminases contain at least one conserved deaminase domain (DM), either active (dark blue) or inactive (light green). Asterisks and circles represent the zinc-binding and active sites,

respectively. In the dual-domain CDs, the N-terminal DM (light green) assists in nucleic acid binding, viral encapsidation and subcellular localization, while the C-terminal DM (dark blue) largely determines catalytic activity, oligomerization, and substrate specificity. In the single-domain CDs, all these functions are served by a single DM. The DM of A3s is diversified into Z1, Z2, and Z3, based on sequence homology[23]. In prokaryotes, except from DM, RibD additionally contains a reductase domain, while CDAT8 has a ferredoxin-like domain (FLD; yellow) and a THUMP domain (orange) for tRNA recognition[30]. Last, Sare_4829 bears a WXG domain (light brown) for secretion, two transmembrane domains (TM; dark brown), and a DOC domain (light purple) for post-translational modification of other proteins. Interestingly, the catalytically inactive ADAR3 shows an intact ADA domain, with unknown to date reason of its inactivity. In addition, ADARs and ADADs bear at least one double-stranded RNA binding domain (dsRBD; light blue). ADAR1 also carries two Z-DNA binding domains (ZBDs), called Zα and Zβ (pink), for binding to Z-conformation duplexes at transcription sites and probably for interaction with other proteins. Moreover, ADAR1 has an arginine-glycine or arginine-rich domain (R/G; dark purple), ADAR2 has a phosphorus-rich domain (P; dark purple), and ADAR3 has an arginine-rich, ssRNA-binding domain (R; dark purple). Eukaryotic or prokaryotic ADATs and ADAs lack a nucleic acid-binding domain. However, ADATs bind specific sites in tRNAs via certain, unidentified residues in their DM. ADAT3 is catalytically inactive, due a mutation in the active site residue. Eukaryotic ADA1 has a Dipeptidyl peptidase-4 (DPP4)-binding domain (BD; light pink), while ADA2 has a domain for homodimerization and interaction with proteoglycans (DD; very light blue) and a putative receptor-binding (PRB) domain (PRBD; dark green).

Base-editing enzymes: Classification, mechanisms and biological roles

In this section, we describe the classification, the mode of action, and the biological functions of CDs and ADs in both prokaryotes and eukaryotes.

1. Cytidine deaminases (CDs)

Prokaryotic CDs

Prokaryotes express cytidine deaminases that convert cytosine bases, either in the monomeric state (free nucleobases, or as part of nucleosides/nucleotides), or in multimeric nucleotides (ssDNA/ dsDNA, mRNA, tRNA, non-coding RNA). These prokaryotic enzymes are involved in metabolic conversions (CDD/CDA-like deaminases) or act as toxins (BTCDAs) (Fig. 1b, 2a; Table 1).

1.1 CDD/CDA-like family

Bacterial and sporadically archaeal CDAs have been reported to play a pivotal role in the pyrimidine salvage pathway, either maintaining the nucleotide pool for RNA and DNA synthesis or degrading pyrimidines to provide carbon and nitrogen to the cell[27]. However, prokaryotic CDAs are not crucial for cell survival, as they do not participate in the standard pyrimidine biosynthesis pathway. This explains why CDAs are absent in the genomes of some bacteria, including cyanobacteria[28,29].

Several other deaminases with the same active site have also been identified, including the blasticidin S-deaminase (BSD), the deoxycytidylate monophosphate (dCMP) deaminases, and the RibD-like deaminases. BSD, in actinobacteria and firmicutes, generates a modified base comprised in the antibiotic blasticidin S, while the panbacterial (and rarely archaeal) dCMP and RibD-like deaminases participate in the uracil and riboflavin biosynthetic pathway, respectively[1]. Moreover, the homodimeric CDAT8 specifically deaminates the acceptor stem of several archaeal tRNAs (at position 8) to ensure proper folding and functionality. The size and shape of this stem is crucial for CDAT8 binding[30].

1.2 Bacterial toxin cytidine deaminases (BTCDAs)

Bacteria produce toxin deaminases as weapons against different organisms, ranging from closely related bacterial strains to eukaryotic host cells. Depending on the expression organism, bacterial toxin deaminases can be exported through eight different secretory systems[1].

Deaminase genes usually recombine with a neighbouring unrelated toxin gene, forming highly variable (even in inter-strain level) new deaminase toxins, called 'polymorphic toxins'[1]. Toxin genes may code for nucleases, peptidases, and nucleic acid-binding domains for even higher diversity. Sometimes multiple unrelated toxin domains with distinct functions recombine with a deaminase domain, generating 'polytoxins'[31,32]. Polymorphic toxin and polytoxin genes are flanked by genes that code for immune proteins that protect the toxin-production cells from self-intoxication and foreign toxins[33]. Alternatively, polymorphic toxins and polytoxins are involved in self-non self cell discrimination during formation of biofilm or in multicellular colonies. Polymorphic toxins (such as from the BURPS668_1122, YwqJ, MafB19, DYW-like, SCP1.201, XOO_2897, and OTT_1508 clades) are exported via the Type V and VII secretory systems as well as the PrsW-like intramembrane peptidase-dependent mechanism[34-37].

Deaminase toxin genes that are not flanked by immunity genes ('non-polymorphic toxins') mainly belong to the prokaryotic AID/APOBEC-like, OTT_1508-like, and XOO_2897- like clades. Non-polymorphic toxin genes are present in the genomes (frequently in prophage regions) of endosymbiotic or endoparasitic bacteria and attack either their eukaryotic hosts (by targeting their RNAs) or invading mobile genetic elements (MGEs) such as viruses (by hypermutation). Moreover, non-polymorphic toxin genes are usually found in extracellular pathogens of eukaryotes, and in free-living bacteria[1]. They are expressed in various Gram-negative bacteria, and are secreted to their environment as anti-bacterial toxins against other Gram-negative bacteria[1,38,39]. Contrary to all other DNA CDs that predominantly act on ssDNA, DNA CDs from Gram-negative bacteria act solely on dsDNA ('double-stranded DNA deaminase toxin A, DddA'), causing chromosome degradation and DNA replication arrest of their bacterial competitors[40,41]. Notably, these dsDNA mutagenic toxins also trigger cellular aggregation, alterations in colony morphology, and enhanced antibiotic tolerance in toxin-resistant cells, promoting their adaptation and evolution[41]. Non-polymorphic toxins are excreted via the T6SS, the Type II and VI secretory systems as well as the PVC and TcdB/TcaC-like export pathways[1,33,38,39,42-44]. Intriguingly, a more rare secretion system that mimics phage transduction has been detected in *Clostridium*

perfringens, where the deaminase toxin gene is flanked by genes homologous to DNA-packaging system of caudate phages that encapsulate the toxin in a phage capsid for delivery[45].

Eukaryotic CDs

Eukaryotes express CDs that act on free nucleobases/nucleosides/nucleotides, and on nucleotides in DNA or RNA, participating in host defence against foreign invaders as well as in diverse essential biological processes (Fig. 1b, 2a; Table 1). Hence, their dysregulation may lead to diverse diseases (Fig. 3). Eukaryotic CDs have been classified as either AID/APOBEC enzymes or non-AID/APOBEC enzymes.

1.3 Activation-induced deaminase and Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (AID/APOBEC) family

The AID/APOBEC family members edit ssRNA and/or ssDNA (e.g. replication forks) for antibody diversification (AID), lipid metabolism (A1), tissue development (A2), regulation of gene expression (A4) as well as protection against DNA/RNA viruses, (endogenous) retroviruses, retrotransposons, and cancer development (AID, A1, A3s, and A4)[46]. Cells control APOBEC activity through regulation of their (a) expression, (b) protein stability, (c) cellular localization, and (d) oligomerization, as their ability to mutate ssDNA may endanger the genomic DNA[47].

AID

AID is essential for diversification ('somatic hypermutation, SHM'), gene conversion (GCV), and recombination ('class-switch recombination, CSR') of immunoglobulin (*igm*) genes in antigen-activated B cells of vertebrates[48]. AID enhances antibody affinity (via SHM-induced point mutations in *igm* and affinity maturation), and it alters the antibody specificity (via GCV-induced ssDNA and dsDNA breaks in *igm* that trigger homologous recombination) or function (through CSR-induced deletional recombination). In this way, it plays a pivotal role in adaptive immune response against pathogenic bacteria, parasites, viruses, and toxins[49]. After viral infection, AID deamination directly causes hypermutation

Table 1. Characteristics of CD families found in prokaryotes and eukaryotes.

CD family[a]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[b]	Deamination Substrate	Recognition motif[c]	Biological role[e]
PROKARYOTES							
1. Cytidine deaminase domain/Cytidine deaminase-like family (CDD/CDA-like family)							
CDA-like clade	294	Widespread in bacteria (incl. *E. coli*) except cyanobacteria, rarely in archaea	n.a.	C	Free nucleotides	-	Participation in pyrimidine salvage pathway for nucleic acid synthesis and formation of carbon/nitrogen source.
BSD	130	Actinobacteria, firmicutes	n.a.	C	Blasticidin S	-	Resistance to the antibiotic blasticidin S.
dCMP deaminases	136	All bacteria, rarely in archaea	n.a.	C	dCMP	-	Participation in uracil biosynthetic pathway.
RibD-like clade	367	All bacteria, rarely in archaea	n.a.	C	Unknown	-	Participation in riboflavin biosynthetic pathway.
CDAT8	278	Widespread in archaea (incl. *Methanopyrus kandleri*)	n.a.	C	tRNA	Variable	Proper folding and functionality of several tRNAs.
2. Bacterial toxin cytidine deaminases (BTCDAs)							
Polymorphic toxins	359	Proteobacteria, actinobacteria, firmicutes	n.a.	E	ssDNA or ssRNA	-	Defence against foreign nucleic acids. Kin-non kin cell discrimination during formation of biofilm or in multicellular colonies.
Polytoxins	5,185	Widespread in bacteria	n.a.	E	ssDNA or ssRNA	-	Defence against foreign nucleic acids. Kin-non kin cell discrimination during formation of biofilm or in multicellular colonies.
Non-polymorphic toxins	160	Endosymbiotic/ endoparasitic/ free-living bacteria, extracellular pathogens	n.a.	E	mRNA, tRNA, short non-coding RNA, ssDNA	-	Activity against eukaryotic hosts or distantly related environmental competitors.
		Gram$^-$ bacteria (incl. *E. coli*)	n.a.	E	Only dsDNA	-	Activity against other Gram$^-$ bacteria. Induction of chromosome degradation and DNA replication arrest of bacterial competitors.

Table 1. Characteristics of CD families found in prokaryotes and eukaryotes. *(continuation)*

CD family[a]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[b]	Deamination Substrate	Recognition motif[c]	Biological role[e]
EUKARYOTES							
1. Activation-induced deaminase and Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (AID/APOBEC) family							
AID ***[2 isoforms: AID-1/2]***	198	Vertebrates (incl. human)	Activated B-cells	C (active, inactive) >N (active)	mRNA and ssDNA (transcription bubble, stem-loop) of immunoglobulin genes	5'-WRCH-3' (at variable region); 5'-AGCT-3' (at switch region); G_4 sites[b] (strong binding)	Diversification, gene conversion, and recombination of immunoglobulin genes. Enhancement of antibody affinity. Alteration of antibody specificity and function. Host defence against bacteria, parasites, toxins, viruses (HBV), and retrotransposons (L1). Diversification of lymphocyte receptors. Regulation of methylation patterns. Reprogramming of induced pluripotent stem cells.
APOBEC1 (A1)	236	All mammals (incl. human)	Small intestine, liver (in certain species)	C>N	(Non)-coding mRNA; ssDNA (growth/ differentiation genes); L1 retrotransposons	5'-ACW$_{4\text{-}6}$UGAUW$_2$GW$_4$-3'[b]	Truncates *apo-B100* mRNA (to reduce LDL and VLDL levels in blood), *nf1* mRNA (to inhibit tumour suppression), *nat1* mRNA (to inhibit translational repression), and eif4g2 mRNA (to prevent translation initiation). mRNA instability (induction of cancer development). Anti-viral activity (HSV-1, hepatitis, MLV, HIV-1).
APOBEC2 (A2)	224	Jawed vertebrates (incl. human)	Skeletal/cardiac muscles. Traces in liver, lung, GI tract, spleen, kidney	C=N	mRNA	-	Muscle function/ differentiation/ development/ post-injury regeneration. Demethylation of myogenic and myoblast differentiation genes.
		Zebrafishes, frogs	Proliferating/ Dedifferentiated Müller glia	-	-	-	Regulation of retina regeneration. Optic nerve regeneration. Control of left-right axis specification.
		Goats, chickens	-	-	-	-	Lipid metabolism.
APOBEC3A (A3A) ***[2 isoforms: L and S]***	199	Most mammals (incl. human)	Monocytes, macrophages, dendritic cells, keratinocytes, hepatocytes	N>C	ssDNA (replication forks, ssDNA gaps, replication fork mimics, transcriptional bubble structures, and linear ssDNA, mDNA). SINE/ LINE. (Non)-coding mRNA.	5'-VC-3' or 5'-TCW-3' sites (especially 5'-YTCA-3') in short (4-8 nt) ssDNA hairpin loops. 5'-WTC-3' sites in RNA stem-loops	Promotion/ Suppression of cancer development, diversification, and resistance to therapies. Anti-viral (SeV, SARS-CoV-2, HCoV-NL63, HTLV-1, HCV, RSV, PERV, EBV, KSHV, HPV, HSV-1, HBV, CMV, HIV-$1_{\Delta Vif}$, SIVmac$_{\Delta Vpx}$, MLVΔgGag, AAV-2, parvovirus, IFV, MV) activity. Anti-retrotransposon (LTR, non-LTR) activity. Participation in mDNA catabolism. Demethylation of cancer-related genes. High impact on transcriptome and pro-inflammatory/metabolic responses.

Table 1. Characteristics of CD families found in prokaryotes and eukaryotes. *(continuation)*

CD family[a]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[b]	Deamination Substrate	Recognition motif[c]	Biological role[e]
APOBEC3B (A3B) ***[3 isoforms: A3B-1/2/3]***	382	Human, hominids	Leukocytes, spleen, testes, heart, thymus, prostate, ovaries.	N	ssDNA (replication, transcription, DSB-repair, stalled/ collapsed replication forks); RNA	5′-TCW-3′ sites (especially 5′-RTCA-3′) in ssDNA and RNA	Anti-viral (HPV, SIV, SIV_{cpz}, SIV_{gor}, HBV, HCV, HPV, adenovirus A12/B3/C2, JC, MCV, BK, HTLV-1, and parvovirus) activity. Demethylation of genes. Promotion of tumour diversity and cancer progression.
APOBEC3C (A3C)	190	Human, hominids	Lymphoid and liver cells	C>N	ssDNA (viral cDNA or host genome)	5′-TC-3′ (especially 5′-TTC-3′, 5′-TCA-3′) Strong RNA binding	Weak anti-viral (SIV, SIV_{agm}, SIV_{mac}, SIV_{cpz}, SIV_{gor}, MLV, HSV-1, HPV, HTLV-1, $FV_{\Delta bet}$, HBV, ZIKV, HIV-$1_{\Delta Vif}$, HCoV-NL63) activity. Weak anti-retrotransposon (LINE, SINE) activity. Promotion/ Suppression of cancer development and invasiveness. Cell cycle progression and cell proliferation.
APOBEC3DE (A3DE, A3D) ***[4 Isoforms: A3Dv1/2/6/7]***	386	Human, hominids	T/B/dendritic cells, macrophages, hepatocytes	C (A3Dv1/7) N (A3Dv2/6)	ssDNA (viral cDNA or host genome)	5′-TC-3′; 5′-CC-3′ (less preferred); RNA binding	Weak anti-viral (SIV_{smm}, HBV, HCoV-NL63, HIV-$1_{\Delta Vif}$) activity. Weak anti-retrotransposon (LINE, SINE) activity. Promotion/ Suppression of cancer development and progression.
APOBEC3F (A3F) ***[3 Isoforms: A3F-1/2/3]***	373	Human, hominids	T/B/dendritic cells, macrophages, hepatocytes	C	ssDNA (viral cDNA or host genome)	5′-TC-3′ (especially 5′-TTCR-3′) sites and long (30 – 85 nt) ssDNA stretches; RNA binding	Moderate anti-viral (HIV-$1_{\Delta Vif}$, HIV-$2_{\Delta Vif}$, SIV, SIV_{smm}, SIV_{cpz}, SIV_{gor}, HBV, HCoV-NL63, HTLV-1, RV) activity. Anti-retrotransposon (LINE, SINE) activity. Anti-ERV and anti-ERV-1 element activities. Promotion/ Suppression of cancer development and progression.
APOBEC3G (A3G) ***[2 Isoforms: A3F-1/3]***	384	Human, hominids	T/B/dendritic cells, macrophages, hepatocytes	C	ssDNA (viral cDNA or host genome); ssRNA (host mRNA)	5′-CC-3′ (preferably 5′-(C)CC(A)-3′) and 5′-TC-3′ (less preferred) sites on ssDNA; 5′-CCG-3′ sites flanked by inverted repeats in RNA stem-loop structures	Strong anti-viral (HIV-$1_{\Delta Vif}$, HIV-$2_{\Delta Vif}$, SIV, SIV_{smm}, SIV_{cpz}, SIV_{gor}, HBV, HCV, HTLV-1, MLV, HPV, MuV, MV, RSV, MOV10, HSV-1, EV71, PERV, MMTV) activity. Anti-retrotransposon (LINE, SINE) activity. Anti-ERV activity. Promotion/Suppression of cancer progression and metastasis. Cell stress response and promotion of cell proliferation under hypoxia and cellular crowding conditions.
APOBEC3H (A3H) ***[7 Isoforms: I-VII]***	200	Human, hominids	Mainly T and liver cells	C>N (II, V, VII); N>C (I)	ssDNA (viral cDNA or host genome)	5′-TC(W)-3′ (cDNA or overhangs of RNA/ DNA hybrids); Binding to dsRNA, ssRNA, and RNA/DNA heteroduplexes (short A-form RNA duplexes)	Moderate anti-viral (HIV-$1_{\Delta Vif}$, HIV-$2_{\Delta Vif}$, SIV_{smm}, HCoV-NL63, HBV, SIV_{cpz}, SIV_{gor}, HTLV-1) activity. Anti-retrotransposon (L1) activity. Demethylation of genes. Promotion/ Suppression of cancer progression and metastasis.
APOBEC4 (A4)	367	Jawed vertebrates (incl. human), reptiles, fish	Testicles, bronchiolar/pulmonary/tracheal/nasal epithelial cells	C/N	-	-	Participation in spermatogenesis. Pro-viral (HIV-1) activity.

Table 1. Characteristics of CD families found in prokaryotes and eukaryotes. *(continuation)*

CD family[a]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[b]	Deamination Substrate	Recognition motif[c]	Biological role[e]
APOBEC4 (A4)	367	Jawed vertebrates (incl. human), reptiles, fish, amphibians	Testicles, bronchiolar/pulmonary/tracheal/nasal epithelial cells	C/N	-	-	Participation in spermatogenesis. Pro-viral (HIV-1) activity.
		Chicken	Testicles, thymus, kidney, duodenum (mainly bursa of Fabricius, B/T cells, macrophages)	-	-	-	Anti-viral (NDV, IBV) activity.
APOBEC5 (A5)	152	Non-mammalian tetrapods	-	-	-	-	-
SNAD clade	255	Poikilothermic and low body temperature vertebrates, sponges	-	-	-	-	Secreted probably for activity against pathogens, parasites, and virus-infected cells.
2. Non-AID/APOBEC proteins							
OTT_1508-like clade	540	Protozoa, metazoans, parasites, fungi, kinetoplastids, lycophytes.	-	-	mRNA; Short non-coding RNA	-	Anti-viral and anti-parasitic activity. Editing of transcripts from selfish or repetitive elements and highly variable effector genes. Inhibition of heterokaryon formation in filamentous fungi. Protein recoding.
		Protozoan parasites	-	-	Anticodon $tRNA^{Trp}$	5′-CCA-3′	-
DYW-like clade	129	Plants, amoeba, mushrooms, rotifers, fungi, ascomycete oomycetes	-	-	Mitochondrial and chloroplast mRNA	-	Restoration of terminated ORFs and missense codons.
CDA-like clade	146	Ubiquitous	Liver, white blood cells	C>N	Cytidine	-	Participation in pyrimidine salvage pathway for nucleic acid synthesis and formation of carbon/nitrogen source.
		Jawless vertebrates	-	-	-	-	Role in somatic diversification of antibody genes during B lymphocyte development.

Table 1. Characteristics of CD families found in prokaryotes and eukaryotes. *(continuation)*

CD family	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[b]	Deamination Substrate	Recognition motif[c]	Biological role[e]
		Chlorophyte algae, kinetoplastids, stramenopiles, alveolates	-	-	Probably ssDNA and ssRNA	-	-
3. Unidentified fungal cytidine deaminases acting on coding mRNA (UCDARs)							
-	-	Basidiomycetes	-	-	Coding mRNA	-	Adaptation at different substrate availabilities.
		Trypanosoma	-	C	tRNA	Variable	Possible role in tRNA folding and functionality.
		High eukaryotes	-	O	tRNA	Variable	Possible role in tRNA folding and functionality.
		Plants	-	M	tRNA	Variable	Possible role in tRNA folding and functionality.

[a]In eukaryotes, ' isoforms' and ' size' refer to the human homologue (when applicable), except from OTT_1508-like clade (homologue from *Leishmania tarentolae*), and DYW-like clade (homologue from *Naegleria gruberi*). In prokaryotes, they refer to the homologue of the representative organism of the group (*E. coli* for CDA-like clade and RibD-like clade; *Aspergillus terreus* for BSD; *Bacillus subtilis* for dCMP deaminases; *Methanopyrus kandleri* for CDAT8; *Angomonas deanei* for polymorphic toxins; *Salinispora arenicola* for polytoxins; *Amoebophilus asiaticus* for non-polymorphic toxins). ' Size' also refers to the unedited isoform, unless stated otherwise.

[b]Synonyms of AID = AICDA; CDAT8 = MK0935.

[c]C = cytoplasm; N = nucleus; C/N = cytoplasm or nucleus; C>N = mainly cytoplasm; N>C = mainly nucleus; C=N = equal distribution in cytoplasm and nucleus; E = extracellular environment; O = organelles; M = mitochondria

[d]W = A or T; R = A or G; H = A, C or T; V = A, C or G

[e]HBV = Hepatitis B Virus; HSV-1 = Herpes Simplex Virus 1; MLV = Murine Leukemia Virus; HIV-1 = Human Immunodeficiency Virus 1; SeV = Sendai Virus; SARS-CoV-2 = Severe Acute Respiratory Syndrome Coronavirus 2; HCoV-NL63 = ; HTLV-1 = Human T-cell Leukaemia Virus type 1; HCV = Hepatitis C Virus; RSV = Respiratory Syncytial Virus; PERV = Porcine Endogenous Retrovirus; EBV = Epstein-Barr Virus; KSHV = Kaposi Sarcoma-associated Herpes Virus; HPV = Human papillomavirus; CMV = Cytomegalovirus; SIV = Simian Immunodeficiency Virus; SIVmacΔVpx = SIV from rhesus macaques; MLVΔgGag = Murine Leukemia Virus; AAV-2 = Adeno-associated Virus 2; IFV = Influenza Virus; MV = Measles Virus; SIVagm = SIV from African Green Monkeys; SIVcpz = SIV from chimpanzees; SIVgor = SIV from gorillas; FVΔbet = Foamy Virus; ZIKV = Zika Virus; SIVsmm = SIV from sooty mangabeys; HIV-2ΔVif = Human Immunodeficiency Virus type 2; RV = Rubella Virus; MuV = Mumps Virus; MOV10 = Moloney Leukemia Virus 10; EV71 = Enterovirus 71; MMTV = Mouse Mammary Tumor-like Virus; NDV = Newcastle Disease Virus; IBV = Infectious Bronchitis Virus.

All references used for this table are mentioned in the main text

of the viral RNA genome and it recruits RNA exosomes into viral ribonucleoprotein complexes for degradation, in a transcription-coupled manner[50,51]. In addition, AID is able to restrict L1 retrotransposition[52]. In fish, AID is involved in diversification of lymphocyte receptors[53,54].

APOBEC1 (A1)

A1 is expressed in the small intestine of all mammals and the liver of certain species (e.g. rodents), participating in lipid metabolism. It specifically edits the *apo-B100* mRNA, generating a stop codon that results in the formation of a truncated (48 amino acids) protein, called Apo-B48, as a means to decrease levels of lipids (low-density lipoprotein, LDL; very-low-density lipoprotein, VLDL)[55]. Alternatively, A1 may trigger tumorigenesis, for instance through induction of C-to-U hypermutations in the 3'-UTR of cancer-related mRNAs that affect their stability (Fig. 3a)[56-62]. Tumorigenesis is also provoked by A1-mediated ssDNA hypermutations in growth and differentiation genes[55,63]. Moreover, human A1-mediated editing has been observed in viral DNA genomes after infection of human cells, as a means to restrict the infection, while rodent A1 has been reported to edit DNA and RNA of viruses and transposons[52,64].

APOBEC2 (A2)

A2 is predominantly expressed in muscles of jawed vertebrates for normal muscle function, differentiation, development and post-injury regeneration[65-70]. In zebrafish and frogs, A2 controls left-right axis specification[70], while in goats it may play a role in lipid metabolism, like its close homologue A1[71]. A2 is also related to chicken body weight with unknown to date physiology[72]. Contrary to other AID/APOBEC family members, the substrate and enzymatic activity of A2 still remain elusive.

APOBEC3s (A3s)

A3A is widely expressed in most mammals, while in humans it is mainly found in hematopoietic cells (monocytes, macrophages, and dendritic cells)[73-78]. A3A hypermutates the ssDNA lagging strand template of replication forks, inducing C-to-U

deamination which is generally followed by U-to-T conversion via DNA mismatch repair and DNA replication. This activity may sometimes lead to cancer progression or diversification and resistance to drug therapies due to the formation of localised hypermutations that are associated with some cancer types (kataegis)[79]. However, too high A3 expression may suppress cancer progression through excess hypermutations that lead to the generation of double-stranded DNA breaks (DSDBs), DNA degradation, G1 cell cycle arrest, production of tumorigenic reactive oxygen species, and thus cell apoptosis[80-83]. Repeating cycles of stress-dependent induction and suppression of A3A expression provide cancer cells a sophisticated way to protect themselves from excessive A3A-mediated genomic instability and genotoxicity[84]. In addition, A3A is able to edit intracytoplasmic mitochondrial DNA (mDNA), nascent complementary DNAs from short interspersed repetitive elements (SINE-1; S1) or L1, and viral ssDNA[73,85,86]. Specifically, A3A is crucial in mDNA catabolism, as it obviates dangerous signal accumulation and inflammation. However, upregulation of A3A can cause mitochondrial membrane depolarization and apoptosis[81,87]. Moreover, A3A inhibits both long terminal repeat (LTR)-retrotransposons (in a deaminase-independent manner) and non-LTR retrotransposons (in a deaminase-dependent manner), diminishing retrotransposition frequencies of L1 and *Alu* elements[88]. High A3A levels are also observed upon infection by (cancer-related) RNA or DNA viruses, due to replication stress and/or interferon response[73,76,84,88-98]. A3A blocks viral integration through host or viral genome hypermutation but some viruses encode ribonucleotide reductase (RNR) large subunits that bind, inhibit, and relocalize A3A[89,99]. Generally, the deamination activity of A3A is more efficient in short (4-8 nt) ssDNA hairpin loops (usually found in 'passenger' genes) than in hairpins with longer loops, ssDNA gaps, replication fork mimics, (transcriptional) bubble structures, and completely linear ssDNA, which are usually occupied by RNA polymerase and RPA[100]. Moreover, high stem strength structures are non-editable by RPA and thus preferred by A3A[101]. Except from cytidine, A3A is also able to deaminate 5-mC and 5-oxidized deoxycytidine derivatives, which are cancer mutation hotspots[80,83,102]. A3A additionally mediates C-to-U editing in the stem-loop of certain (non)-coding mRNAs of monocytes and macrophages, in a sequence-specific manner[103-105]. The stability of the stem and the size of the loop are also crucial for the RNA editing efficiency, as tri- or tetra-loops are highly preferred[104]. A3A-mediated RNA editing is highly triggered by interferon release, hypoxia and cellular crowding, having a high impact on the transcriptome as well as on metabolic, pro-inflammatory, and anti-viral responses in macrophages[105-107]. Altogether, A3A shows the strongest, cancer-related deaminase activity compared to all the rest APOBECs, as well as unique RNA editing ability.

A3B is expressed in most tissues, and is the second strongest cancer genome mutator[79,108,109]. A3B is less catalytically active than A3A but it is expressed in higher levels in tumours, enhancing tumour diversity and cancer progression (Table 3a)[46,79,80,84,103,109-119]. Specifically, except from RNA, A3B also deaminates ssDNA in the lagging strand (during replication), the non-template strand (during transcription), the stretches exposed during the resection stage of DSB-repair, and the stalled/collapsed replication forks[46,84,103,109,120]. In this way, A3B induces genotoxic stress, promoting or inhibiting cancer progression[46,79,103,121]. Moreover, similar to A3A, A3B is able to deaminate the cancer mutation hotspot 5-MeC, though with much lower efficiency[78,111,122]. Furthermore, A3B induces expression of chemokines and cytokines, leading to tumour cell survival and immune evasion[123]. Upon viral infection, A3B expression is upregulated in order to mutagenise the viral genome or complementary DNA (cDNA) intermediates and thus inhibit the infection (Table 3a)[109,124-127]. However, some viruses escape A3B-immunity by either modifying the A3B viral target sites or relocalising A3B from the nucleus to perinuclear and cytoplasmic aggregates[89,112,127-131]. Except from its anti-viral activity, A3B also inhibits retrotransposition (L1; S1)[132].

A3C is expressed predominantly in lymphoid and liver cells[108,125,133-135]. A3C has weak activity against RNA/DNA viruses and endogenous retroelements as well as for host genome mutagenesis, thus requiring RNA-binding, homodimerization, or heterooligomerization with other A3s to rich high activities[50,94,109,125,128,129,131,132,134,136-138]. A3C edits solely ssDNA, and usually acts in concert with ADAR1 for missense editing and transcript instability of the cell cycle control kinase CDK13[134,139]. Together, these deaminases also edit STAT3 to prevent phosphorylation and β-catenin degradation[140]. A3C is overexpressed in several cancers, either supporting invasiveness or providing a strong anti-cancer immune response (Table 3a)[133,141-144]. Specifically, A3C exhibits strong RNA binding, and interacts with nucleolar, DNA repair, and RNA-binding proteins, supporting cell proliferation[145]. However, if A3C reaches too high levels, it leads to cell cycle arrest, DNA damage and growth inhibition of cancer cells[138]. High levels of A3C are also observed in HBV-related cirrhosis[50].

A3D, A3F, A3G, and A3H compose an anti-retroviral armada of APOBECs that synergistically hypermutate viral genomes, and secondarily show anti-retrotransposition activity and promoter or block cancer development[50,94,108,109,120,125,126,132,137,141,144,146-153]. All proteins are expressed in T/B/dendritic cells, macrophages, and hepatocytes[108,137,151], exhibiting both ssDNA-editing and RNA-binding abilities[151]. Specifically, they convert C-to-U in viral cDNA during reverse transcription, which results in either cDNA degradation

(through UDG-mediated excision of Us and endonuclease-mediated cleavage of abasic-sites) or non-synonymous mutations (missense and, to a lesser extent, nonsense substitutions) in the viral genome, blocking replication and transmission[94,154-157]. In addition, A3F, A3G and A3H strongly bind viral RNA to delay initiation of primer extension, diminish processivity, sterically hinder RT and DNA elongation, and promote lethal RT template switching[151,154-158]. The ability of A3F to create a roadblock for the polymerase derives from its ability to form large and stable oligomers that tightly bind nucleic acids, especially compared to A3G that forms more unstable and concentration-dependent oligomers. Although this deaminase-independent mechanism shows weaker anti-viral effect compared to hypermutation, the delay in polymerization offers more time for viral DNA deamination, providing synergistic anti-viral activity[154,159]. A3F exhibits higher ssDNA affinity than A3G, and thus it inhibits RT more efficiently. Therefore, A3F heterooligomerizes with A3G to improve its ability to slow down RT as well as to jump over RNA/DNA hybrids for higher processivity[159]. An alternative deaminase-independent mechanisms of viral restriction, A3F and A3H bind viral nucleoproteins to hinder replication[94]. Also, A3G binds directly to the RT to hamper RT template switching, proviral DNA synthesis, and viral replication. Moreover, A3G binds to viral RNA helicases to decrease their activity or upregulates the expression of the cellular metabolism modulators hampering protein translation[64,160]. A3F shows preference for long (30 – 85 nt) ssDNA stretches, as it lands, scans and edits distantly spaced targets in ssDNA, through jumping or intersegmental transfer movements[161]. This limitation renders it less mutagenic than A3G and A3H, which additionally have short-range sliding ability to edit closely spaced targets (up to 20 nt)[135]. However, when ssDNA is present in much higher concentrations than the enzyme, A3F preferably deaminates short ssDNA[162]. A3G and A3H additionally edit host ssRNA, preferably in stem-loop structures[160]. Although A3G has similar affinity for both substrates, it edits multiple target sites on ssDNA but often a single target site on ssRNA. The editing efficiency is positively correlated to the stem stability and negatively correlated to the loop size (maximum efficiency with up to four loops). mRNA editing may lead to protein diversity or affect stability, turnover, translatability or efficiency of subsequent editing at a neighbouring target site[104]. A3G introduces non-synonymous mutations in the mRNA of many ribosomal and translational genes in order to rearrange the cellular translation apparatus. Moreover, A3G-mediated RNA editing promotes glycolysis over mitochondrial respiration, enhancing cell proliferation[150].

APOBEC4 (A4) and APOBEC5 (A5)

A4 is detected in jawed vertebrates, reptiles, amphibians and some fish, while A4 of chicken (chA4) is constitutively expressed mainly in hematopoietic stem cells[51,52,111,163]. In mammals, A4 is restricted to the testicles, probably participating in spermatogenesis[51]. Human A4 usually forms dimers and interacts weakly with ssDNA, showing undetectable deaminase activity. However, it enhances viral replication by regulating both host and endogenous LTR promoters, in a deaminase-independent manner[52]. Except from multiple myeloma, it is also overexpressed in epithelial cells of the respiratory system upon SARS-CoV-2 infection, with an elusive biological role[51,52,164-166]. Similarly, chA4 is overexpressed in immune cells (B/T cells, macrophages) and infected organs (testicles, thymus, kidney, duodenum) upon viral infection, decreasing viral RNA accumulation[51]. Generally, the biological role of A4 still remains largely untapped. A5 is found in non-mammalian tetrapods with a currently unknown biological role[167].

Secreted Novel AID/APOBEC-like Deaminases (SNADs)

SNADs are secreted by metazoa, and dictyostelium-like slime molds. Metazoan SNADs are categorised into four subclades, SNADs1-4. SNADs1-3 are found in poikilothermic and low body temperature vertebrates, while SNAD4 is found in sponges. A role in defence against viruses and/or other pathogens has been proposed, but this remains to be experimentally verified[16].

1.4 Non-AID/APOBEC proteins

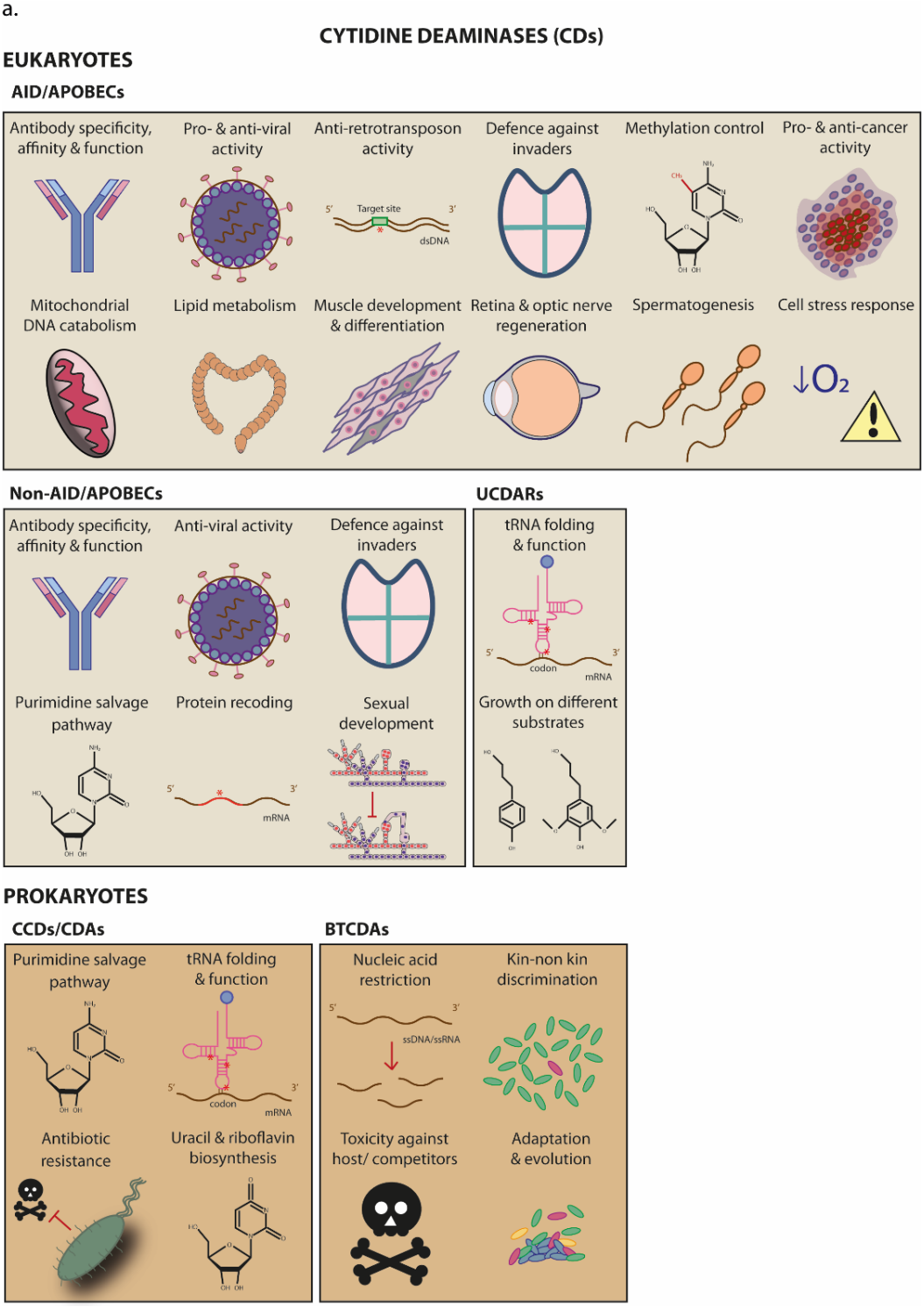
Apart from some members of the well-studied AID/APOBEC family, also the OTT_1508-like, the DYW-like, and the CDA-like clades are crucial for protection of the eukaryotic host from viruses and other parasites. In addition, they play a pivotal role in cytidine deamination either at the level of monomeric substrates (e.g. pyrimidine salvage pathway) or at the level of nucleic acids (e.g. modifying DNA and/or RNA for protein recoding and antibody diversification, repair of mitochondrial and chloroplast mRNA). The eukaryotic deaminases of the OTT_1508-like clade show a wide phylogenetic distribution, ranging from fungi and plants to protozoan and metazoan animals[1]. They protect the host's genomic integrity through hypermutation of transcripts from selfish or repetitive

elements (to block heterokaryon formation), endogenous defensive molecules (for broader recognition of invading parasite moieties), and mRNAs or short non-coding RNAs (to provide advantageous functionalities)[1,168,169]. Remarkably, the deaminase from a protozoan parasite edits the anticodon of $tRNA^{Trp}$ ([5'-CCA]-to-[5'-UCA]) which has been correlated with alternative genetic codes[170]. In land plants, CDs of the DYW-like clade are guided by RNA-binding pentatricopeptide repeat (PPR) proteins to edit mitochondrial and chloroplast mRNA in order to restore terminated ORFs or missense codons[1,171]. This is of high importance, as genetic material in these organelles generally accumulate deleterious mutations, due to the low recombination events and the lack of repair mechanisms[172]. DYW deaminases are also present in amoeba, rotifers, and fungi[173]. Human CDA1, mainly expressed in liver and white blood cells, participates in the pyrimidine salvage pathway to recycle free pyrimidines, either for synthesis of nucleic acids or to supply carbon and nitrogen for biosynthesis purposis[174,175]. CDA2 induces somatic diversification of antibody genes during B lymphocyte development in jawless vertebrates[176]. Furthermore, other deaminases that pertain to the CDD/CDA-like clade (for example *Leishmania* LmjF36.5940) show potential ssRNA and ssDNA editing ability in algae, kinetoplastids, stramenopiles and alveolates[177]. Dysregulation of these enzymes is linked to inflammation and diverse cancers in higher eukaryotes (Table 3a; Fig. 3a)[178-181].

1.5 Unidentified fungal cytidine deaminases acting on coding mRNA (UCDARs)

In basidiomycetes, UCDARs act on coding mRNAs of wood-degrading enzymes to adapt at different substrate availabilities. In kinetoplastids, eukaryotic organelles, and plant mitochondria, UCDARs edit tRNAs probably for optimal folding. However, many details of these enzymes still remain unidentified[182].

a.


CYTIDINE DEAMINASES (CDs)
EUKARYOTES
AID/APOBECs
Antibody specificity, affinity & function
Pro- & anti-viral activity
Anti-retrotransposon activity
Target site
dsDNA
Defence against invaders
Methylation control
Pro- & anti-cancer activity
Mitochondrial DNA catabolism
Lipid metabolism
Muscle development & differentiation
Retina & optic nerve regeneration
Spermatogenesis
Cell stress response
↓O2
Non-AID/APOBECs
Antibody specificity, affinity & function
Anti-viral activity
Defence against invaders
Purimidine salvage pathway
Protein recoding
mRNA
Sexual development
UCDARs
tRNA folding & function
codon
mRNA
Growth on different substrates
PROKARYOTES
CCDs/CDAs
Purimidine salvage pathway
tRNA folding & function
codon
mRNA
Antibiotic resistance
Uracil & riboflavin biosynthesis
BTCDAs
Nucleic acid restriction
ssDNA/ssRNA
Kin-non kin discrimination
Toxicity against host/ competitors
Adaptation & evolution

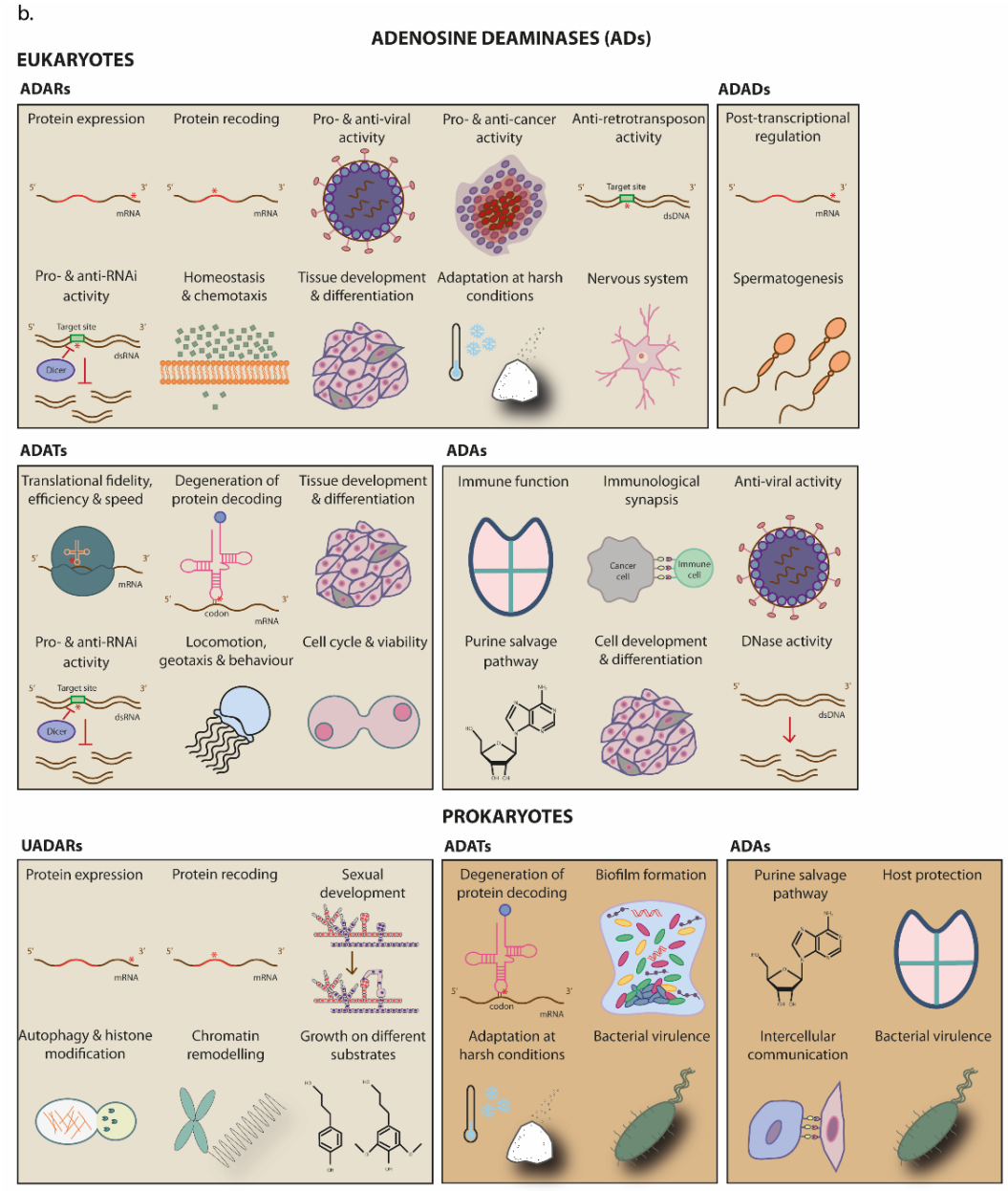


Fig. 2 Biological role of CDs and ADs. Mechanisms of action of (**a**) CDs and (**b**) ADs in eukaryotic and prokaryotic organisms. Boxes represent different subgroups of each deaminase category.

2. Adenosine deaminases (ADs)

Prokaryotic ADs

Prokaryotes express adenosine deaminases that act either on free nucleobases/ nucleosides/ nucleotides (ADAs) or on mRNA/ tRNA (TadA) (Fig. 1b), and they are essential for survival (Fig. 1b, 2b; Table 2).

2.1 Adenosine deaminases acting on nucleosides (ADAs)

Prokaryotic ADAs are widespread in bacteria and archaea, and they exist mainly in two isoforms, the intracellular ADA1 and the secreted ADA2. ADA1 balances the nucleotide pool in the cell through the purine salvage pathway. Some bacterial symbionts complement the lack of ADAs in their eukaryotic host[183,184]. Moreover, ADA1-variants of endophytic bacteria decrease the levels of pterin, an adenosine analogue and a biochemical compound that activates an immune response pathway[185]. In contrast, ADA2 participates in a 'purinergic signalling' pathway that mediates communication between prokaryotic cells (for example to stimulate bacterial proliferation), and protects bacterial cells or their eukaryotic hosts from immunosuppressive (methyl)adenosine overproduced by pathogenic bacteria[183,186-191]. Numerous other bacterial ADAs from *Streptomyces* species (such as SanADA1-4, ScoADA, and SauADA1-5) convert the intermediate adenosine in the biosynthesis pathway of the antibiotic and broad-spectrum antiviral arabinofuranosyladenine (Ara-A) compound to produce arabinofuranosylhypoxanthine (Ara-I), which has significantly lower anti-viral bioactivity[186,187,192,193].

2.2 tRNA adenosine deaminase A (TadA)

TadA is the most widespread adenosine deaminase in bacteria and essential for cell survival under harsh conditions (antibiotic presence, starvation, oxidative stress)[194-196]. TadA is the first identified RNA editing enzyme that alters not only tRNAs but also mRNAs[194,197]. With respect to tRNA editing, it acts as an homodimer on the adenine at the first anticodon position (A_{34}) of $tRNA^{Arg}_{ACG}$ and rarely of $tRNA^{Leu}_{AAG}$[6,7,197-200]. The conversion of A-to-I allows a single tRNA species to translate multiple synonymous codons with U, A or C in the third position. Recognition of the tRNA anticodon stem–loop is sufficient and required for efficient site-specific deamination[197]. In addition, TadA expressed from

pathogenic strains of *E. coli*, *Klebsiella pneumonia* and *Yersinia enterocolitica* edits coding mRNAs of self-killing toxin genes from the evolutionarily conserved Host-Killing (Hok) toxin-antitoxin family. Indicatively, TadA edits *hokB* toxin mRNA either in the presence of antibiotics or upon cell density increase in order to inhibit cell growth, as a survival response to the antibiotic activity or to starvation, respectively[194,195]. Furthermore, TadA of *Xanthomonas oryzae pv. oryzicola* and *Pseudomonas putida* recodes the mRNA of the flagellar filament protein FliC, after exposure to oxidative stress (H_2O_2). This A-to-I editing event results in Ser-to-Pro in the FliC protein, and in a modified flagellar filament structure that increases motility, and contributes to tolerance to oxidative stress, biofilm formation and bacterial virulence[196]. The TadA edited sites of these mRNA substrates have a secondary structure that resembles the afore-mentioned $tRNA^{Arg}_{ACG}$ anticodon stem–loop structure[194,196,197].

Eukaryotic ADs

Eukaryotes express ADs that act on (m)RNA (ADARs, ADADs, UADARs), tRNA (ADATs), or on free nucleobases/ nucleosides/ nucleotides (ADAs), playing an essential role in health and survival (Fig. 1b, 2b, 3; Table 2).

2.3 Adenosine deaminases acting on nucleosides (ADAs)

Eukaryotic ADAs are present in (in)vertebrates and plants, exhibiting a universal role in the purine salvage pathway. Similar to prokaryotic ADs, human ADA mainly exists in two isoforms, ADA1 and ADA2, with complementary roles in a range of biological processes[201,202]. ADA1 is expressed in macrophages and lymphoid cells both intracellularly (nucleus or cytoplasm) and extracellularly (cell surface-bound), while ADA2 is predominantly secreted by myeloid cells[202-206]. Intracellular ADA1 maintains the deoxyadenosine pool in the purine salvage pathway during cell differentiation and selection performed at the bone marrow or thymus[207]. When bound to the T cell–surface, extracellular ADA1 (ecto-ADA1, eADA1) metabolises adenosine into the immunologically inert inosine[208,209]. Contrary to ADA1, ADA2 is mainly an ecto-enzyme (cell surface-bound), and exhibits lower expression levels and enzymatic activity[210,211]. Except from acting as a growth factor for the development and differentiation of haematopoietic and endothelial cells[183,203,204], ADA2 also decreases excess extracellular deoxyadenosine to prevent uptake

and conversion to deoxyinosine by endocellular ADA1 that would cause spontaneous interferon and inflammatory cytokine release and thus (auto)inflammatory diseases (Table 1)[212,213]. Except from its extracellular adenosine deaminase activity, ADA2 additionally acts an intracellular lysosomal DNase that degrades ligands sensed by the cytosolic receptor Cyclic Guanosine monophosphate–Adenosine monophosphate Synthetase (cGAS) that would otherwise trigger interferon release[214]. Except from their enzymatic activity, ADA1 and ADA2 also present a non-enzymatic activity when bound to the cell surface, which mainly serves in the regulation of deoxyadenosine receptors and the creation of immunological synapses for extracellular communication ('purinergic signalling') in proliferation, differentiation, migration, apoptosis and other physiological processes[202,215]. ADAs play important roles, as reflected by the fact that misregulation of ADAs' activity causes a wide spectrum of hematologic, inflammatory, immune, metabolic, neurological, and skin dysfunctions as well as various cancers (Fig. 3)[183,216].

Table 2. Characteristics of AD families found in prokaryotes and eukaryotes.

AD family[a,b]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[c]	Deamination Substrate	Recognition motif[d]	Biological role[e]
PROKARYOTES							
1. Adenosine deaminases acting on tRNA (ADATs)							
TadA	167	Widespread in bacteria (incl. *E. coli*)	n.a.	C, E	A_{34} of $tRNA^{Arg}_{ACG}$ and $tRNA^{Leu}_{AAG}$, coding mRNA (!)	5'-UACG-3' (or rarely 5'-VACG-3') within a stem-loop structure	Codon decoding. Survival response to antibiotics, starvation, and oxidative stress. Induction of biofilm formation, bacterial virulence, siderophore biosynthesis and Fe^{3+} uptake.
2. Adenosine deaminases acting on nucleosides (ADAs)							
ADA1	396	Widespread in bacteria/archaea (incl. *Streptomyces coelicolor*)	n.a.	C	(2'-deoxy)adenosine, Ara-A	-	Balance of adenosine levels in the purine salvage pathway. Complementation of ADA1 lack in eukaryotic symbionts. Interference with pterin biochemical compounds. Decrease of anti-viral activity of purine nucleoside antibiotics.
ADA2	Variable	Gram-negative *Bacteroidetes* (incl. *Streptomyces coelicolor*)	n.a.	E	(Methyl)adenosine	-	Participation in a 'purinergic signalling' system that mediates communication among prokaryotic cells. Degradation of extracellular (methyl)adenosine to protect the eukaryotic host from infection.
EUKARYOTES							
1. Adenosine deaminases acting on mRNA (ADARs)							
ADAR1 ***[5 isoforms]***	1226 *[p150]*; 931 *[p110, edited]*	Mammals (incl. humans)	All	N>C *(p110)* C>N *(p150)*	(Non)-coding dsRNA. Viral dsRNA.	5'-UAG-3' *(-1: U > A > C > G;* *+1: G > C = A > U)*; 5'-UA-3' (viral UTRs)	Regulation of protein expression. Protein recoding. Pro-viral (HHV8, EBV, MV)/ anti-viral (HCV, HDV, LCMV, IFV-A, MV), pro/anti-cancer, anti-apoptotic, anti-inflammatory, anti-retrotranscriptional (L1), pro-/anti-RNAi activity. Homeostasis. Organ development. Keratinocytes migration/ differentiation.
		Mouse	Skeletal muscles	-	-	-	Skeletal myogenesis. Myoblast morphogenesis. Anti-apoptotic activity.
		Earthworm	Ventral nerve cord, dorsal vessel, gut epithelia	-	-	-	Chemotaxis. Homeostasis. Muscle redifferentiation/regeneration.
		Newt	Heart, limbs	C>N (uninjured) N>C (injured)	-	-	Cell proliferation/remodelling. Tissue regeneration.
		Oyster, abalone	-	-	Viral non-coding dsRNA.	-	Anti-viral (MHV) activity.
		Coleoids	Neurons	-	-	-	-

Table 2. Characteristics of AD families found in prokaryotes and eukaryotes. *(continuation)*

AD family[a,b]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[c]	Deamination Substrate	Recognition motif[d]	Biological role[e]
ADAR2 ***[6 isoforms]***	741 *[2b];* 701 *[2a, edited]*	All vertebrates (incl. humans), invertebrates.	CNS, many organs (incl. pancreas, testis)	N *(2a; 2b)*	Coding dsRNA. Onco-(pre)miRs.	5'-UAG-3' *(-1: U > A > C > G, +1: G > C = A > U);* Q/R site	Protein recoding during developmental stages. Differentiation of germ cells in spermatogenesis. Anti-cancer activity.
		Poikilothermic animals	Brain, heart, limbs	-	-	-	Adaptation to harsh environments.
		Newt	Brain, heart, limbs	N (uninjured) C>N (injured)	-	-	Tissue regeneration.
		Larvae and flies	CNS, few other tissues in embryos	N *(3/4 S; 3/4 G)*	Coding dsRNA. dFMR1 substrates.	5'-HA-3'; Q/R-N, Y/C, S/S, M/V, E/E, N/D, K/R, N/S, S/G, M/I, N/S, N/GorD, N/GorS, R/G, K/K, I/V sites	Protein recoding. Nervous system functionality. Anti-viral (SV) activity.
		Coleoids	Neurons	-	(Non)-coding dsRNA.	5'-AAG-3' *(-1: A > U > C > G; +1: G > A = T = C)*	Neuronal morphology/ excitability/ diversification. Adaptation to low temperatures and high salinity.
		Lophotrochozoans	-	-	-	-	-
ADAR3 ***[2 isoforms]***	739	Mammals (incl. humans)	Brain	N	ADAR 1/2 substrates. (Non)-coding dsRNA. ssRNA (!)	5'-UAG-3' *(-1: U > A > C = G; +1: G > C > A = U)*	Inhibition of ADAR1 and 2 activity. Protein recoding. Transcription regulation. mRNA decay. Cell/neuronal signalling. Cell autophagy. Development of nervous system, memory, neuronal plasticity. cognition, and behaviour. Pro/anti-cancer activity, anti-inflammatory, and anti-apoptotic activity.
ceADR1 ***[5 isoforms]***	964 *[C]*	*C. elegans*	Neural cells, vulva	-	dsRNA	-	Transcription regulation. Facilitation of ceADR2-mediated editing. Development. Neuronal function. Locomotion.
ceADR2 ***[no isoforms]***	495	*C. elegans*	Neural cells	C (inactive) N (active)	Non-coding dsRNA	5'-UUAA-3' *(-2, -1: U > A > C > G; 3' nt: A > G > U > C)*	Development. Neuronal function. Locomotion.
ADAR-like	335	Invertebrates (incl. *Octopus bimaculoides*)	-	-	-	-	Probably role in development.
2. Adenosine deaminase domain-containing proteins (ADADs)							
ADAD1 ***[3 isoforms]***	576	Male vertebrates (incl. humans)	Testis	N	Non-coding dsRNA.	Unknown	Post-transcriptional regulation. Spermatogenesis.

Table 2. Characteristics of AD families found in prokaryotes and eukaryotes. *(continuation)*

AD family[a,b]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[c]	Deamination Substrate	Recognition motif[d]	Biological role[e]
ADAD2 ***[2 isoforms]***	583	Male vertebrates (incl. humans)	Testis	N, C	dsRNA	Unknown	Post-transcriptional regulation. Spermatogenesis.
3. Adenosine deaminases acting on tRNA (ADATs)							
ADAT1 ***[2 isoforms]***	502	Humans, mice	Most tissues	C	A_{37} of $tRNA^{Ala}_{AGC}$	5′-GC<u>A</u>UGC-3′	Increase of base-stacking. Stabilization of codon–anticodon interactions. Block of translational frameshifts. Enhancement of translational fidelity.
		Yeast	-	C	-	-	Increase of base-stacking. Stabilization of codon–anticodon interactions. Block of translational frameshifts. Enhancement of translational fidelity.
		Male pig	Testis	C	-	-	Fertility.
		Earthworm	Muscles	C	-	-	Partial restoration of the injured region. Tissue redifferentiation.
		D. melanogaster	Mainly CNS	C	-	-	Development of brain and/or ventral nerve cord.
		Ctenophora	Aboral organs	C	-	-	Locomotion. Geotaxis. Behavior. Tissue complexity. Regeneration.
		Cridaria, porifera, placozoa	-	-	-	-	-
ADAT2-3 complex (hetADAT) ***[ADAT2: 2 isoforms; ADAT3: no isoforms]***	191 (ADAT2) 351 (ADAT3)	All (incl. humans)	-	C	A_{34} of $tRNA^{Thr}_{AGU}$, $tRNA^{Ala}_{AGC}$, $tRNA^{Pro}_{AGG}$, $tRNA^{Ser}_{AGA}$, $tRNA^{Leu}_{AAG}$, $tRNA^{Ile}_{AAU}$, $tRNA^{Val}_{AAC}$, $tRNA^{Arg}_{ACG}$	5′-YU<u>A</u>-3′	Degeneration of codon decoding. Influence on codon bias and tRNA gene copy number. Increase of translational efficiency and speed. Prevention of translational frameshift, mistranslation, ribosome stalling and translational arrest of low complexity, TAPSLIVR-rich proteins.
		Yeast	-	C	A_{34} of $tRNA^{Thr}_{AGU}$, $tRNA^{Ala}_{AGC}$, $tRNA^{Pro}_{AGG}$, $tRNA^{Ser}_{AGA}$, $tRNA^{Ile}_{AAU}$, $tRNA^{Val}_{AAC}$, $tRNA^{Arg}_{ACG}$	-	Cell cycle. Viability.
		Trypanosomes	-	-	C-to-U deamination on ssDNA (!)	-	-
		Ctenophora, cridaria, porifera, placozoa	-	-	-	-	-
AtTadA ***[no isoforms]***	1307	*A. thaliana*, castor bean, rice, poplar, alga	Chloroplast	-	A_{34} of cp-$tRNA^{Arg}_{ACG}$	5′-CU<u>A</u>-3′	Increase of chloroplast translation efficiency and yields of chloroplast-encoded proteins. Photosynthesis and plant growth.

Table 2. Characteristics of AD families found in prokaryotes and eukaryotes. *(continuation)*

AD family[a,b]	Size (aa) [isoforms][a]	Organism	Tissue of expression	Cellular localisation[c]	Deamination Substrate	Recognition motif[d]	Biological role[e]
AtTadA ***[no isoforms]***	1307	*A. thaliana*, castor bean, rice, poplar, alga	Chloroplast	-	A_{34} of cp-$tRNA^{Arg}_{ACG}$	5′-CUA-3′	Increase of chloroplast translation efficiency and yields of chloroplast-encoded proteins. Photosynthesis and plant growth.
4. Adenosine deaminases acting on nucleosides (ADAs)							
ADA1 ***[no isoforms]***	363	All (incl. humans) except plants, most insects, low fungi and several pathogens (some higher fungi and protists)	All (mainly B and T lymphocytes, and macrophages)	C/N>E	(2′-deoxy)adenosine	n.a.	Enzymatic and non-enzymatic regulation of immune function. Maintenance of the 2′-deoxyadenosine pool in the purine salvage pathway during cell differentiation and selection (in bone marrow and thymus). Enhancement of germinal centre formation/ anti-HIV humoral immunity. Metabolism of adenosine into the immunologically inert inosine. Induction of intercellular communication, B/T lymphocyte differentiation or maturation, Th2 immunity. Formation of immunological synapsis (by bridging dendritic with T cells).
		Dogs, horses, pigs, sows, cows	-	-	(2′-deoxy)adenosine	n.a.	-
ADA2 ***[2 isoforms]***	511	Most vertebrates (incl. humans), invertebrates, high fungi/protista	Mainly activated monocytes, macrophages, and dendritic cells	E>C/N	(2′-deoxy)adenosine	n.a.	Growth factor for development and differentiation of endothelial and haematopoietic cells. Lysosomal DNase activity. Induction of B/T lymphocyte differentiation and maturation. Balance of 2′-deoxyadenosine levels on the surface of endothelial cells to prevent imflammation.
		Dogs, horses, pigs, sows	-	-	(2′-deoxy)adenosine	n.a.	-
		D. melanogaster	Haemocytes	-	(2′-deoxy)adenosine	n.a.	-
		Frogs	-	-	(2′-deoxy)adenosine	n.a.	-
5. Unidentified fungal adenosine deaminases acting on coding mRNA (UADARs)							
-	-	Fungi (incl. filamentous ascomycetes, wood-decaying basidiomycetes)	-	-	Coding mRNA	5′-YUARGR-3′ *(-2: C = U > A = G; -1: U > A = C > G; +1: A = G > U > C; +2: G > U = C > A; +3: G = A > Y)* in hairpin loops	Protein recoding. Regulation of gene expression. **Filamentous ascomycetes:** Formation of sexual fruiting bodies. Ascus/ascospore/sexual development. Autophagy. Histone modification,. Chromatin remodelling. Signalling. **Basidomycetes:** Substrate specificity of wood-decaying fungi (growth on different substrates).

[a]isoforms' and ' size' refer to the human homologue (when applicable) or the representative organism of the group (for example *C. iana*, *Octopus bimaculoides*). In prokaryotes, they refer to *E. coli*, except from ADA1 and ADA2 (*Streptomyces coelicolor*). ' Size' also refers isoform, unless stated otherwise.

[a]In eukaryotes, isoforms' and 'size' refer to the human homologue (when applicable) or the representative organism of the group (for example *C. elegans*, *A. thaliana*, *Octopus bimaculoides*). In prokaryotes, they refer to *E. coli*, except from ADA1 and ADA2 (*Streptomyces coelicolor*). 'Size' also refers to the unedited isoform, unless stated otherwise.
[b]Synonyms of ADAR1 = ADAR, DSRAD, G1P1, IFI4; p110 = ADAR1-S; p150 = ADAR1-L; ADAR2 = ADARB1, DRADA2, RED1; ADAR2a = ADAR2S, DRADA2a, RED1-S; ADAR2b = ADAR2L, DRADA2B, RED1-L; 3/4 S = 3b, ¾; 3/4 G = 3α; ADAR3 = ADARB2, RED2; ceADR1 = ADR-1; ceADR2 = ADR-2, ceADAR; ADADs = TENRs; ADAD1 = TENR; ADAD2 = TENRL; ADATs = TADs; ADAT1 = Tad1; scADAT1 = scTad1p; ADAT2 = Tad2, DEADC1; ADAT3 = Tad3; hetADAT = ADAT; ADA2 = ADGF, IDGFL, CECR1; TadA = Tad1; ADA2 = ADGF.

2.4 Adenosine deaminases acting on RNA (ADARs)

ADARs are present only in metazoans (both vertebrates and invertebrates)[197,217,218]. They are predominantly expressed in the nervous system to ensure correct neural development and function[219]. Secondarily, some are involved in tissue generation and organ differentiation as well as in anti/pro-viral mechanisms[220-224]. Three members of the ADAR family have been characterised in high metazoans (ADAR1-3) and two in lower metazoans (ceADR1 and 2), all presenting a high variety of dsRNA substrates and expression patterns[225]. Of them, ADAR3 and ceADR1 are catalytically inactive, although they still bind to dsRNA substrates for regulation of various cellular processes, including the editing activity of other ADAR variants (ADAR1 and 2, ceADR2)[226-228].

Catalytically active ADARs (ADAR1, ADAR2, ceADR1) convert A-to-I in dsRNA structures of coding and non-coding regions, conferring mutagenic and duplex-melting activity, respectively[229,230]. Long (>100 bp), almost fully-complementary dsRNA regions (formed by repetitive elements) are edited non-specifically at multiple sites (50% of As; 'hyperediting'), while long, imperfect dsRNAs with mismatches/ bulges/ interior loops or short (>15 bp) dsRNA structures are edited specifically at certain sites ('site-specific editing')[226,231-233]. In coding regions, A-containing sites that form duplexes with sequences of the same exon or the flanking intron are selectively deaminated[226]. Upon A-to-I deamination of mRNAs, the inosine is recognised as guanosine by ribosomes. Hence, editing may result in non-synonymous substitutions, generating protein variants from a single pre-mRNA ('protein recoding')[234]. This type of RNA editing renders ADARs essential for genetic and functional diversity in eukaryotes[197]. Moreover, A-to-I deamination of mRNAs may sometimes generate truncated proteins (either with adjusted or lost functionality) because of inosine-mediated ribosome stalling in the edited mRNA sites[235]. In non-coding regions, in case of secondary structure, the replacement of a canonical Watson-Crick A:U base pair by a non-canonical wobble I:U base pair interaction weakens

or partially melts the dsRNA structure, rendering it vulnerable to degradation[229,230,236]. Most of the edited sites are located in non-coding RNA associated with retrotransposons, i.e. in 5' and 3' untranslated regions (UTRs) and in introns of mRNA transcripts, and in long non-coding RNAs (lncRNAs). UTR-editing concerns *Alu* repeats that flank the coding region of transcripts in opposite orientation, hence allowing for the formation of dsRNA structures after transcription[237,238]. Intron editing generates or destroys splice acceptor sites to regulate differential/alternative splicing, while lncRNA editing impacts infection by dsDNA viruses[239,240]. Apart from editing retrotransposon-associated RNAs, ADARs are also able to edit and thus downregulate miRNA precursors during RNA interference (RNAi). This is achieved by recognition of edited I-containing dsRNA structures of miRNA precursors by the RNAse III Tudor Staphylococcal Nuclease (Tudor-SN) that degrades them, limiting their availability[241]. Alternatively, ADAR binding to the miRNA precursors can sterically inhibit Dicer cleavage ability and RISC loading, suppressing miRNA processing[242-244]. ADARs also edit miRNA targeting sites in mRNA 3'UTRs, affecting the miRNA-mediated post-transcriptional repression of the encoded gene[245]. Altogether, ADAR-mediated RNA editing can regulate mRNA translation, pre-mRNA splicing, RNA stability, mRNA and viral genetic stability, and RNA structure-dependent activities, like miRNA biogenesis and targeting or protein-RNA interactions. In humans and primates, ADAR-mediated editing is abundant and it almost absolutely concerns non-coding dsRNAs (*Alu* repeats represent ~10% of the genome)[246,247]. In contrast, cephalopods show massive non-synonymous recoding, due to their need to diversify their proteome for adaptation in harsh environments[182,246,248].

ADARs are regulated by post-transcriptional RNA modifications (ADAR pre-mRNA splicing and self-editing), post-translational processing (SUMOylation, ubiquitination, and phosphorylation), alteration of protein levels, protein–protein interactions, substrate RNA competition with other proteins or RNAs, and colocalization with substrate RNAs[217]. Mis-regulation of ADAR levels and activity is linked to different neuropathological and psychiatric diseases as well as several cancers (Fig. 3)[249-252].

2.5 Adenosine deaminase domain-containing proteins (ADADs)

ADADs are ADAR-like proteins with dsRNA-binding potential but without deaminase activity[249]. They are expressed almost solely in the spermatid cells of male vertebrates during meiotic and post-meiotic spermatogenesis, participating in testis-specific post-

transcriptional processes, such as mRNA transport, alternative splicing, and packing of heterogeneous nuclear RNA (hnRNA)[253,254]. In humans and mice, only two ADADs are expressed, ADAD1 and ADAD2[249]. Both ADADs are crucial regulators of male germ cell differentiation, acting at different stages of spermatogenesis with currently unknown target specificity[253]. Deficiency of these enzymes may lead to infertility[253,255]. ADADs have also been found in the invertebrate lophotrochozoans, but their molecular mechanism still remains to be elucidated[256].

2.6 Adenosine deaminases acting on tRNA (ADATs)

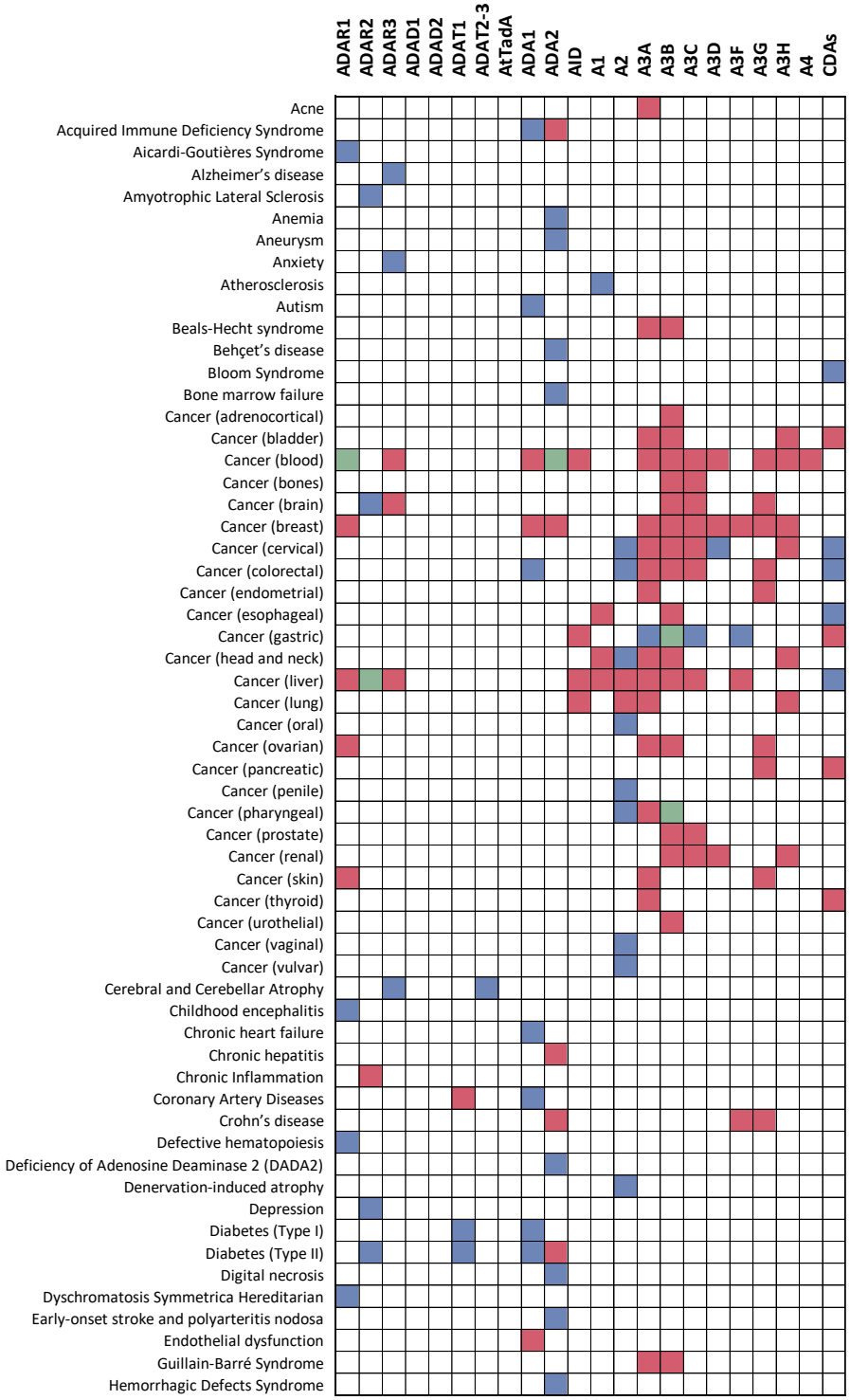
Eukaryotic ADATs play a pivotal role in tissue development/ redifferentiation, locomotion, and viability. These essential enzymes edit tRNAs mainly at two positions: (i) the wobble position-34 that corresponds to the first nucleotide of the anticodon (that base pairs with the third nucleotide of the cognate codon), and (ii) the position-37 that follows the anticodon. The adenosine at position-34 is edited either by a heterodimeric complex composed of ADAT2 and ADAT3 (called hetADAT complex), or by the ADAT2-orthologue tRNA adenosine deaminase arginine (TADA) in chloroplasts. Position-37 is edited by ADAT1. Like their prokaryotic counterparts, editing of tRNAs by ADATs and their subfamily TADAs allows for the translation of multiple codons by a single tRNA[197,257-259]. The activity of these deaminases highly depends on the nucleotide sequence and the three-dimensional structure of the entire tRNA, in particular of the anticodon loop[12,198,257]. Lack of ADAT2 and 3 can lead to cell death, while ADAT1 has been reported not to be essential (Fig. 3)[12].

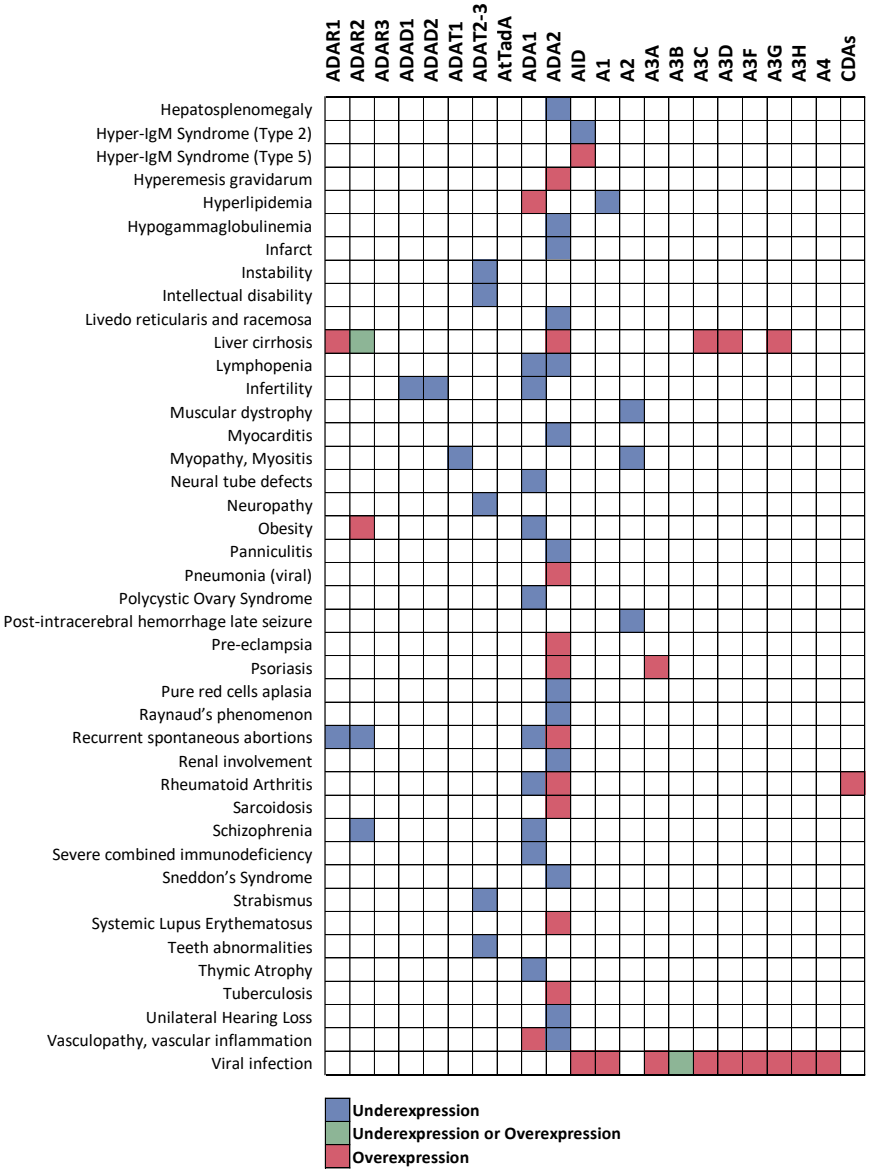
2.7 Unidentified fungal adenosine deaminases acting on coding mRNA (UADARs)

Fungi express ADs that participate in sexual reproduction and wood degradation[182,260-264]. These deaminases act on coding mRNA of genes related to these biological processes, sometimes generating synonymous codons but mostly non-synonymous mutations (premature stop codon correction and single amino acid variations)[182,260,261,265]. Although generally considered negligible, synonymous mutations can strongly affect co-translational protein folding, mRNA structure/stability and association with other RNAs or proteins[182,266]. Non-synonymous mutations alter the protein sequences, providing subtle

modifications or even novel functions that confer adaptive advantages, especially in the case of editing of mRNAs of essential genes[265]. In filamentous ascomycetes, high A-to-I editing in mRNAs involved in sexual reproduction is observed during formation of sexual fruiting bodies, and ascus/ascospore/sexual development[260-263]. These are mainly transcripts of genes involved in autophagy, histone modification, chromatin remodeling, development, and pheromone signalling[182,265,267]. The deaminations mainly concern repeat-induced point mutations before meiosis, and meiotic silencing by unpaired DNA[268,269]. Lack of these deamination events results in defects in morphology or formation of ascospores, even in their discharge[260,262,267,270]. In basidomycetes (*Polyporales*), low mRNA editing is reported, and only in transcripts of enzymes crucial for wood degradation, enabling growth on different substrates[182,264]. The identity and exact biological relevance of the enzymes that catalyse these A-to-I deaminations in fungi still need to be elucidated.

a.


ADAR1
ADAR2
ADAR3
ADAD1
ADAD2
ADAT1
ADAT2-3
AtTadA
ADA1
ADA2
AID
A1
A2
A3A
A3B
A3C
A3D
A3F
A3G
A3H
A4
CDAs
Acne
Acquired Immune Deficiency Syndrome
Aicardi-Goutières Syndrome
Alzheimer's disease
Amyotrophic Lateral Sclerosis
Anemia
Aneurysm
Anxiety
Atherosclerosis
Autism
Beals-Hecht syndrome
Behçet's disease
Bloom Syndrome
Bone marrow failure
Cancer (adrenocortical)
Cancer (bladder)
Cancer (blood)
Cancer (bones)
Cancer (brain)
Cancer (breast)
Cancer (cervical)
Cancer (colorectal)
Cancer (endometrial)
Cancer (esophageal)
Cancer (gastric)
Cancer (head and neck)
Cancer (liver)
Cancer (lung)
Cancer (oral)
Cancer (ovarian)
Cancer (pancreatic)
Cancer (penile)
Cancer (pharyngeal)
Cancer (prostate)
Cancer (renal)
Cancer (skin)
Cancer (thyroid)
Cancer (urothelial)
Cancer (vaginal)
Cancer (vulvar)
Cerebral and Cerebellar Atrophy
Childhood encephalitis
Chronic heart failure
Chronic hepatitis
Chronic Inflammation
Coronary Artery Diseases
Crohn's disease
Defective hematopoiesis
Deficiency of Adenosine Deaminase 2 (DADA2)
Denervation-induced atrophy
Depression
Diabetes (Type I)
Diabetes (Type II)
Digital necrosis
Dyschromatosis Symmetrica Hereditarian
Early-onset stroke and polyarteritis nodosa
Endothelial dysfunction
Guillain-Barré Syndrome
Hemorrhagic Defects Syndrome


ADAR1
ADAR2
ADAR3
ADAD1
ADAD2
ADAT1
ADAT2-3
AtTadA
ADA1
ADA2
AID
A1
A2
A3A
A3B
A3C
A3D
A3F
A3G
A3H
A4
CDAs
Hepatosplenomegaly
Hyper-IgM Syndrome (Type 2)
Hyper-IgM Syndrome (Type 5)
Hyperemesis gravidarum
Hyperlipidemia
Hypogammaglobulinemia
Infarct
Instability
Intellectual disability
Livedo reticularis and racemosa
Liver cirrhosis
Lymphopenia
Infertility
Muscular dystrophy
Myocarditis
Myopathy, Myositis
Neural tube defects
Neuropathy
Obesity
Panniculitis
Pneumonia (viral)
Polycystic Ovary Syndrome
Post-intracerebral hemorrhage late seizure
Pre-eclampsia
Psoriasis
Pure red cells aplasia
Raynaud's phenomenon
Recurrent spontaneous abortions
Renal involvement
Rheumatoid Arthritis
Sarcoidosis
Schizophrenia
Severe combined immunodeficiency
Sneddon's Syndrome
Strabismus
Systemic Lupus Erythematosus
Teeth abnormalities
Thymic Atrophy
Tuberculosis
Unilateral Hearing Loss
Vasculopathy, vascular inflammation
Viral infection
Underexpression
Underexpression or Overexpression
Overexpression

b.

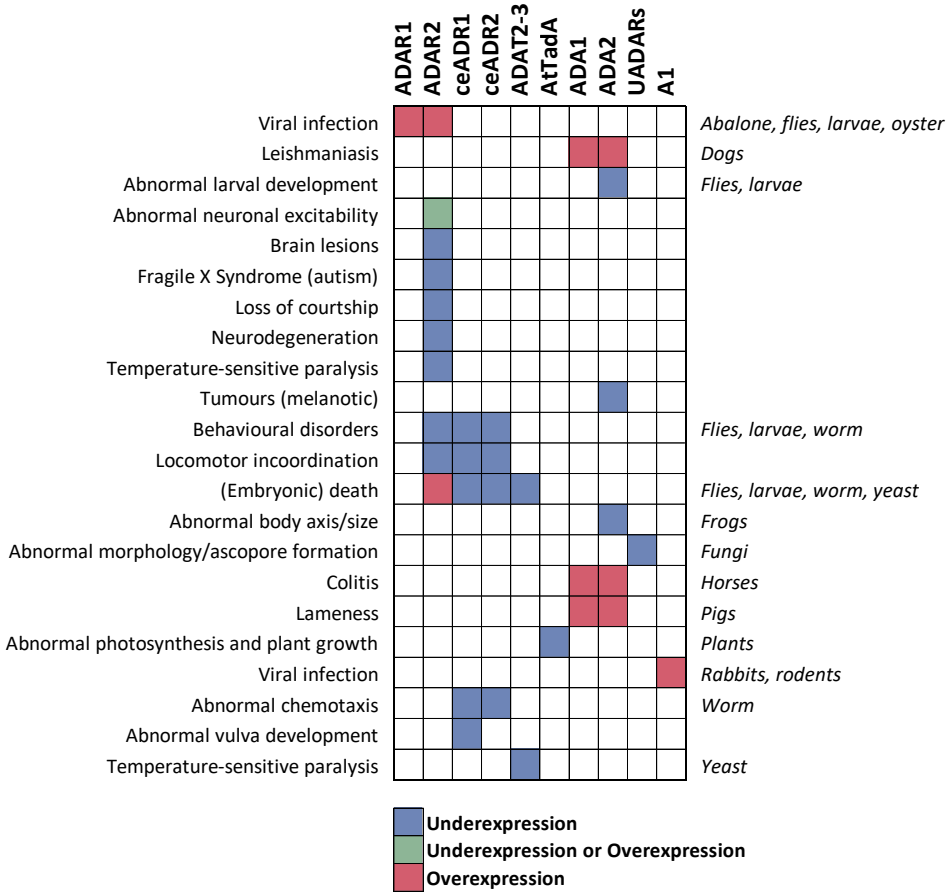


Fig. 3 Diseases related to dysregulation of ADs and CDs. Several disorders of (**a**) humans or (**b**) other organisms are correlated to unbalanced expression levels of ADs and CDs. These enzymes are underexpressed (blue), overexpressed (red), or present either of the two expression states depending on the stage of the disease (green).

BASE-EDITING TECHNOLOGY

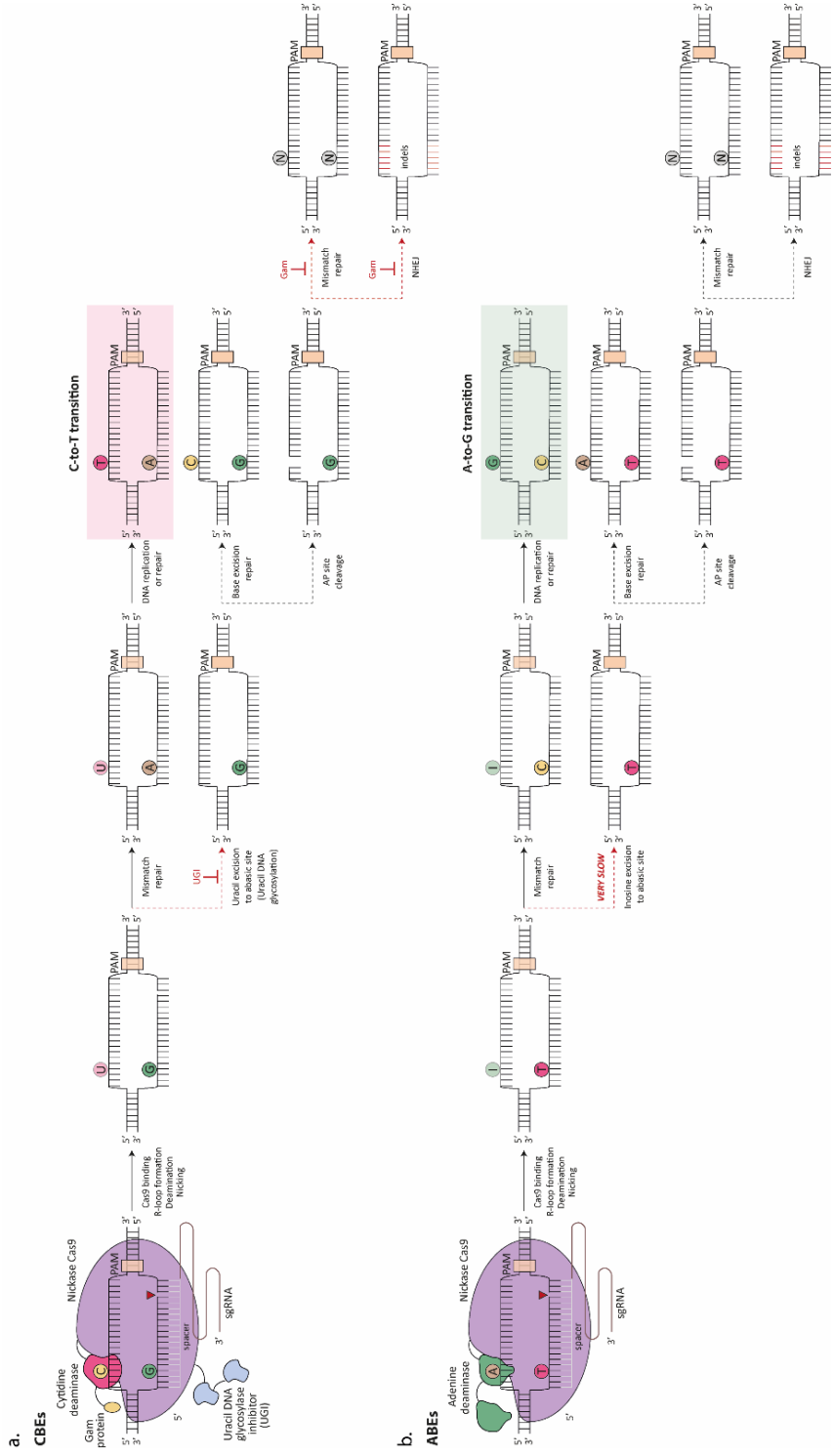
During the last decade, the discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) genes has enabled their exploitation as next-generation genetic engineering tools, potentially revolutionising the development of applications in medicine, agriculture, and biotechnology[271-273]. The CRISPR-Cas technology employs sequence-specific, RNA-guided nucleases that introduce a double-stranded DNA break (DSDB) in the target DNA (protospacer)[274]. Subsequent repair of the DSDB can lead to either precise modifications through template-dependent homology directed repair (HDR), or to insertions/deletions (indels) via imprecise non-homologous end joining (NHEJ) repair[274-276]. Recently, several deaminase enzymes have been fused to partially (nickase) or completely (dead) catalytically inactive Cas enzymes to introduce precise single nucleotide substitutions on the target DNA without generation of DSDBs (Fig. 4, 5)[277-378]. In this way, precision editing depends neither on the generation of DSDBs (which may result in undesired recombination artefacts), nor on the HDR system (which often has a relatively low efficiency, only active during part of the cell cycle), and hence the editing does not require supply of donor DNA fragments.

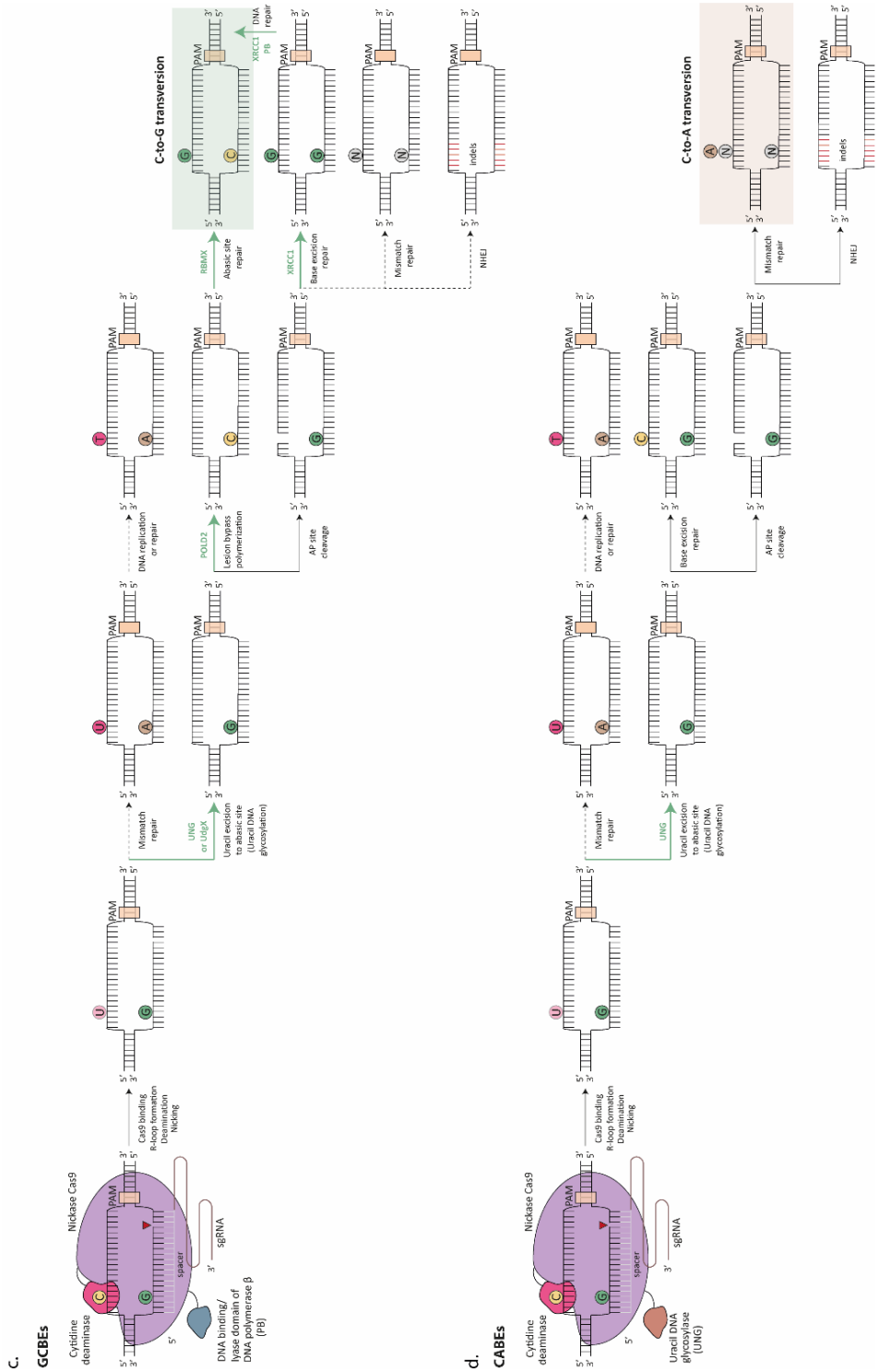
Base-editing proceeds through the step-wise Cas-mediated recognition of the target site: check for an appropriate protospacer adjacent motif (PAM), followed by local unwinding of the dsDNA through base pairing between the RNA guide and the target DNA strand after which deaminase-mediated single base modification(s) occur(s) in a certain region (editing window) of the displaced non-target DNA strand[277,320,342] (Fig. 4). To date, most base-editors (BEs) contain a nickase variant of the Cas9 protein from *Streptococcus pyogenes* (nSpCas9) that recognizes 5'-NGG-3' PAM (Fig. 5). To enable base-editing at alternative PAM sites, numerous engineered nSpCas9 variants (nVQR; nVRER; nEQR; nVRQR; nNRCH; nNRTH; nNRRH; nNG; nSpG; nSpRY; n(i)SpymacCas9)[280,282,284,285,289-291,293,294,298,308-312,319,324,328,321-323,333,345,347-349,353,356,357,361,362,369,372,374,375] have been described, as well as naturally existing Cas9 orthologues (from *Staphylococcus aureus*, nSaCas9; *Neisseria meningitidis*, nNm2Cas9; *Streptococcus canis*, nScCas9; *Campylobacter jejuni*, nCjCas9; *Streptococcus thermophilus*, nStCas9; *Staphylococcus auricularis*; nSauriCas9)[280,286,295-297,304,325,332,345,347,350-353,358,361,362], and nThermoCas9 (**Chapter 5**). Likewise, base editors of Cas12a proteins (from *Lachnospiraceae* sp., dLbCas12a; *Acidaminococcus* sp., dAsCas12a)[301,353] and engineered variants thereof (nSaKKH; dLbCas12a-RVRR; dimpLbCas12a; d(en)AsCas12a; dCasMINI)[287,298,302,303,325,334,335,346-348,353] have been developed (Fig. 5). For applications that require high precision, native or

engineered narrow-window (nYE1; nYE2; nYEE; nEE; nFE1; nYFE; nSECURE)[280,287,290,291,297,304,306] and sequence context-specific (most A1 variants)[280,287,290-292,297,304-306,308,309] deaminases have been used to lower undesired editing of neighbouring bases (bystander edits) (Fig. 5). In contrast, in gene inactivation and mutagenesis applications, Cas9 variants recognizing long protospacer regions (nSt1Cas9; nNm2Cas9)[296,297,358], altered base-editor protein structures (circular permutants; BE-PIGS system; inlaid base-editors)[290,298,299,353,354], and native or engineered deaminases with broad-window and high activity (A3A; PmCDA1; evoCDA; PmCDA1Δ; AID and its engineered variants)[282,289,297,307,308,310,311,321,325,326,328,331] have been adopted, sometimes in multiple copies (BE-PLUS and CRISPR-X systems)[230,328] (Fig. 5). To limit other types of undesired on-target base-editing, such as the generation of transversion mutations or indels, additional proteins (UGI; Mu Gam) have been fused to the base-editing chimeras[277,287,288] (Fig. 5). In addition, undesired off-target base-editing usually occurs both on DNA and RNA level, given the native ability of deaminases to act on both substrates[306,313,359,360]. To minimize Cas-dependent and Cas-independent off-targets, base-editors with high-fidelity nSpCas9/dCas12a variants (neSpCas9(1.1); nSpCas9-HF1/2; nHiFiCas9; nHypaCas9?; nevoCas9; nSniper-Cas9; denAsCas12a; denAsCas12a-HF1)[277,279,283,303,310,327,343,344,353] and precise deaminase enzymes (YE1; RrA3F(F130L); AmA1; SsA3B(R54Q); PpA1(H122A); PpA1(R33A); TadA*8e)[280,287,290-292,304,312,322,333,353,362] have been generated, respectively (Fig. 5). Apart from using RNA-dependent Cas effectors, mitochondrial and chloroplast DNA base-editing has also been achieved by combining a CD (DddA; DddA-A6; DddA-A11)[336-340] and/or an AD (Tad8*e)[363] with guide-independent TALEs (Fig. 5a, b).

BEs can also be programmed to target RNA substrates. Most RNA BEs contain a dCas13b protein (from *Riemerella anatipestifer*, dRanCas13b; *Prevotella* sp., dPspCas13b) fused to a catalytically enhanced or modified ADAR2 AD domain, acting on dsRNA[339,341,367]. Like Cas13, these Cas13-based RNA BEs do not require a PAM. Alternatively, native ADAR1 or ADAR2 are recruited by engineered antisense RNA molecules for targeted RNA editing[365,366].

In the next section, we describe the types of DNA and RNA BEs developed to date, classified by the nature of their deaminase enzyme.


a.
CBEs
Nickase Cas9
Cytidine deaminase
Gam protein
Uracil DNA glycosylase inhibitor (UGI)
sgRNA
spacer
PAM
Cas9 binding
R-loop formation
Deamination
Nicking
DNA replication or repair
Mismatch repair
UGI
Uracil excision to abasic site (Uracil DNA glycosylation)
Base excision repair
AP site cleavage
Gam
Mismatch repair
NHEJ
indels
C-to-T transition
b.
ABEs
Adenine deaminase
VERY SLOW
Inosine excision to abasic site
A-to-G transition


c.
GCBEs
Cytidine deaminase
Nickase Cas9
PAM
spacer
sgRNA
DNA binding/ lyase domain of DNA polymerase β (PB)
Cas9 binding
R-loop formation
Deamination
Nicking
Mismatch repair
UNG or UdgX
Uracil excision to abasic site (Uracil DNA glycosylation)
DNA replication or repair
POLD2
Lesion bypass polymerization
AP site cleavage
RBMX
Abasic site repair
XRCC1
PB
DNA repair
Base excision repair
Mismatch repair
NHEJ
indels
C-to-G transversion
C-to-A transversion
d.
CABEs
Cytidine deaminase
Nickase Cas9
sgRNA
Uracil DNA glycosylase (UNG)
Cas9 binding
R-loop formation
Deamination
Nicking
Mismatch repair
UNG
Uracil excision to abasic site (Uracil DNA glycosylation)
DNA replication or repair
Base excision repair
AP site cleavage
Mismatch repair
NHEJ
indels

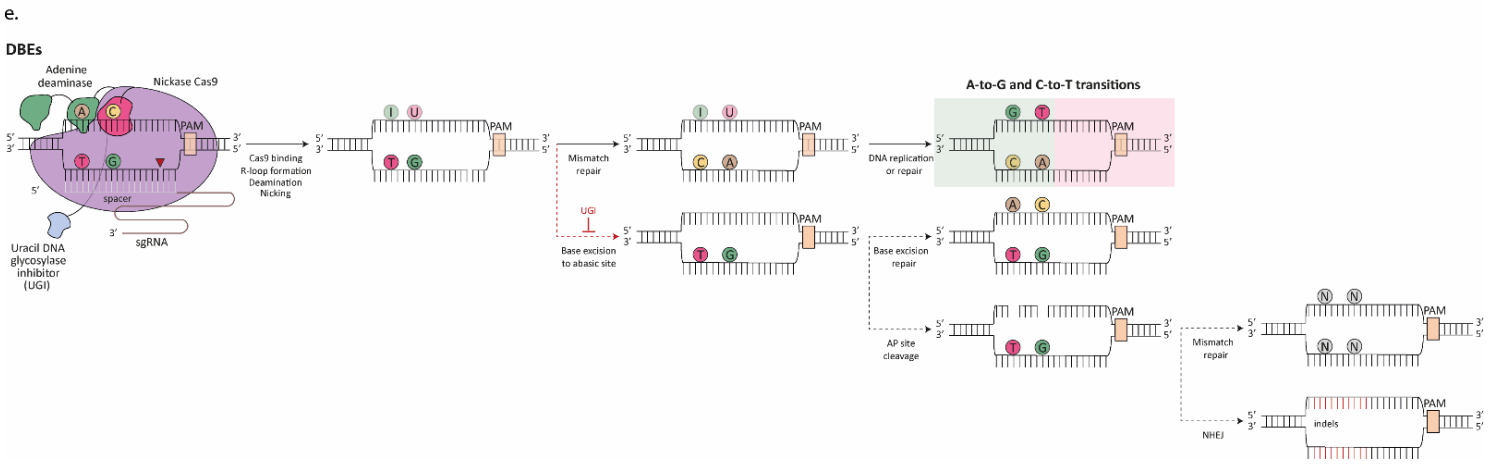


Fig. 4 Base conversion mechanism for different types of BEs. **a.** CBEs deaminate C•G to U•G, which is partially converted into T•A via mismatch repair and DNA replication/repair. Addition of UGI and Gam weaken the formation of undesired by-products. **b.** ABEs deaminate A•T to I•T, which is predominantly converted into G•C via mismatch repair and DNA replication/repair. The very slow kinetics of inosine excision imply negligible formation of by-products. **c.** CGBEs deaminate C•G to U•G and enhance the formation of G•C through the addition of UNG, UdgX, POLD2, RBMX, XRCC1 and/or PB. **d.** CABEs deaminate C•G to U•G and enhance the formation of A•T through the addition of UNG. **e.** DBEs deaminate C•G to U•G and A•T to I•T simultaneously, leading to the generation of T•A and G•C via mismatch repair and DNA replication/repair. Addition of UGI weaken the formation of undesired by-products.

1. Cytosine base-editors (CBEs)

CBEs are the first synthetic, progammable DNA BEs established, and they catalyse C•G to T•A transition on the target DNA (Fig. 4a). Specifically, CBEs convert C•G to U•G, which is subsequently replaced by U•A through cellular mismatch repair and, finally, by T•A via DNA replication/repair[277,320]. However, the uracil base of the U•G deamination product may be excised to abasic site through uracil DNA glycosylation, leading to regeneration of the original pair (C•G) or unintended base-scrambling (such as C•G to G•C or C•G to A•T conversions) and indel formation[277,320,379,380]. Fusion of one or multiple copies of uracil DNA glycosylation inhibitor (UGI) from *Bacillus subtilis* bacteriophage PBS1 and/or from the bacteriophage Mu Gam to CBEs has massively increased product purity and editing efficiency[277,287,320,342]. To further increase C•G to T•A editing efficiency, the selection of proper CD variants is crucial. As mentioned in the first section of this review, the only characterised CDs that act on DNA originate from the eukaryotic AID/APOBEC family and from the prokaryotic BTCDA family (Table 1). Current CBEs predominantly employ the eukaryotic AID/APOBEC family members with the highest catalytic activity on ssDNA substrates (A1; A3A; A3B; A3F; A3G; A3H; AID; PmCDA1), while the prokaryotic DddA toxin from the BTCDA family was recently harnessed (in a split form) for nuclear and organellar DNA base-editing because of its unique ability to deaminate dsDNA substrates (Table 1; Fig. 5a)[336-339]. Alternatively, evolved ADAR variants are fused to dCas13b for cytidine deamination on RNA substrates (Fig. 5a)[341].

1.1 A1-based CBEs

Rat APOBEC (rA1) (expressed as homodimer) was the first CD applied in CBEs because of its higher ssDNA deaminase activity compared to other CDs, such as hA3G, hAID, and PmCDA1[277]. rA1 was initially fused to a catalytically inactive SpCas9 variant (dSpCas9), forming the BE1 system with low editing efficiencies[277]. After the addition of UGI (BE2), the replacement of dSpCas9 with nickase SpCas9 (nSpCas9; BE3) as well as the fusion of a second copy of UGI (BE4) and the Gam protein (BE4-Gam), base-editing efficiencies and purities were increased[277,287,288]. To expand CBE-mediated editing also to non-NGG sites, engineered SpCas9 variants with altered PAM preferences (nxCas9; nVQR; nVRER; nEQR; nNRCH; nNRTH; nNRRH; nNG; nSpG; nSpRY; n(i)SpymacCas9)[280282,289-291,293,294,298,308-312,321-324,328,333] as well as native or engineered SpCas9 orthologues (nSaCas9; nSaKKHCas9; nScCas9; nSt1Cas9; nSt1Cas9-LMG18311; nSt1Cas9-CNRZ1066; nNm2Cas9;

nSauriCas9)[280,286,287,295-297,304,325,326,334-336] and Cas12 variants (dLbCas12a; dLbCas12a-RVRR; dimpLbCas12a; dAsCas12a; denAsCas12a)[301-303] were applied. The editing window of most A1-based CBEs was approximately 5 to 8 nucleotides (Fig. 5a). Nonetheless, to increase the efficiency of gene disruption, the editing window has also been extended to up to 18 nucleotides through (a) recognition of GCN4 peptides by multiple copies of rA1-UGI (BE-PLUS)[300], (b) integration of the rA1 gene in the PAM-interacting domain of SpCas9 (BE-PIGS)[299], (c) circular permutation of SpCas9 (nCP1012; nCP1028; nCP1041; nCP1249)[290,298], (d) utilisation of SpCas9 orthologues (nSt1Cas9; nNm2Cas9, ThermoCas9)[296,297] (see also **Chapter 6**) or (e) implementation of the highly expressed ancestrally reconstructed APOBEC1 node 689 (Anc689)[309]. In contrast, to optimise CBEs for precision base editing other engineering has been executed. More narrow editing windows as well as less by-stander editing and Cas-independent off-targeting were observed with native (from *Alligator mississippiensis*, AmA1; *Pongo pygmaeus*, PpA1)[292] or engineered A1 variants (PpA1 H122A; YE1; YE2; YEE; EE; R33A; A1 R33A/K34A; FE1; YFE; AALN; FERNY)[280,287,290,291,297,304,306,307]. Moreover, Cas-dependent off-targeting was minimised using more accurate Cas variants (nSpCas9-HF1/2, nxCas9, nSniper-Cas9 and the catalytically inactive denAsCas12a)[277-279,281,283,289,310,311]. In general, wild-type and engineered A1 variants preferably edit 5'-HC-3' sites (T<u>C</u> ≥ C<u>C</u> ≥ A<u>C</u>), presenting poor activity on 5'-G<u>C</u>-3' regions[277,381]. To overcome this limitation, the engineered A1 variants evoA1 and (evo)FERNY[307,308] as well as alternative CDs were later adopted in CBEs.

1.2 A3A-based CBEs

A3A (expressed mainly as monomer) is the strongest ssDNA mutator of the AID/APOBEC family[318]. A3A-based editors show improved editing efficiencies (especially in the 5'-G<u>C</u>-3' context), and broader editing window[307,383] compared to A1-based editors[307,331,381-383]. A3A has been combined with engineered SpCas9 variants (nxCas9; nVRQR; nSpCas9-HF1; nRY; nNG; nG; nNRRH; nNRCH)[310,312], nNm2Cas9[297], and native or engineered Cas12a proteins (dLbCas12a; dAsCas12a; denCas12a)[301,303] to enable recognition of alternative PAM sites. High RNA off-target editing has initially been observed with A3A-based editors[360], which has recently been abolished with engineered A3A variants (R128A; Y130F)[313]. Moreover, CBEs with narrower editing windows and fewer bystander mutations have been achieved with engineered A3A variants (eA3A; Y130F; Y132D; N57A; QF; Δ182; Δ186)[310,311,314], especially with eA3A that presents high sequence context dependence on 5'-TC-3'

motifs[360]. On the other hand, larger editing windows were reached with four UGI copies[312].

1.3 A3B-based CBEs

After A3A, the truncated human A3B (hA3Bctd) is the second strongest ssDNA mutator of the AID/APOBEC family, and it predominantly recognizes 5'-TC-3' sites[318]. Because of their lower editing efficiencies compared to A3A-based CBEs, only few A3B-based CBEs have been developed to date. CBEs containing hA3Bctd present a larger editing window than those expressing the native/engineered A3B orthologue from *Sus scrofa* (naturally expressed as monomer)[292,317]. The latter together with engineering hA3Bctd variants (A3Bctd-VHM; A3Bctd-KKR) reduced Cas9-independent off-target editing, while maintaining on-target efficiencies[292,318].

1.4 A3F-based CBEs

Despite the high ssDNA affinity of A3F, A3F (expressed as monomer) exhibits low deaminase activity and thus the A3F-based editors have been reported to be the weakest of all AID/APOBEC-derived base editors[318,319]. Hence, only two A3F-based CBEs have been created, expressing SpCas9 fused to native or engineered (F130L) A3F protein from *Rhinopithecus roxellana*. RrA3F preferably recognizes 5'-TC-3' sites, like most A3s, and it shares a similar editing window with A1-based CBEs[292]. As expected from their low on-target activity, both A3F-based editors also have decreased Cas9-indepenent off-targeting levels[292,318].

1.5 A3G-based CBEs

After A3A, A3B, A1 and AID, A3G (expressed as monomer) presents the fifth highest deaminase activity among AID/APOBECs in the CBE context[277,292,318]. Unlike the other A3 variants, A3G shows preference for 5'-CC-3' motifs, allowing for deamination of sites that were previously considered 'hard to get'[314,317,319]. The (e)A3G-based editors exhibit reduced bystander mutations[384]. Hence, native and engineered (oA3G, NL hA3G, D316R/D317R, eA3G) A3G variants have been combined with nSpCas9, nSpCas9-NG, and

nNm2Cas9 to broaden the targeting range, presenting activity windows similar to A1-based CBEs[291,297,311,314,317,319].

1.6 A3H-based CBEs

A3H is expressed mainly as monomer. A3H-based editors are the sixth strongest among AID/APOBEC-based CBEs, followed by the A3F editors[318,319]. Like A3F, only two A3H-based CBEs have been generated, expressing SpCas9 fused to native or engineered (R175/R176E) human A3H-II. The editing window of both versions is similar to that of A1-based CBEs[317].

1.7 AID-PmCDA1-based CBEs

AID and PmCDA1 are expressed either as monomers or homodimers. AID/PmCDA1-based editors exhibit the fourth highest deaminase activity among AID/APOBECs CBEs, and they provide wider editing windows and more relaxed target site specificity[307,383]. Especially large windows are reached in the presence of eAID[297,328], while hypermutagenesis (at a narrower window) occurs in the presence of AIDx[334,335]. Intriguingly, in the CRISPR-X system, four copies of AIDx (a fusion between AID and the MS2 coat protein) are recruited in couples by two MS2 RNA hairpins fused to the Cas9 RNA guide, creating an editing window of 100 nucleotides with limited off-target editing but low product purity[328]. Wild-type hAID and PmCDA1 present negligible RNA off-target activity[360]. Reduced DNA off-target editing has also been observed when high-fidelity SpCas9 (HF2, eSpCas9, HypaCas9)[327] or precise AID/PmCDA1 (tCDA1EQ, CDA1 Δ188/Δ190/Δ192/Δ193/Δ194, PmCDA1Δ, eCDA, AIDmono, hAID*Δ, NL hAID, AICDA)[311,326,328,330-333] variants were applied. Notably, hAID*Δ and PmCDA1, like A3A, presented preference for 5'-GC-3' sites[320,331-333], while evoCDA1 was able to edit 5'-NC-3' sites[307]. AID/PmCDA1-based editors have also been combined with numerous engineered SpCas9 variants (nNG, nRY, nVQR, niSpyMac, nVRER, nxCas9)[282,289,308,311,321-324,328,333] and native/engineered Cas9 orthologues (nSaCas9, nSaKKH)[297,334,335] with distinct PAM preferences.

1.8 DddA-based CBEs

Whereas most DNA-specific CDs only edit ssDNA, the DddA toxin has been demonstrated to deaminate dsDNA[40,41]. Synthetic fusions have been constructed of a single, split DddA domain (from *Burkholderia cenocepacia,* native or engineered[40]) and a pair of Cas9 proteins (dSpCas9, nSaKKHCas9)[336]. Using two guides, a pair of closely spaced protospacers are targeted by the Cas9 effectors, allowing for the split DddA domains to dimerize, and edit cytidine(s) in the region between the two protospacers[336]. Using these monomeric DddA-constructs may prevent toxicity and minimise off-targeting. Wild-type and engineered DddA split variants have also been combined with TALEs, enabling base-editing in mitochondrial and chloroplast dsDNA. All DddA-based CBEs preferably edit 5'-TC-3' sites, like most APOBECs[336-340].

1.9 ADAR2-based RNA CBEs

RNA-targeting CBEs have also been developed. Although CDs acting on RNA substrates could theoretically be used for the construction of RNA CBEs, the high RNA affinity of these enzymes would probably lead to transcriptome-wide off-target activity. However, this has never been tested. Instead, the AD domain of the human ADAR2 (ADAR2DD) has been evolved into a CD (ADAR2DD-r16) and fused to the catalytically inactive dCas13b orthologue from *Riemerella anatipestifer* (dRanCas13b) in order to form the first RNA CBE (RESCUEr16). However, RESCUEr16 exhibited substantial A-to-I RNA editing. Hence, ADAR2DD-r16 was further engineered to a higher fidelity variant (ADAR2DD-r16/S375A), giving rise to a RNA CBE system with improved specificity but with slightly lower C-to-U deamination activity (RESCUE-S)[341].

2. Adenine base-editors (ABEs)

Inspired by CBEs, ABEs were subsequently developed to catalyse A•T (or A•U) to G•C transition on DNA (or RNA) substrates (Fig. 4b). On DNA level, ADs convert A•T to I•T (A-to-I on displaced, non-target strand), which is subsequently replaced by I•C through cellular mismatch repair (in case of nCas9, repair of nicked target strand) and, finally, by G•C via DNA replication/repair. The inosine base of the I•T deamination product may also be excised to abasic site, resulting in regeneration of the original pair (A•T) or in

unintended base-scrambling (T•A or C•G) and indel formation. Nevertheless, the first step of this secondary pathway (inosine excision to abasic site) is very slow, implying that the favourable equilibrium results in high G•C product purity[342]. Contrary to CDs, none of the characterised ADs is able to naturally catalyse A•T to G•C deamination on dsDNA substrates. Therefore, current ABEs employ nCas9 or dCas12a proteins fused to evolved TadA variants with the acquired ability to act on ssDNA. Evolved TadA enzymes are also combined with inactive or split DddA and TALE proteins for mitochondrial dsDNA unwinding and targeting, respectively. Alternatively, native ADAR proteins are recruited to the target site by antisense RNA molecules, or engineered ADAR2DD variants are fused to dCas13b for adenosine deamination on RNA substrates (Fig. 5b).

2.1 TadA-based ABEs

Initial attempts of DNA adenosine deamination employed fusion of *Escherichia coli* TadA (naturally expressed as homodimer) with nSpCas9 (ABE0.1), resulting in negligible editing efficiencies[342]. TadA was selected because it shares homology with the APOBEC enzymes that bind ssDNA in a way that resembles TadA binding to the single-stranded anticodon loop of $tRNA^{Arg385}$. Moreover, contrary to ADARs, TadA does not require small-molecule activators for activity[386]. TadA acts as an homodimer in nature, with one monomer binding to the tRNA substrate and the other monomer catalysing the deamination reaction[259]. Hence, seven new generations of ABEs (ABE1 – ABE7) were subsequently developed, all containing two fused copies of TadA; a wild-type and a mutant variant. The mutant variant derived from several cycles of *in vitro* evolution and protein engineering for increased affinity to DNA substrates. ABE7.10 was the most efficient editor, exhibiting a small editing window (around 4 nucleotides), minimum indel rates, almost absolute product purity, and low off-target editing[342]. Cas9-dependent off-targeting was minimized using precise Cas variants (SpCas9-HF1/2; xCas9; eCas9; HypaCas9; Sniper-Cas9; evoCas9)[281,343,344]. Notably, a minimum editing window (2 nucleotides) was achieved with further engineered ABE7.10 variants[313]. In contrast, large editing windows were observed with inlaid base editors (IBEs)[354], containing engineered tadA gene embedded into the cas9 gene, as well as with circularly permutated SpCas9 variants (nCP1012; nCP1028; nCP1041; nCP1249)[298,343,353] and engineered dCas12f (dCasMINI)[346], due to altered ABE architectures. To expand the targeting scope of ABEs, engineered SpCas9 variants with altered PAM preferences (nxCas9; nSpG; nNGA; nNGC(*); nNGX(*); nNGX-NGA; nNGX-NGC; nRY; nVQR; nVRER; nEQR; nVRQR; nNRRH; nNRTH; nNRCH;

n(i)SpymacCas9)[280,281,284,293,294,298,312,322,324,333,345,347-349,353,356,357,361,362] as well as native or engineered SpCas9 homologues (nSaCas9; nSaKKHCas9; nScCas9; nCjCas9)[296,298,332,345,347,348-353,358,361,362] and Cas12 variants (dLbCas12a; dAsCas12a; dCas12f)[346,353] were applied. Compact ABEs (miniABEs) were also constructed, containing a single copy of ABE7.10 that was further engineered for optimal on-target and minimal off-target editing efficiencies[345,360]. Of them, miniABE(GG)-NGA and SECURE-ABE were the best performing variants. ABE7.10 was further evolved into several ABE8 variants, including the powerful ABE8e editors that contain a single TadA8e domain that dimerizes with the TadA8e of another ABE8 molecule, presenting robust on-target activity[312,322,333,353,362]. The ABE8e has adopted several Cas variants (NG; RY; SpG; NRTH; NRRH; NRCH; SaCas9; SaKKHCas9; ScCas9; LbCas12a; enCas12a) for recognition of alternative PAMs[312,322,333,353,362]. Finally, ninth generation ABEs (ABE9s) were generated with similar or even higher on-target activity[362]. Interestingly, none of the TadA-derived ABEs showed the sequence context specificity of TadA variants (Fig. 5b).

2.2 TadA-DddA-based ABEs

To enable adenosine deamination of mitochondrial dsDNA, the Tad*8e AD was combined with TALEs and a split DddA CD (for dsDNA unwinding), forming an alternative dipartite ABE called sTALED. The split DddA was later replaced by an inactive DddA mutant, creating either a dimeric (dTALED) or a monomeric (mTALED) editor. Like DdCBEs, editing occurs in the region on the dsDNA between the two TALE-binding sites[363].

2.3 ADAR-based RNA ABEs

The first RNA ABE was developed by fusing the λ-phage N protein-BoxB to ADAR2DD, and, separately, the boxB hairpin to an antisense guide RNA (λN-BoxB)[322]. However, high on-target activity came with the cost of high off-target activity. This limitation was later circumvented using engineered antisense oligonucleotides that recruit endogenous ADAR2 (RESTORE)[365], and short engineered RNAs that recruit endogenous ADAR1 and ADAR2 (LEAPER)[366]. Both ADAR1 and ADAR2 are naturally expressed as homodimers. An engineered hyperactive ADAR2DD variant has been tethered to catalytically inactive Cas13b from *Prevotella* sp. (dPspCas13b), generating the REPAIRv1 system. Further engineering of ADAR2DD did massively decrease the transcriptome-wide off-targeting of

REPAIRv1, creating REPAIRv2. Nevertheless, REPAIRv2 presented lower on-target activity than REPAIRv1[339,367]. Therefore, the REPAIRx system was finally developed, by embedding the engineered ADAR2DD of REPAIRv1into the dPspCas13b protein. REPAIRx exhibited retained the on-target efficiency of REPAIRv1 and the specificity of REPAIRv2[368].

3. Cytidine-to-guanine base-editors (CGBEs)

To expand the base-editing scope beyond transition mutations, UGI-less CBEs were fused to proteins that enhance the formation of secondary products of the cytidine deamination pathway (Fig. 4c). Specifically, cytidine-to-guanine base-editors (CGBEs) catalyse C•G to G•C transversion in eukaryotes, while cytidine-to-adenosine base-editors (CABEs; discussed in the next section) catalyse C•G to A•T transversion in *E. coli*[370]. Six different auxiliary proteins have been implemented in CGBEs to date (Fig. 5c). After initial deamination of C•G to U•G by CD, CGBEs-mediated expression of uracil DNA glycosylase (UNG) or its orthologue from *Mycobacterium smegmatis* (UdgX) enhances uracil excision from the U•G pair to an abasic site (unpaired G)[369,370,372]. DNA polymerase D2 (POLD2) subsequently catalyses lesion bypass polymerization, converting the unpaired G into an unpaired C[369]. As a next step, RNA-binding motif protein X-linked (RBMX) assists abasic site repair, generating the desired G•C pair[369]. Only two CD enzymes (A1; A3A) have been used in the CGBE context to date (Fig. 5c).

3.1 A1-based CGBEs

rA1 was the first CD applied in CGBEs because of the well-studied behaviour of rA1-based CBEs. Two main categories of proteins are tethered to UGI-less CBEs, those concerning base excision initiation (UNG; UdgX) [369,370,372] and those related to base excision repair (POLD2; RBMX; XRCC1; PB)[369,381]. rA1-based CGBEs combining both approaches exhibited the highest editing efficiencies and purities[369]. Engineered nSpCas9 variants have been applied either to broaden the targeting scope (NG; VRQR)[369,372] or to decrease off-target editing (HF1)[369]. CGBEs generally present an exceptionally narrow activity window (Fig. 5c). To this end, the engineered R33A and Anc689 rA1 variants have been used to further narrow or enlarge the editing window, respectively[369,372].

3.2 A3A-based CGBEs

The high catalytic activity of A3A was also harnessed for the development of CGBEs. As such, hA3A-based CGBEs exhibited high editing efficiency and purity, by combining the base excision initiation UdgX enzyme with the base excision repair RBMX enzyme. Native and engineered (NG; HF1) SpCas9 variants were also applied.

4. Cytidine-to-adenosine base-editors (CABEs)

The first steps of C-to-A transversion catalyzed by cytidine-to-adenosine base-editors (CABEs) are similar to those of the aforementioned cytidine-to-guanine base-editors. Hence, upon initial deamination of C•G to U•G by CD, uracil DNA glycosylase catalyzes uracil excision from the U•G pair, resulting in an abasic site (unpaired G)[369,370,372]. Subsequently, in case of CABEs, the unpaired G is cleaved off by AP lyase. X-ray repair cross-complementing protein 1 (XRCC1), a base-excision repair protein, then facilitates the incorporation of G at the abasic site. The formed G•G pair is, finally, converted by XRCC1 or DNA polymerase β (PB) into the desired G•C pair[371]. CABEs are currently under-developed because of their high toxicity in both prokaryotes and eukaryotes. A single CABE has been reported to date (AID-nCas9-UNG), presenting high C•G to A•T transversion efficiency and negligible off-targeting editing in *E. coli*[370] (Fig. 4d, 5d). Interestingly, CABEs have never been successfully applied in eukaryotes.

5. Dual base-editors (DBEs)

DBEs constitute a combination of CBEs with ABEs, and they perform simultaneous C•G to T•A and A•T to G•C transitions on the same protospacer, using a single nCas9 protein (Fig. 4e). Similar to CBEs, DBEs employ highly efficient CD enzymes (A1; A3A; AID; PmCDA1; DddA), and at least one UGI protein for optimal base-editing efficiency and purity. In parallel, both DNA and RNA DBEs adopt some of the most popular AD variants of their corresponding ABEs. Specifically, all DNA DBEs express highly efficient, evolved TadA variants together with eukaryotic AID/APOBEC family members or with the prokaryotic DddA (Fig. 5e).

5.1 A1-TadA-based DBEs

Combination of ABE7.10 with BE3 created the only A1-TadA-based DBE published to date. ACBEmax exhibits small editing windows as well as relatively low on-target and off-target efficiencies compared to other DBEs (such as Target-ACE and Target-ACEmax)[373].

5.2 A3A-TadA-based DBEs

Combination of ABE7.10 with A3A-BE3 created the STEME-1 system, presenting similar efficiencies to ACBEmax but remarkably wider cytidine deamination window, due to its A3A domain[374]. Fusion of ABE8eAW to A3A-Y130F-CBE and replacement of nSpCas9 with nSpCas9-RY generated the CABE-RY editor, which showed shorter editing windows and lower RNA off-targeting than its corresponding single BEs[375].

5.3 AID/PmCDA1-TadA-based DBEs

Tethering of ABE7.10 to Target-AID resulted in ACBE, Target-ACE, and Target-ACEmax. All systems show similar on-target and off-target activities compared to their corresponding single-function BEs[373,376]. Engineering of the linker between the TadA*7.10 and nSpCas9 (ACBE-16N system) further increased the on-target editing efficiency[376]. In addition, combination of miniABEmax(V82G) with Target-AID and an additional copy of UGI created the SPACE editor with slightly reduced activity on adenine, similar activity on cytidine, minimal off-targeting, slightly narrowed editing windows, and similar or improved product purities and indel frequencies compared to their single BEs[377]. Last, fusion of ABE7.10 to AID-BE3 and to an additional copy of UGI generated the A&C-BEmax, with the largest cytidine deaminase editing window reported for DBEs. Similar to SPACE, A&C-BEmax presented slightly lower or similar activity on adenine and cytidine, respectively, as well as remarkably reduced off-target editing than their single BEs[378].

5.4 DddA-TadA-based DBEs

Addition of UGI at the sTALED system enabled dual base-editing on mitochondrial DNA. Shorter cytidine deaminase window was observed for this DBE compared to its cognate ABE system (sTALED without UGI), the product purity was notably high[363].

5.5 ADAR2-based RNA DBEs

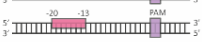
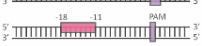
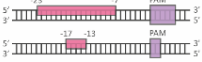
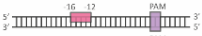
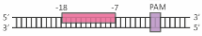
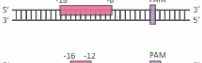
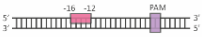
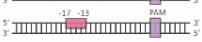
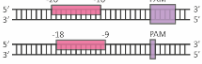
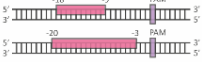
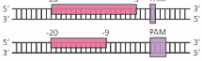
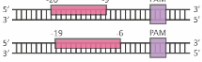
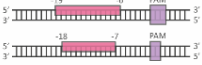
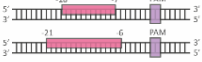
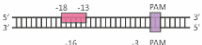
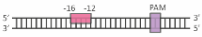
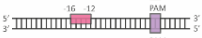
Except from DNA DBEs, RNA DBEs have also been developed. The previously described RESCUE and RESCUE-S systems were further exploited for multiplex C•G to T•A and A•T to G•C RNA editing by providing a pre-crRNA guide array that is programmed to target C and A bases simultaneously[341]. Hence, the relaxed ability of the engineered ADAR2DD variants extends the RNA editing toolbox, conferring novel base-editing approaches.

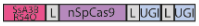
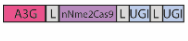
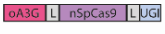
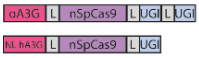
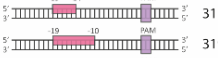
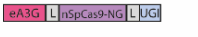
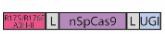
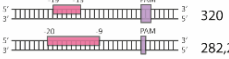
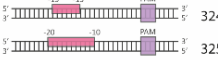
a.

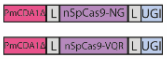
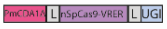
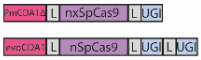
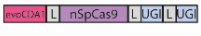
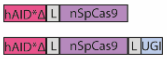
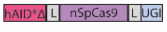
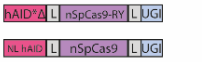
CYTOSINE BASE-EDITORS (CBEs)

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
A1 base-editors					
BE1	A1 L dSpCas9	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	277
BE2	A1 L dSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	277
HF2-BE2	A1 L dSpCas9-HF2 L UGI	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	278
BE3	A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	277
HF-BE3	A1 L nSpCas9-HF1 L UGI	5′-HC-3′	5′-NGG-3′	-19 -9 PAM	277,279
VQR-BE3	A1 L nSpCas9-VQR L UGI	5′-HC-3′	5′-NGAN-3′	-17 -10 PAM	280
VRER-BE3	A1 L nSpCas9-VRER L UGI	5′-HC-3′	5′-NGCG-3′	-18 -13 PAM	280
EQR-BE3	A1 L nSpCas9-EQR L UGI	5′-HC-3′	5′-NGNG-3′	-17 -13 PAM	280
xBE3	A1 L nxSpCas9 L UGI	5′-HC-3′	5′-NG-3′	-19 -10 PAM	281
NG-BE3	A1 L nSpCas9-NG L UGI	5′-HC-3′	5′-NG-3′	-19 -14 PAM	282
Sniper-Cas9-BE3	A1 L nSniperCas9 L UGI	5′-HC-3′	5′-NGG-3′	-18 -13 PAM	283
iSpyMac-BE3	A1 L niSpyMacnCas9 L UGI	5′-HC-3′	5′-NAAN-3′	-19 -10 PAM	284,285
ScBE3	A1 L nScCas9 L UGI	5′-HC-3′	5′-NNG-3′	-17 -12 PAM	286
SaBE3	A1 L nSaCas9 L UGI	5′-HC-3′	5′-NNGRRT-3′	-19 -10 PAM	280
Sa(KKH)BE3	A1 L nSaKKHCas9 L UGI	5′-HC-3′	5′-NNNRRT-3′	-18 -10 PAM	287
BE4	A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-18 -12 PAM	287,288
xCas9-BE4	A1 L nxCas9 L UGI L UGI	5′-HC-3′	5′-NG-3′	-20 -7 PAM	289
NG-BE4	A1 L nSpCas9-NG L UGI L UGI	5′-HC-3′	5′-NG-3′	-17 -7 PAM	290,291
SaBE4	A1 L nSaCas9 L UGI L UGI	5′-HC-3′	5′-NNGRRT-3′	-19 -10 PAM	287
BE4-AmAPOBEC1	AmA1 L nSpCas9 L UGI L UGI	5′-DC-3′	5′-NGG-3′	-17 -13 PAM	292
BE4-PpAPOBEC1	PpA1 L nSpCas9 L UGI L UGI	5′-WC-3′	5′-NGG-3′	-17 -15 PAM	292
BE4-Gam	Gam L A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	287
SaBE4-Gam	Gam L A1 L nSaCas9 L UGI L UGI	5′-HC-3′	5′-NNGRRT-3′	-19 -10 PAM	287
BE4max	A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-18 -12 PAM	288
G-BE4max	A1 L nSpCas9-G L UGI L UGI	5′-HC-3′	5′-NG-3′	-18 -12 PAM	293
RY-BE4max	A1 L nSpCas9-RY L UGI L UGI	5′-HC-3′	5′-NR/Y-3′	-18 -12 PAM	293
NRRH-BE4max	A1 L nSpCas9-NRRH L UGI L UGI	5′-HC-3′	5′-NRRH-3′	-17 -13 PAM	294
NRTH-BE4max	A1 L nSpCas9-NRTH L UGI L UGI	5′-HC-3′	5′-NRTH-3′	-17 -13 PAM	294
NRCH-BE4max	A1 L nSpCas9-NRCH L UGI L UGI	5′-HC-3′	5′-NRCH-3′	-17 -13 PAM	294
SauriBE4max	A1 L nSauriCas9 L UGI L UGI	5′-HC-3′	5′-NNGG-3′	-15 -12 PAM	295
St1BE4max	A1 L nSt1Cas9 L UGI L UGI	5′-HC-3′	5′-NNRGAA-3′	-21 -4 PAM	296
St1BE4max-LMG18311	A1 L nSt1Cas9 LMG18311 L UGI L UGI	5′-HC-3′	5′-NNGCAA-3′	-21 -9 PAM	296
St1BE4max- CNRZ1066	A1 L nSt1Cas9 CNRZ1066 L UGI L UGI	5′-HC-3′	5′-NNACAA-3′	-18 -10 PAM	296
St1BE4max-TH1477	A1 L nSt1Cas9 TH1477 L UGI L UGI	5′-HC-3′	5′-NNGAAA-3′	-19 -9 PAM	296
rA1-dNme2-CBE	A1 L dNme2Cas9 L UGI L UGI	5′-HC-3′	5′-NNNNCC-3′	-20 -10 PAM	297
rA1-nNme2-CBE	A1 L nNme2Cas9 L UGI L UGI	5′-HC-3′	5′-NNNNCC-3′	-20 -10 PAM	297
CP1012-CBEmax	A1 L nSpCas9 CP1012 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -10 PAM	298
CP1028-CBEmax	A1 L nSpCas9 CP1028 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -10 PAM	298
CP1041-CBEmax	A1 L nSpCas9 CP1041 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -11 PAM	298
CP1249-CBEmax	A1 L nSpCas9 CP1249 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -11 PAM	298
BE-RuvCGE	nSpCas9 A1 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-15 -9 PAM	299

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
BE-PLUS	10xGCN4 nSpCas9 ScFv A1 UGI	5′-HC-3′	5′-NGG-3′	-18 -5 PAM	300
BE-PIGS	nSpCas9 A1 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -7 PAM	299
dLbCpf1-BE	A1 L dLbCas12a L UGI	5′-HC-3′	5′-TTTV-3′	PAM +8 +13	301
dLbCas12a-RVRR-BE	A1 L dLbCas12a RVRR L UGI	5′-HC-3′	5′-TNTN-3′	PAM +8 +13	302
impLbCas12a-BE	A1 L impLbCas12a L UGI	5′-HC-3′	5′-TGTV-3′	PAM +8 +13	302
dAsCas12a-BE	A1 L dAsCas12a L UGI	5′-HC-3′	5′-TTTV-3′	PAM +8 +13	301
enAsBE(enAsCas12a)	A1 L enAsCas12a L UGI	5′-HC-3′	5′-NTTV-3′	PAM +8 +13	303
Mutant A1 base-editors					
YE1-BE3	YE1 A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-17 -14 PAM	280
VQR-YE1-BE3	YE1 A1 L nSpCas9-VQR L UGI	5′-HC-3′	5′-NGAN-3′	-16 -13 PAM	280
YE1-Sa(KKH)-BE3	YE1 A1 L SaKKHCas9 L UGI	5′-HC-3′	5′-NNNRRT-3′	-18 -10 PAM	287
YE1-BE3-nNme2Cas9	YE1 A1 L nNme2Cas9 L UGI	5′-HC-3′	5′-NNNNCC-3′	-23 -13 PAM	304
YE1-BE4	YE1 A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -10 PAM	290
YE1-BE4-NG	YE1 A1 L nSpCas9-NG L UGI L UGI	5′-HC-3′	5′-NG-3′	-16 -13 PAM	290
YE1-BE4-CP1028	YE1 A1 L nSpCas9 CP1028 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -9 PAM	290
YE1-BE4-Gam	Gam L YE1 A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-18 -15 PAM	291
YE1-BE4max	YE1 A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-17 -15 PAM	291
YE2-BE3	YE2 A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-16 -15 PAM	280
YE2-BE4-NG	YE2 A1 L nSpCas9-NG L UGI L UGI	5′-HC-3′	5′-NG-3′	-16 -13 PAM	290
YE2-BE4-CP1028	YE1 A1 L nSpCas9 CP1028 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-16 -10 PAM	290
YEE-BE3	YEE A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-16 -15 PAM	280
YEE-BE4-NG	YEE A1 L nSpCas9-NG L UGI L UGI	5′-HC-3′	5′-NG-3′	-16 -14 PAM	290
YEE-BE4-CP1028	YEE A1 L nSpCas9 CP1028 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-16 -10 PAM	290
YEE-BE4-Gam	Gam L YEE A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-16 -15 PAM	291
YEE-BE4max	YEE A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-16 -15 PAM	291
EE-BE3	EE A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-16 -15 PAM	280
VQR-EE-BE3	EE A1 L nSpCas9-VQR L UGI	5′-HC-3′	5′-NGAN-3′	-16 -13 PAM	280
EE-Sa(KKH)-BE3	EE A1 L SaKKHCas9 L UGI	5′-HC-3′	5′-NNNRRT-3′	-18 -10 PAM	287
EE-BE4-NG	EE A1 L nSpCas9-NG L UGI L UGI	5′-HC-3′	5′-NG-3′	-16 -13 PAM	290
EE-BE4-CP1028	EE A1 L nSpCas9 CP1028 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-16 -10 PAM	290
FE1-BE3	FE1 A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-16 -14 PAM	280
RA-BE3	RA A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	305
YFE-BE4max	YFE A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-16 -15 PAM	291
AALN-BE4	AALN A1 L nSpCas9 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-10 -8 PAM	290
BE3-R33A (SECURE)	R33A A1 L nSpCas9 L UGI	5′-HC-3′	5′-NGG-3′	-17 -13 PAM	306
BE3-R33A/K34A (SECURE)	R33A+K34A A1 L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′	-17 -13 PAM	306
R33A+K34A-BE4	R33A+K34A A1 L nSpCas9 L UGI L UGI	5′-TC-3′	5′-NGG-3′	-13 -9 PAM	290,292
R33A+K34A-BE4-NG	R33A+K34A A1 L nSpCas9-NG L UGI L UGI	5′-HC-3′	5′-NG-3′	-16 -14 PAM	290
R33A+K34A-BE4-CP1028	R33A+K34A A1 L nSpCas9 CP1028 L UGI L UGI	5′-HC-3′	5′-NGG-3′	-10 PAM	290
evoBE4max	evoA1 L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′	-18 -13 PAM	307
PevorAC1-NG	evoA1 L nSpCas9-NG L UGI L UGI	5′-BC-3′	5′-NG-3′	-18 -11 PAM	308
Anc689BE4max-SpCas9	ancA1 L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′	-17 -6 PAM	309
Anc689BE4max-NG	ancA1 L nSpCas9-NG L UGI	5′-TC-3′	5′-NG-3′	-17 -6 PAM	309
AncBE4max	ancA1 L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′	-18 -12 PAM	288,307

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
FERNY-BE4max	FERNY A1 L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′		307
evoFERNY-BE4max	evoFERNY A1 L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′		307
PevoFERNY-NG	evoFERNY L nSpCas9-NG L UGI L UGI	5′-BC-3′	5′-NG-3′		308
YE1-nNme2-CBE	YE1 A1 L nNme2Cas9 L UGI L UGI	5′-NC-3′	5′-NNNNCC-3′		297
eA1-nNme2-CBE	eA1 L nNme2Cas9 L UGI L UGI	5′-NC-3′	5′-NNNNCC-3′		297
BE4-PpAPOBEC1 R33A	PpA1 R33A L nSpCas9 L UGI L UGI	5′-TC-3′	5′-NGG-3′		292
BE4-PpAPOBEC1 H122A	PpA1 H122A L nSpCas9 L UGI L UGI	5′-KC-3′	5′-NGG-3′		292
A3A base-editors					
A3A-BE3	A3A L nSpCas9 L UGI	5′-KC-3′	5′-NGG-3′		310
A3A-BE4max	A3A L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′		307
hA3A-BE3	hA3A L nSpCas9 L UGI	5′-GC-3′	5′-NGG-3′		310,311
PhA3A-max-NG	hA3A L nSpCas9-NG L UGI L UGI	5′-BC-3′	5′-NG-3′		308
Mutant A3A base-editors					
eA3A(N57G)-BE3	eA3A L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′		310
eA3A-BE3(xCas9)	eA3A L nxCas9 L UGI	5′-TC-3′	5′-NG-3′		310
eA3A-BE3(VRQR)	eA3A L nSpCas9-VRQR L UGI	5′-TC-3′	5′-VRQR-3′		310
eA3A-HF1-BE3-2xUGI	eA3A L nSpCas9-HF1 L UGI L UGI	5′-TC-3′	5′-NGG-3′		310
eA3A-HypaBE3-2xUGI	eA3A L nHypaCas9 L UGI L UGI	5′-TC-3′	5′-NGG-3′		310
eA3A(N57G)-BE4	eA3A L nSpCas9 L UGI L UGI	5′-TC-3′	5′-NGG-3′		310
eA3A-BE4max	eA3A L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′		307
eA3A-nNme2-CBE	eA3A L nNme2Cas9 L UGI L UGI	5′-NC-3′	5′-NNNNCC-3′		297
eA3A-SpRY	eA3A L nSpCas9-RY L UGI L UGI L UGI L UGI	5′-TC-3′	5′-NR/Y-3′		312
eA3A-SpG	eA3A L nSpCas9-G L UGI L UGI L UGI L UGI	5′-TC-3′	5′-NG-3′		312
eA3A-NRRH	eA3A L nSpCas9-NRRH L UGI L UGI L UGI L UGI	5′-TC-3′	5′-NRRH-3′		312
eA3A-NRCH	eA3A L nSpCas9-NRCH L UGI L UGI L UGI L UGI	5′-TCR-3′	5′-NRCH-3′		312
A3A-W98Y-BE4max	W98Y A3A L nSpCas9 L UGI L UGI	5′-NC-3′	5′-NGG-3′		307
hA3A-NL-BE3	NL A3A L nSpCas9 L UGI	5′-NC-3′	5′-NGG-3′		311
hA3A-BE3-R128A	R128A hA3A L nSpCas9 L UGI	5′-NC-3′	5′-NGG-3′		311,313
hA3A-eBE3-Y130F	Y130F A3A L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′		314
A3A-Y130F-CBE	Y130F A3A L nSpCas9 L UGI L UGI	5′-TC-3′	5′-NGG-3′		315,316
hA3A-eBE3-Y132D	Y132D A3A L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′		314
A3A(N57A)-BE3	N57A A3A L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′		310
A3A QF(N57Q/Y130F)-BE3	QF A3A L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′		310
A3AΔ182-BE3	A3AΔ182 L nSpCas9 L UGI	5′-NC-3′	5′-NGG-3′		311
A3AΔ186-BE3	A3AΔ186 L nSpCas9 L UGI	5′-NC-3′	5′-NGG-3′		311
hA3A(Y130F)Δ186-BE3	hA3A Y130F Δ186 L nSpCas9 L UGI	5′-NC-3′	5′-NGG-3′		311
BEACON1	hA3A W104A Y132D L dAsCas12a L UGI	5′-HC-3′	5′-TTTV-3′		301
BEACON2	hA3A W90Y W104A,Y130F L denAsCas12a L UGI	5′-HC-3′	5′-NTTV-3′		303
A3B base-editors					
A3Bctd-BE3	A3Bctd L nSpCas9 L UGI	5′-TC-3′	5′-NGG-3′		317
BE4-SsAPOBEC3B	SsA3B L nSpCas9 L UGI L UGI	5′-BC-3′	5′-NGG-3′		292
Mutant A3B base-editors					
hA3B-NL-BE3	NL A3B L nSpCas9 L UGI	5′-NC-3′	5′-NGG-3′		311
A3Bctd-VHM-BE3	A3Bctd VHM L nSpCas9 L UGI	5′-NCN-3′	5′-NGG-3′	*Not determined*	318

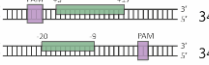
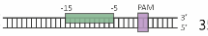
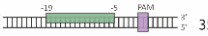
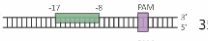
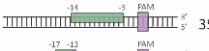
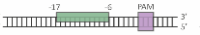
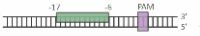
Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
A3Bctd-KKR-BE3		5′-NCN-3′	5′-NGG-3′	*Not determined*	318
BE4-SsAPOBEC3B R54Q		5′-BC-3′	5′-NGG-3′		292
A3F base-editors					
BE4-RrA3F		5′-DC-3′	5′-NGG-3′		292
Mutant A3F base-editors					
BE4-RrA3F F130L		5′-BC-3′	5′-NGG-3′		292
A3G base-editors					
A3G-BE3		5′-CC-3′	5′-NGG-3′		314,317,319
A3G-BE4max		5′-CC-3′	5′-NGG-3′		319
A3G-nNme2-CBE		5′-NC-3′	5′-NNNNCC-3′		297
Mutant A3G base-editors					
oA3G-BE3		5′-(C)CC-3′	5′-NGG-3′		319
oA3G-BE4max		5′-TC-3′	5′-NGG-3′		319
hA3G-NL-BE3		5′-NC-3′	5′-NGG-3′		311
A3G(D316R/D317R)-BE4max		5′-CC-3′	5′-NGG-3′		319
eA3G-BE4max		5′-CC-3′	5′-NGG-3′		291
eA3Gctd-nSpyCas9-NG		5′-CC-3′	5′-NG-3′		319
A3H base-editors					
A3H-II-BE3		5′-TC-3′	5′-NGG-3′		317
Mutant A3H base-editors					
A3H-II(R175/R176E)-BE3		5′-TC-3′	5′-NGG-3′		317
AID/PmCDA1 base-editors					
Target-AID		5′-WRC-3′	5′-NGG-3′		320
TargetAID-NG		5′-WRC-3′	5′-NG-3′		282,289,321
TargetAID-SpRY		5′-WRC-3′	5′-NR/Y-3′		322
TargetAID-VQR		5′-WRC-3′	5′-NGAN-3′		323
TargetAID-iSpyMacCas9		5′-WRC-3′	5′-NAAN-3′		324
TargetAID-SaCas9		5′-WRC-3′	5′-NNGRRT-3′		325,326
TargetAID-SaKKH		5′-WRC-3′	5′-NNNRRT-3′		325
SpyCas9-HF2-pBE		5′-WRC-3′	5′-NGG-3′		327
eSpyCas9(1.1)-pBE		5′-WRC-3′	5′-NGG-3′		327
HypaCas9-pBE		5′-WRC-3′	5′-NGG-3′		327
cCDA1-BE3		5′-WRC-3′	5′-NGG-3′		311
cCDA1-NGBE3		5′-WRC-3′	5′-NG-3′		311
nCDA1-BE3		5′-WRC-3′	5′-NGG-3′		311
nCDA1-NGBE3		5′-WRC-3′	5′-NG-3′		282,311
nCDA1-VQR-BE3		5′-WRC-3′	5′-NGAN-3′		311,323
nCDA1-VRERBE3		5′-WRC-3′	5′-NGCG-3′		311
nCDA1-xBE3		5′-WRC-3′	5′-NG-3′		311
CDA1-BE4max		5′-WRC-3′	5′-NGG-3′		307
AID-BE3		5′-WRC-3′	5′-NGG-3′		328
Mutant AID/ PmCDA1 base-editors					
AID-2S		5′-WRC-3′	5′-NGG-3′		326
Sa-AID-2S		5′-WRC-3′	5′-NNGRRT-3′		326
AID-3S		5′-WRC-3′	5′-NGG-3′		326
SaAID-3S		5′-WRC-3′	5′-NGG-3′		326
nCDA1Δ188-NGBE3		5′-NC-3′	5′-NG-3′		311
nCDA1Δ190-NGBE3		5′-NC-3′	5′-NG-3′		311
nCDA1Δ192-NGBE3		5′-NC-3′	5′-NG-3′		311

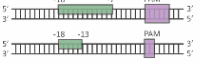
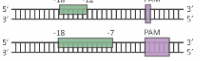
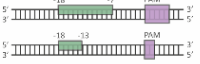
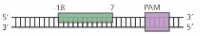
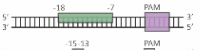
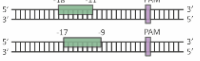
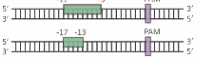
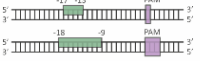
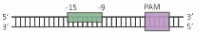
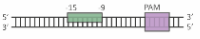
Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
nCDA1Δ193-NGBE3		5′-NC-3′	5′-NG-3′		311
nCDA1Δ194-NGBE3		5′-NC-3′	5′-NG-3′		311
nCDA1Δ-NGBE3		5′-NC-3′	5′-NG-3′		311
nCDA1Δ-VQR-BE3		5′-NC-3′	5′-NGAN-3′		311
nCDA1Δ-VRERBE3		5′-NC-3′	5′-NGCG-3′		311
nCDA1Δ-xBE3		5′-NC-3′	5′-NG-3′		311
evoCDA1-BE4max (CDA F23S/A123V/I195F)		5′-NC-3′	5′-NGG-3′		307
PevoCDA1-NG		5′-BC-3′	5′-NG-3′		308
PevoCDA1-eNG		5′-BC-3′	5′-NG-3′		308
eCDA		5′-WRC-3′	5′-NGG-3′		329
AIDmono		5′-WRC-3′	5′-NGG-3′		330
hAID*Δ-SpCas9 (rBE5)		5′-AC-3′	5′-NGG-3′		331
hAID*Δ-SpCas9 (rBE9)		5′-MC-3′	5′-NGG-3′		331
hAID*Δ-ScCas9		5′-WRC-3′	5′-NNG-3′		332
hAID*Δ-SpRY		5′-WRC-3′	5′-NR/Y-3′		333
hAID-NL-BE3		5′-NC-3′	5′-NGG-3′		311
mAID-NL-BE3		5′-NC-3′	5′-NGG-3′		311
cAICDA-NL-BE3		5′-NC-3′	5′-NGG-3′		311
eAID-BE3		5′-NC-3′	5′-NGG-3′		328
eAID-BE4max (K10E/T82I/E156G/P182X)		5′-NC-3′	5′-NGG-3′		328
eAID(N51G)-BE4max		5′-NC-3′	5′-NGG-3′		328
eAID-NG-BE4max		5′-NC-3′	5′-NG-3′		328
eAID-nNme2-CBE		5′-NC-3′	5′-NNNNCC-3′		297
AIDx (dCas9-AID-P182X)		5′-WGC-3′	5′-NGG-3′		334,335
AIDx-nCas9-UGI(TAM)		5′-WGC-3′	5′-NGG-3′		334
AIDx-nCas9(SaKKH)-UGI(TAM)		5′-WGC-3′	5′-NNNRRT-3′		334,335
CRISPR-X		5′-WRC-3′	5′-NGG-3′		328
DddA base-editors					
DdCBE (Aureus C-orientation) (nuclear DNA base-editing)		5′-TC-3′	5′-NGG-3′ and 5′-NNNRRT-3′	*(DNA base-editing in the region between the two protospacers)*	336
DdCBE (Aureus N-orientation) (nuclear DNA base-editing)		5′-TC-3′	5′-NGG-3′ and 5′-NNNRRT-3′	*(DNA base-editing in the region between the two protospacers)*	336
DdCBE (nuclear DNA base-editing)		5′-TC-3′	-		336
DdCBE (mitochondrial DNA base-editing)		5′-TC-3′	-		336
cp-DdCBE (chloroplast DNA base-editing)		5′-TC-3′	-		337-339
Mutant DddA base-editors					
T7-DdCBE-DddA6 (nuclear DNA base-editing)		5′-TC-3′	-		340
T7-DdCBE-DddA11 (nuclear DNA base-editing)		5′-HC-3′	-		340
DdCBE-DddA6 (mitochondrial DNA base-editing)		5′-TC-3′	-		340
DdCBE-DddA11 (mitochondrial DNA base-editing)		5′-TC-3′	-		340
Mutant Adar2 RNA base-editors					
RESCUEr16		-	-	*(RNA base-editing)*	341
RESCUE-S		-	-	*(RNA base-editing)*	341

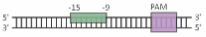
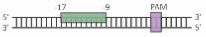
b.

ADENINE BASE-EDITORS (ABEs)

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
TadA base-editors					
ABE0.1	TadA L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
Mutant TadA base-editors					
ABE1.2	TadA*1.2 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE2.1	TadA*2.1 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE2.9	TadA*2.1 L TadA*2.1 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE3.1	TadA*3.1 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE4.3	TadA*3.1 L TadA*3.1 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE5.1	TadA*5.1 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE5.3	TadA L TadA L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE6.3	TadA L TadA*6.3 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE6.4	TadA L TadA*6.4 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE7.4	TadA L TadA*7.4 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE7.9	TadA L TadA*7.9 L nSpCas9	None	5′-NGG-3′	-17 -12 PAM	342
ABE7.10	TadA L TadA*7.10 L nSpCas9	None	5′-NGG-3′	-17 -14 PAM	342
ABE7.10-xCas9 (xABE)	TadA L TadA*7.10 L nxSpCas9	None	5′-NG-3′	-17 -14 PAM	281
ABE7.10-CP1028	TadA L TadA*7.10 L nSpCas9 CP1028	None	5′-NGG-3′	-17 9 PAM	298
ABE7.10-CP1041	TadA L TadA*7.10 L nSpCas9 CP1041	None	5′-NGG-3′	-17 9 PAM	298
ABE7.10-HF1	TadA L TadA*7.10 L nxSpCas9	None	5′-NGG-3′	-17 -14 PAM	343
ABE7.10-HF2	TadA L TadA*7.10 L nSpCas9 CP1041	None	5′-NGG-3′	-17 -14 PAM	343
ABE7.10-eCas9	TadA L TadA*7.10 L neSpCas9	None	5′-NGG-3′	-17 -14 PAM	343
ABE7.10-HypaCas9	TadA L TadA*7.10 L nHypaCas9	None	5′-NGG-3′	-17 -14 PAM	343
ABE7.10-SniperCas9	TadA L TadA*7.10 L nSniperCas9	None	5′-NGG-3′	-17 -14 PAM	344
ABE7.10-evoCas9	TadA L TadA*7.10 L nevoCas9	None	5′-NGG-3′	-17 -14 PAM	343
ABE7.10-G (SpG-ABE)	TadA L TadA*7.10 L nSpCas9-G	None	5′-NG-3′	-17 -14 PAM	293
ABE7.10-NG	TadA L TadA*7.10 L nSpCas9-NG	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGA	TadA L TadA*7.10 L nSpCas9-NGA	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGC	TadA L TadA*7.10 L nSpCas9-NGC	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGC-loop	TadA L TadA*7.10 L nSpCas9-NGC*	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGX	TadA L TadA*7.10 L nSpCas9-NGX	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGX-NGA	TadA L TadA*7.10 L nSpCas9-NGX NGA	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGX-NGC	TadA L TadA*7.10 L nSpCas9-NGX NGC	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-NGX-loop	TadA L TadA*7.10 L nSpCas9-NGX*	None	5′-NG-3′	-17 -15 PAM	345
ABE7.10-RY (SpRY-ABE)	TadA L TadA*7.10 L nSpCas9-RY	None	5′-NR/Y-3′	-17 -14 PAM	293,322
ABE7.10-VQR (VQR-ABE)	TadA L TadA*7.10 L nSpCas9-VQR	None	5′-NGA-3′	-17 -13 PAM	280,347-349
ABE7.10-VRER (VRER-ABE)	TadA L TadA*7.10 L nSpCas9-VRER	None	5′-NGCG-3′	-17 -13 PAM	280,347,349
ABE7.10-EQR (EQR-ABE)	TadA L TadA*7.10 L nSpCas9-EQR	None	5′-NGNG-3′	-17 -14 PAM	280,349
ABE7.10-VRQR (VRQR-ABE)	TadA L TadA*7.10 L nSpCas9-VRQR	None	5′-NGAH-3′	-17 -14 PAM	280,349
ABE7.10-NRRH (NRRH-ABE)	TadA L TadA*7.10 L nSpCas9-NRRH	None	5′-NRRH-3′	-17 -14 PAM	294
ABE7.10-NRTH (NRTH-ABE)	TadA L TadA*7.10 L nSpCas9-NRTH	None	5′-NRTH-3′	-17 -14 PAM	294
ABE7.10-NRCH (NRCH-ABE)	TadA L TadA*7.10 L nSpCas9-NRCH	None	5′-NRCH-3′	-17 -14 PAM	294
ABE7.10-iSpyMacCas9	TadA L TadA*7.10 L niSpMacCas9	None	5′-TAAA-3′	-12 PAM	324

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
ABE7.10-SaCas9 (ABESa)	TadA L TadA*7.10 L nSaCas9	None	5′-NNGRRT-3′		347,350
ABE7.10-SaKKH (SaKKH-ABE)	TadA L TadA*7.10 L nSaKKHCas9	None	5′-NNNRRT-3′		347,348
ABE7.10-ScCas9	TadA L TadA*7.10 L nScCas9	None	5′-NNG-3′	Not determined	332,345
ABE7.10-CjCas9 (CjABE)	TadA L TadA*7.10 L nCjCas9	None	5′-NNNNRYAC-3′		351,352
ABE7.10-LbCas12a	TadA L TadA*7.10 L dLbCas12a	None	5′-TTTV-3′		353
ABE7.10-AsCas12a	TadA L TadA*7.10 L dAsCas12a	None	5′-TTTV-3′		353
ABE7.10-Cas12f (engineered) (dCasMINI-ABE)	TadA L TadA*7.10 L dCasMINI	None	5′-TTTR-3′		346
ABE7.10-S (miniABE)	TadA*7.10 L nSpCas9	None	5′-NGG-3′		345
MiniABE-NG	TadA*7.10 L nSpCas9-NG	None	5′-NG-3′		345
MiniABE(V82G)-NG	TadA*7.10 V82G L nSpCas9-NG	None	5′-NG-3′		345
MiniABE(GG)-NG	TadA*7.10 GG L nSpCas9-NG	None	5′-NG-3′		345
MiniABE(A56G)-NG	TadA*7.10 A56G L nSpCas9-NG	None	5′-NG-3′		345
MiniABE(GG)-NGA	TadA*7.10 GG L nSpCas9-NGA	None	5′-NGA-3′		345
IBE02	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE10	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE11	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE12	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE12.0	nSpCas9 TadA	None	5′-NGC-3′	Not determined	354
IBE12.1	nSpCas9 TadA*8.5	None	5′-NGC-3′	Not determined	354
IBE12.2	nSpCas9 TadA* V82A/Y147T/Q154S	None	5′-NGC-3′	Not determined	354
IBE12.3	nSpCas9 TadA* I76Y/V82S/Y147T/Q154S	None	5′-NGC-3′	Not determined	354
IBE12.4	nSpCas9 TadA* V82A/Y147R/Q154R	None	5′-NGC-3′	Not determined	354
IBE12.5	nSpCas9 TadA* I76Y/V82S/Y147R/Q154R	None	5′-NGC-3′	Not determined	354
IBE13	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE15	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE16	TadA L nSpCas9 TadA*7.10	None	5′-NGG-3′		354
IBE23	TadA L nSpCas9 dRNH TadA*7.10	None	5′-NGG-3′		354
ABE7.10max	TadA L TadA*7.10 L nSpCas9	None	5′-NGG-3′		355
xABEmax	TadA L TadA*7.10 L nxSpCas9	None	5′-NG-3′		298
ABEmax-NG	TadA L TadA*7.10 L nSpCas9-NG	None	5′-NG-3′		356
ABEmax-VRQR	TadA L TadA*7.10 L nSpCas9-VRQR	None	5′-NGAH-3′		298
ABEmax-VRER	TadA L TadA*7.10 L nSpCas9-VRER	None	5′-NGCG-3′		298
ABEmax-CP1012	TadA L TadA*7.10 L nSpCas9 CP1012	None	5′-NGG-3′		298
ABEmax-CP1028	TadA L TadA*7.10 L nSpCas9 CP1028	None	5′-NGG-3′		298
ABEmax-CP1041	TadA L TadA*7.10 L nSpCas9 CP1041	None	5′-NGG-3′		298
ABEmax-CP1249	TadA L TadA*7.10 L nSpCas9 CP1249	None	5′-NGG-3′		298
SaABEmax	TadA L TadA*7.10 L nSaCas9	None	5′-NNGRRT-3′		298
SaKKH-ABEmax	TadA L TadA*7.10 L nSaKKHCas9	None	5′-NNNRRT-3′		298
SpyMacABEmax	TadA L TadA*7.10 L nSpMacCas9	None	5′-NAAR-3′		284,324,357

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
ABEmaxSC	TadA L tadA*7.10 L nScCas9	None	5′-NNG-3′		345
SauriABEmax	TadA L tadA*7.10 L nSauriCas9	None	5′-NNGG-3′		295
St1ABEmax	TadA L tadA*7.10 L nSt1Cas9	None	5′-NNRGAA-3′		296,358
ABEmaxAW	TadA L tadA*7.10 V106W L nSpCas9	None	5′-NGG-3′		359
miniABEmax	TadA*7.10 L nSpCas9	None	5′-NGG-3′		360
miniABEmax(K20A/R21A) [SECURE-ABE]	TadA*7.10 K20A/R21A L nSpCas9	None	5′-NGG-3′		360
miniABEmax(V82G)	TadA*7.10 V82G L nSpCas9	None	5′-NGG-3′		360
ABE7.10(D53E)	TadA L tadA*7.10 Y123E L nSpCas9	None	5′-NGG-3′		313
ABE7.10(F148A)	TadA L tadA*7.10 F148A L nSpCas9	None	5′-NGG-3′		313
ABE7.10(E50A,V106W)	TadA L tadA*7.10 E50A V106W L nSpCas9	None	5′-NGG-3′		313
dABE8.8	TadA*8.8 L dSpCas9	None	5′-NGG-3′		361
ABE8.8	TadA*8.8 L nSpCas9	None	5′-NGG-3′		361
ABE8.8-NG	TadA*8.8 L nSpCas9-NG	None	5′-NG-3′		361
ABE8.8-SaCas9	TadA*8.8 L nSaCas9	None	5′-NNGRRT-3′		361
dABE8.13	TadA*8.13 L dSpCas9	None	5′-NGG-3′		361
ABE8.13	TadA*8.13 L nSpCas9	None	5′-NGG-3′		361
ABE8.13-NG	TadA*8.13 L nSpCas9-NG	None	5′-NG-3′		361
ABE8.13-SaCas9	TadA*8.13 L nSaCas9	None	5′-NNGRRT-3′		361
dABE8.17	TadA*8.17 L dSpCas9	None	5′-NGG-3′		361
ABE8.17	TadA*8.17 L nSpCas9	None	5′-NGG-3′		361,362
ABE8.17-NG	TadA*8.17 L nSpCas9-NG	None	5′-NG-3′		361,362
ABE8.17-SaCas9	TadA*8.17 L nSaCas9	None	5′-NNGRRT-3′		361
ABE8.17-ScCas9	TadA*8.17 L nScCas9	None	5′-NNG-3′		362
ABE8.17(V106W)	TadA*8.17 V106W L nSpCas9	None	5′-NGG-3′		361
dABE8.20	TadA*8.20 L dSpCas9	None	5′-NGG-3′		361
ABE8.20	TadA*8.20 L nSpCas9	None	5′-NGG-3′		361,362
ABE8.20-NG	TadA*8.20 L nSpCas9-NG	None	5′-NG-3′		361,362
ABE8.20-SaCas9	TadA*8.20 L nSaCas9	None	5′-NNGRRT-3′		361
ABE8.20-ScCas9	TadA*8.20 L nScCas9	None	5′-NNG-3′		362
ABE8e	TadA*8e L nSpCas9	None	5′-NGG-3′		353
ABE8e-CP1028	TadA*8e L nSpCas9 CP1028	None	5′-NGG-3′		353
ABE8e-CP1041	TadA*8e L nSpCas9 CP1041	None	5′-NGG-3′		353
ABE8e-SpRY	TadA*8e L nSpCas9-RY	None	5′-NR/Y-3′		322,333
ABE8e-SpG	TadA*8e L nSpCas9-G	None	5′-NG-3′		312
ABE8e-NG	TadA*8e L nSpCas9-NG	None	5′-NG-3′		353,362
ABE8e-NRTH	TadA*8e L nSpCas9 NRTH	None	5′-NRTH-3′		312
ABE8e-NRRH	TadA*8e L nSpCas9 NRTH	None	5′-NRRH-3′		312
ABE8e-NRCH	TadA*8e L nSpCas9 NRTH	None	5′-NRCH-3′		312
ABE8e-SaCas9	TadA*8e L nSaCas9	None	5′-NNGRRT-3′		353
ABE8e-SaKKHCas9	TadA*8e L nSaKKHCas9	None	5′-NNNRRT-3′		353
ABE8e-ScCas9	TadA*8e L nScCas9	None	5′-NNG-3′		362

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
ABE8e-LbCas12a	TadA*7.10 L dLbCas12a	None	5′-TTTV-3′		353
ABE8e-enAsCas12a	TadA*7.10 L denAsCas12a	None	5′-TTYN-3′		353
ABE8eAW	TadA*8e V106W L nSpCas9	None	5′-NGG-3′		353
ABE8eAW-SaCas9	TadA*8e V106W L nSaCas9	None	5′-NNGRRT-3′		353
ABE8eAW-LbCas12a	TadA*8e V106W L dLbCas12a	None	5′-TTTV-3′		353
ABE9	TadA*9 L nSpCas9	None	5′-NGG-3′		362
ABE9-NG	TadA*9 L nSpCas9-NG	None	5′-NG-3′		362
ABE9-RY	TadA*9 L nSpCas9-RY	None	5′-NR/Y-3′		362
ABE9-ScCas9	TadA*9 L nScCas9	None	5′-NNG-3′		362
Mutant TadA-DddA base-editors					
sTALED without UGI (mitochondrial DNA base-editing)	Left TALE array … Split DddA L TadA*8e Split DddA L … Right TALE array	None	-		363
Mutant TadA-mutant DddA base-editors					
dTALED (mitochondrial DNA base-editing)	Left TALE array … L Inactive DddA E1347A TadA*8e L … Right TALE array	None	-		363
mTALED (mitochondrial DNA base-editing)	Left TALE array … L TadA*8e L Inactive DddA E1347A	None	-		363
ADAR base-editors					
λN-BoxB	λN ADAR2DD	None	-	*(RNA base-editing)*	322
RESTORE	Engineered antisense oligonucleotides that recruite ADARs	-	-	*(RNA base-editing)*	365
LEAPER	Engineered antisense RNA that recruites ADARs	-	-	*(RNA base-editing)*	366
Mutant ADAR2 base-editors					
REPAIRv1	dPspCas13b L ADAR2DD E488Q	None	-	*(RNA base-editing)*	339,367
REPAIRv2	dPspCas13b L ADAR2DD E488Q T375G	None	-	*(RNA base-editing)*	339,367
REPAIRx	ADAR2DD E488Q Split dCas13b	None	-	*(RNA base-editing)*	368

c.

C-TO-G BASE-EDITORS (CGBEs)

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
A1 base-editors					
APOBEC1-nCas9	A1 L nSpCas9	5′-RCTA-3′	5′-NGG-3′		369
APOBEC1-nCas9-NG	A1 L nSpCas9-NG	5′-RCTA-3′	5′-NG-3′		369
APOBEC-nCas9-Ung	A1 L nSpCas9 L UNG	5′-HC-3′	5′-NGG-3′		370
ACX, rPB(8kD)	A1 L nSpCas9 L PB	5′-WCW-3′ 5′-ACC-3′ 5′-GCT-3′	5′-NGG-3′		371
ACX, rXRCC1	A1 L nSpCas9 L XRCC1	5′-WCW-3′ 5′-ACC-3′ 5′-GCT-3′	5′-NGG-3′		371
POLD2-APOBEC1-UdgX-nCas9-UdgX	POLD2 L A1 L UdgX L nSpCas9 L UdgX	5′-RCTA-3′	5′-NGG-3′		369
UdgX-APOBEC1-UdgX-HF-nCas9	UdgX L A1 L UdgX L nSpCas9-HF1	5′-RCTA-3′	5′-NGG-3′		369
UdgX-APOBEC1-UdgX-nCas9-NG	UdgX L A1 L UdgX L nSpCas9-NG	5′-RCTA-3′	5′-NGG-3′		369
Mutant A1 base-editors					
EE-nCas9	EE A1 L nSpCas9	5′-RCTA-3′	5′-NGG-3′		369
UdgX-EE-UdgX-nCas9-UdgX	UdgX L EE A1 L UdgX L nSpCas9 L UdgX	5′-TCT-3′	5′-NGG-3′		369
CGBE1	UNG L R33A A1 L nSpCas9	5′-HC-3′	5′-NGG-3′		372
CGBE1-NG	UNG L R33A A1 L nSpCas9-NG	5′-HC-3′	5′-NG-3′		372
CGBE1-VRQR	UNG L R33A A1 L nSpCas9-VRQR	5′-HC-3′	5′-NGA-3′		372
miniCGBE1	R33A A1 L nSpCas9	5′-HC-3′	5′-NGG-3′		372
miniCGBE1-NG	R33A A1 L nSpCas9-NG	5′-HC-3′	5′-NG-3′		372
miniCGBE1-VRQR	R33A A1 L nSpCas9-VRQR	5′-HC-3′	5′-NGA-3′		372
Anc689-nCas9	Anc689 A1 L nSpCas9	5′-CTA-3′	5′-NGG-3′		369
Anc689-nCas9-NG	Anc689 A1 L nSpCas9-NG	5′-CTA-3′	5′-NG-3′		369
UdgX-Anc689-UdgX-nCas9-RBMX	UdgX L Anc689 A1 L UdgX L nSpCas9 L RBMX	5′-RCTA-3′	5′-NGG-3′		369
UdgX-Anc689-UdgX-nCas9-NG-RBMX	UdgX L Anc689 A1 L UdgX L nSpCas9-NG L RBMX	5′-RCTA-3′	5′-NGG-3′		369
Mutant A3A base-editors					
eA3A-nCas9	eA3A L nSpCas9	5′-TC-3′	5′-NGG-3′		369
eA3A-nCas9-NG	eA3A L nSpCas9-NG	5′-TC-3′	5′-NG-3′		369
RBMX-eA3A-UdgX-nCas9	RBMX L eA3A L UdgX L nSpCas9	5′-CTC-3′	5′-NGG-3′		369
RBMX-eA3A-UdgX-HF-nCas9	RBMX L eA3A L UdgX L nSpCas9-HF1	5′-RCTA-3′	5′-NGG-3′		369

d.

C-TO-A BASE-EDITORS (CABEs)

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
AID base-editors					
AID-nCas9-Ung	hAID L nSpCas9 L UNG	Not determined	5′-NGG-3′		370

e.

DUAL BASE-EDITORS (DBEs)

Base-editor	Architecture	Deaminase specificity	PAM specificity	Editing window	References
A1 - mutant TadA dual base-editors					
ACBEmax	TadA L TadA*7.10 L nSpCas9 L A1 L UGI	Not determined	5′-NGG-3′		373
A3A - mutant TadA dual base-editors					
STEME-1	A3A L TadA L TadA*7.10 L nSpCas9 L UGI	Not determined	5′-NGG-3′		374
STEME-NG	A3A L TadA L TadA*7.10 L nSpCas9-NG L UGI	Not determined	5′-NG-3′		374

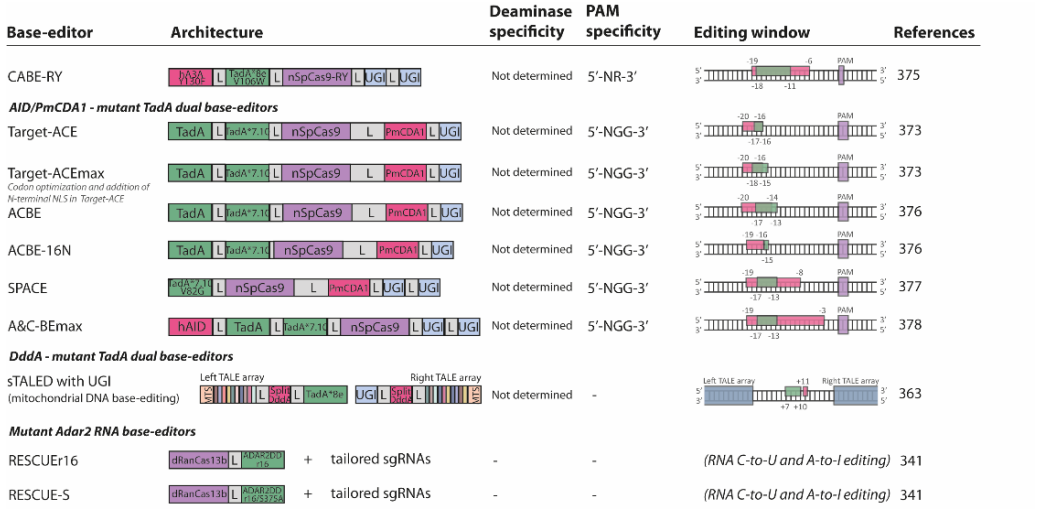


Fig. 5 Classification and characteristics of current base-editing tools. BEs are categorized into **A.** Cytosine base-editors (CBEs) that catalyse C•G to T•A transitions; **B.** Adenine base-editors (ABEs) that catalyse A•T [or A•U] to G•C transitions; **C.** Cytosine-to-guanine base-editors (CGBEs) that catalyse C•G to G•C transversions; **D.** Cytosine-to-adenine base-editors (CABEs) that catalyse C•G to A•T transversions; **E.** Dual base-editors (DBEs) that catalyse simultaneous C•G to T•A and A•T [or A•U] to G•C transitions.

Applications & challenges of base-editing technology

The base-editing technology was first established in 2016[277,320] and in the years thereafter it has already been applied in a wide range of organisms, including mammals[381], plants[387], and bacteria[388]. Most popular applications concern (a) gene inactivation/activation (via generation/correction of start and stop codons)[388-390]; (b) gene expression modification (via editing of cis-regulatory elements)[391]; (c) alternative splicing, exon skipping, and intron retention (via modification of splice sites)[392,393]; (d) genomic diversification, directed protein evolution, and identification of drug targets or drug resistance mutations (via directed mutagenesis)[394]; (e) editing of single-nucleotide polymorphisms (SNPs) in dividing and non-diving cells, identification of conserved functional amino acids, allele pyramiding, crop improvement, and de novo domestication (via precise editing)[395]; (f) construction of combinatorial mutation library (via multiplexing)[396]; and (g) molecular recording of chemical or physical stimuli (via inducible control of precise editing)[397,398]. Adjusting to the needs of each application, a plethora of base-editing tools with distinct characteristics (type of modification, nucleotide context, PAM compatibility, editing window, efficiency, specificity, product purity) has been generated, as described above.

Current BEs are highly dependent on the PAM-specificity of their Cas component. Although the availability of several native and engineered Cas proteins did broaden the targeting scope of CBEs and ABEs, very limited targetable sites are available for CABEs, CGBEs, and DBEs (Fig. 5c-e). Moreover, BEs comprising alternative Cas variants are usually less efficient than their nSpCas9 BEs. Therefore, examination of additional Cas nucleases (such as members of the rapidly expanding Cas12 family) as well as discovery of novel Cas enzymes and further engineering of existing variants would increase the PAM flexibility of all types of CRISPR BEs. Notably, compact Cas proteins with alternative PAM preferences (like Cas12m[399], Cas12e[400]) would be of special interest because they fit into AAV or HDAd5/35++ adenovirus capsids, as recently demonstrated for base-editors containing the relatively small variants SauriCas9, CjCas9, and Nme2Cas9[295]. Intriguingly, AAV delivery of the compact Cas12f1 nuclease has already been applied for genome editing purposes[401]. The large size of current BEs has implied their delivery required splitting their gene into two separate AAV vectors and reconstitution by trans-splicing inteins, decreasing the on-target editing efficiency[402-404]. Furthermore, lentiviral and plasmid delivery of BEs has resulted in high off-target editing because of prolonged BE

expression[279]. Temporal control of BE activity[279,395,405], utilisation of small-molecule-controlled editing agents[406-409], or sgRNA truncations[283] may decrease undesired off-target effects. Alternatively, more transient delivery systems, such as mRNA and ribonucleoprotein (RNP) delivery, have been successfully applied to achieve higher on-target/off-target ratio, with the former outperforming the latter in terms of on-target efficiency[279,290,292,310,361,395,404,410-414]. An additional round of BE RNP electroporation is sometimes required to reach high efficiencies, which may significantly reduce cell viability[410]. Hence, chemical modification of sgRNA or of BE mRNA as well as codon optimization of the BE mRNA may be used to drastically increase on-target editing efficiency[412,415]. Finally, further diversification and optimization of current BEs may expand their applicability. The natural diversity of deaminase enzymes may be harnessed for the development of novel BEs with improved on-target efficiencies, minimum off-target activities, and relaxed sequence specificities. Indicatively, human A2 and CDAT8 from *Methanopyrus kandleri* are compact (224 and 278 amino acids, respectively) CDs that act with high efficiency exclusively on RNA substrates without any sequence context requirement, suggesting their potential as RNA CBEs, for example in the context of the RESCUE system. However, optimisation of the catalytic activity of CDAT8 at mesophilic temperatures may be required, due to its thermophilic origin. Moreover, the NAEGRDRAFT_32387 CD enzyme from *Naegleria gruberi* naturally acts on mitochondrial and chloroplast RNA, implying its possible fusion to TALEs for the generation of the first programmable organellar RNA CBE. Last, ceADR2 from *C. elegans* could also be tested as an alternative to ADAR2DD engineered variants in the REPAIR system, potentially forming a next generation RNA ABE with no sequence context requirements.

Altogether, deaminase enzymes span across all kingdoms of life, exhibiting an impressive variety of biological roles and biochemical characteristics. Their combination predominantly with CRISPR-Cas systems has given rise to the nascent base-editing technology that (together with the recently developed prime editing approach that also does not require the generation of DSDBs[416]) has high potential especially in human therapeutics.

References

1. Iyer, L. M., Zhang, D., Rogozin, I. B., & Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic acids research, 39(22), 9473-9497.
2. Ambroggio, X. I., Rees, D. C., & Deshaies, R. J. (2004). JAMM: a metalloprotease-like zinc site in the proteasome and signalosome. PLoS biology, 2(1), e2.
3. Greasley, S. E., Horton, P., Ramcharan, J., Beardsley, G. P., Benkovic, S. J., & Wilson, I. A. (2001). Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis. Nature structural biology, 8(5), 402-406.
4. Glaser, P., Danchin, A., Kunst, F., Zuber, P., & Nakano, M. M. (1995). Identification and isolation of a gene required for nitrate assimilation and anaerobic growth of Bacillus subtilis. Journal of bacteriology, 177(4), 1112-1115.
5. Xu, Q., Kozbial, P., McMullan, D., Sri Krishna, S., Brittain, S. M., Ficarro, S. B., ... & Wilson, I. A. (2008). Crystal structure of an ADP-ribosylated protein with a cytidine deaminase-like fold, but unknown function (TM1506), from Thermotoga maritima at 2.70 Å resolution. Proteins: Structure, Function, and Bioinformatics, 71(3), 1546-1552.
6. Kuratani, M., Ishii, R., Bessho, Y., Fukunaga, R., Sengoku, T., Shirouzu, M., ... & Yokoyama, S. (2005). Crystal structure of tRNA adenosine deaminase (TadA) from Aquifex aeolicus. Journal of Biological Chemistry, 280(16), 16002-16008.
7. Kim, J., Malashkevich, V., Roday, S., Lisbin, M., Schramm, V. L., & Almo, S. C. (2006). Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. Biochemistry, 45(20), 6407-6416.
8. Nygaard, P. (1986). On the role of cytidine deaminase in cellular metabolism. In Purine and Pyrimidine Metabolism in Man V (pp. 415-420). Springer, New York, NY.
9. Pospisilova, H., & Frebort, I. (2007). Aminohydrolases acting on adenine, adenosine and their derivatives. Biomedical Papers of the Medical Faculty of Palacky University in Olomouc, 151(1).
10. Gerber, A. P., & Keller, W. (2001). RNA editing by base deamination: more enzymes, more targets, new mysteries. Trends in biochemical sciences, 26(6), 376-384.
11. Diwan, G. D., & Agashe, D. (2018). Wobbling forth and drifting back: the evolutionary history and impact of bacterial tRNA modifications. Molecular biology and evolution, 35(8), 2046-2059.
12. Gerber, A. P., & Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science, 286(5442), 1146-1149.
13. Grice, L. F., & Degnan, B. M. (2015). The origin of the ADAR gene family and animal RNA editing. BMC evolutionary biology, 15(1), 1-7.
14. Sawyer, S. L., Emerman, M., & Malik, H. S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS biology, 2(9), e275.
15. LaRue, R. S., Jónsson, S. R., Silverstein, K. A., Lajoie, M., Bertrand, D., El-Mabrouk, N., ... & Harris, R. S. (2008). The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. BMC molecular biology, 9(1), 1-20.
16. Krishnan, A., Iyer, L. M., Holland, S. J., Boehm, T., & Aravind, L. (2018). Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. Proceedings of the National Academy of Sciences, 115(14), E3201-E3210.
17. Liao, W., Hong, S. H., Chan, B. H. J., Rudolph, F. B., Clark, S. C., & Chan, L. (1999). APOBEC-2, a cardiac-and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochemical and biophysical research communications, 260(2), 398-404.
18. Severi, F., Chicca, A., & Conticello, S. G. (2011). Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. Molecular biology and evolution, 28(3), 1125-1129.
19. Rogozin, I. B., Basu, M. K., Jordan, I. K., Pavlov, Y. I., & Koonin, E. V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases predicted by computational analysis. Cell cycle, 4(9), 1281-1285.
20. Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K., & Neuberger, M. S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. Molecular biology and evolution, 22(2), 367-377.
21. Blanc, V., & Davidson, N. O. (2010). APOBEC-1-mediated RNA editing. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2(5), 594-602.

22. Ito, J., Gifford, R. J., & Sato, K. (2020). Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. Proceedings of the National Academy of Sciences, 117(1), 610-618.
23. LaRue, R. S., Andrésdóttir, V., Blanchard, Y., Conticello, S. G., Derse, D., Emerman, M., ... & Harris, R. S. (2009). Guidelines for naming nonprimate APOBEC3 genes and proteins. Journal of virology, 83(2), 494-497.
24. Münk, C., Beck, T., Zielonka, J., Hotz-Wagenblatt, A., Chareza, S., Battenberg, M., ... & Yuhki, N. (2008). Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. Genome biology, 9(3), 1-20.
25. Münk, C., Willemsen, A., & Bravo, I. G. (2012). An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC evolutionary biology, 12(1), 1-16.
26. Esnault, C., Millet, J., Schwartz, O., & Heidmann, T. (2006). Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. Nucleic acids research, 34(5), 1522-1531.
27. Frances, A., & Cordelier, P. (2020). The emerging role of cytidine deaminase in human diseases: a new opportunity for therapy?. Molecular Therapy, 28(2), 357-366.
28. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., ... & Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science, 273(5278), 1058-1073.
29. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., ... & Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA research, 3(3), 109-136.
30. Randau, L., Stanley, B. J., Kohlway, A., Mechta, S., Xiong, Y., & Söll, D. (2009). A cytidine deaminase edits C to U in transfer RNAs in Archaea. Science, 324(5927), 657-659.
31. Anantharaman, V., & Aravind, L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. Genome biology, 4(12), 1-15.
32. Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., & Orth, K. (2009). AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. Science, 323(5911), 269-272.
33. Zhang, D., Iyer, L. M., & Aravind, L. (2011). A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. Nucleic acids research, 39(11), 4532-4552.
34. Hayes, C. S., Aoki, S. K., & Low, D. A. (2010). Bacterial contact-dependent delivery systems. Annual review of genetics, 44(1), 71-90.
35. Pallen, M. J. (2002). The ESAT-6/WXG100 superfamily–and a new Gram-positive secretion system?. Trends in microbiology, 10(5), 209-212.
36. Simeone, R., Bottai, D., & Brosch, R. (2009). ESX/type VII secretion systems and their role in host–pathogen interaction. Current opinion in microbiology, 12(1), 4-10.
37. Ellermeier, C. D., & Losick, R. (2006). Evidence for a novel protease governing regulated intramembrane proteolysis and resistance to antimicrobial peptides in Bacillus subtilis. Genes & development, 20(14), 1911-1922.
38. Coulthurst, S. (2019). The Type VI secretion system: a versatile bacterial weapon. Microbiology, 165(5), 503-515.
39. Hood, R. D., Singh, P., Hsu, F., Güvener, T., Carl, M. A., Trinidad, R. R., ... & Mougous, J. D. (2010). A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. Cell host & microbe, 7(1), 25-37.
40. Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., ... & Liu, D. R. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature, 583(7817), 631-637.
41. de Moraes, M. H., Hsu, F., Huang, D., Bosch, D. E., Zeng, J., Radey, M. C., ... & Mougous, J. D. (2021). An interbacterial DNA deaminase toxin directly mutagenizes surviving target populations. Elife, 10, e62967.
42. Hurst, M. R., Glare, T. R., & Jackson, T. A. (2004). Cloning Serratia entomophila antifeeding genes—a putative defective prophage active against the grass grub Costelytra zealandica. Journal of bacteriology, 186(15), 5116-5128.
43. Bowen, D., Rocheleau, T. A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R., & ffrench-Constant, R. H. (1998). Insecticidal toxins from the bacterium Photorhabdus luminescens. Science, 280(5372), 2129-2132.
44. Hurst, M. R., Glare, T. R., Jackson, T. A., & Ronson, C. W. (2000). Plasmid-located pathogenicity determinants of Serratia entomophila, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of Photorhabdus luminescens. Journal of Bacteriology, 182(18), 5127-5138.
45. Burroughs, A. M., Iyer, L. M., & Aravind, L. (2007). Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. Gene and protein evolution, 3, 48-65.
46. Nikkilä, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H. N., Wallberg, F., ... & Lord, C. J. (2017). Elevated APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. British journal of cancer, 117(1), 113-123.

47. Li, J., Chen, Y., Li, M., Carpenter, M. A., McDougle, R. M., Luengas, E. M., ... & Mueller, J. D. (2014). APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. Journal of molecular biology, 426(6), 1296-1307.
48. Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell, 102(5), 553-563.
49. Feng, Y., Seija, N., Di Noia, J. M., & Martin, A. (2020). AID in antibody diversification: there and back again. Trends in immunology, 41(7), 586-600.
50. Kanagaraj, A., Sakamoto, N., Que, L., Li, Y., Mohiuddin, M., Koura, M., ... & Kitamura, K. (2019). Different antiviral activities of natural APOBEC3C, APOBEC3G, and APOBEC3H variants against hepatitis B virus. Biochemical and Biophysical Research Communications, 518(1), 26-31.
51. Shi, M., Tan, L., Zhang, Y., Meng, C., Wang, W., Sun, Y., ... & Ding, C. (2020). Characterization and functional analysis of chicken APOBEC4. Developmental & Comparative Immunology, 106, 103631.
52. Marino, D., Perković, M., Hain, A., Jaguva Vasudevan, A. A., Hofmann, H., Hanschmann, K. M., ... & Münk, C. (2016). APOBEC4 Enhances the Replication of HIV-1. PLoS One, 11(6), e0155422.
53. Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., ... & Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nature immunology, 8(6), 647-656.
54. Guo, J. U., Su, Y., Zhong, C., Ming, G. L., & Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell, 145(3), 423-434.
55. Siriwardena, S. U., Chen, K., & Bhagwat, A. S. (2016). Functions and malfunctions of mammalian DNA-cytosine deaminases. Chemical reviews, 116(20), 12688-12710.
56. Fossat, N., Tourle, K., Radziewic, T., Barratt, K., Liebhold, D., Studdert, J. B., ... & Tam, P. P. (2014). C to U RNA editing mediated by APOBEC 1 requires RNA-binding protein RBM 47. EMBO reports, 15(8), 903-910.
57. Sowden, M. P., Ballatori, N., Jensen, K. L. D. M., Reed, L. H., & Smith, H. C. (2002). The editosome for cytidine to uridine mRNA editing has a native complexity of 27S: identification of intracellular domains containing active and inactive editing factors. Journal of cell science, 115(5), 1027-1039.
58. Anant, S., MacGinnitie, A. J., & Davidson, N. O. (1995). apobec-1, the Catalytic Subunit of the Mammalian Apolipoprotein B mRNA Editing Enzyme, Is a Novel RNA-binding Protein∗. Journal of Biological Chemistry, 270(24), 14762-14767.
59. Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., … & Chiba, T. (2012). Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. International journal of cancer, 130(6), 1294-1301.
60. Cappione, A. J., French, B. L., & Skuse, G. R. (1997). A potential role for NF1 mRNA editing in the pathogenesis of NF1 tumors. American journal of human genetics, 60(2), 305.
61. Yamanaka, S., Poksay, K. S., Arnold, K. S., & Innerarity, T. L. (1997). A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. Genes & development, 11(3), 321-333.
62. Rosenberg, B. R., Hamilton, C. E., Mwangi, M. M., Dewell, S., & Papavasiliou, F. N. (2011). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. Nature structural & molecular biology, 18(2), 230-236.
63. Saraconi, G., Severi, F., Sala, C., Mattiuz, G., & Conticello, S. G. (2014). The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. Genome biology, 15(7), 1-10.
64. Michalski, D., Ontiveros, J. G., Russo, J., Charley, P. A., Anderson, J. R., Heck, A. M., ... & Wilusz, J. (2019). Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection. Journal of Biological Chemistry, 294(44), 16282-16296.
65. Ohtsubo, H., Sato, Y., Suzuki, T., Mizunoya, W., Nakamura, M., Tatsumi, R., & Ikeuchi, Y. (2017). APOBEC2 negatively regulates myoblast differentiation in muscle regeneration. The International Journal of Biochemistry & Cell Biology, 85, 91-101.

66. Liao, W., Hong, S. H., Chan, B. H. J., Rudolph, F. B., Clark, S. C., & Chan, L. (1999). APOBEC-2, a cardiac-and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochemical and biophysical research communications, 260(2), 398-404.
67. Mikl, M. C., Watt, I. N., Lu, M., Reik, W., Davies, S. L., Neuberger, M. S., & Rada, C. (2005). Mice deficient in APOBEC2 and APOBEC3. Molecular and cellular biology, 25(16), 7270-7277.
68. Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., & Chen, X. S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature, 445(7126), 447-451.
69. Ohtsubo, H., Sato, Y., Suzuki, T., Mizunoya, W., Nakamura, M., Tatsumi, R., & Ikeuchi, Y. (2017). Data supporting possible implication of APOBEC2 in self-renewal functions of myogenic stem satellite cells: toward understanding the negative regulation of myoblast differentiation. Data in brief, 12, 269-273.
70. Vonica, A., Rosa, A., Arduini, B. L., & Brivanlou, A. H. (2011). APOBEC2, a selective inhibitor of TGFβ signaling, regulates left–right axis specification during early embryogenesis. Developmental biology, 350(1), 13-23.
71. Ianni, A., Bennato, F., Martino, C., Colapietro, M., & Martino, G. (2021). Whole Blood Transcriptome Profiling Reveals Positive Effects of Olive Leaves-Supplemented Diet on Cholesterol in Goats. Animals, 11(4), 1150.
72. Tarsani, E., Kranis, A., Maniatis, G., Avendano, S., Hager-Theodorides, A. L., & Kominakis, A. (2019). Discovery and characterization of functional modules associated with body weight in broilers. Scientific reports, 9(1), 1-12.
73. Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., & Harris, R. S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nature structural & molecular biology, 17(2), 222-229.
74. Koning, F. A., Newman, E. N., Kim, E. Y., Kunstman, K. J., Wolinsky, S. M., & Malim, M. H. (2009). Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. Journal of virology, 83(18), 9474-9485.
75. Mussil, B., Suspène, R., Aynaud, M. M., Gauvrit, A., Vartanian, J. P., & Wain-Hobson, S. (2013). Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. PloS one, 8(8), e73641.
76. Berger, G., Durand, S., Fargier, G., Nguyen, X. N., Cordeil, S., Bouaziz, S., ... & Cimarelli, A. (2011). APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. PLoS pathogens, 7(9), e1002221.
77. Caval, V., Suspene, R., Vartanian, J. P., & Wain-Hobson, S. (2014). Orthologous mammalian APOBEC3A cytidine deaminases hypermutate nuclear DNA. Molecular biology and evolution, 31(2), 330-340.
78. Laude, H. C., Caval, V., Bouzidi, M. S., Li, X., Jamet, F., Henry, M., ... & Vartanian, J. P. (2018). The rabbit as an orthologous small animal model for APOBEC3A oncogenesis. Oncotarget, 9(45), 27809.
79. Siriwardena, S. U., Perera, M. L., Senevirathne, V., Stewart, J., & Bhagwat, A. S. (2019). A tumor-promoting phorbol ester causes a large increase in APOBEC3A expression and a moderate increase in APOBEC3B expression in a normal human keratinocyte cell line without increasing genomic uracils. Molecular and cellular biology, 39(1), e00238-18.
80. Law, E. K., Levin-Klein, R., Jarvis, M. C., Kim, H., Argyris, P. P., Carpenter, M. A., ... & Harris, R. S. (2020). APOBEC3A catalyzes mutation and drives carcinogenesis in vivo. Journal of Experimental Medicine, 217(12).
81. Caval, V., Suspène, R., Khalfi, P., Gaillard, J., Caignard, G., Vitour, D., ... & Wain-Hobson, S. (2021). Frame-shifted APOBEC3A encodes two alternative proapoptotic proteins that target the mitochondrial network. Journal of Biological Chemistry, 297(3).
82. Suspène, R., Aynaud, M. M., Guétard, D., Henry, M., Eckhoff, G., Marchio, A., ... & Wain-Hobson, S. (2011). Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proceedings of the National Academy of Sciences, 108(12), 4858-4863.
83. Caval, V., Suspène, R., Shapira, M., Vartanian, J. P., & Wain-Hobson, S. (2014). A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3' UTR enhances chromosomal DNA damage. Nature communications, 5(1), 1-7.
84. Oh, S., Bournique, E., Bowen, D., Jalili, P., Sanchez, A., Ward, I., ... & Buisson, R. (2021). Genotoxic stress and viral infection induce transient expression of APOBEC3A and pro-inflammatory genes through two distinct pathways. Nature communications, 12(1), 1-17.
85. Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcia-Perez, J. L., O'Shea, K. S., Moran, J. V., & Cullen, B. R. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proceedings of the National Academy of Sciences, 103(23), 8780-8785.
86. Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., ... & Weitzman, M. D. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Current Biology, 16(5), 480-485.

87. Suspene, R., Mussil, B., Laude, H., Caval, V., Berry, N., Bouzidi, M. S., ... & Vartanian, J. P. (2017). Self-cytoplasmic DNA upregulates the mutator enzyme APOBEC3A leading to chromosomal DNA damage. Nucleic acids research, 45(6), 3231-3241.
88. Vasudevan, A. A. J., Balakrishnan, K., Franken, A., Krikoni, A., Häussinger, D., Luedde, T., & Münk, C. (2021). Murine leukemia virus resists producer cell APOBEC3A by its Glycosylated Gag but not target cell APOBEC3A. Virology, 557, 1-14.
89. Cheng, A. Z., Moraes, S. N., Shaban, N. M., Fanunza, E., Bierle, C. J., Southern, P. J., ... & Harris, R. S. (2021). APOBECs and herpesviruses. Viruses, 13(3), 390.
90. Land, A. M., Law, E. K., Carpenter, M. A., Lackey, L., Brown, W. L., & Harris, R. S. (2013). Endogenous APOBEC3A DNA cytosine deaminase is cytoplasmic and nongenotoxic. Journal of Biological Chemistry, 288(24), 17253-17260.
91. Samanta, M., Iwakiri, D., Kanda, T., Imaizumi, T., & Takada, K. (2006). EB virus-encoded RNAs are recognized by RIG-I and activate signaling to induce type I IFN. The EMBO journal, 25(18), 4207-4214.
92. Schilling, E. M., Scherer, M., & Stamminger, T. (2021). Intrinsic immune mechanisms restricting human cytomegalovirus replication. Viruses, 13(2), 179.
93. Leonardi, A., Rosani, U., & Brun, P. (2020). Ocular surface expression of SARS-CoV-2 receptors. Ocular Immunology and Inflammation, 28(5), 735-738.
94. Milewska, A., Kindler, E., Vkovski, P., Zeglen, S., Ochman, M., Thiel, V., ... & Pyrc, K. (2018). APOBEC3-mediated restriction of RNA virus replication. Scientific reports, 8(1), 1-12.
95. Santoro, M. G., Rossi, A., & Amici, C. (2003). NF-κB and virus infection: who controls whom. The EMBO journal, 22(11), 2552-2560.
96. Narvaiza, I., Linfesty, D. C., Greener, B. N., Hakata, Y., Pintel, D. J., Logue, E., ... & Weitzman, M. D. (2009). Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. PLoS pathogens, 5(5), e1000439.
97. Warren, C. J., Xu, T., Guo, K., Griffin, L. M., Westrich, J. A., Lee, D., ... & Pyeon, D. (2015). APOBEC3A functions as a restriction factor of human papillomavirus. Journal of virology, 89(1), 688-702.
98. Kano, M., Kondo, S., Wakisaka, N., Wakae, K., Aga, M., Moriyama-Kita, M., ... & Yoshizaki, T. (2019). Expression of estrogen receptor alpha is associated with pathogenesis and prognosis of human papillomavirus-positive oropharyngeal cancer. International Journal of Cancer, 145(6), 1547-1557.
99. Kondo, S., Wakae, K., Wakisaka, N., Nakanishi, Y., Ishikawa, K., Komori, T., ... & Yoshizaki, T. (2017). APOBEC3A associates with human papillomavirus genome integration in oropharyngeal cancers. Oncogene, 36(12), 1687-1697.
100. Brown, A. L., Collins, C. D., Thompson, S., Coxon, M., Mertz, T. M., & Roberts, S. A. (2021). Single-stranded DNA binding proteins influence APOBEC3A substrate preference. Scientific reports, 11(1), 1-13.
101. Langenbucher, A., Bowen, D., Sakhtemani, R., Bournique, E., Wise, J. F., Zou, L., ... & Lawrence, M. S. (2021). An extended APOBEC3A mutation signature in cancer. Nature communications, 12(1), 1-11.
102. Suspène, R., Aynaud, M. M., Vartanian, J. P., & Wain-Hobson, S. (2013). Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase. PloS one, 8(6), e63461.
103. Jalili, P., Bowen, D., Langenbucher, A., Park, S., Aguirre, K., Corcoran, R. B., ... & Buisson, R. (2020). Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. Nature communications, 11(1), 1-13.
104. Sharma, S., & Baysal, B. E. (2017). Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. PeerJ, 5, e4136.
105. Sharma, S., Patnaik, S. K., Thomas Taggart, R., Kannisto, E. D., Enriquez, S. M., Gollnick, P., & Baysal, B. E. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nature communications, 6(1), 1-15.
106. Baysal, B. E., De Jong, K., Liu, B., Wang, J., Patnaik, S. K., Wallace, P. K., & Taggart, R. T. (2013). Hypoxia-inducible C-to-U coding RNA editing downregulates SDHB in monocytes. PeerJ, 1, e152.
107. Alqassim, E. Y., Sharma, S., Khan, A. N. M., Emmons, T. R., Cortes Gomez, E., Alahmari, A., ... & Baysal, B. E. (2021). RNA editing enzyme APOBEC3A promotes pro-inflammatory M1 macrophage polarization. Communications biology, 4(1), 1-11.
108. Lackey, L., Law, E. K., Brown, W. L., & Harris, R. S. (2013). Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. Cell Cycle, 12(5), 762-772.
109. Warren, C. J., Westrich, J. A., Van Doorslaer, K., & Pyeon, D. (2017). Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. Viruses, 9(8), 233.

110. Heller, M., Prigge, E. S., Kaczorowski, A., von Knebel Doeberitz, M., Hohenfellner, M., & Duensing, S. (2018). APOBEC3A expression in penile squamous cell carcinoma. Pathobiology, 85(3), 169-178.
111. Zou, J., Wang, C., Ma, X., Wang, E., & Peng, G. (2017). APOBEC3B, a molecular driver of mutagenesis in human cancers. Cell & bioscience, 7(1), 1-7.
112. He, X., Xu, H., Wang, X., Wu, J., Niu, J., & Gao, P. (2019). Associations between the single nucleotide polymorphisms of APOBEC3A, APOBEC3B and APOBEC3H, and chronic hepatitis B progression and hepatocellular carcinoma in a Chinese population. Molecular Medicine Reports, 20(3), 2177-2188.
113. Gara, S. K., Tyagi, M. V., Patel, D. T., Gaskins, K., Lack, J., Liu, Y., & Kebebew, E. (2020). GATA3 and APOBEC3B are prognostic markers in adrenocortical carcinoma and APOBEC3B is directly transcriptionally regulated by GATA3. Oncotarget, 11(36), 3354.
114. Feng, C., Zheng, Q., Yang, Y., Xu, M., Lian, Y., Huang, J., & Jiang, Y. (2019). APOBEC3B high expression in gastroenteropancreatic neuroendocrine neoplasms and association with lymph metastasis. Applied Immunohistochemistry & Molecular Morphology, 27(8), 599-605.
115. Xu, L., Chang, Y., An, H., Zhu, Y., Yang, Y., & Xu, J. (2015, August). High APOBEC3B expression is a predictor of recurrence in patients with low-risk clear cell renal cell carcinoma. In Urologic Oncology: Seminars and Original Investigations (Vol. 33, No. 8, pp. 340-e1). Elsevier.
116. Kim, H., Kim, O., Lee, M. A., Lee, J. Y., Hong, S. H., Ha, U. S., ... & Kim, I. H. (2021). Prognostic Impact of APOBEC3B Expression in Metastatic Urothelial Carcinoma and Its Association with Tumor-Infiltrating Cytotoxic T Cells. Current Oncology, 28(3), 1652-1662.
117. Xia, S., Gu, Y., Zhang, H., Fei, Y., Cao, Y., Fang, H., ... & Zhang, W. (2021). Immune inactivation by APOBEC3B enrichment predicts response to chemotherapy and survival in gastric cancer. Oncoimmunology, 10(1), 1975386.
118. Fan, Q., Huang, T., Sun, X., Wang, Y. W., Wang, J., Liu, Y., ... & Wang, Y. D. (2021). HPV-16/18 E6-induced APOBEC3B expression associates with proliferation of cervical cancer cells and hypomethylation of Cyclin D1. Molecular Carcinogenesis, 60(5), 313-330.
119. Sieuwerts, A. M., Doebar, S. C., de Weerd, V., Verhoef, E. I., Beauford, C. M., Agahozo, M. C., ... & van Deurzen, C. H. (2019). APOBEC3B gene expression in ductal carcinoma in situ and synchronous invasive breast cancer. Cancers, 11(8), 1062.
120. Seplyarskiy, V. B., Andrianova, M. A., & Bazykin, G. A. (2017). APOBEC3A/B-induced mutagenesis is responsible for 20% of heritable mutations in the TpCpW context. Genome research, 27(2), 175-184.
121. Kostrzak, A., Caval, V., Escande, M., Pliquet, E., Thalmensi, J., Bestetti, T., ... & Langlade-Demoyen, P. (2017). APOBEC3A intratumoral DNA electroporation in mice. Gene therapy, 24(2), 74-83.
122. Fu, Y., Ito, F., Zhang, G., Fernandez, B., Yang, H., & Chen, X. S. (2015). DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. Biochemical Journal, 471(1), 25-35.
123. Wang, D., Li, X., Li, J., Lu, Y., Zhao, S., Tang, X., ... & Yang, Y. (2019). APOBEC3B interaction with PRC2 modulates microenvironment to promote HCC progression. Gut, 68(10), 1846-1857.
124. Vieira, V. C., Leonard, B., White, E. A., Starrett, G. J., Temiz, N. A., Lorenz, L. D., ... & Harris, R. S. (2014). Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. MBio, 5(6), e02234-14.
125. Yu, Q., Chen, D., König, R., Mariani, R., Unutmaz, D., & Landau, N. R. (2004). APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. Journal of Biological Chemistry, 279(51), 53379-53386.
126. Bonvin, M., Achermann, F., Greeve, I., Stroka, D., Keogh, A., Inderbitzin, D., ... & Greeve, J. (2006). Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication. Hepatology, 43(6), 1364-1374.
127. Lejeune, N., Poulain, F., Willemart, K., Blockx, Z., Mathieu, S., & Gillet, N. A. (2021). Infection of Bronchial Epithelial Cells by the Human Adenoviruses A12, B3, and C2 Differently Regulates the Innate Antiviral Effector APOBEC3B. Journal of Virology, 95(13), e02413-20.
128. Hernandez-Pacheco, N., Farzan, N., Francis, B., Karimi, L., Repnik, K., Vijverberg, S. J., ... & Pino-Yanes, M. (2019). Genome-wide association study of inhaled corticosteroid response in admixed children with asthma. Clinical & Experimental Allergy, 49(6), 789-798.
129. Leal, F. E., Menezes, S. M., Costa, E. A., Brailey, P. M., Gama, L., Segurado, A. C., ... & Van Weyenbergh, J. (2018). Comprehensive antiretroviral restriction factor profiling reveals the evolutionary imprint of the ex Vivo and in vivo IFN-β response in HTLV-1-associated neuroinflammation. Frontiers in microbiology, 9, 985.

130. Rüder, U., Denkert, C., Alisa Kunze, C., Jank, P., Lindner, J., Jöhrens, K., ... & Darb Esfahani, S. (2019). APOBEC3B protein expression and mRNA analyses in patients with high-grade serous ovarian carcinoma.
131. Nakano, Y., Yamamoto, K., Ueda, M. T., Soper, A., Konno, Y., Kimura, I., ... & Sato, K. (2020). A role for gorilla APOBEC3G in shaping lentivirus evolution including transmission to humans. PLoS pathogens, 16(9), e1008812.
132. Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., ... & Tokunaga, K. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic acids research, 35(9), 2955-2964.
133. Kawahara, R., Recuero, S., Nogueira, F. C., Domont, G. B., Leite, K. R., Srougi, M., ... & Palmisano, G. (2019). Tissue proteome signatures associated with five grades of prostate cancer and benign prostatic hyperplasia. Proteomics, 19(21-22), 1900174.
134. Vasudevan, A. A. J., Hofmann, H., Willbold, D., Häussinger, D., Koenig, B. W., & Münk, C. (2017). Enhancing the catalytic deamination activity of APOBEC3C is insufficient to inhibit Vif-deficient HIV-1. Journal of molecular biology, 429(8), 1171-1191.
135. Adolph, M. B., Ara, A., Feng, Y., Wittkopp, C. J., Emerman, M., Fraser, J. S., & Chelico, L. (2017). Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. Nucleic acids research, 45(6), 3378-3394.
136. Chen, X. S. (2021). Insights into the Structures and Multimeric Status of APOBEC Proteins Involved in Viral Restriction and Other Cellular Functions. Viruses, 13(3), 497.
137. Hultquist, J. F., Lengyel, J. A., Refsland, E. W., LaRue, R. S., Lackey, L., Brown, W. L., & Harris, R. S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. Journal of virology, 85(21), 11220-11234.
138. Tao, L., Jiang, Z., Xu, M., Xu, T., & Liu, Y. (2019). Induction of APOBEC3C facilitates the genotoxic stress-mediated cytotoxicity of artesunate. Chemical research in toxicology, 32(12), 2526-2537.
139. Khalfi, P., Suspène, R., Caval, V., Thiers, V., Beauclair, G., Marchio, A., ... & Vartanian, J. P. (2022). APOBEC3C S188I Polymorphism Enhances Context-Specific Editing of Hepatitis B Virus Genome. The Journal of Infectious Diseases, 226(5), 891-895.
140. Jiang, Q., Isquith, J., Ladel, L., Mark, A., Holm, F., Mason, C., ... & Jamieson, C. (2021). Inflammation-driven deaminase deregulation fuels human pre-leukemia stem cell evolution. Cell reports, 34(4), 108670.
141. Asaoka, M., Patnaik, S. K., Ishikawa, T., & Takabe, K. (2021). Different members of the APOBEC3 family of DNA mutators have opposing associations with the landscape of breast cancer. American journal of cancer research, 11(10), 5111.
142. Wang, K., Li, L., Fu, L., Yuan, Y., Dai, H., Zhu, T., ... & Yuan, F. (2019). Integrated bioinformatics analysis the function of RNA binding proteins (RBPs) and their prognostic value in breast cancer. Frontiers in pharmacology, 10, 140.
143. Zhang, M., Wang, X., Chen, X., Guo, F., & Hong, J. (2020). Prognostic value of a stemness index-associated signature in primary lower-grade glioma. Frontiers in genetics, 11, 441.
144. Went, M., Kinnersley, B., Sud, A., Johnson, D. C., Weinhold, N., Försti, A., ... & Houlston, R. S. (2019). Transcriptome-wide association study of multiple myeloma identifies candidate susceptibility genes. Human genomics, 13(1), 1-8.
145. Constantin, D., Dubuis, G., Conde-Rubio, M. D. C., & Widmann, C. (2022). APOBEC3C, a nucleolar protein induced by genotoxins, is excluded from DNA damage sites. The FEBS Journal, 289(3), 808-831.
146. Zhu, Z., Ji, X., Zhu, W., Cai, T., Xu, C., Huang, C., ... & Zhou, L. (2021). Comprehensive bioinformatics analyses of APOBECs family and identification of APOBEC3D as the unfavorable prognostic biomarker in clear cell renal cell carcinoma. Journal of Cancer, 12(23), 7101.
147. Yang, Z., Tao, Y., Xu, X., Cai, F., Yu, Y., & Ma, L. (2018). Bufalin inhibits cell proliferation and migration of hepatocellular carcinoma cells via APOBEC3F induced intestinal immune network for IgA production signaling pathway. Biochemical and Biophysical Research Communications, 503(3), 2124-2131.
148. Anwar, F., Davenport, M. P., & Ebrahimi, D. (2013). Footprint of APOBEC3 on the genome of human retroelements. Journal of virology, 87(14), 8195-8204.
149. Xu, F., Liu, T., Zhou, Z., Zou, C., & Xu, S. (2021). Comprehensive Analyses Identify APOBEC3A as a Genomic Instability-Associated Immune Prognostic Biomarker in Ovarian Cancer. Frontiers in immunology, 12.
150. Sharma, S., Wang, J., Alqassim, E., Portwood, S., Cortes Gomez, E., Maguire, O., ... & Baysal, B. E. (2019). Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. Genome biology, 20(1), 1-17.
151. Ikeda, T., Molan, A. M., Jarvis, M. C., Carpenter, M. A., Salamango, D. J., Brown, W. L., & Harris, R. S. (2019). HIV-1 restriction by endogenous APOBEC3G in the myeloid cell line THP-1. The Journal of General Virology, 100(7), 1140.

152. Han, W., Xu, J., & Shen, G. L. (2020). Prognostic implication and functional annotations of APOBEC3G expression in patients with Melanoma. Journal of Cancer, 11(17), 5245.
153. Luo, Y., Yan, B., Liu, L., Yin, L., Ji, H., An, X., ... & Herr, I. (2021). Sulforaphane inhibits the expression of long noncoding RNA H19 and its target APOBEC3G and thereby pancreatic cancer progression. Cancers, 13(4), 827.
154. Adolph, M. B., Ara, A., & Chelico, L. (2019). APOBEC3 host restriction factors of HIV-1 can change the template switching frequency of reverse transcriptase. Journal of molecular biology, 431(7), 1339-1352.
155. Wang, J., Shaban, N. M., Land, A. M., Brown, W. L., & Harris, R. S. (2018). Simian immunodeficiency virus Vif and human APOBEC3B interactions resemble those between HIV-1 Vif and human APOBEC3G. Journal of virology, 92(12), e00447-18.
156. Salamango, D. J., McCann, J. L., Demir, Ö., Becker, J. T., Wang, J., Lingappa, J. R., ... & Harris, R. S. (2020). Functional and structural insights into a Vif/PPP2R5 complex elucidated using patient HIV-1 isolates and computational modeling. Journal of virology, 94(21), e00631-20.
157. Wan, L., Kamba, K., Nagata, T., & Katahira, M. (2020). An insight into the dependence of the deamination rate of human APOBEC3F on the length of single-stranded DNA, which is affected by the concentrations of APOBEC3F and single-stranded DNA. Biochimica et Biophysica Acta (BBA)-General Subjects, 1864(2), 129346.
158. Shaban, N. M., Shi, K., Lauer, K. V., Carpenter, M. A., Richards, C. M., Salamango, D., ... & Harris, R. S. (2018). The antiviral and cancer genomic DNA deaminase APOBEC3H is regulated by an RNA-mediated dimerization mechanism. Molecular cell, 69(1), 75-86.
159. Ara, A., Love, R. P., Follack, T. B., Ahmed, K. A., Adolph, M. B., & Chelico, L. (2017). Mechanism of enhanced HIV restriction by virion coencapsidated cytidine deaminases APOBEC3F and APOBEC3G. Journal of virology, 91(3), e02230-16.
160. Solomon, W. C., Myint, W., Hou, S., Kanai, T., Tripathi, R., Kurt Yilmaz, N., ... & Matsuo, H. (2019). Mechanism for APOBEC3G catalytic exclusion of RNA and non-substrate DNA. Nucleic acids research, 47(14), 7676-7689.
161. Wan, L., Nagata, T., & Katahira, M. (2018). Influence of the DNA sequence/length and pH on deaminase activity, as well as the roles of the amino acid residues around the catalytic center of APOBEC3F. Physical Chemistry Chemical Physics, 20(5), 3109-3117.
162. Wan, L., Kamba, K., Nagata, T., & Katahira, M. (2020). An insight into the dependence of the deamination rate of human APOBEC3F on the length of single-stranded DNA, which is affected by the concentrations of APOBEC3F and single-stranded DNA. Biochimica et Biophysica Acta (BBA)-General Subjects, 1864(2), 129346.
163. Li, A., Wu, J., Zhai, A., Qian, J., Wang, X., Qaria, M. A., ... & Zhang, F. (2019). HBV triggers APOBEC2 expression through miR-122 regulation and affects the proliferation of liver cancer cells. International Journal of Oncology, 55(5), 1137-1148.
164. Liu, R., Wu, P., Ogrodzki, P., Mahmoud, S., Liang, K., Liu, P., ... & Zaher, W. A. (2021). Genomic epidemiology of SARS-CoV-2 in the UAE reveals novel virus mutation, patterns of co-infection and tissue specific host immune response. Scientific Reports, 11(1), 1-14.
165. Meshcheryakova, A., Pietschmann, P., Zimmermann, P., Rogozin, I. B., & Mechtcheriakova, D. (2021). AID and APOBECs as Multifaceted Intrinsic Virus-Restricting Factors: Emerging Concepts in the Light of COVID-19. Frontiers in Immunology, 12, 2429.
166. Walker, B. A., Wardell, C. P., Murison, A., Boyle, E. M., Begum, D. B., Dahir, N. M., ... & Morgan, G. J. (2015). APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. Nature communications, 6(1), 1-11.
167. Feng, Y., Goubran, M. H., Follack, T. B., & Chelico, L. (2017). Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. Scientific reports, 7(1), 1-11.
168. Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell, 135(7), 1201-1212.
169. Saupe, S. J. (2000). Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. Microbiology and molecular biology reviews, 64(3), 489-502.
170. Alfonzo, J. D., Blanc, V., Estevez, A. M., Rubio, M. A. T., & Simpson, L. (1999). C to U editing of the anticodon of imported mitochondrial tRNATrp allows decoding of the UGA stop codon in Leishmania tarentolae. The EMBO journal, 18(24), 7056-7062.
171. Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A., & Takenaka, M. (2011). PPR proteins network as site-specific RNA editing factors in plant organelles. RNA biology, 8(1), 67-70.

172. Neiman, M., & Taylor, D. R. (2009). The causes of mutation accumulation in mitochondrial genomes. Proceedings of the Royal Society B: Biological Sciences, 276(1660), 1201-1209.
173. Knoop, V. (2011). When you can't trust the DNA: RNA editing changes transcript sequences. Cellular and Molecular Life Sciences, 68(4), 567-586.
174. Watanabe, S. I., & Uchida, T. (1996). Expression of cytidine deaminase in human solid tumors and its regulation by 1 α, 25-dihydroxyvitamin D3. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1312(2), 99-104.
175. Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R., & Jensen, L. J. (2014). COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database, 2014.
176. Morimoto, R., O'Meara, C. P., Holland, S. J., Trancoso, I., Souissi, A., Schorpp, M., ... & Boehm, T. (2020). Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys. Science Immunology, 5(45), eaba0925.
177. Brown, R. S. (2005). Zinc finger proteins: getting a grip on RNA. Current opinion in structural biology, 15(1), 94-98.
178. Gemble, S., Ahuja, A., Buhagiar-Labarchède, G., Onclercq-Delic, R., Dairou, J., Biard, D. S., ... & Amor-Guéret, M. (2015). Pyrimidine pool disequilibrium induced by a cytidine deaminase deficiency inhibits PARP-1 activity, leading to the under replication of DNA. PLoS genetics, 11(7), e1005384.
179. Mameri, H., Bièche, I., Meseure, D., Marangoni, E., Buhagiar-Labarchède, G., Nicolas, A., ... & Amor-Guéret, M. (2017). Cytidine Deaminase Deficiency Reveals New Therapeutic Opportunities against CancerCDA Deficiency in Cancer. Clinical Cancer Research, 23(8), 2116-2126.
180. Somasekaram, A., Jarmuz, A., How, A., Scott, J., & Navaratnam, N. (1999). Intracellular Localization of Human Cytidine Deaminase: identification of a functional nuclear localization signal. Journal of Biological Chemistry, 274(40), 28405-28412.
181. Ciccolini, J., Dahan, L., André, N., Evrard, A., Duluc, M., Blesius, A., ... & Mercier, C. (2010). Cytidine deaminase residual activity in serum is a predictive marker of early severe toxicities in adults after gemcitabine-based chemotherapies. Journal of clinical oncology, 28(1), 160-165.
182. Teichert, I. (2020). Fungal RNA editing: who, when, and why?. Applied Microbiology and Biotechnology, 104(13), 5689-5695.
183. Skaldin, M., Tuittila, M., Zavialov, A. V., & Zavialov, A. V. (2018). Secreted bacterial adenosine deaminase is an evolutionary precursor of adenosine deaminase growth factor. Molecular Biology and Evolution, 35(12), 2851-2861.
184. Ramsey, J. S., MacDonald, S. J., Jander, G., Nakabachi, A., Thomas, G. H., & Douglas, A. E. (2010). Genomic evidence for complementary purine metabolism in the pea aphid, Acyrthosiphon pisum, and its symbiotic bacterium Buchnera aphidicola. Insect Molecular Biology, 19, 241-248.
185. Kathiresan, K., Saravanakumar, K., Sahu, S. K., & Sivasankaran, M. (2014). Adenosine deaminase production by an endophytic bacterium (Lysinibacillus sp.) from Avicennia marina. 3 Biotech, 4(3), 235-239.
186. Sagar, S., Kaur, M., & Minneman, K. P. (2010). Antiviral lead compounds from marine sponges. Marine drugs, 8(10), 2619-2638.
187. Showalter, H. H., Bunge, R. H., French, J. C., Hurley, T. R., Leeds, R. L., Leja, B., ... & Edmunds, C. R. (1992). Improved production of pentostatin and identification of fermentation cometabolites. The Journal of Antibiotics, 45(12), 1914-1918.
188. Burnstock, G., & Verkhratsky, A. (2009). Evolutionary origins of the purinergic signalling system. Acta physiologica, 195(4), 415-447.
189. Thammavongsa, V., Kern, J. W., Missiakas, D. M., & Schneewind, O. (2009). Staphylococcus aureus synthesizes adenosine to escape host immune responses. Journal of Experimental Medicine, 206(11), 2417-2427.
190. Thammavongsa, V., Missiakas, D. M., & Schneewind, O. (2013). Staphylococcus aureus degrades neutrophil extracellular traps to promote immune cell death. Science, 342(6160), 863-866.
191. Verkhratsky, A., & Burnstock, G. (2014). Biology of purinergic signalling: its ancient evolutionary roots, its omnipresence and its multiple functional significance. Bioessays, 36(7), 697-705.
192. Wu, P., Wan, D., Xu, G., Wang, G., Ma, H., Wang, T., ... & Chen, W. (2017). An unusual protector-protégé strategy for the biosynthesis of purine nucleoside antibiotics. Cell chemical biology, 24(2), 171-181.
193. Mayer, A. M., Glaser, K. B., Cuevas, C., Jacobs, R. S., Kem, W., Little, R. D., ... & Shuster, D. E. (2010). The odyssey of marine pharmaceuticals: a current pipeline perspective. Trends in pharmacological sciences, 31(6), 255-265.
194. Bar-Yaacov, D., Mordret, E., Towers, R., Biniashvili, T., Soyris, C., Schwartz, S., ... & Pilpel, Y. (2017). RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. Genome research, 27(10), 1696-1703.

195. Verstraeten, N., Knapen, W. J., Kint, C. I., Liebens, V., Van den Bergh, B., Dewachter, L., ... & Michiels, J. (2015). Obg and membrane depolarization are part of a microbial bet-hedging strategy that leads to antibiotic tolerance. Molecular cell, 59(1), 9-21.
196. Nie, W., Wang, S., He, R., Xu, Q., Wang, P., Wu, Y., ... & Chen, G. (2020). A-to-I RNA editing in bacteria increases pathogenicity and tolerance to oxidative stress. PLoS pathogens, 16(8), e1008740.
197. Wolf, J., Gerber, A. P., & Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. The EMBO journal, 21(14), 3841-3851.
198. Elias, Y., & Huang, R. H. (2005). Biochemical and structural studies of A-to-I editing by tRNA: A34 deaminases at the wobble position of transfer RNA. Biochemistry, 44(36), 12057-12065.
199. Rafels-Ybern, À., Torres, A. G., Camacho, N., Herencia-Ropero, A., Roura Frigolé, H., Wulff, T. F., ... & Ribas de Pouplana, L. (2019). The expansion of inosine at the wobble position of tRNAs, and its role in the evolution of proteomes. Molecular biology and evolution, 36(4), 650-662.
200. Bar-Yaacov, D., Pilpel, Y., & Dahan, O. (2018). RNA editing in bacteria: occurrence, regulation and significance. RNA biology, 15(7), 863-867.
201. Moriwaki, Y., Yamamoto, T., & Higashino, K. (1999). Enzymes involved in purine metabolism-a review of histochemical localization and functional implications. Histology and histopathology, 14(4), 1321-1340.
202. Kaljas, Y., Liu, C., Skaldin, M., Wu, C., Zhou, Q., Lu, Y., ... & Zavialov, A. V. (2017). Human adenosine deaminases ADA1 and ADA2 bind to different subsets of immune cells. Cellular and Molecular Life Sciences, 74(3), 555-570.
203. Meyts, I., & Aksentijevich, I. (2018). Deficiency of adenosine deaminase 2 (DADA2): updates on the phenotype, genetics, pathogenesis, and treatment. Journal of clinical immunology, 38(5), 569-578.
204. Moens, L., Hershfield, M., Arts, K., Aksentijevich, I., & Meyts, I. (2019). Human adenosine deaminase 2 deficiency: A multi-faceted inborn error of immunity. Immunological Reviews, 287(1), 62-72.
205. Hirschhorn, R., & Ratech, H. (1980). Isozymes of adenosine deaminase. Isozymes, 4, 131-157.
206. Rao, K. S., Kumar, H. A., Rudresh, B. M., Srinivas, T., & Bhat, K. H. (2010). A Comparative study and evaluation of serum adenosine deaminase activity in the diagnosis of pulmonary tuberculosis. Biomed Res, 21(2), 184-8.
207. Sauer, A. V., Brigida, I., Carriglio, N., & Aiuti, A. (2012). Autoimmune dysregulation and purine metabolism in adenosine deaminase deficiency. Frontiers in Immunology, 3, 265.
208. Cekic, C., & Linden, J. (2016). Purinergic regulation of the immune system. Nature Reviews Immunology, 16(3), 177-192.
209. Cassani, B., Mirolo, M., Cattaneo, F., Benninghoff, U., Hershfield, M., Carlucci, F., ... & Aiuti, A. (2008). Altered intracellular and extracellular signaling leads to impaired T-cell functions in ADA-SCID patients. Blood, The Journal of the American Society of Hematology, 111(8), 4209-4219.
210. Kutryb-Zajac, B., Harasim, G., Jedrzejewska, A., Krol, O., Braczko, A., Jablonska, P., ... & Smolenski, R. T. (2021). Macrophage-derived adenosine deaminase 2 correlates with m2 macrophage phenotype in triple negative breast cancer. International journal of molecular sciences, 22(7), 3764.
211. Franco, R., Casadó, V., Ciruela, F., Saura, C., Mallol, J., Canela, E. I., & Lluis, C. (1997). Cell surface adenosine deaminase: much more than an ectoenzyme. Progress in neurobiology, 52(4), 283-294.
212. Zhou, Q., Yang, D., Ombrello, A. K., Zavialov, A. V., Toro, C., Zavialov, A. V., ... & Aksentijevich, I. (2014). Early-onset stroke and vasculopathy associated with mutations in ADA2. New England Journal of Medicine, 370(10), 911-920.
213. Dhanwani, R., Takahashi, M., Mathews, I. T., Lenzi, C., Romanov, A., Watrous, J. D., ... & Sharma, S. (2020). Cellular sensing of extracellular purine nucleosides triggers an innate IFN-β response. Science advances, 6(30), eaba3688.
214. Greiner-Tollersrud, O. K., Boehler, V., Bartok, E., Krausz, M., Polyzou, A., Schepp, J., ... & Proietti, M. (2020). ADA2 is a lysosomal DNase regulating the type-I interferon response. bioRxiv.
215. Huang, Z., Xie, N., Illes, P., Di Virgilio, F., Ulrich, H., Semyanov, A., ... & Tang, Y. (2021). From purines to purinergic signalling: molecular functions and human diseases. Signal Transduction and Targeted Therapy, 6(1), 1-20.
216. Ebrahim-Rad, M., Khatami, S., Ansari, S., Jalylfar, S., Valadbeigi, S., & Saghir, R. (2018). Adenosine deaminase 1 as a biomarker for diagnosis and monitoring of patients with acute lymphoblastic leukemia. Journal of Medical Biochemistry, 37(2), 128-133.
217. Wang, Y., Zheng, Y., & Beal, P. A. (2017). Adenosine deaminases that act on RNA (ADARs). The Enzymes, 41, 215-268.
218. Jin, Y., Zhang, W., & Li, Q. (2009). Origins and evolution of ADAR-mediated RNA editing. IUBMB life, 61(6), 572-578.
219. Li, X., Overton, I. M., Baines, R. A., Keegan, L. P., & O'Connell, M. A. (2013). The ADAR RNA editing enzyme controls neuronal excitability in Drosophila melanogaster. Nucleic acids research, 42(2), 1139-1151.

220. Wagner, R. W., Yoo, C., Wrabetz, L., Kamholz, J., Buchhalter, J., Hassan, N. F., ... & McMorris, F. A. (1990). Double-stranded RNA unwinding and modifying activity is detected ubiquitously in primary tissues and cell lines. Molecular and cellular biology, 10(10), 5586-5590.
221. Kumar, M., & Carmichael, G. G. (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proceedings of the National Academy of Sciences, 94(8), 3542-3547.
222. Cao, Y., Cao, R., Huang, Y., Zhou, H., Liu, Y., Li, X., ... & Hao, P. (2018). A comprehensive study on cellular RNA editing activity in response to infections with different subtypes of influenza a viruses. BMC genomics, 19(1), 43-56.
223. Cuevas, J. M., Combe, M., Torres-Puente, M., Garijo, R., Guix, S., Buesa, J., ... & Sanjuán, R. (2016). Human norovirus hyper-mutation revealed by ultra-deep sequencing. Infection, Genetics and Evolution, 41, 233-239.
224. Samuel, C. E. (2011). Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. Virology, 411(2), 180-193.
225. Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., ... & Li, J. B. (2017). Dynamic landscape and regulation of RNA editing in mammals. Nature, 550(7675), 249-254.
226. Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annual review of biochemistry, 79, 321.
227. Wang, Y., Chung, D. H., Monteleone, L. R., Li, J., Chiang, Y., Toney, M. D., & Beal, P. A. (2019). RNA binding candidates for human ADAR3 from substrates of a gain of function mutant expressed in neuronal cells. Nucleic acids research, 47(20), 10801-10814.
228. Washburn, M. C., Kakaradov, B., Sundararaman, B., Wheeler, E., Hoon, S., Yeo, G. W., & Hundley, H. A. (2014). The dsRBP and inactive editor ADR-1 utilizes dsRNA binding to regulate A-to-I RNA editing across the C. elegans transcriptome. Cell reports, 6(4), 599-607.
229. Walkley, C. R., & Li, J. B. (2017). Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. Genome biology, 18(1), 1-13.
230. Slotkin, W., & Nishikura, K. (2013). Adenosine-to-inosine RNA editing and human disease. Genome Med. 5, 105.
231. Bass, B. L. (2002). RNA editing by adenosine deaminases that act on RNA. Annual review of biochemistry, 71(1), 817-846.
232. Deffit, S. N., & Hundley, H. A. (2016). To edit or not to edit: regulation of ADAR editing specificity and efficiency. Wiley Interdisciplinary Reviews: RNA, 7(1), 113-127.
233. Wahlstedt, H., & f# x2013; hman, M. (2011). Site-selective versus promiscuous A-to-I editing. Wiley Interdisciplinary Reviews: RNA, 2(6), 761-771.
234. Zinshteyn, B., & Nishikura, K. (2009). Adenosine-to-inosine RNA editing. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 1(2), 202-209.
235. Licht, K., Hartl, M., Amman, F., Anrather, D., Janisiw, M. P., & Jantsch, M. F. (2019). Inosine induces context-dependent recoding and translational stalling. Nucleic acids research, 47(1), 3-14.
236. Wang, I. X., So, E., Devlin, J. L., Zhao, Y., Wu, M., & Cheung, V. G. (2013). ADAR regulates RNA editing, transcript stability, and gene expression. Cell reports, 5(3), 849-860.
237. Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., ... & Jantsch, M. F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nature biotechnology, 22(8), 1001-1005.
238. Blow, M., Futreal, P. A., Wooster, R., & Stratton, M. R. (2004). A survey of RNA editing in human brain. Genome research, 14(12), 2379-2387.
239. Cui, Y., Huang, T., & Zhang, X. (2015). RNA editing of microRNA prevents RNA-induced silencing complex recognition of target mRNA. Open biology, 5(12), 150126.
240. Pujantell, M., Riveira-Muñoz, E., Badia, R., Castellví, M., Garcia-Vidal, E., Sirera, G., ... & Ballana, E. (2017). RNA editing by ADAR1 regulates innate and antiviral immune functions in primary macrophages. Scientific reports, 7(1), 1-14.
241. Li, C. L., Yang, W. Z., Chen, Y. P., & Yuan, H. S. (2008). Structural and functional insights into human Tudor-SN, a key component linking RNA interference and editing. Nucleic acids research, 36(11), 3579-3589.
242. Iizasa, H., Wulff, B. E., Alla, N. R., Maragkakis, M., Megraw, M., Hatzigeorgiou, A., ... & Nishikura, K. (2010). Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency. Journal of Biological Chemistry, 285(43), 33358-33370.
243. Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., & Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science, 315(5815), 1137-1140.

244. Yang, W., Wang, Q., Howell, K. L., Lee, J. T., Cho, D. S. C., Murray, J. M., & Nishikura, K. (2005). ADAR1 RNA deaminase limits short interfering RNA efficacy in mammalian cells. Journal of Biological Chemistry, 280(5), 3946-3953.
245. Zhang, L., Yang, C. S., Varelas, X., & Monti, S. (2016). Altered RNA editing in 3′ UTR perturbs microRNA-mediated regulation of oncogenes and tumor-suppressors. Scientific reports, 6(1), 1-13.
246. Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., Ziv, T., ... & Eisenberg, E. (2017). Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell, 169(2), 191-202.
247. Ramaswami, G., & Li, J. B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic acids research, 42(D1), D109-D113.
248. Yablonovitch, A. L., Deng, P., Jacobson, D., & Li, J. B. (2017). The evolution and adaptation of A-to-I RNA editing. PLoS genetics, 13(11), e1007064.
249. Mannion, N., Arieti, F., Gallo, A., Keegan, L. P., & O'Connell, M. A. (2015). New insights into the biological role of mammalian ADARs; the RNA editing proteins. Biomolecules, 5(4), 2338-2362.
250. Bajad, P., Jantsch, M. F., Keegan, L., & O'Connell, M. (2017). A to I editing in disease is not fake news. RNA biology, 14(9), 1223-1231.
251. Tomaselli, S., Locatelli, F., & Gallo, A. (2014). The RNA editing enzymes ADARs: mechanism of action and human disease. Cell and tissue research, 356(3), 527-532.
252. Gallo, A., Vukic, D., Michalík, D., O'connell, M. A., & Keegan, L. P. (2017). ADAR RNA editing in human disease; more to it than meets the I. Human genetics, 136(9), 1265-1278.
253. Snyder, E., Chukrallah, L., Seltzer, K., Goodwin, L., & Braun, R. E. (2020). ADAD1 and ADAD2, testis-specific adenosine deaminase domain-containing proteins, are required for male fertility. Scientific reports, 10(1), 1-14.
254. Schumacher, J. M., Lee, K., Edelhoff, S., & Braun, R. E. (1995). Distribution of Tenr, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. Biology of reproduction, 52(6), 1274-1283.
255. Connolly, C. M., Dearth, A. T., & Braun, R. E. (2005). Disruption of murine Tenr results in teratospermia and male infertility. Developmental biology, 278(1), 13-21.
256. Rosani, U., Bai, C. M., Maso, L., Shapiro, M., Abbadi, M., Domeneghetti, S., ... & Venier, P. (2019). A-to-I editing of Malacoherpesviridae RNAs supports the antiviral role of ADAR1 in mollusks. BMC evolutionary biology, 19(1), 1-18.
257. Gerber, A., Grosjean, H., Melcher, T., & Keller, W. (1998). Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. The EMBO journal, 17(16), 4780-4789.
258. Grosjean, H., Auxilien, S., Constantinesco, F., Simon, C., Corda, Y., Becker, H. F., ... & Fourrey, J. L. (1996). Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs: a review. Biochimie, 78(6), 488-501.
259. Losey, H. C., Ruthenburg, A. J., & Verdine, G. L. (2006). Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. Nature structural & molecular biology, 13(2), 153-159.
260. Liu, H., Wang, Q., He, Y., Chen, L., Hao, C., Jiang, C., ... & Xu, J. R. (2016). Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. Genome research, 26(4), 499-509.
261. Blank-Landeshammer, B., Teichert, I., Märker, R., Nowrousian, M., Kück, U., & Sickmann, A. (2019). Combination of Proteogenomics with peptide de novo sequencing identifies new genes and hidden posttranscriptional modifications. MBio, 10(5), e02367-19.
262. Liu, H., Li, Y., Chen, D., Qi, Z., Wang, Q., Wang, J., ... & Xu, J. R. (2017). A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in Neurospora crassa. Proceedings of the National Academy of Sciences, 114(37), E7756-E7765.
263. Teichert, I., Dahlmann, T. A., Kück, U., & Nowrousian, M. (2017). RNA editing during sexual development occurs in distantly related filamentous ascomycetes. Genome biology and evolution, 9(4), 855-868.
264. Zhu, Y., Luo, H., Zhang, X., Song, J., Sun, C., Ji, A., ... & Chen, S. (2014). Abundant and selective RNA-editing events in the medicinal mushroom Ganoderma lucidum. Genetics, 196(4), 1047-1057.
265. Bian, Z., Ni, Y., Xu, J. R., & Liu, H. (2019). A-to-I mRNA editing in fungi: occurrence, function, and evolution. Cellular and Molecular Life Sciences, 76(2), 329-340.
266. Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015). Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Molecular cell, 59(5), 744-754.
267. Wang, Q., Liu, H., Xu, H., Hei, R., Zhang, S., Jiang, C., & Xu, J. R. (2019). Independent losses and duplications of autophagy-related genes in fungal tree of life. Environmental microbiology, 21(1), 226-243.
268. Gladyshev, E. (2017). Repeat-induced point mutation and other genome defense mechanisms in fungi. The fungal kingdom, 687-699.

269. Wang, C., Xu, J. R., & Liu, H. (2016). A-to-I RNA editing independent of ADARs in filamentous fungi. RNA biology, 13(10), 940-945.
270. Cao, S., He, Y., Hao, C., Xu, Y., Zhang, H., Wang, C., ... & Xu, J. R. (2017). RNA editing of the AMD1 gene is important for ascus maturation and ascospore discharge in Fusarium graminearum. Scientific reports, 7(1), 1-11.
271. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science, 315(5819), 1709-1712.
272. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., ... & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science, 321(5891), 960-964.
273. Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., ... & Tripathi, V. (2021). Novel CRISPR–Cas systems: an updated review of the current achievements, applications, and future research perspectives. International journal of molecular sciences, 22(7), 3327.
274. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. science, 337(6096), 816-821.
275. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. Science, 339(6121), 823-826.
276. Rouet, P., Smih, F., & Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proceedings of the National Academy of Sciences, 91(13), 6064-6068.
277. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature, 533(7603), 420-424.
278. Liang, P., Sun, H., Sun, Y., Zhang, X., Xie, X., Zhang, J., ... & Songyang, Z. (2017). Effective gene editing by high-fidelity base editor 2 in mouse zygotes. Protein & cell, 8(8), 601-611.
279. Rees, H. A., Komor, A. C., Yeh, W. H., Caetano-Lopes, J., Warman, M., Edge, A. S., & Liu, D. R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nature communications, 8(1), 1-10.
280. Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nature biotechnology, 35(4), 371-376.
281. Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., ... & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature, 556(7699), 57-63.
282. Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., ... & Toki, S. (2019). Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. Nature plants, 5(1), 14-17.
283. Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., ... & Kim, J. S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. Nature communications, 9(1), 1-10.
284. Chatterjee, P., Lee, J., Nip, L., Koseki, S. R., Tysinger, E., Sontheimer, E. J., ... & Jakimo, N. (2020). A Cas9 with PAM recognition for adenine dinucleotides. Nature communications, 11(1), 1-6.
285. Jakimo, N., Chatterjee, P., Nip, L., & Jacobson, J. M. (2018). A Cas9 with complete PAM recognition for adenine dinucleotides. BioRxiv, 429654.
286. Chatterjee, P., Jakimo, N., & Jacobson, J. M. (2018). Minimal PAM specificity of a highly similar SpCas9 ortholog. Science advances, 4(10), eaau0766.
287. Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., ... & Liu, D. R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C: G-to-T: A base editors with higher efficiency and product purity. Science advances, 3(8), eaao4774.
288. Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., ... & Liu, D. R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nature biotechnology, 36(9), 843-846.
289. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., ... & Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science, 361(6408), 1259-1262.
290. Doman, J. L., Raguram, A., Newby, G. A., & Liu, D. R. (2020). Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nature biotechnology, 38(5), 620-628.
291. Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., ... & Li, Z. (2020). Efficient base editing with high precision in rabbits using YFE-BE4max. Cell death & disease, 11(1), 1-11.
292. Yu, Y., Leete, T. C., Born, D. A., Young, L., Barrera, L. A., Lee, S. J., ... & Gaudelli, N. M. (2020). Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. Nature communications, 11(1), 1-10.

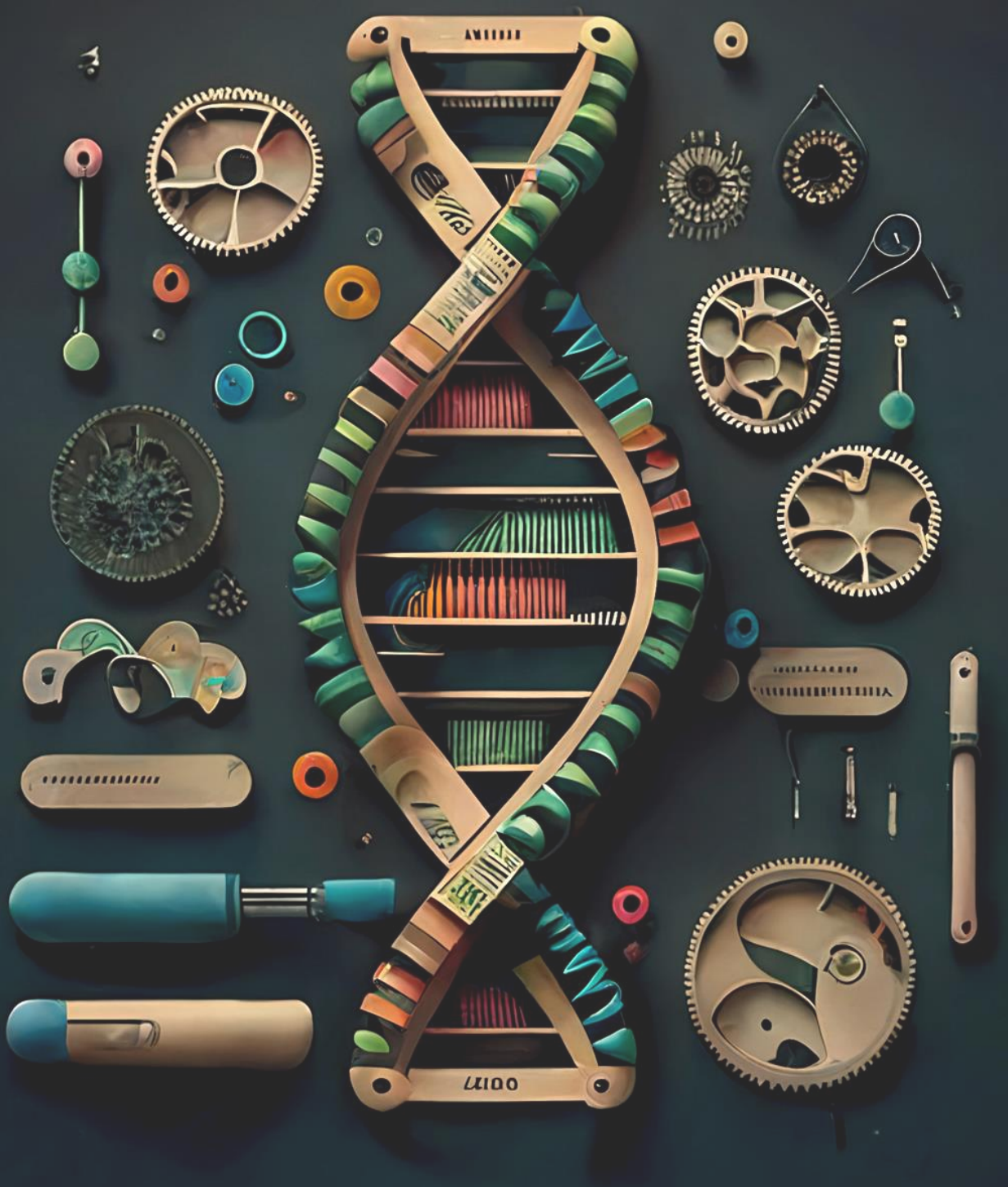
293. Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science, 368(6488), 290-296.
294. Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., ... & Liu, D. R. (2020). Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nature biotechnology, 38(4), 471-481.
295. Davis, J. R., Wang, X., Witte, I. P., Huang, T. P., Levy, J. M., Raguram, A., ... & Liu, D. R. (2022). Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. Nature Biomedical Engineering, 6(11), 1272-1283.
296. Agudelo, D., Carter, S., Velimirovic, M., Duringer, A., Rivest, J. F., Levesque, S., ... & Doyon, Y. (2020). Versatile and robust genome editing with Streptococcus thermophilus CRISPR1-Cas9. Genome research, 30(1), 107-117.
297. Liu, Z., Chen, S., Jia, Y., Shan, H., Chen, M., Song, Y., ... & Li, Z. (2021). Efficient and high-fidelity base editor with expanded PAM compatibility for cytidine dinucleotide. Science China Life Sciences, 64(8), 1355-1367.
298. Huang, T. P., Zhao, K. T., Miller, S. M., Gaudelli, N. M., Oakes, B. L., Fellmann, C., ... & Liu, D. R. (2019). Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. Nature biotechnology, 37(6), 626-631.
299. Wang, Y., Zhou, L., Liu, N., & Yao, S. (2019). BE-PIGS: a base-editing tool with deaminases inlaid into Cas9 PI domain significantly expanded the editing scope. Signal transduction and targeted therapy, 4(1), 1-3.
300. Jiang, W., Feng, S., Huang, S., Yu, W., Li, G., Yang, G., ... & Huang, X. (2018). BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. Cell research, 28(8), 855-861.
301. Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., ... & Chen, J. (2018). Base editing with a Cpf1–cytidine deaminase fusion. Nature biotechnology, 36(4), 324-327.
302. Tóth, E., Varga, E., Kulcsár, P. I., Kocsis-Jutka, V., Krausz, S. L., Nyeste, A., ... & Welker, E. (2020). Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. Nucleic acids research, 48(7), 3722-3733.
303. Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., ... & Joung, J. K. (2019). Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nature biotechnology, 37(3), 276-282.
304. Sontheimer, E. J., Gao, X., Mir, A., Edraki, A., Wolfe, S. A., & Liu, P. (2022). U.S. Patent Application No. 17/285,440.
305. Zafra, M. P., Schatoff, E. M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A. Y., ... & Dow, L. E. (2018). Optimized base editors enable efficient editing in cells, organoids and mice. Nature biotechnology, 36(9), 888-893.
306. Grünewald, J., Zhou, R., Garcia, S. P., Iyer, S., Lareau, C. A., Aryee, M. J., & Joung, J. K. (2019). Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature, 569(7756), 433-437.
307. Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W. H., Zheng, C., Newby, G. A., ... & Liu, D. R. (2019). Continuous evolution of base editors with expanded target compatibility and improved activity. Nature biotechnology, 37(9), 1070-1079.
308. Zeng, D., Liu, T., Tan, J., Zhang, Y., Zheng, Z., Wang, B., ... & Zhu, Q. (2020). PhieCBEs: plant high-efficiency cytidine base editors with expanded target range. Molecular Plant, 13(12), 1666-1669.
309. Wang, M., Wang, Z., Mao, Y., Lu, Y., Yang, R., Tao, X., & Zhu, J. K. (2019). Optimizing base editors for improved efficiency and expanded editing scope in rice. Plant Biotechnology Journal, 17(9), 1697.
310. Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., ... & Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nature biotechnology, 36(10), 977-982.
311. Tan, J., Zhang, F., Karcher, D., & Bock, R. (2020). Expanding the genome-targeting scope and the site selectivity of high-precision base editors. Nature communications, 11(1), 1-11.
312. Li, J., Xu, R., Qin, R., Liu, X., Kong, F., & Wei, P. (2021). Genome editing mediated by SpCas9 variants with broad non-canonical PAM compatibility in plants. Molecular plant, 14(2), 352-360.
313. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., ... & Yang, H. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature, 571(7764), 275-278.
314. Wang, X., Li, J., Wang, Y., Yang, B., Wei, J., Wu, J., ... & Yang, L. (2018). Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. Nature biotechnology, 36(10), 946-949.
315. Li, G., Sretenovic, S., Eisenstein, E., Coleman, G., & Qi, Y. (2021). Highly efficient C-to-T and A-to-G base editing in a Populus hybrid. Plant Biotechnology Journal, 19(6), 1086.
316. Randall, L. B., Sretenovic, S., Wu, Y., Yin, D., Zhang, T., Eck, J. V., & Qi, Y. (2021). Genome-and transcriptome-wide off-target analyses of an improved cytosine base editor. Plant Physiology, 187(1), 73-87.

317. Martin, A. S., Salamango, D. J., Serebrenik, A. A., Shaban, N. M., Brown, W. L., & Harris, R. S. (2019). A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. Scientific reports, 9(1), 1-8.
318. Jin, S., Fei, H., Zhu, Z., Luo, Y., Liu, J., Gao, S., ... & Gao, C. (2020). Rationally designed APOBEC3B cytosine base editors with improved specificity. Molecular cell, 79(5), 728-740.
319. Yu, W., Li, J., Huang, S., Li, X., Li, P., Li, G., ... & Huang, X. (2021). Harnessing A3G for efficient and selective C-to-T conversion at C-rich sequences. BMC biology, 19(1), 1-11.
320. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., ... & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science, 353(6305), aaf8729.
321. Ren, B., Liu, L., Li, S., Kuang, Y., Wang, J., Zhang, D., ... & Zhou, H. (2019). Cas9-NG greatly expands the targeting scope of the genome-editing toolkit by recognizing NG and other atypical PAMs in rice. Molecular plant, 12(7), 1015-1026.
322. Ren, Q., Sretenovic, S., Liu, S., Tang, X., Huang, L., He, Y., ... & Zhang, Y. (2021). PAM-less plant genome editing using a CRISPR–SpRY toolbox. Nature Plants, 7(1), 25-33.
323. Wu, Y., Xu, W., Wang, F., Zhao, S., Feng, F., Song, J., ... & Yang, J. (2019). Increasing cytosine base editing scope and efficiency with engineered Cas9-PmCDA1 fusions and the modified sgRNA in rice. Frontiers in genetics, 10, 379.
324. Sretenovic, S., Yin, D., Levav, A., Selengut, J. D., Mount, S. M., & Qi, Y. (2021). Expanding plant genome-editing scope by an engineered iSpyMacCas9 system that targets A-rich PAM sequences. Plant communications, 2(2), 100101.
325. Qin, R., Li, J., Li, H., Zhang, Y., Liu, X., Miao, Y., ... & Wei, P. (2019). Developing a highly efficient and wildly adaptive CRISPR-SaCas9 toolset for plant genome editing. Plant Biotechnology Journal, 17(4), 706.
326. Li, A., Mitsunobu, H., Yoshioka, S., Suzuki, T., Kondo, A., & Nishida, K. (2022). Cytosine base editing systems with minimized off-target effect and molecular size. Nature communications, 13(1), 1-8.
327. Xu, W., Song, W., Yang, Y., Wu, Y., Lv, X., Yuan, S., ... & Yang, J. (2019). Multiplex nucleotide editing by high-fidelity Cas9 variants with improved efficiency in rice. BMC plant biology, 19(1), 1-10.
328. Liu, Z., Shan, H., Chen, S., Chen, M., Zhang, Q., Lai, L., & Li, Z. (2019). Improved base editor for efficient editing in GC contexts in rabbits with an optimized AID-Cas9 fusion. The FASEB Journal, 33(8), 9210-9219.
329. Qin, R., Liao, S., Li, J., Li, H., Liu, X., Yang, J., & Wei, P. (2020). Increasing fidelity and efficiency by modifying cytidine base-editing systems in rice. The Crop Journal, 8(3), 396-402.
330. Liu, L. D., Huang, M., Dai, P., Liu, T., Fan, S., Cheng, X., ... & Meng, F. L. (2018). Intrinsic nucleotide preference of diversifying base editors guides antibody ex vivo affinity maturation. Cell Reports, 25(4), 884-892.
331. Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., ... & Zhou, H. (2018). Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. Molecular plant, 11(4), 623-626.
332. Wang, M., Xu, Z., Gosavi, G., Ren, B., Cao, Y., Kuang, Y., ... & Zhou, H. (2020). Targeted base editing in rice with CRISPR/ScCas9 system. Plant Biotechnology Journal, 18(8), 1645.
333. Xu, Z., Kuang, Y., Ren, B., Yan, D., Yan, F., Spetz, C., ... & Zhou, H. (2021). SpRY greatly expands the genome editing scope in rice with highly flexible PAM recognition. Genome biology, 22(1), 1-15.
334. Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., & Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nature methods, 13(12), 1029-1035.
335. Yuan, J., Ma, Y., Huang, T., Chen, Y., Peng, Y., Li, B., ... & Chang, X. (2018). Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. Molecular cell, 72(2), 380-394.
336. Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., ... & Liu, D. R. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature, 583(7817), 631-637.
337. Kang, B. C., Bae, S. J., Lee, S., Lee, J. S., Kim, A., Lee, H., ... & Kim, J. S. (2021). Chloroplast and mitochondrial DNA editing in plants. Nature Plants, 7(7), 899-905.
338. Nakazato, I., Okuno, M., Yamamoto, H., Tamura, Y., Itoh, T., Shikanai, T., ... & Arimura, S. I. (2021). Targeted base editing in the plastid genome of Arabidopsis thaliana. Nature Plants, 7(7), 906-913.
339. Li, R., Char, S. N., Liu, B., Liu, H., Li, X., & Yang, B. (2021). High-efficiency plastome base editing in rice with TAL cytosine deaminase. Molecular Plant, 14(9), 1412-1414.
340. Mok, B. Y., Kotrys, A. V., Raguram, A., Huang, T. P., Mootha, V. K., & Liu, D. R. (2022). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. Nature Biotechnology, 1-10.
341. Abudayyeh, O. O., Gootenberg, J. S., Franklin, B., Koob, J., Kellner, M. J., Ladha, A., ... & Zhang, F. (2019). A cytosine deaminase for programmable single-base RNA editing. Science, 365(6451), 382-386.
342. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. Nature, 551(7681), 464-471.

343. Hong, R., Ma, S., & Wang, F. (2019). Improving the specificity of adenine base editor using high-fidelity Cas9. bioRxiv, 712109.
344. Kim, D., Kim, D. E., Lee, G., Cho, S. I., & Kim, J. S. (2019). Genome-wide target specificity of CRISPR RNA-guided adenine base editors. Nature biotechnology, 37(4), 430-435.
345. Xu, L., Zhang, C., Li, H., Wang, P., Gao, Y., Mokadam, N. A., ... & Han, R. (2021). Efficient precise in vivo base editing in adult dystrophic mice. Nature communications, 12(1), 1-14.
346. Xu, X., Chemparathy, A., Zeng, L., Kempton, H. R., Shang, S., Nakamura, M., & Qi, L. S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Molecular Cell, 81(20), 4333-4345.
347. Hua, K., Tao, X., & Zhu, J. K. (2019). Expanding the base editing scope in rice by using Cas9 variants. Plant biotechnology journal, 17(2), 499-504.
348. Yang, L., Zhang, X., Wang, L., Yin, S., Zhu, B., Xie, L., ... & Li, D. (2018). Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. Protein & cell, 9(9), 814-819.
349. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., ... & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature, 523(7561), 481-485.
350. Hua, K., Tao, X., Yuan, F., Wang, D., & Zhu, J. K. (2018). Precise A· T to G· C base editing in the rice genome. Molecular plant, 11(4), 627-630.
351. Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., ... & Kim, J. S. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nature communications, 8(1), 1-12.
352. Li, X., Qian, X., Wang, B., Xia, Y., Zheng, Y., Du, L., ... & Lu, Z. (2020). Programmable base editing of mutated TERT promoter inhibits brain tumour growth. Nature cell biology, 22(3), 282-288.
353. Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., ... & Liu, D. R. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nature biotechnology, 38(7), 883-891.
354. Chu, S. H., Packer, M., Rees, H., Lam, D., Yu, Y., Marshall, J., ... & Slaymaker, I. M. (2021). Rationally designed base editors for precise editing of the sickle cell disease mutation. The CRISPR Journal, 4(2), 169-177.
355. Qin, W., Lu, X., Liu, Y., Bai, H., Li, S., & Lin, S. (2018). Precise A• T to G• C base editing in the zebrafish genome. BMC biology, 16(1), 1-8.
356. Schene, I. F., Joore, I. P., Oka, R., Mokry, M., van Vugt, A. H., van Boxtel, R., ... & Fuchs, S. A. (2020). Prime editing for functional repair in patient-derived disease models. Nature communications, 11(1), 1-8.
357. Liu, Z., Shan, H., Chen, S., Chen, M., Song, Y., Lai, L., & Li, Z. (2019). Efficient base editing with expanded targeting scope using an engineered Spy-mac Cas9 variant. Cell discovery, 5(1), 1-4.
358. Garneau, J. E., Dupuis, M. È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., ... & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature, 468(7320), 67-71.
359. Rees, H. A., Wilson, C., Doman, J. L., & Liu, D. R. (2019). Analysis and minimization of cellular RNA editing by DNA adenine base editors. Science advances, 5(5), eaax5717.
360. Grünewald, J., Zhou, R., Iyer, S., Lareau, C. A., Garcia, S. P., Aryee, M. J., & Joung, J. K. (2019). CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nature biotechnology, 37(9), 1041-1048.
361. Gaudelli, N. M., Lam, D. K., Rees, H. A., Solá-Esteves, N. M., Barrera, L. A., Born, D. A., ... & Ciaramella, G. (2020). Directed evolution of adenine base editors with increased activity and therapeutic application. Nature biotechnology, 38(7), 892-900.
362. Yan, D., Ren, B., Liu, L., Yan, F., Li, S., Wang, G., ... & Zhou, H. (2021). High-efficiency and multiplex adenine base editing in plants using new TadA variants. Molecular Plant, 14(5), 722-731.
363. Cho, S. I., Lee, S., Mok, Y. G., Lim, K., Lee, J., Lee, J. M., ... & Kim, J. S. (2022). Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. Cell, 185(10), 1764-1776.
364. Montiel-González, M. F., Vallecillo-Viejo, I. C., & Rosenthal, J. J. (2016). An efficient system for selectively altering genetic information within mRNAs. Nucleic acids research, 44(21), e157-e157.
365. Merkle, T., Merz, S., Reautschnig, P., Blaha, A., Li, Q., Vogel, P., ... & Stafforst, T. (2019). Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. Nature biotechnology, 37(2), 133-138.
366. Qu, L., Yi, Z., Zhu, S., Wang, C., Cao, Z., Zhou, Z., ... & Wei, W. (2019). Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. Nature biotechnology, 37(9), 1059-1069.
367. Cox, D. B., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., & Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science, 358(6366), 1019-1027.

368. Liu, Y., Mao, S., Huang, S., Li, Y., Chen, Y., Di, M., ... & Chi, T. (2020). REPAIR x, a specific yet highly efficient programmable A> I RNA base editor. The EMBO Journal, 39(22), e104748.
369. Koblan, L. W., Arbab, M., Shen, M. W., Hussmann, J. A., Anzalone, A. V., Doman, J. L., ... & Liu, D. R. (2021). Efficient C• G-to-G• C base editors developed using CRISPRi screens, target-library analysis, and machine learning. Nature biotechnology, 39(11), 1414-1425.
370. Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., ... & Zhang, X. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. Nature biotechnology, 39(1), 35-40.
371. Chen, L., Park, J. E., Paa, P., Rajakumar, P. D., Prekop, H. T., Chew, Y. T., ... & Chew, W. L. (2021). Programmable C: G to G: C genome editing with CRISPR-Cas9-directed base excision repair proteins. Nature communications, 12(1), 1-7.
372. Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L. M., ... & Joung, J. K. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nature biotechnology, 39(1), 41-46.
373. Sakata, R. C., Ishiguro, S., Mori, H., Tanaka, M., Tatsuno, K., Ueda, H., ... & Yachie, N. (2020). Base editors for simultaneous introduction of C-to-T and A-to-G mutations. Nature biotechnology, 38(7), 865-869.
374. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., ... & Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. Nature biotechnology, 38(7), 875-882.
375. Tao, W., Liu, Q., Huang, S., Wang, X., Qu, S., Guo, J., ... & Huang, X. (2021). CABE-RY: A PAM-flexible dual-mutation base editor for reliable modeling of multi-nucleotide variants. Molecular Therapy-Nucleic Acids, 26, 114-121.
376. Xie, J., Huang, X., Wang, X., Gou, S., Liang, Y., Chen, F., ... & Lai, L. (2020). ACBE, a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems. BMC biology, 18(1), 1-14.
377. Grünewald, J., Zhou, R., Lareau, C. A., Garcia, S. P., Iyer, S., Miller, B. R., ... & Joung, J. K. (2020). A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. Nature biotechnology, 38(7), 861-864.
378. Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., ... & Li, D. (2020). Dual base editor catalyzes both cytosine and adenine base conversions in human cells. Nature biotechnology, 38(7), 856-860.
379. Wood, R. D. (1996). DNA repair in eukaryotes. Annual review of biochemistry, 65(1), 135-167.
380. Sancar, A. (1996). DNA excision repair. Annual review of biochemistry, 65(1), 43-81.
381. Molla, K. A., & Yang, Y. (2019). CRISPR/Cas-mediated base editing: technical considerations and practical applications. Trends in biotechnology, 37(10), 1121-1142.
382. Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., ... & Gao, C. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nature Biotechnology, 36(10), 950-953.
383. Rees, H. A., & Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nature reviews genetics, 19(12), 770-788.
384. Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., ... & Li, Z. (2020). Precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. BMC biology, 18(1), 1-14.
385. Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., ... & Aihara, H. (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nature structural & molecular biology, 24(2), 131-139.
386. Macbeth, M. R., Schubert, H. L., VanDemark, A. P., Lingam, A. T., Hill, C. P., & Bass, B. L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. Science, 309(5740), 1534-1539.
387. Azameti, M. K., & Dauda, W. P. (2021). Base editing in plants: applications, challenges, and future prospects. Frontiers in Plant Science, 1531.
388. Wang, Y., Liu, Y., Zheng, P., Sun, J., & Wang, M. (2021). Microbial base editing: a powerful emerging technology for microbial genome engineering. Trends in Biotechnology, 39(2), 165-180.
389. Wang, X., Liu, Z., Li, G., Dang, L., Huang, S., He, L., ... & Ma, X. (2020). Efficient gene silencing by adenine base editor-mediated start codon mutation. Molecular Therapy, 28(2), 431-440.
390. Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., ... & Gao, C. (2018). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. Genome biology, 19(1), 1-9.
391. Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., & Gao, C. (2020). Fine-tuning sugar content in strawberry. Genome biology, 21(1), 1-14.
392. Li, Z., Xiong, X., Wang, F., Liang, J., & Li, J. F. (2019). Gene disruption through base editing-induced messenger RNA missplicing in plants. New Phytologist, 222(2), 1139-1148.
393. Xue, C., Zhang, H., Lin, Q., Fan, R., & Gao, C. (2018). Manipulating mRNA splicing by base editing in plants. Science China Life Sciences, 61(11), 1293-1300.

394. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., ... & Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. Nature biotechnology, 38(7), 875-882.
395. Yeh, W. H., Chiang, H., Rees, H. A., Edge, A. S., & Liu, D. R. (2018). In vivo base editing of post-mitotic sensory cells. Nature communications, 9(1), 1-10.
396. Wang, Y., Liu, Y., Liu, J., Guo, Y., Fan, L., Ni, X., ... & Ma, Y. (2018). MACBETH: multiplex automated Corynebacterium glutamicum base editing method. Metabolic Engineering, 47, 200-210.
397. Tang, W., & Liu, D. R. (2018). Rewritable multi-event analog recording in bacterial and mammalian cells. Science, 360(6385), eaap8992.
398. Farzadfard, F., Gharaei, N., Higashikuni, Y., Jung, G., Cao, J., & Lu, T. K. (2019). Single-nucleotide-resolution computing and memory in living cells. Molecular cell, 75(4), 769-780.
399. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., ... & Koonin, E. V. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nature reviews microbiology, 15(3), 169-182.
400. Liu, J. J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L., ... & Doudna, J. A. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature, 566(7743), 218-223.
401. Kim, D. Y., Lee, J. M., Moon, S. B., Chin, H. J., Park, S., Lim, Y., ... & Kim, Y. S. (2022). Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. Nature biotechnology, 40(1), 94-102.
402. Winter, J., Luu, A., Gapinske, M., Manandhar, S., Shirguppe, S., Woods, W. S., ... & Perez-Pinera, P. (2019). Targeted exon skipping with AAV-mediated split adenine base editors. Cell discovery, 5(1), 1-12.
403. Chen, Y., Zhi, S., Liu, W., Wen, J., Hu, S., Cao, T., ... & Huang, J. (2020). Development of Highly Efficient Dual-AAV Split Adenosine Base Editor for In Vivo Gene Therapy. Small Methods, 4(9), 2000309.
404. Levy, J. M., Yeh, W. H., Pendse, N., Davis, J. R., Hennessey, E., Butcher, R., ... & Liu, D. R. (2020). Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. Nature biomedical engineering, 4(1), 97-110.
405. Kim, K., Ryu, S. M., Kim, S. T., Baek, G., Kim, D., Lim, K., ... & Kim, J. S. (2017). Highly efficient RNA-guided base editing in mouse embryos. Nature biotechnology, 35(5), 435-437.
406. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A., & Liu, D. R. (2015). Small molecule–triggered Cas9 protein with improved genome-editing specificity. Nature chemical biology, 11(5), 316-318.
407. Gao, X., Tao, Y., Lamas, V., Huang, M., Yeh, W. H., Pan, B., ... & Liu, D. R. (2018). Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. Nature, 553(7687), 217-221.
408. Tang, W., Hu, J. H., & Liu, D. R. (2017). Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. Nature communications, 8(1), 1-8.
409. Kundert, K., Lucas, J. E., Watters, K. E., Fellmann, C., Ng, A. H., Heineike, B. M., ... & Kortemme, T. (2019). Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. Nature communications, 10(1), 1-11.
410. Zeng, J., Wu, Y., Ren, C., Bonanno, J., Shen, A. H., Shea, D., ... & Bauer, D. E. (2020). Therapeutic base editing of human hematopoietic stem cells. Nature Medicine, 26(4), 535-541.
411. Wang, L., Li, L., Ma, Y., Hu, H., Li, Q., Yang, Y., ... & Li, D. (2020). Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. Cell research, 30(3), 276-278.
412. Webber, B. R., Lonetree, C. L., Kluesner, M. G., Johnson, M. J., Pomeroy, E. J., Diers, M. D., ... & Moriarity, B. S. (2019). Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. Nature communications, 10(1), 1-10.
413. Park, D. S., Yoon, M., Kweon, J., Jang, A. H., Kim, Y., & Choi, S. C. (2017). Targeted base editing via RNA-guided cytidine deaminases in Xenopus laevis embryos. Molecules and cells, 40(11), 823.
414. Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., ... & Joung, J. K. (2019). Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nature biotechnology, 37(3), 276-282.
415. Jiang, T., Henderson, J. M., Coote, K., Cheng, Y., Valley, H. C., Zhang, X. O., ... & Xue, W. (2020). Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nature communications, 11(1), 1-9.
416. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., ... & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 576(7785), 149-157.



Chapter 3

Characterizing a type III CRISPR-Cas-associated adenosine deaminase

Despoina Trasanidou*, Victor Pool*, Jurre Steens, Carl Salazar, John van der Oost & Raymond H. J. Staals**

Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

*Equal contribution

**Corresponding author

Abstract

Type III CRISPR-Cas systems provide multi-faceted defense against invading viruses not only via degradation of viral RNA and DNA but also through generation of cyclic oligoadenylate (cOA) signaling molecules that bind CARF proteins to allosterically activate diverse effector domains. Despite their widespread presence and high variability, the vast majority of CARF proteins has remained unexplored. Here, we demonstrate that a predicted CARF-linked adenosine deaminase enzyme from *Methylacidiphilum fumariolicum* (named mfCARF-AD) is a putative cOA-degrading ring-nuclease and cOA-activated deaminase with no, or at best weak, activity on free adenosine nucleotides. We show that mfCARF-AD does not act on free adenine and AMP, creating the hypothesis of alternative substrate specificity. Finally, we exhibit that mfCARF-AD forms oligomers, is associated to a type III-B CRISPR-Cas system, and interacts with purine salvage and carbon metabolism proteins, probably to block biological processes crucial for viral replication.

Keywords

Adenosine deaminase, CRISPR-Cas, type III, CARF

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins are adaptive immune systems that mediate antiviral response in bacteria and archaea[1-3]. CRISPR-Cas systems are divided into two classes. Class 1 systems (types I, III, and IV) comprise a multi-subunit effector complex, while Class 2 systems (type II, V, and VI) contain a single multidomain effector protein[4]. Type III systems are of particular interest because of their unique ability to cleave both the viral mRNA specifically and the adjacent non-template ssDNA of the transcription bubble non-specifically[5-8]. In addition, mRNA binding may trigger the production of cyclic oligoadenylate molecules (cOAs) that allosterically activate type III-associated proteins, by binding to their CRISPR associated Rossman fold (CARF) domain[9,10]. CARF activation can be temporarily controlled due to the capability of some CARF domains to degrade bound cOA-molecules, which is termed ring nuclease activity[11]. CARF-containing proteins are typically RNases that degrade phage- or host-derived mRNA transcripts in a sequence-nonspecific fashion[12], and DNases that either nick supercoiled DNA to collapse DNA replication forks of viral genes[13] or introduce lethal, double-stranded DNA (dsDNA) breaks at the bacterial genome (or plasmids) to block phage replication[14]. Alternatively, CARF proteins may act as transcriptional regulators, membrane proteins, proteases, nitrilases, adenylate cyclases, or even adenosine deaminases with hitherto unknown biological roles in host defense[15].

Adenosine deaminases (ADs) are enzymes with the ability to convert adenosine (A) to inosine (I)[16]. ADs are present in both prokaryotes and eukaryotes, and act either on free nucleotides/ nucleosides/ nucleobases or on nucleic acids (mainly dsRNA; rarely ssRNA), depending on their biological role. Especially, in prokaryotes, ADs acting on free nucleotides play a crucial role in the purine salvage pathway, intercellular communication, host protection, and bacterial virulence, while ADs acting on nucleic acids that participate in the degeneration of protein decoding, biofilm formation, bacterial virulence, and adaptation at harsh conditions[17]. Hence, the predicted CARF ADs may be activated by cOAs produced by type III CRISPR-Cas systems to protect the bacterial host against viral infection. However, their deamination ability, substrate specificity, and biological role still remain to be elucidated.

During the last decade, CRISPR-Cas systems have been exploited as genetic engineering tools in a wide variety of organisms[18-22]. In 2017, ADs were combined for first time with the CRISPR-Cas technology in order to introduce precise genomic point mutations (A•T to

G•C), circumventing the need for double-strand breaks (DSBs) and donor DNA templates[23]. The first generation of adenine base editor (called ABE7.10) comprised of a catalytically impaired Cas9 protein (D10A nickase Cas9; nCas9) fused to an evolved and engineered variant of the tRNA adenosine deaminase A from *Escherichia coli* (TadA*7.10) that overcomes the inability of ADs to edit DNA[17,23]. To optimize its DNA editing efficiency, the *tadA*7.10* gene was further evolved (*tadA-8e*), generating the ABE8e editor that exhibits up to 1,100 times faster deamination activity[24,25]. However, increased off-target editing has also been observed on both the DNA and RNA level (even with the *tada-8e* V106W mutant version)[24], hampering the safe use of ABE8e for clinical applications.

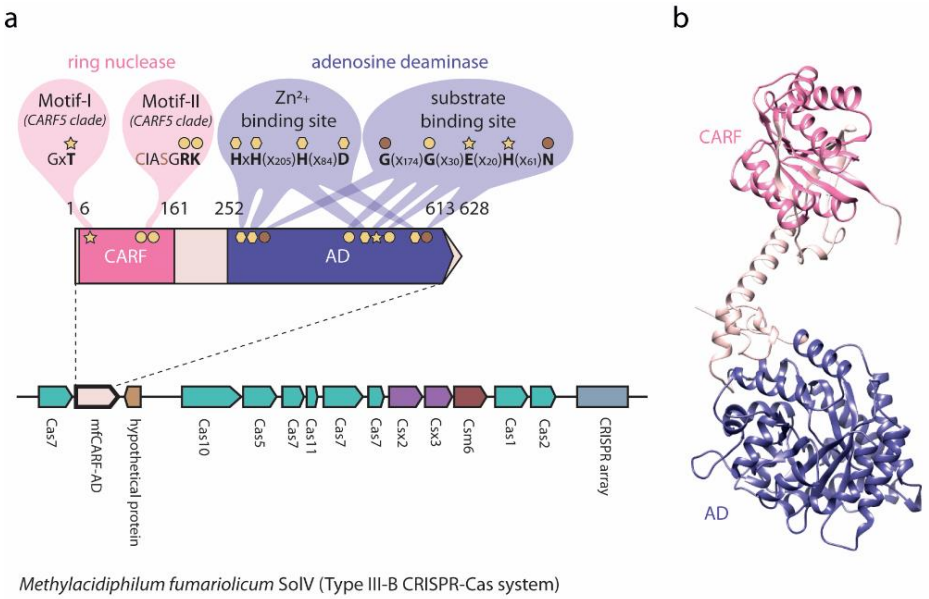
Here, we characterize a predicted CARF-AD protein from the thermophilic bacterium *Methylacidiphilum fumariolicum* SolV (MFUM_700012; herein referred to as 'mfCARF-AD')[26]. We investigate its deaminase activity, substrate specificity, cOA-dependence, and oligomeric state, aiming not only to reveal its biological role, but also to investigate its potential for the development of a cOA-activated, type III ABE with strictly controlled catalytic activity and thus negligible off-target editing.

Results

In silico prediction of mfCARF-AD enzymatic activity and substrate specificity

Analysis of the protein domain architecture of the mfCARF-AD protein has been performed in order to understand its enzymatic activity and substrate specificity. mfCARF-AD (MFUM_700012, WP_009060269.1) is a 628 amino acid protein that contains an N-terminal CARF domain and a C-terminal AD domain (Fig. 1a, b). mfCARF-AD has been previously assigned to the CARF5-clade[26]. Members of this clade have two conserved ring nuclease motifs (motif-I: Gx(T/S); motif-II: S(I/L)AGGRK) in the CARF domain, suggesting their ability to degrade cOAs[26]. Motif-II (mainly residues R and K) is crucial for cOA binding, while motif-I (mainly residue T or S) is largely responsible for cOA degradation[26]. As both CARF motifs are present in mfCARF-AD, it is concluded that all residues essential for ligand (cOA) binding and ring nuclease activity are conserved (Fig. 1a; Supplementary Fig. 1). So, this may allow for temporal control of the cOA-based activation of its C-terminal deaminase effector domain.

Like several other CARF5-clade members, the effector domain of mfCARF-AD is a predicted AD domain[26] (Fig. 1a, b). The annotation of mfCARF-AD as a deaminase was confirmed by a Phyre2-based structural analysis that revealed three ADs as the closest structural homologs of mfCARF-AD (Supplementary Fig. 2, 3). To examine its substrate specificity and thus its enzymatic activity, we compared the zinc-binding, the substrate-binding, and the active sites of mfCARF-AD with those of adenosine deaminases (ADs) that convert adenosine into inosine, acting either on nucleic acids (ADNs)[27-35] or on free adenosines (ADAs)[36-39]. In addition, we compared mfCARF-AD with other members of the amidohydrolase superfamily, such as adenine deaminases (ADEs) that convert free adenine into hypoxanthine[40-42], and adenosine monophosphate (AMP) deaminases (AMPDs) that convert free AMP into inosine monophosphate (IMP)[43-45]. Our analyses on the primary amino acid level showed that mfCARF-AD is highly divergent from ADNs (Supplementary Fig. 4, 5). Moreover, mfCARF-AD differed from ADEs and AMPDs, missing essential residues involved in ribose occlusion and phosphate binding, respectively[46,47] (Supplementary Fig. 6, 7). In contrast, high similarity between the AD domain of mfCARF-AD and ADAs was observed, presenting identical residues for the zinc-binding site (H266, H268, H473, D558) and the catalytic site (E476, H497) (Fig. 1a; Supplementary Fig. 8). However, only the second (G445) of the three substrate-binding residues of mfCARF-AD was identical to that of ADAs, while the first (G270) and the third (N559) substrate-binding residues of mfCARF-AD were distinct from those of most ADAs that contained an aspartate residue at both positions, implying different substrate affinity and/or specificity (Fig. 1a; Supplementary Fig. 8). The replacement of the aspartate by the asparagine at the third substrate-binding position has been previously reported to cause 99% decrease of substrate affinity in mouse ADA (mADA), as aspartate is necessary for appropriate positioning of the substrate to the active site[48]. The N559 mutation is conserved among some ADAs and the vast majority of CARF-ADs from the CARF5-clade (Supplementary Fig. 1, 8), suggesting that mfCARF-AD and these homologs share similar catalytic efficiency and/or substrate specificity. One of these proteins, the ADA from *Streptomyces coelicolor* (scADA1), has been experimentally characterized[48,49]. Like the deactivated mADA, ScADA1 has a very low turnover rate of adenosine[48,49]. In contrast, the ADA from *Escherichia coli* (ecADA1) contains the conserved to most ADAs aspartate residue, and exhibits much higher affinity to adenosine[50]. Therefore, we conclude that mfCARF-AD is probably not an ADN, an ADE, or an AMPD, but rather an AD with a different substrate specificity and/or catalytic activity than most ADAs.



Methylacidiphilum fumariolicum SolV (Type III-B CRISPR-Cas system)

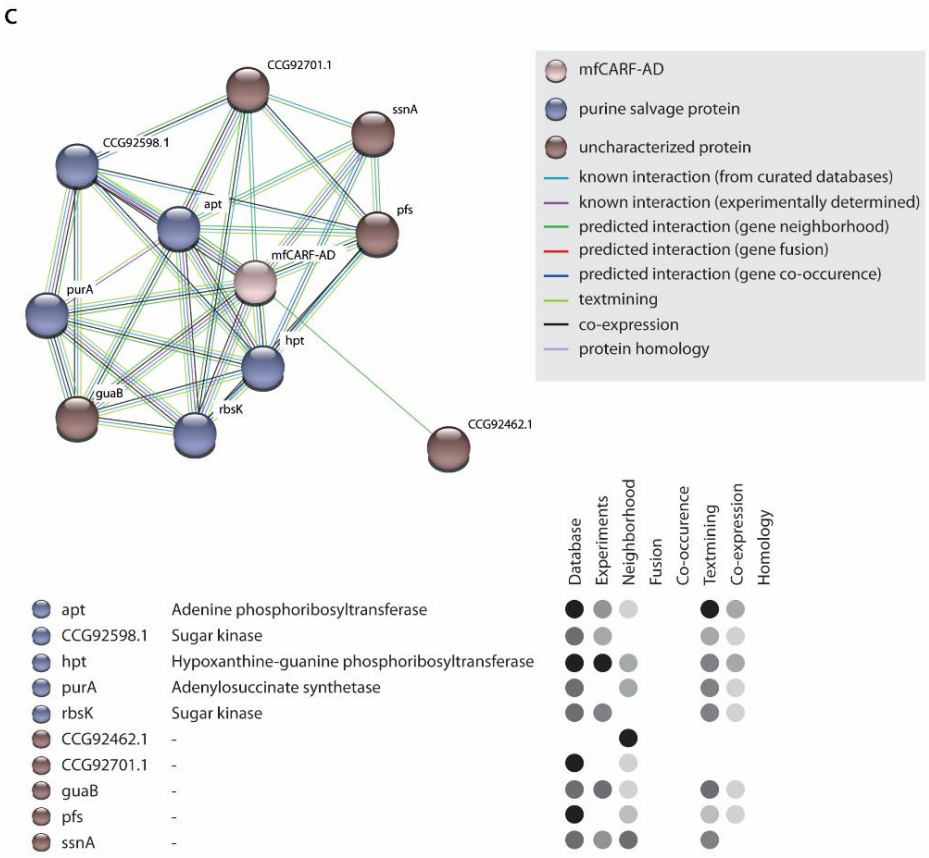


Fig. 1 *In silico* characterization of mfCARF-AD. **a.** Domain architecture and genomic locus of mfCARF-AD. The CARF and AD domains are shown as pink and blue boxes, respectively. Yellow symbols indicate the presence of conserved site residues: active site (stars), substrate binding site (circles), and zinc binding site (hexagons). Brown circles indicate substrate binding residues that are not conserved between mfCARF-AD (and homologs) and other adenosine deaminases. **b.** Three dimensional structure of mfCARF-AD predicted by AlphaFold. Pink, and blue colors represent the CARF and AD domains, respectively. **c.** Functional protein association network of mfCARF-AD using STRING analysis. Blue and brown circles indicate purine salvage and uncharacterized proteins, respectively, while colored lines show known or predicted interactions, text mining, co-expression, and protein homology.

Altogether, our *in silico* study suggests that mfCARF-AD is a putative cOA-degrading ring-nuclease and a predicted cOA-activated AD with probably weak, or no, activity on free adenosines.

In silico prediction of mfCARF-AD protein interactions

To understand the biological role of mfCARF-AD, we subsequently investigated its genomic context. We found that the *mfcarf-ad* gene is located adjacent to the type III-B CRISPR-Cas system of *M. fumariolicum* SolV (Fig. 1a). To examine if this pattern is conserved among the CARF5-clade members, we additionally studied the genomic context of ten other CARF5 proteins with a similar domain organization and motifs (i.e. CARF-AD) as mfCARF-AD (Supplementary Fig. 1). In agreement with a previous study[26], the vast majority of the CARF5 genes were found to reside in the proximity of gene clusters encoding type III-A and III-B CRISPR-Cas systems (Supplementary Fig. 9), strongly suggesting a functional link of the CARF5 proteins with type III defense. This hypothesis was confirmed via STRING analysis that indicated direct interaction of CARF5 proteins with gene clusters encoding type III CRISPR-Cas complexes (Supplementary Fig. 10).

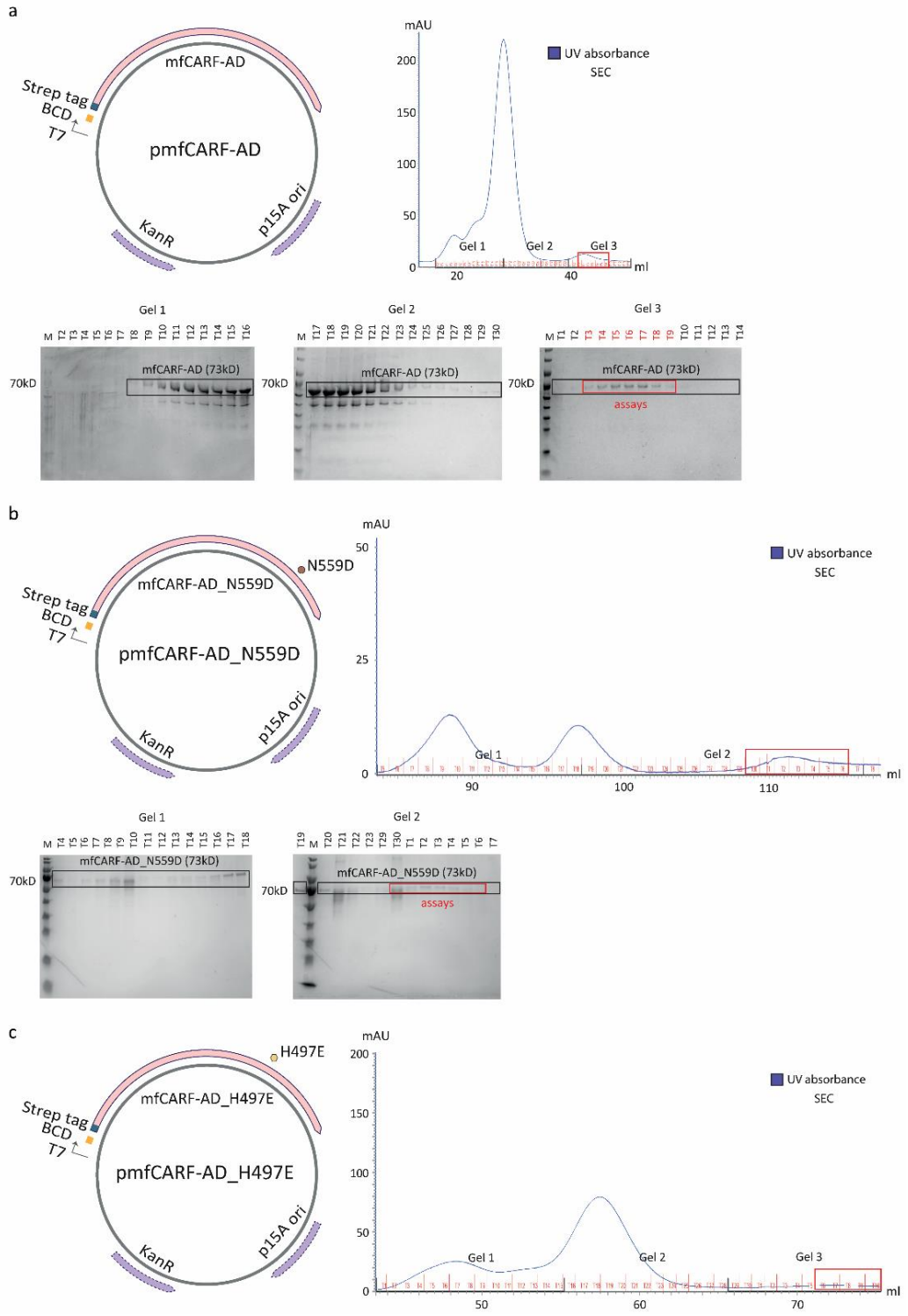
Furthermore, we examined the presence of additional, “non-Cas” genes in the same genomic loci. Although the *mfcarf-ad* gene was flanked only by Cas genes (Fig. 1a), other CARF5 genes analyzed were flanked predominantly either by genes that encode proteins with potential immunity-related functions (such as nucleases, peptidases, transposases, toxins/antitoxins, antibiotic resistance proteins) that sometimes also contained a CARF domain, or by purine salvage genes (like phosphoribosyl-transferases) (Supplementary Fig. 9). Intriguingly, in almost half of the tested CARF5 genomic loci, the CARF5 genes appeared to be positioned in close proximity to a DUF1887 domain-containing gene (also

known as 'CRISPR ancillary nuclease 2'; 'Can2'), which encodes a CARF-nuclease that degrades dsDNA[51]. Given the consistency of this observation (Supplementary Fig. 9), we speculate that CARF5 proteins may be co-expressed with DUF1887 proteins in the presence of type III CRISPR-Cas-generated cOAs, either to combine their enzymatic activities against viral infection or to suppress the DUF1887 CARF-nuclease activity through CARF5 ring nuclease-mediated cOA degradation, reducing collateral damage and allowing the cell to recover through waking up from a dormant state[52]. STRING analysis further supports this idea by suggesting a functional interaction between CARF5s and DUF1887 proteins (Supplementary Fig. 10). Given the absence of *duf1887* genes in the remaining genomic loci, including that of *mfcarf-ad*, it is also possible that CARF5s may alternatively interact with diverse other immunity or purine salvage proteins upon viral infection, or it may even act alone. In this line, STRING analysis showed that all tested CARF5s interact with multiple purine salvage proteins (Fig. 1c; Supplementary Fig. 10). Especially mfCARF-AD, like most deaminases acting on free adenosines, was found to associate with adenine phosphoribosyltransferase, hypoxanthine-guanine phosphoribosyltransferase, adenylosuccinate synthetase, and sugar kinases that are involved in purine biosynthesis and carbon metabolism[53] (Fig. 1c).

Overall, similar to all CARF5s, mfCARF-AD is tightly linked to a type III CRISPR-Cas system, and is predicted to interact with purine salvage and carbon metabolism proteins. However, to reveal details of the functionality of the AD domain of mfCARF-AD an experimental analysis is required.

Purification of wild-type and mutant mfCARF-AD variants

To experimentally investigate the enzymatic activity and substrate specificity of mfCARF-AD, we generated the pmfCARF-AD expression plasmid including the codon harmonized *mfcarf-ad* gene fused to an N-terminal Strep-tag and expressed under the control of the T7 promoter and the BCD system[54] (Fig. 2a). In parallel, we substituted the theoretically defective N559 residue of mfCARF-AD with an aspartate residue to generate the mfCARF-AD_N559D mutant variant with theoretically restored adenosine deaminase activity, based on the aforementioned *in silico* analysis and previous experimental studies[48-50] (Supplementary Fig. 11a). As negative control, we replaced one of the crucial for zinc binding (H497) residues of mfCARF-AD with a glutamate residue (mfCARF-AD_H497E; Supplementary Fig. 11b), previously reported to strongly reduce deaminase activity[46,55].


a
mfCARF-AD
Strep tag
BCD
T7
pmfCARF-AD
KanR
p15A ori
mAU
UV absorbance
SEC
Gel 1
Gel 2
Gel 3
ml
Gel 1
Gel 2
Gel 3
70kD
mfCARF-AD (73kD)
assays
b
mfCARF-AD_N559D
N559D
pmfCARF-AD_N559D
mfCARF-AD_N559D (73kD)
c
mfCARF-AD_H497E
H497E
pmfCARF-AD_H497E

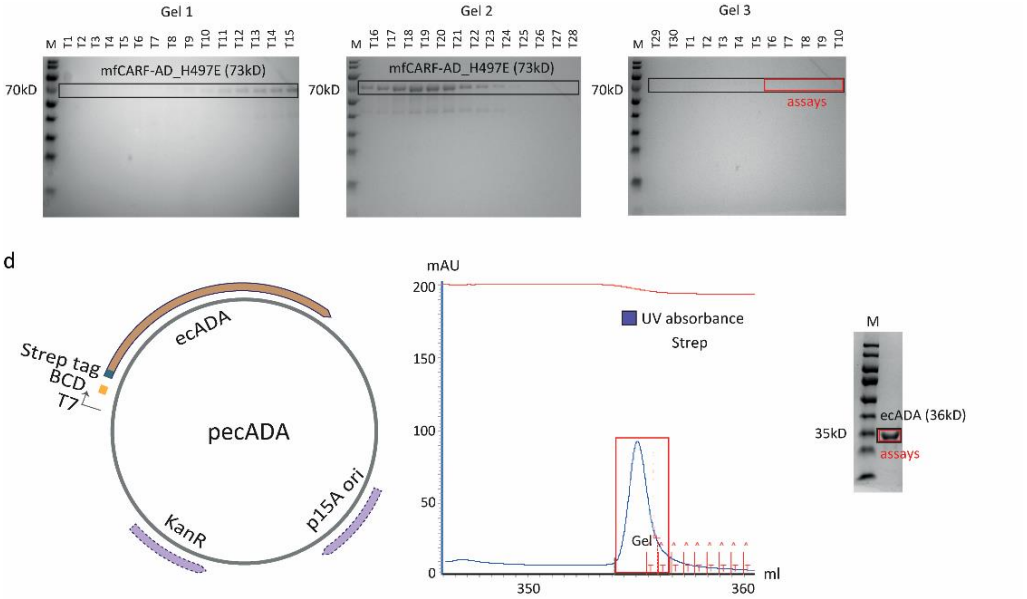


Fig. 2 Purification of mfCARF-AD variants and ecADA1. Plasmid map, chromatograph and SDS-PAGE gels from the purification of (**a**) mfCARF-AD, (**b**) mfCARF-AD_N559D, (**c**) mfCARF-AD_H497E, and (**d**) ecADA1. The wild-type and mutant mfCARF-AD variants were purified using both a Strep and a SEC column, while ecADA1 was purified only with Strep column. Red boxes indicate the samples used for the *in vitro* assays.

As positive control, the experimentally characterized ecADA1 was selected because of its high deaminase activity[50]. We replaced the *mfcarf-ad* gene in pmfCARF-AD with the codon harmonized *mfcarf-ad_n559d*, *mfcarf-ad_h497e*, and *ecada1* genes, creating the pmfCARF-AD_N559D, the pmfCARF-AD_ H497E, and the pecADA1 constructs, respectively (Fig. 2b-d). After transformation of each expression plasmid in the *E. coli* BL21(DE3) strain and overexpression of the corresponding proteins, we performed a two-step protein purification (Strep-tag, Supplementary Fig. 12; Size Exclusion Chromatography, Fig. 2a-c) for the wild-type/mutant mfCARF-AD variants and a one-step purification (Strep-tag, Fig. 2d) for ecADA1. Fractions of high purity (Fig. 2) were used for subsequent *in vitro* experiments.

Deaminase activity of wild-type and mutant mfCARF-AD variants on free adenosine and derivatives

We continued by evaluating the deaminase activity of mfCARF-AD, and of its theoretically activated (mfCARF-AD_N559D) and deactivated (mfCARF-AD_ H497E) variants for deamination of free adenosine, the *in silico* predicted substrate of these enzymes. *In vitro* assays based on the difference in absorbance at 265nm between the deamination substrate adenosine and the product inosine demonstrated negligible deaminase activity for all mfCARF-AD variants, either in the presence or in the absence of signaling $cOA_{2\text{-}6}$ molecules (Fig. 3a). In contrast, the positive deaminase control ecADA1 showed complete conversion of adenosine into inosine (Fig. 3a). Next, we examined whether the catalytic activity of the mfCARF-AD protein depends on the concentration of protein, cOAs, or zinc. However, also under these conditions, no deaminase activity was detected in any of the mfCARF-AD variants (Supplementary Fig. 13). Likewise, in agreement with our *in silico* prediction, no activity was detected on adenosine-derived substrates variation of the substrate (adenine, AMP, and ATP) (Fig. 3b). The slight change in absorbance for these compounds was similar to that observed for adenosine and inosine in the presence of any protein (Fig. 3a; Supplementary Fig. 13). In the absence of protein, this effect was not observed, suggesting that proteins may affect the measured absorbance in this assay. Moreover, mock controls indicated no correlation between buffer and absorbance change (Supplementary Fig. 14). On the whole, our findings indicate that mfCARF-AD, even after restoration of its deaminase activity, is not able to deaminate adenine, neither the stand-alone base nor the base as part of the nucleoside (adenosine) or the nucleotide (AMP, ATP).

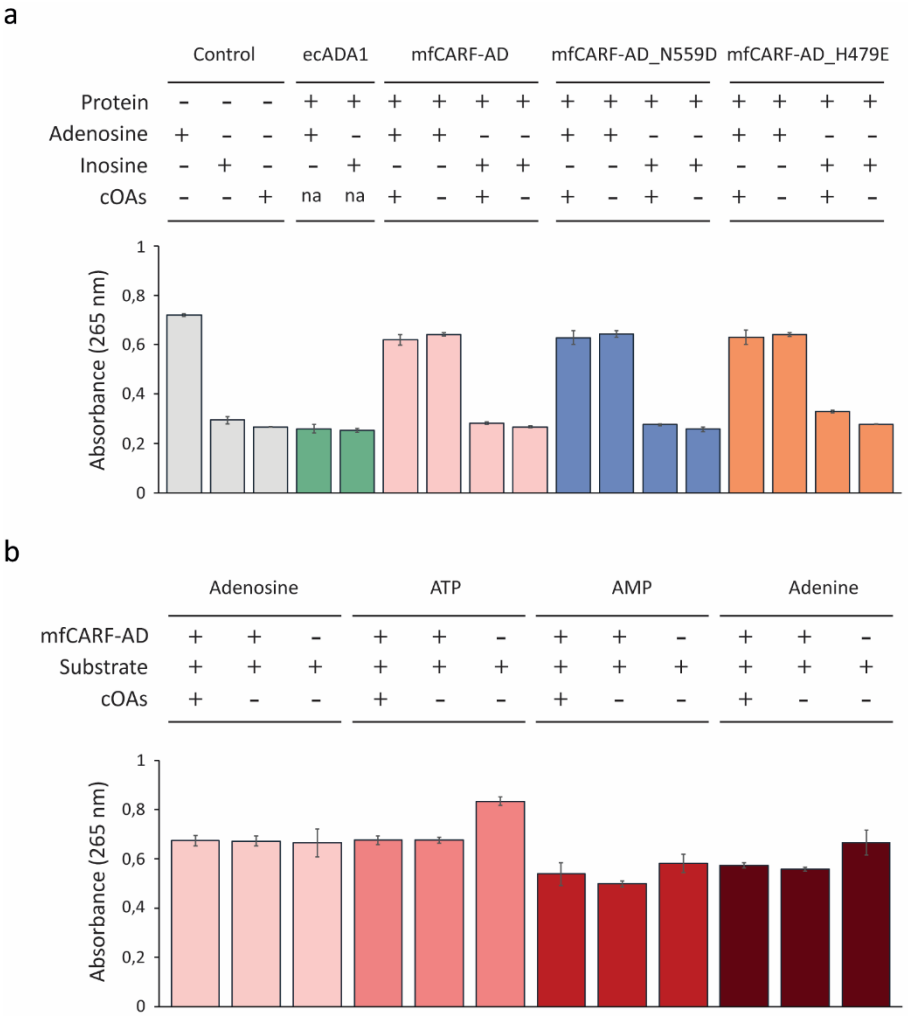


Fig. 3 *In vitro* assays for deamination of free adenosine, adenine, AMP, and ATP. **a.** Absorbance measured at 265 nm after 3h of reaction in a mixture of protein (ecADA1 in green, mfCARF-AD in pink, mfCARF-AD_N559D in blue, or mfCARF-AD_H479E in orange) with adenosine or inosine, in the presence or the absence of cOAs, when applicable. Symbols '+', '-', and 'na' indicate presence, absence, and non-applicable element, respectively. Bar graphs were created based on results from three independent replicates. **b.** Absorbance measured at 265 nm after 3h of reaction in a mixture of mfCARF-AD with adenosine, ATP, AMP, or adenine, in the presence or the absence of cOAs. Different shades of pink indicate different deamination substrates. Bar graphs were created based on results from three independent replicates.

Deaminase activity of mfCARF-AD on nucleic acids

We further examined the ability of mfCARF-AD to act on nucleic acids, and especially RNA substrates as adenosine deamination has never been observed on DNA in nature[17]. We designed a 5'-Cy5-labeled, 36 nucleotide (nt) target RNA substrate containing a central single adenine base, flanked by multiple uracil and cytosine bases (Fig. 4a). After incubation with mfCARF-AD, we subjected the reaction to RNase T1 digestion that specifically cleaves at the 3' end of a guanine or an inosine residue. Hence, successful deamination of the single adenine base into inosine would result in two cleavage products (a 13 nt 5'-Cy5-labeled product, and a 23 nt 3' unlabeled product). As a positive control for RNase T1-mediated cleavage, the same 5'-Cy5-labeled target RNA substrate was used, with the only difference that instead of a single adenine base it contained a single guanine base. Successful cleavage and thus formation of the 13 nt 5'-Cys-labeled product was observed for the positive control RNA substrate (called 'PC RNA'), but not for the target RNA substrate (called 'target RNA') (Fig. 4a), implying that mfCARF-AD was not able to convert adenine into inosine. Instead, a smear of unspecific degradation products was observed in the presence of mfCARF-AD, and RNaseAlert fluorimetric assay indicated that the mfCARF-AD purification sample was contaminated with unspecific RNase enzyme(s) (Fig. 4b). Hence, screening for deaminase activity of mfCARF-AD on RNA was inconclusive.

Oligomeric state of mfCARF-AD

We continued by evaluating the oligomeric state of mfCARF-AD and the possible enhancement of its activated conformation in the presence of cOAs, inspired by previous findings on cOA-mediated activation of CARF proteins[11,56-60]. We incubated mfCARF-AD with no or different cOA species (cA_3, cA_4, cA_6), and subjected the reactions for analysis on BN/SDS-PAGE. We consistently observed three differently-migrating bands, regardless of cOA addition (Fig. 5). This is in agreement with several previously characterized CARF proteins/ring nucleases which exist in solution as oligomers (homodimers, homotetramers, or homohexamers), independently of the presence of cOAs that solely stabilize pre-existing oligomers into their conformationally activated form[11,56-60]. The mfCARF-AD seems to exist predominantly as an oligomer of high molecular weight (complete complex), and, to a lesser extent, as an oligomer of low molecular weight (partial complex), while the monomeric state is less abundant (Fig. 5). Hence, it is possible that mfCARF-AD has a quaternary structure similar to the homohexameric (complete

complex) Csx1 CARF protein from *Sulfolobus islandicus* that is composed of a trimer of dimers (partial complexes)[58]. Taken together, mfCARF-AD forms high molecular weight complexes in solution independently of cOAs, and cOAs do not stabilize these complexes.

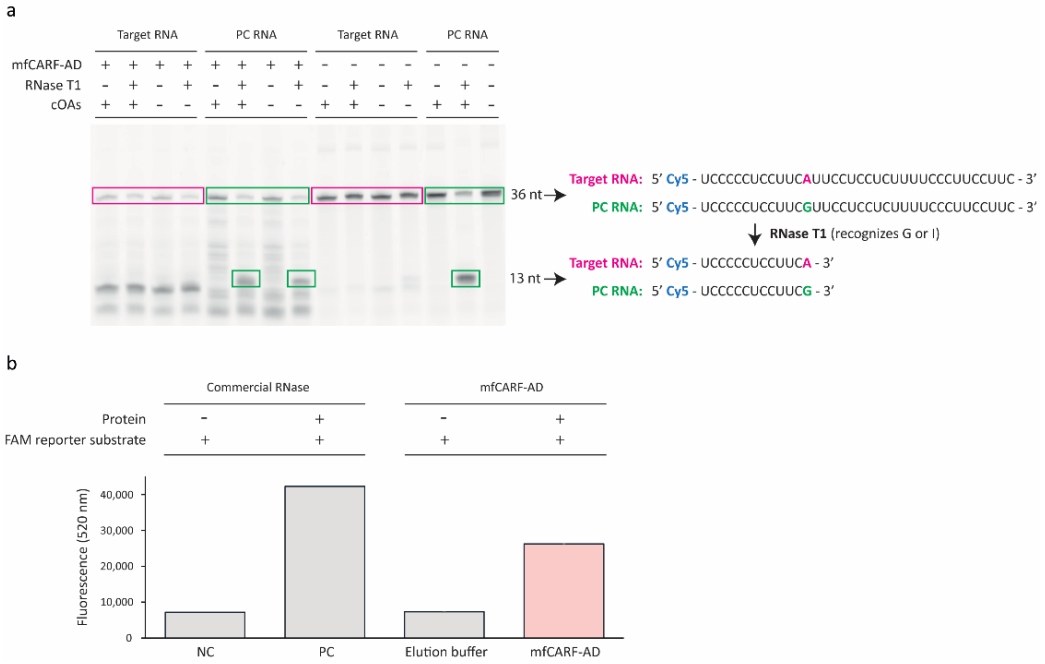


Fig. 4 *In vitro* assays for deamination of RNA. **a.** 20% Urea-PAGE gel of target (pink) and positive control ('PC'; green) RNA products generated after 16 h of incubation with mfCARF-AD and RNase 1 treatment. Uncleaved RNA (36 nt; big pink and green boxes) indicates unsuccessful deamination of adenine into inosine (target RNA) and/or inefficient RNase-mediated recognition of inosine (target RNA) or guanine (PC) followed by cleavage. Cleaved RNA (13 nt; small green boxes) show efficient RNase-mediated recognition and cleavage of the guanine site (PC). **b.** Fluorescence signal measured at 520 nm after incubation of commercial RNase or mfCARF-AD with FAM reporter commercial substrates, during the RNaseAlert assay.

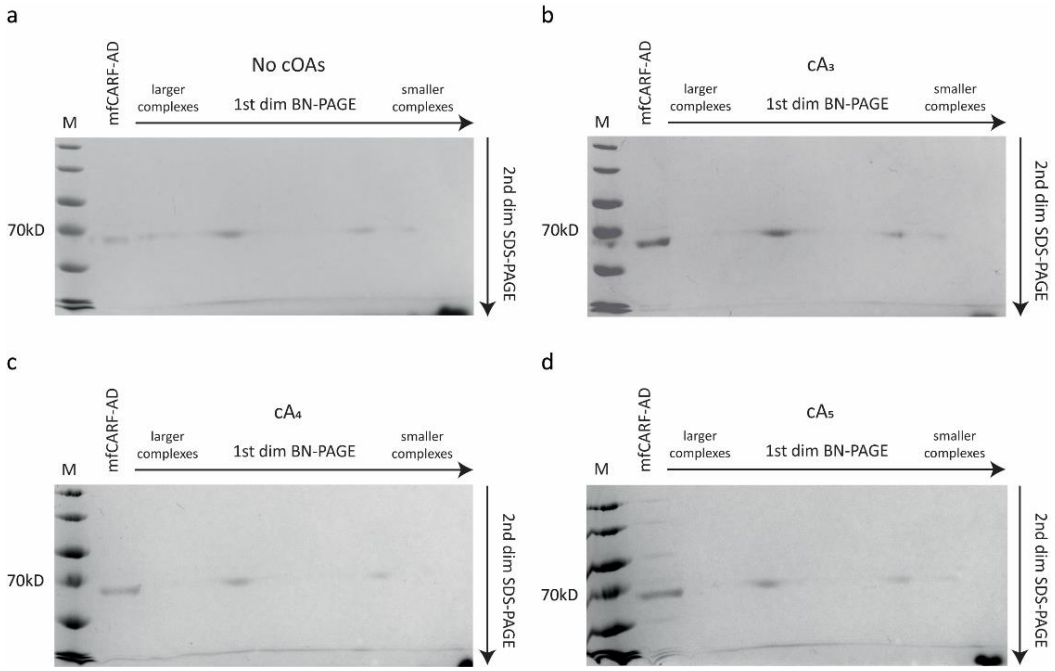


Fig. 5 *In vitro* assays for oligomerization of mfCARF-AD. Two-dimensional BN/SDS-PAGE gels indicating formation of mfCARF-AD complexes (**a**) in the absence of cOAs or in the presence of (**b**) cA_3, (**c**) cA_4, and (**d**) cA_6.

Discussion

In this study, we investigated the enzymatic activity, substrate specificity and biological role (*in silico* and *in vitro*) of mfCARF-AD, a member of the experimentally uncharacterized CARF5-clade. We predict that mfCARF-AD and other CARF5 proteins are CARF ring nucleases with weak or no ability to deaminate free adenosines, probably due to deficient substrate affinity. This is explained by the lack of conservation of crucial residues for substrate-binding (G270, N559) that can be observed in other ADAs. Specifically, replacement of the second aspartate residue with asparagine (DGD > GGN) has been shown to dramatically decrease the affinity for adenosine[48]. In agreement with our prediction, no activity was observed for mfCARF-AD, even at high protein, cOA, or zinc concentrations. In an attempt to restore adenosine deaminase activity, we created the mfCARF-AD_N559D mutant. However, no activity was detected, implying that not only the N559 but also one or more residues (including the G270 residue) should be replaced

by asparagine to render mfCARF-AD active, as observed in most ADAs, including the efficient ecADA1 and mADA[48,50]. In addition, selective recognition of single cOA species (cA3, cA4, cA6), rapid cOA-degradation by the CARF ring nuclease domain of mfCARF-AD, or contamination with unspecific RNase enzyme(s) may have obscured the results. Finally, we showed both bioinformatically and experimentally that mfCARF-AD does not act on free adenine, AMP, and ATP. Moreover, the RNA deamination assay was not conclusive because of RNase activity in the absence of cOAs, most likely suggesting RNase contamination of the mfCARF-AD sample. Hence, re-evaluation of the RNA deaminase activity as well as examination of alternative adenine-containing substrates, such as cOAs, (d)ADP, dATP, DNA, NAD(P), FAD, and co-enzyme A, would shed light on the specificity of mfCARF-AD.

The *in silico* analysis revealed that the *mfcarf-ad* gene is flanked by the type III-B CRISPR-Cas system of *Methylacidiphilum fumariolicum*, suggesting a possible role in defense against invaders. In addition, we found that several CARF5 genes are located adjacent to type III-A and III-B CRISPR-Cas systems, strongly suggesting a functional association. Moreover, these CARF5 genes may collaborate, in parallel, with neighboring immunity and/or purine salvage genes, possibly joining forces against viral infections. Verifying our *in silico* prediction, we observed that mfCARF-AD clusters with genes encoding enzymes involved in purine biosynthesis and carbohydrate metabolism, as commonly reported for deaminases acting on free adenosine molecules[53]. Thus, we reason that mfCARF-AD is activated by cOAs generated by the adjacent type III-B CRISPR-Cas system after viral infection in order to hamper essential biological processes for viral DNA transcription/replication. A similar mechanism of action has been previously observed for bacterial cytidine deaminases and dNTPases that deplete the deoxynucleotide pool, as an antiviral defense strategy[61-65]. The activity of mfCARF-AD may become crucial especially during suboptimal type III targeting, usually caused by mutated targets and weakly transcribed plasmid genes or late-expressed viral genes[66]. After clearance of the viral threat, we speculate that mfCARF-AD acts as a bi-functional, self-limiting deaminase that degrades cOAs through its predicted ring nuclease activity in order to prevent cell dormancy or death, as previously reported for other CARF ring nucleases[11,52,56,57,60,67]. This hypothesis is further supported by our observation that mfCARF-AD shares the same ring nuclease motifs and quaternary structure with these nucleases. However, experimental studies on the ring nuclease and the antiviral activity of mfCARF-AD are necessary to confirm this scenario.

Conclusion

In summary, we show that mfCARF-AD is a putative ring nuclease with no detectable deaminase activity on free adenosine, adenine, and AMP. Moreover, we predict that mfCARF-AD interacts with a type III CRISPR-Cas system as well as with purine salvage and carbon metabolism proteins, probably being involved in a multi-faced antiviral defense.

Materials and Methods

In silico prediction of protein domains, structure and interactions

The domain architecture of mfCARF-AD (MFUM_700012, WP_009060269.1) from *Methylacidiphilum fumariolicum* SolV was predicted using the NCBI Conserved Domain Search tool (default settings). The ClustalO tool (default settings) was applied for multiple sequence alignment of its CARF domain with that of other CARF5-clade members, and its AD domain with that of adenosine deaminases acting on free adenosines or nucleic acids as well as adenine and AMP deaminases. Phyre2 and AlphaFold (default settings) were used for mfCARF-AD structure prediction. The genomic locus organization and protein interactions of mfCARF-AD and other CARF5-clade members were predicted using Genbank and STRING, respectively.

Bacterial strains and growth conditions

E. coli DH5α and *E. coli* BL21 (DE3) strains were used in this study for cloning and protein expression purposes, respectively (Supplementary Table 1). Bacterial cells were cultured in Lysogeny broth (LB) or on LB agar plates, supplemented with kanamycin (50 μg/ml), at 37°C for 17 h. Liquid cultures were grown in a shaker incubator at 220 rpm.

Construction of expression plasmids

Plasmids, primers/ synthetic DNA framents/ RNA substrates, deaminase gene and protein sequences used in this study are presented in Supplementary Tables 2, 3, 4, and 5,

respectively. The bacterial plasmids were constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB). The fragments for assembling the plasmids were obtained through PCR with Q5® High-Fidelity 2X Master Mix (NEB). The PCR products were run on a 0.8% agarose gel and were subsequently purified using Zymogen gel DNA recovery kit (Zymo Research). The assembled plasmids were transformed to chemically competent *E. coli* DH5α cells[68] (Supplementary Table 1), which were plated on LB agar containing kanamycin (50 µg/ml) and incubated overnight at 37°C. The next day, single colonies were inoculated in LB medium, grown overnight at 37°C (220 rpm) and the plasmids were isolated using the GeneJet plasmid Miniprep kit (ThermoFisher Scientific). All the constructs were verified using Sanger sequencing (Macrogen). Description of the assembled fragments used for the construction of each plasmid is detailed in Supplementary Table 2.

Protein expression and purification

E. coli BL21 (DE3) cells carrying pmfCARF-AD, pmfCARF-AD_N559D, pmfCARF-AD_H497E, or pecADA1 were grown to an OD600 of 0.5 and cold-shocked on ice for 1 h. After induction of protein expression with 0.1 mM IPTG, the culture was grown overnight at 18°C and 120 rpm. The cells were pelleted at 6,000 × g for 20 minutes at 4°C, resuspended in lysis buffer (0.5 M NaCl, 0.1 M Tris-HCl at pH 8.0) supplemented with cOmplete™ Mini EDTA-free protease inhibitor tablets (Roche), lysed by sonication (1 sec on, 2 sec off) for 20 min using Sonopuls (Bandelin), and the lysate was clarified at 30,000 × g for 45 min at 4°C. The clarified lysate was filtered with a 0.45 µm filter (mdi Membrane Technologies), and loaded onto a StrepTrap™ HP 5mL column (GE Healthcare). The proteins of interest were eluted from the column with elution buffer (lysis buffer supplemented with 2.5mM d-Desthiobiotin). The elution fractions were denatured and run for 30 min on a 10% Mini-PROTEAN® TGX™ precast SDS-PAGE gel (Biorad) with the PageRuler™ Prestained Protein Ladder (10 to 180 kDa; ThermoFisher). Fractions of interest were then pooled and concentrated to 0.5 ml using a 10 kDa molecular weight cut-off Amicon® Ultra centrifugal filter (Merck). As a final step, only the wild-type and mutant mfCARF-AD proteins were purified by size exclusion chromatography in SEC buffer (0.5 M NaCl, 0.1 M Tris-HCl at pH 8.0) using a HiLoad® 16/600 Superdex® 200pg column (GE Healthcare). Elution fractions with high purity protein were identified by SDS-PAGE and concentrated using a 10 kDa molecular weight cut-off centrifugal filter (Merck Millipore) to a final concentration of approximately 400 µg/ml and stored at -80°C until use.

In vitro adenosine/adenine/AMP/ATP deamination assay

The proteins of interest (50 nM final concentration) were incubated in assay buffer (200 µl, final volume; 25mM Tris-Succinate at pH 7.6) with adenosine, inosine, adenine, AMP, or ATP (100 µM) inside an Epoch 2 microplate reader (BioTek) at 37°C (ecADA1) or 55°C (mfCARF-AD, mfCARF-AD_N559D and mfCARF-AD_H497E) for 3 h. Reactions with wild-type or mutant mfCARF-AD protein were supplemented with 100 nM $cA_{2\text{-}6}$ (20 nM each). Absorbance at 265nm was measured after 3 h of incubation using the Epoch 2 microplate reader (BioTek). Increasing concentrations of protein (25 nM, 50 Nm, 100 nM, 200 nM, 400 nM), cOAs (25 nM or 250 nM each) or zinc (25 nM, 250 nM) were also examined. Experiments were performed in triplicate and data represent the mean ± standard error of the mean.

In vitro RNA deamination assay

Cy5-labeled RNA substrates containing C and U nucleotides and a single adenine or a guanine to verify deaminase activity (Supplementary Table 3) were ordered from IDT and incubated (1µM, final concentration) at 55°C for 16 h (to ensure complete conversion) in assay buffer (10 µl, final volume) supplemented with cOA mix [3 µM $cA_{2\text{-}6}$; 500nM each) and wild-type or mutant mfCARF-AD protein (250 nM, final concentration). After incubation, the RNA substrates were incubated with RNase T1 (0.1 U/ µl) in RNase T1 buffer (50 mM Tris-HCl, 2 mM EDTA at pH 7.5) for 15 min at 37°C. The RNAs were denatured with 2x loading dye (0.2M Tris-Hcl, 0.2 M Boric acid, 0.002 M EDTA, 7 M urea, ficol, blue bromophenol) at 95°C for 5 min. All reactions were run on 20% UREA-PAGE gels using 15 mA per gel. Gels were imaged using an Amersham Typhoon and the Cy5 scanning option (Cytiva). Presence of RNases was examined using the RNaseAlert® kit (IDT), according to the manufacturer's protocol. Fluorescence signal (excitation at 490 nm, emission at 520 nm) was measured using Synergy™ Mx plate reader (BioTek).

In vitro oligomerization assay

The oligomeric state of mfCARF-AD in the presence of different cOA species (cA_3, cA_4, cA_6) was examined using two-dimensional native/denatured electrophoresis, as previously described[69]. As control, no cOAs were applied. Briefly, after incubation of mfCARF-AD (2.7

µM, final concentration) with cA_3, cA_4, or cA_6 (50 µM, final concentration) at 55°C, the mixtures were subjected to Blue Native (BN)-PAGE (1st dimension) using 3-12% gradient polyacrylamide gels, and subsequently to SDS-PAGE using 10% SDS-PAGE gels (2nd dimension) with one lane for the molecular weight marker (PageRuler™ Prestained Protein Ladder 10 to 180 kDa; ThermoFisher) and the rest of the gel for the 1st dimension sample.

References

1. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., ... & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960-964.
2. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712.
3. Mohanraju, P., Saha, C., van Baarlen, P., Louwen, R., Staals, R. H., & van der Oost, J. (2022). Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology*, *20*(6), 351-364.
4. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., ... & Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67-83.
5. Staals, R. H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., Van Duijn, E., ... & Shinkai, A. (2013). Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. *Molecular cell*, *52*(1), 135-145.
6. Staals, R. H., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., ... & van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular cell*, *56*(4), 518-530.
7. Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes & development*, *30*(4), 447-459.
8. Estrella, M. A., Kuo, F. T., & Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes & development*, *30*(4), 460-470.
9. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*(6351), 605-609.
10. Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., ... & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, *548*(7669), 543-548.
11. Athukoralage, J. S., Rouillon, C., Graham, S., Grüschow, S., & White, M. F. (2018). Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature*, *562*(7726), 277-280.
12. Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J., & Staals, R. H. (2018). Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Bioscience reports*, *38*(3).
13. McMahon, S. A., Zhu, W., Graham, S., Rambo, R., White, M. F., & Gloster, T. M. (2020). Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nature communications*, *11*(1), 1-11.
14. Lau, R. K., Ye, Q., Birkholz, E. A., Berg, K. R., Patel, L., Mathews, I. T., ... & Corbett, K. D. (2020). Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Molecular cell*, *77*(4), 723-733.
15. Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č., & Koonin, E. V. (2020). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic acids research*, *48*(16), 8828-8847.
16. Pospisilova, H., & Frebort, I. (2007). Aminohydrolases acting on adenine, adenosine and their derivatives. *Biomedical Papers of the Medical Faculty of Palacky University in Olomouc*, *151*(1).
17. Trasanidou, D. van der Oost, J. & Staals, R.H.J. (2022). Adenosine and cytidine deaminases: Classification, mechanisms and biological roles (to be submitted).
18. Mohanraju, P., Saha, C., van Baarlen, P., Louwen, R., Staals, R. H., & van der Oost, J. (2022). Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology*, *20*(6), 351-364.

19. Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096.
20. Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262-1278.
21. Wu, W. Y., Lebbink, J. H., Kanaar, R., Geijsen, N., & Van Der Oost, J. (2018). Genome editing by natural and engineered CRISPR-associated nucleases. *Nature chemical biology*, *14*(7), 642-651.
22. Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., ... & Tripathi, V. (2021). Novel CRISPR–Cas systems: an updated review of the current achievements, applications, and future research perspectives. International journal of molecular sciences, 22(7), 3327.
23. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464-471.
24. Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., ... & Liu, D. R. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature biotechnology*, *38*(7), 883-891.
25. Lapinaite, A., Knott, G. J., Palumbo, C. M., Lin-Shiao, E., Richter, M. F., Zhao, K. T., ... & Doudna, J. A. (2020). DNA capture by a CRISPR-Cas9–guided adenine base editor. *Science*, *369*(6503), 566-571.
26. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V., & Aravind, L. (2014). CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in genetics*, *5*, 102.
27. Nakano, M., Fukami, T., Gotoh, S., Takamiya, M., Aoki, Y., & Nakajima, M. (2016). RNA editing modulates human hepatic aryl hydrocarbon receptor expression by creating microRNA recognition sequence. *Journal of Biological Chemistry*, *291*(2), 894-903.
28. Shimokawa, T., Rahman, M. F. U., Tostar, U., Sonkoly, E., Ståhle, M., Pivarcsi, A., ... & Zaphiropoulos, P. G. (2013). RNA editing of the GLI1 transcription factor modulates the output of Hedgehog signaling. *RNA biology*, *10*(2), 321-333.
29. Chen, C. X., Cho, D. S. C., Wang, Q., Lai, F., Carter, K. C., & Nishikura, K. (2000). A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single-and double-stranded RNA binding domains. *Rna*, *6*(5), 755-767.
30. Washburn, M. C., Kakaradov, B., Sundararaman, B., Wheeler, E., Hoon, S., Yeo, G. W., & Hundley, H. A. (2014). The dsRBP and inactive editor ADR-1 utilizes dsRNA binding to regulate A-to-I RNA editing across the C. elegans transcriptome. *Cell reports*, *6*(4), 599-607.
31. Zhao, H. Q., Zhang, P., Gao, H., He, X., Dou, Y., Huang, A. Y., ... & Wei, L. (2015). Profiling the RNA editomes of wild-type C. elegans and ADAR mutants. *Genome research*, *25*(1), 66-75.
32. Snyder, E., Chukrallah, L., Seltzer, K., Goodwin, L., & Braun, R. E. (2020). ADAD1 and ADAD2, testis-specific adenosine deaminase domain-containing proteins, are required for male fertility. *Scientific reports*, *10*(1), 1-14.
33. Maas, S., Gerber, A. P., & Rich, A. (1999). Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes. *Proceedings of the National Academy of Sciences*, *96*(16), 8895-8900.
34. Torres, A. G., Piñeyro, D., Rodríguez-Escribà, M., Camacho, N., Reina, O., Saint-Léger, A., ... & Ribas de Pouplana, L. (2015). Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic acids research*, *43*(10), 5145-5157.
35. Bar-Yaacov, D., Mordret, E., Towers, R., Biniashvili, T., Soyris, C., Schwartz, S., ... & Pilpel, Y. (2017). RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome research*, *27*(10), 1696-1703.
36. Eltzschig, H. K., Faigle, M., Knapp, S., Karhausen, J., Ibla, J., Rosenberger, P., ... & Colgan, S. P. (2006). Endothelial catabolism of extracellular adenosine during hypoxia: the role of surface adenosine deaminase and CD26. *Blood*, *108*(5), 1602-1610.
37. Zavialov, A. V., & Engström, Å. (2005). Human ADA2 belongs to a new family of growth factors with adenosine deaminase activity. *Biochemical Journal*, *391*(1), 51-57.
38. Chang, Z., Nygaard, P., Chinault, A. C., & Kellems, R. E. (1991). Deduced amino acid sequence of Escherichia coli adenosine deaminase reveals evolutionarily conserved amino acid residues: implications for catalytic function. *Biochemistry*, *30*(8), 2273-2280.
39. Skaldin, M., Tuittila, M., Zavialov, A. V., & Zavialov, A. V. (2018). Secreted bacterial adenosine deaminase is an evolutionary precursor of adenosine deaminase growth factor. *Molecular Biology and Evolution*, 35(12), 2851-2861.
40. Prabhakaran, M., Couger, M. B., Jackson, C. A., Weirick, T., & Fathepure, B. Z. (2015). Genome sequences of the lignin-degrading Pseudomonas sp. strain YS-1p and Rhizobium sp. strain YS-1r isolated from decaying wood. *Genome announcements*, *3*(2), e00019-15.

41. PospíŠilová, H., Šebela, M., Novák, O., & Frébort, I. (2008). Hydrolytic cleavage of N 6-substituted adenine derivatives by eukaryotic adenine and adenosine deaminases. *Bioscience reports*, *28*(6), 335-347.
42. Petersen, C., Møller, L. B., & Valentin-Hansen, P. (2002). The cryptic adenine deaminase gene of Escherichia coli: silencing by the nucleoid-associated DNA-binding protein, H-NS, AND activation by insertion elements. *Journal of Biological Chemistry*, *277*(35), 31373-31380.
43. Sabina, R. L., Morisaki, T., Clarke, P., Eddy, R., Shows, T. B., Morton, C. C., & Holmes, E. W. (1990). Characterization of the human and rat myoadenylate deaminase genes. *Journal of Biological Chemistry*, *265*(16), 9423-9433.
44. Akizu, N., Cantagrel, V., Schroth, J., Cai, N., Vaux, K., McCloskey, D., ... & Gleeson, J. G. (2013). AMPD2 regulates GTP synthesis and is mutated in a potentially treatable neurodegenerative brainstem disorder. *Cell*, *154*(3), 505-517.
45. Mahnke-Zizelman, D. K., D'cunha, J., Wojnar, J. M., Brogley, M. A., & Sabina, R. L. (1997). Regulation of rat AMP deaminase 3 (isoform C) by development and skeletal muscle fibre type. *Biochemical Journal*, *326*(2), 521-529.
46. Goble, A. M., Zhang, Z., Sauder, J. M., Burley, S. K., Swaminathan, S., & Raushel, F. M. (2011). Pa0148 from Pseudomonas aeruginosa catalyzes the deamination of adenine. *Biochemistry*, *50*(30), 6589-6597.
47. Han, B. W., Bingman, C. A., Mahnke, D. K., Bannen, R. M., Bednarek, S. Y., Sabina, R. L., & Phillips, G. N. (2006). Membrane association, mechanism of action, and structure of Arabidopsis embryonic factor 1 (FAC1). *Journal of Biological Chemistry*, *281*(21), 14939-14947.
48. Sideraki, V., Mohamedali, K. A., Wilson, D. K., Chang, Z., Kellems, R. E., Quiocho, F. A., & Rudolph, F. B. (1996). Probing the functional role of two conserved active site aspartates in mouse adenosine deaminase. *Biochemistry*, *35*(24), 7862-7872.
49. Pornbanlualap, S., & Chalopagorn, P. (2011). Adenosine deaminase from Streptomyces coelicolor: recombinant expression, purification and characterization. *Protein Expression and Purification*, *78*(2), 167-173.
50. Nygaard, P. (1978). [68] Adenosine deaminase from Escherichia coli. In *Methods in enzymology* (Vol. 51, pp. 508-512). Academic Press.
51. McMahon, S. A., Zhu, W., Graham, S., Rambo, R., White, M. F., & Gloster, T. M. (2020). Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nature communications*, *11*(1), 1-11.
52. Athukoralage, J. S., Graham, S., Rouillon, C., Grüschow, S., Czekster, C. M., & White, M. F. (2020). The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling. *Elife*, *9*, e55852.
53. Huang, Z., Xie, N., Illes, P., Di Virgilio, F., Ulrich, H., Semyanov, A., ... & Tang, Y. (2021). From purines to purinergic signalling: molecular functions and human diseases. *Signal Transduction and Targeted Therapy*, *6*(1), 1-20.
54. Nieuwkoop, T., Claassens, N. J., & van der Oost, J. (2019). Improved protein production and codon optimization analyses in Escherichia coli by bicistronic design. *Microbial biotechnology*, *12*(1), 173-179.
55. Sideraki, V., Wilson, D. K., Kurz, L. C., Quiocho, F. A., & Rudolph, F. B. (1996). Site-directed mutagenesis of histidine 238 in mouse adenosine deaminase: substitution of histidine 238 does not impede hydroxylate formation. *Biochemistry*, *35*(47), 15019-15028.
56. Garcia-Doval, C., Schwede, F., Berk, C., Rostøl, J. T., Niewoehner, O., Tejero, O., ... & Jinek, M. (2020). Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nature communications*, *11*(1), 1-9.
57. Jia, N., Jones, R., Yang, G., Ouerfelli, O., & Patel, D. J. (2019). CRISPR-Cas III-A Csm6 CARF domain is a ring nuclease triggering stepwise cA4 cleavage with ApA> p formation terminating RNase activity. *Molecular cell*, *75*(5), 944-956.
58. Molina, R., Stella, S., Feng, M., Sofos, N., Jauniskis, V., Pozdnyakova, I., ... & Montoya, G. (2019). Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. *Nature communications*, *10*(1), 1-14.
59. Niewoehner, O., & Jinek, M. (2016). Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *Rna*, *22*(3), 318-329.
60. Athukoralage, J. S., McQuarrie, S., Grüschow, S., Graham, S., Gloster, T. M., & White, M. F. (2020). Tetramerisation of the CRISPR ring nuclease Crn3/Csx3 facilitates cyclic oligoadenylate cleavage. *Elife*, *9*.
61. Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., ... & Sorek, R. (2022). Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nature Microbiology*, *7*(8), 1200-1209.
62. Ayinde, D., Casartelli, N., & Schwartz, O. (2012). Restricting HIV the SAMHD1 way: through nucleotide starvation. *Nature Reviews Microbiology*, *10*(10), 675-680.
63. Kondo, N., Nakagawa, N., Ebihara, A., Chen, L., Liu, Z. J., Wang, B. C., ... & Masui, R. (2007). Structure of dNTP-inducible dNTP triphosphohydrolase: insight into broad specificity for dNTPs and triphosphohydrolase-type hydrolysis. *Acta Crystallographica Section D: Biological Crystallography*, *63*(2), 230-239.
64. Mega, R., Kondo, N., Nakagawa, N., Kuramitsu, S., & Masui, R. (2009). Two dNTP triphosphohydrolases from Pseudomonas aeruginosa possess diverse substrate specificities. *The FEBS Journal*, *276*(12), 3211-3221.
65. Quirk, S. T. E. P. H. E. N., & Bessman, M. J. (1991). dGTP triphosphohydrolase, a unique enzyme confined to members of the family Enterobacteriaceae. *Journal of bacteriology*, *173*(21), 6665-6669.

66. Steens, J. A., Salazar, C. R. P., & Staals, R. H. (2022). The diverse arsenal of type III CRISPR–Cas-associated CARF and SAVED effectors. *Biochemical Society Transactions*, *50*(5), 1353-1364.
67. Athukoralage, J. S., Graham, S., Grüschow, S., Rouillon, C., & White, M. F. (2019). A type III CRISPR ancillary ribonuclease degrades its cyclic oligoadenylate activator. *Journal of Molecular Biology*, *431*(15), 2894-2899.
68. Green, R., & Rogers, E. J. (2013). Transformation of chemically competent E. coli. *Methods Enzymol*, *529*, 329-336.
69. Fiala, G. J., Schamel, W. W., & Blumenthal, B. (2011). Blue native polyacrylamide gel electrophoresis (BN-PAGE) for analysis of multiprotein complexes from cellular lysates. *JoVE (Journal of Visualized Experiments)*, (48), e2164.

Supplementary tables & figures

Supplementary Table 1. Strains used in this study.

Strain	Description	Plasmid	Antibiotic resistance	Reference
E. coli DH5α	-	-	-	Lab stock
E. coli BL21 (DE3)	-	-	-	Lab stock

Supplementary Table 2. Plasmids used in this study.

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
ptn0005_strep_amuc	-	-	PCR template for pmfCARF-AD.	Lab stock
pmfCARF-AD	Gibson assembly	Fragment 1: terminator, backbone, T7 promoter, BCD expression system, Strep-tag, and TEV-site from ptn0005_strep_amuc amplified with primers BG21485 and BG21486.	Strep-TEV-mfCARF-AD protein expression and purification for *in vitro* deamination assays, RNaseAlert fluorimetric assay, and oligomerization test. PCR template for pmfCARF-AD_N559D, pmfCARF-AD_H497E, and pecADA.	This study
		Fragment 2: TEV-site, and *mfcarf-ad* gene from G-block_mfCARF-AD amplified with primers BG21493 and BG21494		
pmfCARF-AD_N559D	Gibson assembly	Fragment 1: partly *mfcarf-ad* gene, terminator, and backbone from pmfCARF-AD amplified with primers BG28563 and BG28459	Strep-TEV-mfCARF-AD_N559D protein expression and purification for in vitro deamination assays.	This study
		Fragment 2: backbone, T7 promoter, BCD expression system, Strep-tag, TEV-site, and partly *mfcarf-ad* gene from pmfCARF-AD amplified with primers BG28458 and BG28564		
pmfCARF-AD_H497E	Gibson assembly	Fragment 1: partly *mfcarf-ad* gene, terminator, and backbone from pmfCARF-AD amplified with primers BG28271 and BG28272	Strep-TEV-mfCARF-AD_H497E protein expression and purification for in vitro deamination assays.	This study
		Fragment 2: backbone, T7 promoter, BCD expression system, Strep-tag, TEV-site, and partly *mfcarf-ad* gene from pmfCARF-AD amplified with primers BG28274 and BG28273		
pecADA1	Gibson assembly	Fragment 1: TEV-site, and *ecada* gene from *E. coli* BL21 (DE) genome amplified with primers BG27763 and BG27764	Strep-TEV-ecADA1 protein expression and purification for in vitro deamination assays.	This study
		Fragment 2: terminator, backbone, T7 promoter, BCD expression system, Strep-tag, and TEV-site from pmfCARF-AD amplified with primers BG27761 and BG27762		

Supplementary Table 3. Primers/G-blocks/RNA substrates used in this study.

	Primer/ G-block/ RNA substrate	Sequence (5' -> 3')	Function
Cloning	BG21485	TGCCGACTCAGTTGCTG	Construction of pmfCARF-AD.
	BG21486	GGATTGGAAGTACAGGTTTTCCG	
	BG21493	cgcggaaaacctgtacttccaatccATGGATAGCAGCCCGC	
	BG21494	agtagaagcagcaactgagtcggcaTTAGCTCGGCTCGTTAAATAACG	
	BG28458	GATGGAGTTCTGAGGTCATTAC	Construction of pmfCARF-AD_N559D.
	BG28459	GTAATGACCTCAGAACTCCATC	
	BG28563	CAATACCGATGACATTGGGATTTC	
	BG28564	GAAATCCCAATGTCATCGGTATTG	
	BG28271	GATCGGGGAAGCACTGCATC	Construction of pmfCARF-AD_H497E.
	BG28272	GAATGAACTGCAGGACGAGGC	
	BG28273	GATGCAGTGCTTCCCCGATC	
	BG28274	GCCTCGTCCTGCAGTTCATTC	
	BG27761	gcgagaaaaagtcgccgcgaagtaaTGCCGACTCAGTTGCTG	Construction of pecADA1.
	BG27762	ttaatggcagggtggtatcaatcatGGATTGGAAGTACAGGTTTTCCG	
	BG27763	cgcggaaaacctgtacttccaatccATGATTGATACCACCCTGCC	
	BG27764	agtagaagcagcaactgagtcggcaTTACTTCGCGGCGACTTTTT	
Sequencing	BG21698	AATTCCGACGTCGAATTC	Sequencing of pmfCARF-AD, pmfCARF-AD_N559D, pmfCARF-AD_H497E, and pecADA1.
	BG21699	GCGTCAGATTTCGTGATG	
	BG21742	CTTCAGCAATATCACGGGTAGC	
	BG21708	CTATGTCAGCTGCGTTACTG	
	BG21709	GTCTATGCAGAAATTCGGACCAG	
	BG21710	GGGTCACTGACTGATAACTTACTG	
G-block	BG21469	ctaagctgattgcgacaacaATGGATAGCAGCCCGCAGAAAAACGTCCTGCTGTGCACTCTGGGGTCGACCCTGTCAAAAACCTGGTTAATTGCGGTCGAGGCTCTGTACTTCGTGAACCACTACCAC CAGAAGCTGGATGAAGTGCACCTGCTGACCACCTGCAACCCGGAAGCGGAAGAATGTTTTCAGAACATTGAGCAATGTTTAAAAATCCATTTTCCGGGGTTAATCTATACCTTTACCCAGGTCG ATGAATGTGCAGATATTTGTTCGCAAGAGGATTATTTTCTGTTTCAGGAGGTCTTATATCGGTGGTTTCTGAAGCATATACAGCGCGCAACCGTCTATGTCTGCATAGCATCGGGGCGGAAGAC TATGTCAGCTGCGTTACTGAAAGGGGCGAGCCTGTTTGGCGCGAAATATGTCTTTGATGTGCTGGCAGAGGAACCGTTCCCATCGAGCATGGAAGAAGTCAAAGAAGCACTGAAGACCAATAA AATTAAGTTTGTGTCGTTAGGGAACGAAAAGGGCTGGACCCAGTTAAAAAAGTTATCATTCGAAGAATTTCCGTTAAATGAGATTGAGCGCAAACCGCGGCTGTATTTAGCGACCGTCGCGGA CCGGAAACTGCAAGAAAAAATTAATGAAGTGCAGGACTATAGCTTACGGATCCATTACAACTGGGGGCGGCTGTCAGAGTTACCGTTTCCGGTGCTGGCACGGTGGACCCCAAAAGAATTAGA TTGGTTAAACGAGCCATTAGATCCGAAACGCGACAAAGAATGGATCGAAAAGCTGCCGAAGGTCGAATTACACTGCCATTTAGGGGGGTTTGCAACCCATGGCGAGTTATTAGCTCAAGTGCG ATCATCAGCAGAAGCGCCATCGTCATTACCGCATCTGCTGGAGCCGCCGCTGCCGGAGGGGTGGCCGCGGCCGGCGAAGCCGATTGGGCTGGAAACCTACATGGAACTGGGGCGGGCGACC GGCACCGATCTGCTGAAGGACCCGGGGTGCCTGAAAGAACAGTGTCGCCTGCTGTACCAGGCGCTGCAGAAAGACCGGGTGGTCTATGCAGAAATTCGGACCAGCCCGAATAACTACACCAAC ATCTCGAAAGGCCGCAGCGCGTGGACCGTGCTGTGTGAAATTCGAGATCATTTCCAACAGTGTATGAACGAGAACCAGGGGGCGGACACCTACTGCCATGTGAACTTAATCATCATTGTCACCC GAAAGGATAAAGGCGATCTGTCAGATATTTCGCGGCATCTGTCGCTGGCAGTGACCGCGGCTCAGCATTTTCCGCCGAAAGAATCGTATTGCGGGGTCGTCGGCGTCGACTTAGCGGGGCTGG AATCACCAAAAACCCGGGCGAGCTACTTTGCAGAAGATTTTGTCCCGGTGCACCGATGTGGCCTGGCGGTGACCGCACACGCGGGGGAAATCGATGACGTGGAGGGCATATGGCAGGCAATT TTTCGGTTACATGCACGCCGGATCGGGCATGCACTGCATCTGTTAAAAAGCCAGGATTTATTAAATACCGTGATTGACCGGCATATTGGCATCGAAATGTGCCCGCTGGCAAACTATCAGATCCA TGGCTTTAAGCCAATGGAAAATAATAAGAATACCTATCCGCTGAAAGAATACTTAAACCGGGGGGTCTATGTGACTGTCAATACCGATAATATTGGGATTTCGAAGGGGTCACTGACTGATAAC TTACTGTTTTTAGCGGAGCTGTGCCCGGGGATCCGGCGGTTAGATATCTTAAAATTACTGAATCACGCGTGCCAAATAGCGTTTTTACCACATCAACTGCGGACCGAGTTAATTCACAAAATACA GAAAAATTTATTAACCCCGACCAATCTGGCGTTATTTAACGAGCCGAGCTAAccagatctgacgtcagtcaa

Supplementary Table 3. Primers/G-blocks/RNA substrates used in this study. *(continuation)*

RNA	BG28050 (Cy5-labeled)	UCCCCCUCCUUCAUUCCUCCUCUUUUCCCUUCCUUC	Target RNA substrate.
	BG28052 (Cy5-labeled)	UCCCCCUCCUUCGUUCCUCCUCUUUUCCCUUCCUUC	Positive control RNA substrate.

***Lowercase**: overhang sequence for Gibson assembly*
***Uppercase**: sequence annealing to the PCR template DNA*

Supplementary Table 4. DNA sequences of deaminase genes used in this study.

Gene/ Fusion gene	Sequence (5' -> 3')
mfcarf-ad (wild-type) gene [MFUM_700012]	atggattcttctcctcaaaaaaatgtgcttttatgtacgcttggcagtaccctagcaaaacctggctcattgctgtagaggcgctttacttcgttaatcactaccaccaaaagcttgatgaagttcacttacttactacctgtaatccagaagctgaagaatgctttcaaaatattgagcag tgcctaaaaatccatttttcctggcctcatctatacatttacccaagtcgatgaatgcgcagatatttgcagtcaggaggattatttttcttttcaagaggtcctctatcggtggtttcttaagcatatacaaagagccactgtctatgtgtgtatagccagtgggcgaaagacgatgagcgc ggctctcctcttaaaggggcttcttttatttggagctaaatatgtctttgatgtttttagcagaggaacctttccccctcctctatggaagaagtcaaagaagccttaaagactaacaaaaattaagtttgtttccctgggcaatgaaaagggatggacacaactaaaaaagctgtcattcgaaga atttccactgaacgagattgagagaaaacctcgattatatctcgctacagtggctgaccgaaaattacaggaaaaaattaacgaagttcaagactattctctccggatccattacaattgggggccggctttcagagctacctttttccagttttagcccgttggactcccaaagaactcga ttggttgaatgagcccttggatcctaaaagagacaaagaatggatcgaaaagcttccaaaggtagaactgcactgtcatttgggcggctttgccacacatggagagttgctagcgcaggttcgctcaagcgcagaagctccctcctcactgcctcatcttttagagccacctttacctga gggctggcctcggccagctaagcctattggccttgaaacatacatggaattaggtagggctacaggaaccgatcttctttaaggaccctgggtgtcttaaagaacaatgcagacttttatacccaagctcttcaaaaagaccgtgttgtgtatgcagaaattcgaacttctcctaacaatta cactaatatctccaaaggaagatctgcttggactgttctttgcgaaattcgcgatcatttccagcaatgcatgaatgagaatcaaggtgctgacacttactgtcatgttaatctcatcatcattgtcactcgcaaggataaaggagatctttcggatattagtcgtcatttatcccttgccgt tacagctgcgcaacattttccaccaaaagaaagttattgtggcgtggtcggagtggacctagctggccttgaaagccccaaaactcgtgcttcttactttgcagaagattttgtccctgttcaccgctgcggacttgctgttacagcccacgctggtgaaatcgatgacgttgagggaata tggcaagcaatttttcgactccatgccagaaggatcgggcatgcacttcatctttgaaatctcaagatctgctgaacactgttattgaccgacatattggaatcgaaatgtgtccattagccaattatcaaatccatggatttaagccccatggaaaacaacaagaacacatatccatta aaagaatacctaaatcggggcgtctatgttacggtcaacactgataacattggtattagtaagggcagcttaacggataatctcttatttttggctgagctttgtcctgggatccgtcggctcgatatcctcaaactcttaaaccacgcttgtcagatagctttttctcccccatcagcttcga acagagctgattcacaaaatacaaaaaaacctactcactccaaccaaccttgctctctttaatgagcctttcttaa
mfcarf-ad (codon harmonised) gene	atggatagcagcccgcagaaaaacgtcctgctgtgcactctggggtcgaccctgtcaaaaacctggttaattgcggtcgaggctctgtacttcgtgaaccactaccaccagaagctggatgaagtgcacctgctgaccacctgcaacccggaagcggaagaatgttttcagaacattg agcaatgtttaaaaatccattttccggggttaatctatacctttacccaggtcgatgaatgtgcagatatttgttcgcaagaggattattttctgtttcaggaggtcttatatcggtggtttctgaagcatatacagcgcgcaaccgtctatgtctgcatagcatcggggcggaagactatgt cagctgcgttactgaaaggggcgagcctgtttggcgcgaaatatgtctttgatgtgctggcagaggaaccgttcccatcgagcatggaagaagtcaaagaagcactgaagaccaataaaattaagtttgtgtcgttagggaacgaaaagggctggaccccagttaaaaaagttatcat tcgaagaatttccgttaaatgagattgagcgcaaacccgcggctgtatttagcgaccgtcgcggaccggaaactgcaagaaaaaattaatgaagtgcaggactatagcttacggatccattacaactgggggcggctgtcagagttaccgtttccggtgctggcacggtggaccccaa aagaattagattggttaaacgagccattagatccgaaacgcgacaaagaatggatcgaaaagctgccgaaggtcgaattacactgccatttagggggttgcaacccatggcgagttattagctcaagtgcgatcatcagcagaagcgccatcgtcattaccgcatctgctggagc cgccgctgccggaggggtggccgcggccggcgaagccgattgggctggaaacctacatggaactggggcgggcgaccggcaccgatctgctgaaggacccggggtgcctgaaagaacagtgtcgcctgctgtaccaggcgctgcagaaagaccgggtggtctatgcagaaattcg gaccagcccgaataactacaccaacatctcgaaaggccgcagcgcgtggaccgtgctgtgtgaaattcgagatcatttccaacagtgtatgaacgagaaccagggggcggacacctactgccatgtgaacttaatcatcattgtcacccgaaaggataaaggcgatctgtcagatat ttcgcggcatctgtcgctggcagtgaccgcggctcagcattttccgccgaaagaatcgtattgcggggtcgtcggcgtcgacttagcggggctggaatcaccaaaaacccgggcgagctacttttgcagaagattttgtcccggtgcaccgatgtggcctggcggtgaccgcacacgcgg gggaaatcgatgacgtggagggcatatggcaggcaatttttcggttacatgcacgccggatcggcatgcactgcatctgttaaaaagccaggatttattaaataccgtgattgaccggcatattggcatcgaaatgtgcccgctggcaaactatcgatccatggctttaagccaatg gaaaataataagaatacctatccgctgaaagaatacttaaaccgggggtctatgtgactgtcaatacccgataatattgggatttcgaaggggtcactgactgataacttactgttttagcggagctgtgccccggggatccggcggttagatatcttaaaattactgaatcacgcgtgc caaatagcgttttaccacatcaactgcggaccgagttaattcacaaaatacagaaaaatttattaaccccgaccaatctggcgttatttaacgagccgagctaa
mfcarf-ad_n559d (codon harmonised mutant) gene	atggatagcagcccgcagaaaaacgtcctgctgtgcactctggggtcgaccctgtcaaaaacctggttaattgcggtcgaggctctgtacttcgtgaaccactaccagaagctggatgaagtgcacctgctgaccacctgcaacccggaagcggaagaatgttttcagaacattg agcaatgtttaaaaatccattttccggggttaatctatacctttacccaggtcgatgaatgtgcagatatttgttcgcaagaggattattttctgtttcaggaggtcttatatcggtggtttctgaagcatatacagcgcgcaaccgtctatgtctgcatagcatcggggcggaagactatgt cagctgcgttactgaaaggggcgagcctgtttggcgcgaaatatgtctttgatgtgctggcagaggaaccgttcccatcgagcatggaagaagtcaaagaagcactgaagaccaataaaattaagtttgtgtcgttagggaacgaaaagggctggaccccagttaaaaaagttatcat tcgaagaatttccgttaaatgagattgagcgcaaacccgcggctgtatttagcgaccgtcgcggaccggaaactgcaagaaaaaattaatgaagtgcaggactatagcttacggatccattacaactgggggcggctgtcagagttaccgtttccggtgctggcacggtggaccccaa aagaattagattggttaaacgagccattagatccgaaacgcgacaaagaatggatcgaaaagctgccgaaggtcgaattacactgccatttagggggttgcaacccatggcgagttattagctcaagtgcgatcatcagcagaagcgccatcgtcattaccgcatctgctggagc cgccgctgccggaggggtggccgcggccggcgaagccgattgggctggaaacctacatggaactggggcgggcgaccggcaccgatctgctgaaggacccggggtgcctgaaagaacagtgtcgcctgctgtaccaggcgctgcagaaagaccgggtggtctatgcagaaattcg gaccagcccgaataactacaccaacatctcgaaaggccgcagcgcgtggaccgtgctgtgtgaaattcgagatcatttccaacagtgtatgaacgagaaccagggggcggacacctactgccatgtgaacttaatcatcattgtcacccgaaaggataaaggcgatctgtcagatat ttcgcggcatctgtcgctggcagtgaccgcggctcagcattttccgccgaaagaatcgtattgcggggtcgtcggcgtcgacttagcggggctggaatcaccaaaaacccgggcgagctacttttgcagaagattttgtcccggtgcaccgatgtggcctggcggtgaccgcacacgcgg gggaaatcgatgacgtggagggcatatggcaggcaatttttcggttacatgcacgccggatcggcatgcactgcatctgttaaaaagccaggatttattaaataccgtgattgaccggcatattggcatcgaaatgtgcccgctggcaaactatcgatccatggctttaagccaat gaaaataataagaatacctatccgctgaaagaatacttaaaccgggggtctatgtgactgtcaatacccgatgacattgggatttcgaaggggtcactgactgataacttactgttttagcggagctgtgccccggggatccggcggttagatatcttaaaattactgaatcacgcgtgc caaatagcgttttaccacatcaactgcggaccgagttaattcacaaaatacagaaaaatttattaaccccgaccaatctggcgttatttaacgagccgagctaa
mfcarf-ad_h497e (codon harmonised mutant) gene	atggatagcagcccgcagaaaaacgtcctgctgtgcactctggggtcgaccctgtcaaaaacctggttaattgcggtcgaggctctgtacttcgtgaaccactaccaccagaagctggatgaagtgcacctgctgaccacctgcaacccggaagcggaagaatgttttcagaacattg agcaatgtttaaaaatccattttccggggttaatctatacctttacccaggtcgatgaatgtgcagatatttgttcgcaagaggattattttctgtttcaggaggtcttatatcggtggtttctgaagcatatacagcgcgcaaccgtctatgtctgcatagcatcggggcggaagactatgt cagctgcgttactgaaaggggcgagcctgtttggcgcgaaatatgtctttgatgtgctggcagaggaaccgttcccatcgagcatggaagaagtcaaagaagcactgaagaccaataaaattaagtttgtgtcgttagggaacgaaaagggctggaccccagttaaaaaagttatcat tcgaagaatttccgttaaatgagattgagcgcaaacccgcggctgtatttagcgaccgtcgcggaccggaaactgcaagaaaaaattaatgaagtgcaggactatagcttacggatccattacaactgggggcggctgtcagagttaccgtttccggtgctggcacggtggaccccaa aagaattagattggttaaacgagccattagatccgaaacgcgacaaagaatggatcgaaaagctgccgaaggtcgaattacactgccatttagggggttgcaacccatggcgagttattagctcaagtgcgatcatcagcagaagcgccatcgtcattaccgcatctgctggagc cgccgctgccggaggggtggccgcggccggcgaagccgattgggctggaaacctacatggaactggggcgggcgaccggcaccgatctgctgaaggacccggggtgcctgaaagaacagtgtcgcctgctgtaccaggcgctgcagaaagaccgggtggtctatgcagaaattcg gaccagcccgaataactacaccaacatctcgaaaggccgcagcgcgtggaccgtgctgtgtgaaattcgagatcatttccaacagtgtatgaacgagaaccagggggcggacacctactgccatgtgaacttaatcatcattgtcacccgaaaggataaaggcgatctgtcagatat ttcgcggcatctgtcgctggcagtgaccgcggctcagcattttccgccgaaagaatcgtattgcggggtcgtcggcgtcgacttagcggggctggaatcaccaaaaacccgggcgagctacttttgcagaagattttgtcccggtgcaccgatgtggcctggcggtgaccgcacacgcgg gggaaatcgatgacgtggagggcatatggcaggcaatttttcggttacatgcacgccggatcggggaagcactgcatctgttaaaaagccaggatttattaaataccgtgattgaccggcatattggcatcgaaatgtgcccgctggcaaactatcgatccatggctttaagccaat

Supplementary Table 4. DNA sequences of deaminase genes used in this study. *(continuation)*

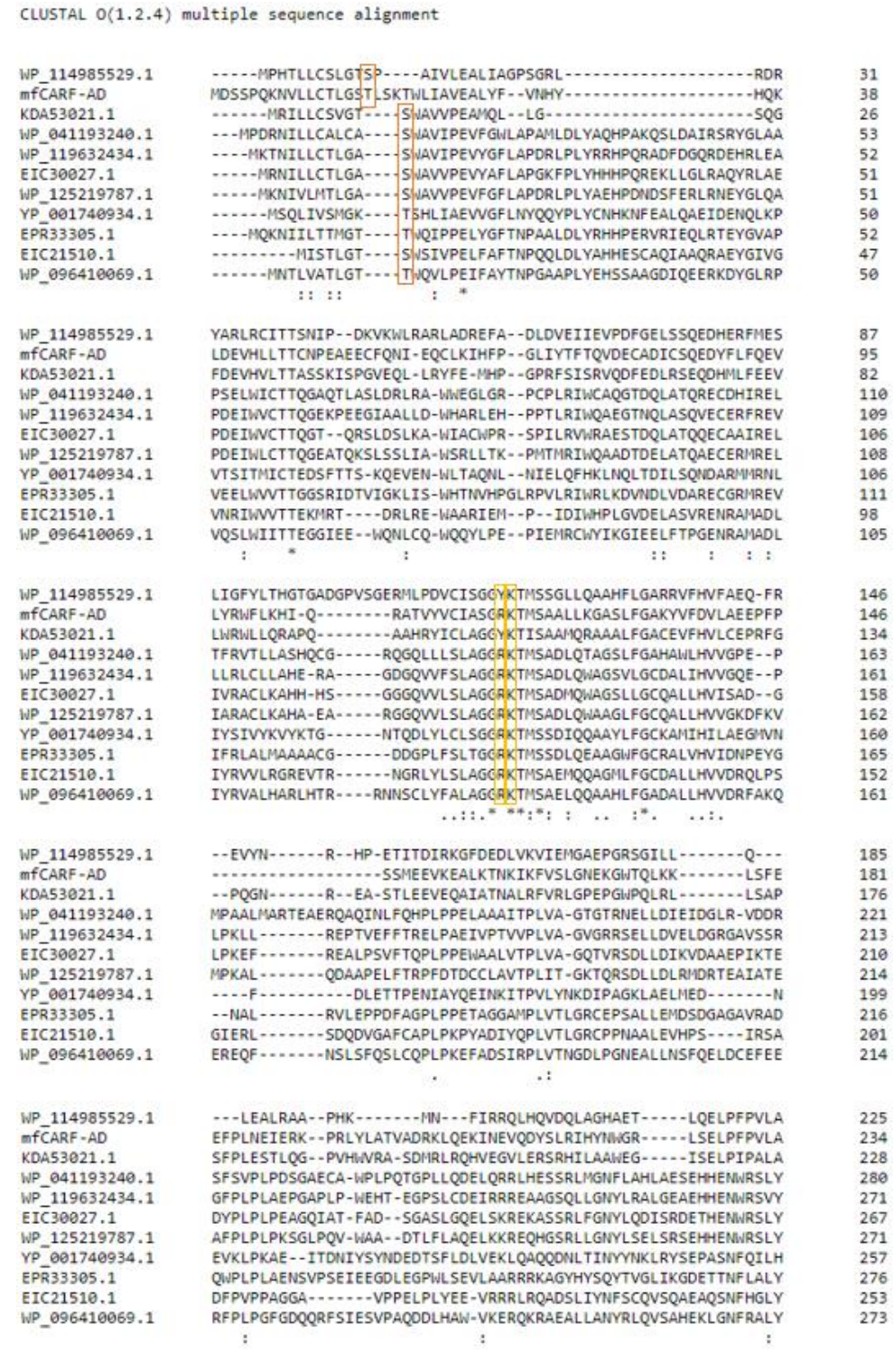
	ggaaaataataagaatacctatccgctgaaagaatacttaaaccgggggtctatgtgactgtcaataccgataatattgggatttcgaaggggtcactgactgataacttactgtttttagcggagctgtgcccggggatccggcggttagatatcttaaaattactgaatcacgcgtg ccaaatagcgtttttaccacatcaactgcggaccgagttaattcacaaaatacagaaaaatttattaaccccgaccaatctggcgttatttaacgagccgagctaa
ecada1 (wild-type) gene	atgattgataccaccctgccattaactgatatccatcgccaccttgatggcaacattcgtccccagaccattcttgaacttggccgccagtataatatctcgcttcctgcacaatccctggaaacactgattccccacgttcaggtcattgccaacgaacccgatctggtgagctttctgac caaacttgactggggcgttaaagttctcgcctctcttgatgcctgtcgccgcgtggcatttgaaaacattgaagatgcagcccgtcacggcctgcactatgtcgagctgcgttttttcaccaggctacatggcaatggcacatcagctgcctgtagcgggtgttgtcgaagcggtgatcgat ggcgtacgtgaaggttgccgcacctttggtgtgcaggcgaagcttatcggcattatgagccggaccttcggcgaagccgcctgtcagcaagagctggaggccttttagcccaccgtgaccagattaccgcacttgatttagccggtgatgaacttggtttcccgggaagtctgttccttt ctcacttcaaccgcgcgcgtgatgcgggctggcatattaccgtccatgcaggcgaagctgccgggccggaaagcatctggcaggcgattcgtgaactgggtgcggagcgtattggacatggcgtaaaagccattgaagatcgggcgctgatggatttttctgcccgagcaacaaattg gtattgaatcctgtctgacctccaatattcagaccagcaccgtagcagagctggctgcacatccgctgaaaacgttccttgagcatggcattcgtgccagcattaacactgacgatcccggcgtacagggagtggatatcattcacgaatataccgttgccgcgccagctgctgggttat cccgcgagcaaatccgccaggcacagattaatggtctggaaatggctttcctcagcgctgaggaaaaacgcgcactgcgagaaaaagtcgccgcgaagtaa

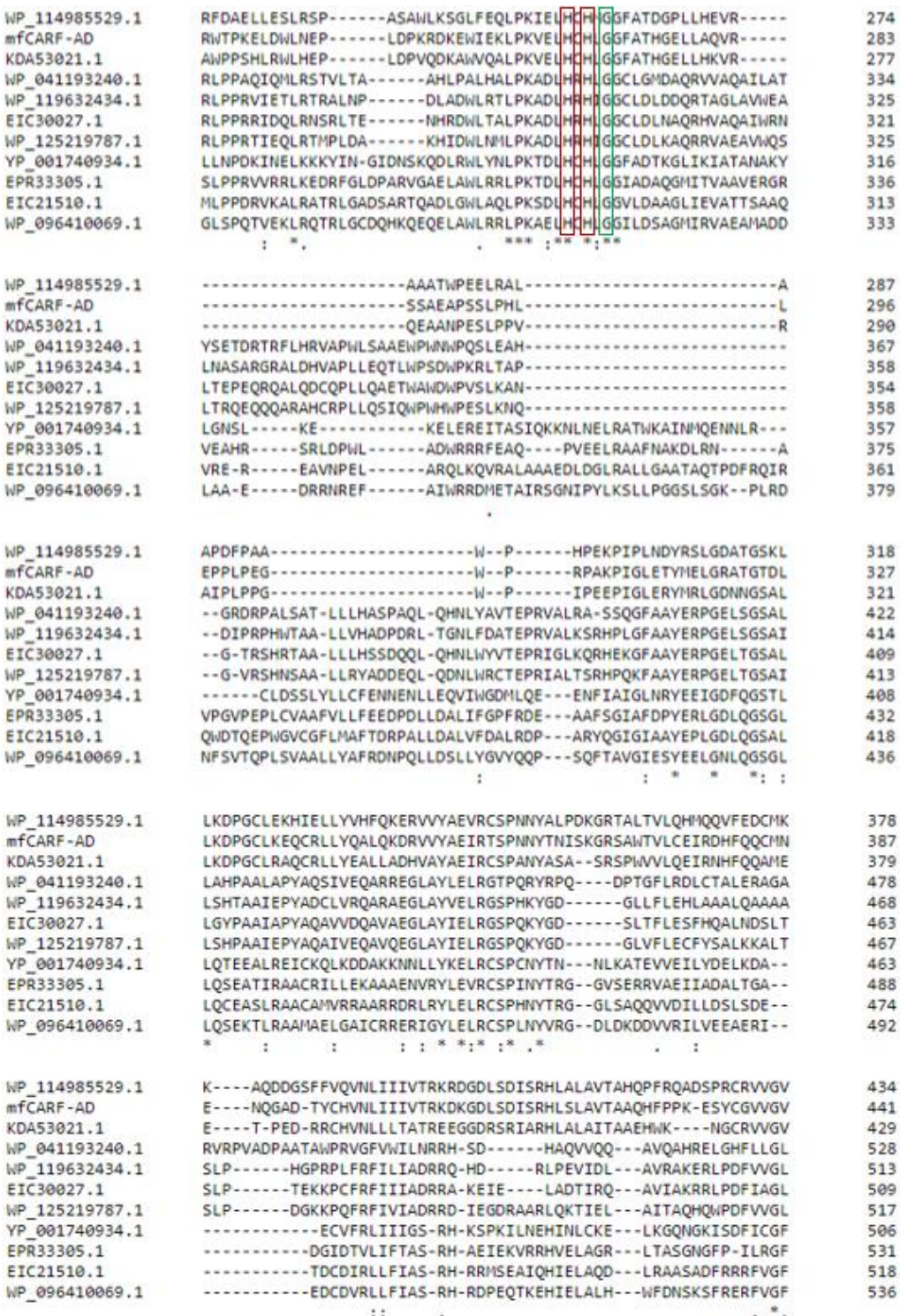
***Red color and underlined**: nucleotide substitution*

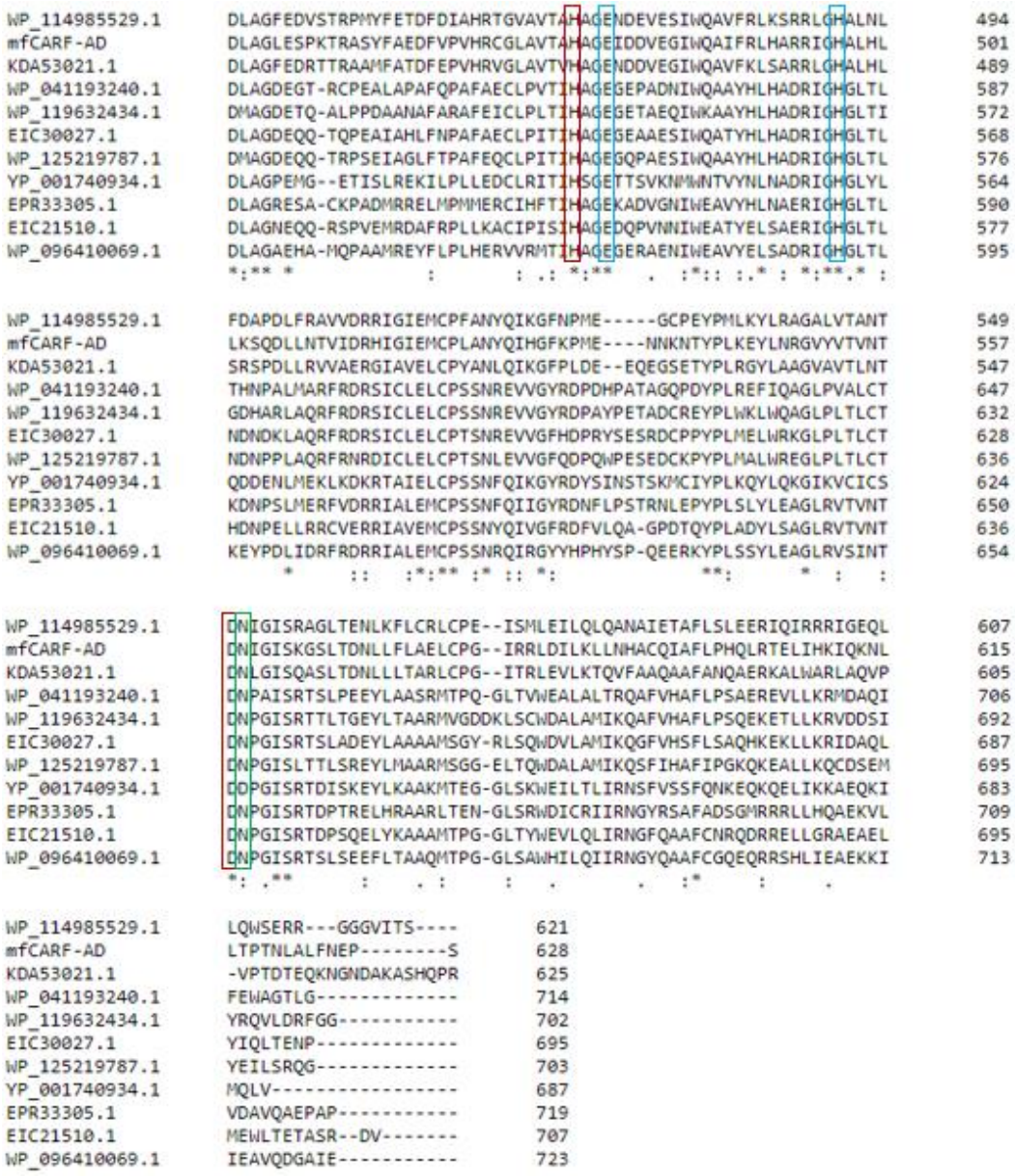
Supplementary Table 5. Amino acid sequences of deaminase proteins used in this study.

Protein/ Fusion protein	Sequence (5' -> 3')
mfCARF-AD	MDSSPQKNVLLCTLGSTLSKTWLIAVEALYFVNHYHQKLDEVHLLTTCNPEAEECFQNIEQCLKIHFPGLIYTFTQVDECADICSQEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTMSAALLKGASLFGAKYVFDVLAEEPFPSSMEE VKEALKTNKIKFVSLGNEKGWTQLKKLSFEEFPLNEIERKPRLYLATVADRKLQEKINEVQDYSLRIHYNWGRLSELPFPVLARWTPKELDWLNEPLDPKRDKEWIEKLPKVELHCHLGGFATHGELLAQVRSSAEAPSSLPHLLEPPLPE GWPRPAKPIGLETYMELGRATGTDLLKDPGCLKEQCRLLYQALQKDRVVYAEIRTSPNNYTNISKGRSAWTVLCEIRDHFQQCMNENQGADTYCHVNLIIIVTRKDKGDLSDISRHLSLAVTAAQHFPPKESYCGVVGVDLAGLESPK TRASYFAEDFVPVHRCGLAVTAHAGEIDDVEGIWQAIFRLHARRIGHALHLLKSQDLLNTVIDRHIGIEMCPLANYQIHGFKPMENNKNTYPLKEYLNRGVYVTVNTDNIGISKGSLTDNLLFLAELCPGIRRLDILKLLNHACQIAFLPH QLRTELIHKIQKNLLTPTNLALFNEPS*
mfCARF-AD_N559D	MDSSPQKNVLLCTLGSTLSKTWLIAVEALYFVNHYHQKLDEVHLLTTCNPEAEECFQNIEQCLKIHFPGLIYTFTQVDECADICSQEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTMSAALLKGASLFGAKYVFDVLAEEPFPSSMEE VKEALKTNKIKFVSLGNEKGWTQLKKLSFEEFPLNEIERKPRLYLATVADRKLQEKINEVQDYSLRIHYNWGRLSELPFPVLARWTPKELDWLNEPLDPKRDKEWIEKLPKVELHCHLGGFATHGELLAQVRSSAEAPSSLPHLLEPPLPE GWPRPAKPIGLETYMELGRATGTDLLKDPGCLKEQCRLLYQALQKDRVVYAEIRTSPNNYTNISKGRSAWTVLCEIRDHFQQCMNENQGADTYCHVNLIIIVTRKDKGDLSDISRHLSLAVTAAQHFPPKESYCGVVGVDLAGLESPK TRASYFAEDFVPVHRCGLAVTAHAGEIDDVEGIWQAIFRLHARRIGHALHLLKSQDLLNTVIDRHIGIEMCPLANYQIHGFKPMENNKNTYPLKEYLNRGVYVTVNTDDIGISKGSLTDNLLFLAELCPGIRRLDILKLLNHACQIAFLPH QLRTELIHKIQKNLLTPTNLALFNEPS*
mfCARF-AD_H497E	MDSSPQKNVLLCTLGSTLSKTWLIAVEALYFVNHYHQKLDEVHLLTTCNPEAEECFQNIEQCLKIHFPGLIYTFTQVDECADICSQEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTMSAALLKGASLFGAKYVFDVLAEEPFPSSMEE VKEALKTNKIKFVSLGNEKGWTQLKKLSFEEFPLNEIERKPRLYLATVADRKLQEKINEVQDYSLRIHYNWGRLSELPFPVLARWTPKELDWLNEPLDPKRDKEWIEKLPKVELHCHLGGFATHGELLAQVRSSAEAPSSLPHLLEPPLPE GWPRPAKPIGLETYMELGRATGTDLLKDPGCLKEQCRLLYQALQKDRVVYAEIRTSPNNYTNISKGRSAWTVLCEIRDHFQQCMNENQGADTYCHVNLIIIVTRKDKGDLSDISRHLSLAVTAAQHFPPKESYCGVVGVDLAGLESPK TRASYFAEDFVPVHRCGLAVTAHAGEIDDVEGIWQAIFRLHARRIGEALHLLKSQDLLNTVIDRHIGIEMCPLANYQIHGFKPMENNKNTYPLKEYLNRGVYVTVNTDNIGISKGSLTDNLLFLAELCPGIRRLDILKLLNHACQIAFLPH QLRTELIHKIQKNLLTPTNLALFNEPS*
ecADA1	MIDTTLPLTDIHRHLDGNIRPQTILELGRQYNISLPAQSLETLIPHVQVIANEPDLVSFLTKLDWGVKVLASLDACRRVAFENIEDAARHGLHYVELRFSPGYMAMAHQLPVAGVVEAVIDGVREGCRTFGVQAKLIGIMSRTFGEAAC QQELEAFLAHRDQITALDLAGDELGFPGSLFLSHFNRARDAGWHITVHAGEAAGPESIWQAIRELGAERIGHGVKAIEDRALMDFLAEQQIGIESCLTSNIQTSTVAELAAHPLKTFLEHGIRASINTDDPGVQGVDIIHEYTVAAPAAG LSREQIRQAQINGLEMAFLSAEEKRALREKVAAK*

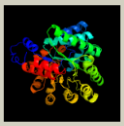
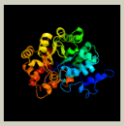
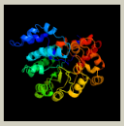
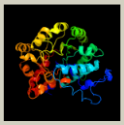
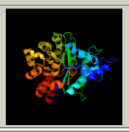
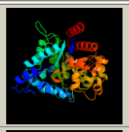
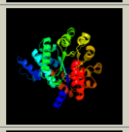
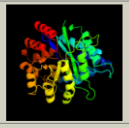
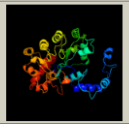
***Red color and underlined**: amino acid substitution*


CLUSTAL O(1.2.4) multiple sequence alignment
WP_114985529.1 -----MPHTLLCSLGTSP----AIVLEALIAGPSGRL--------------------RDR 31
mfCARF-AD MDSSPQKNVLLCTLGST-SKTWLIAVEALYF--VNHY--------------------HQK 38
KDA53021.1 ------MRILLCSVGT----SWAVVPEAMQL--LG----------------------SQG 26
WP_041193240.1 ---MPDRNILLCALCA----SWAVIPEVFGWLAPAMLDLYAQHPAKQSLDAIRSRYGLAA 53
WP_119632434.1 ----MKTNILLCTLGA----SWAVIPEVYGFLAPDRLPLYRRHPQRADFDGQRDEHRLEA 52
EIC30027.1 -----MRNILLCTLGA----SWAVVPEVYAFLAPGKFPLYHHHPQREKLLGLRAQYRLAE 51
WP_125219787.1 -----MKNIVLMTLGA----SWAVVPEVFGFLAPDRLPLYAEHPDNDSFERLRNEYGLQA 51
YP_001740934.1 ------MSQLIVSMGK----TSHLIAEVVGFLNYQQYPLYCNHKNFEALQAEIDENQLKP 50
EPR33305.1 ----MQKNIILTTMGT----TWQIPPELYGFTNPAALDLYRHHPERVRIEQLRTEYGVAP 52
EIC21510.1 ---------MISTLGT----SWSIVPELFAFTNPQQLDLYAHHESCAQIAAQRAEYGIVG 47
WP_096410069.1 ------MNTLVATLGT----TWQVLPEIFAYTNPGAAPLYEHSSAAGDIQEERKDYGLRP 50
WP_114985529.1 YARLRCITTSNIP--DKVKWLRARLADREFA--DLDVEIIEVPDFGELSSQEDHERFMES 87
mfCARF-AD LDEVHLLTTCNPEAEECFQNI-EQCLKIHFP--GLIYTFTQVDECADICSQEDYFLFQEV 95
KDA53021.1 FDEVHVLTTASSKISPGVEQL-LRYFE-MHP--GPRFSISRVQDFEDLRSEQDHMLFEEV 82
WP_041193240.1 PSELWICTTQGAQTLASLDRLRA-WWEGLGR--PCPLRIWCAQGTDQLATQRECDHIREL 110
WP_119632434.1 PDEIWVCTTQGEKPEEGIAALLD-WHARLEH--PPTLRIWQAEGTNQLASQVECERFREV 109
EIC30027.1 PDEIWVCTTQGT--QRSLDSLKA-WIACWPR--SPILRVWRAESTDQLATQQECAAIREL 106
WP_125219787.1 PDEIWLCTTQGEATQKSLSSLIA-WSRLLTK--PMTMRIWQAADTDELATQAECERMREL 108
YP_001740934.1 VTSITMICTEDSFTTS-KQEVEN-WLTAQNL--NIELQFHKLNQLTDILSQNDARMMRNL 106
EPR33305.1 VEELWVVTTGGSRIDTVIGKLIS-WHTNVHPGLRPVLRIWRLKDVNDLVDARECGRMREV 111
EIC21510.1 VNRIWVVTTEKMRT----DRLRE-WAARIEM--P--IDIWHPLGVDELASVRENRAMADL 98
WP_096410069.1 VQSLWIITTEGGIEE--WQNLCQ-WQQYLPE--PIEMRCWYIKGIEELFTPGENRAMADL 105
WP_114985529.1 LIGFYLTHGTGADGPVSGERMLPDVCISGGYKTMSSGLLQAAHFLGARRVFHVFAEQ-FR 146
mfCARF-AD LYRWFLKHI-Q--------RATVYVCIASGRKTMSAALLKGASLFGAKYVFDVLAEEPFP 146
KDA53021.1 LWRWLLQRAPQ--------AAHRYICLAGGYKTISAAMQRAAALFGACEVFHVLCEPRFG 134
WP_041193240.1 TFRVTLLASHQCG-----RQGQLLLSLAGGRKTMSADLQTAGSLFGAHAWLHVVGPE--P 163
WP_119632434.1 LLRLCLLAHE-RA-----GDGQVVFSLAGGRKTMSADLQWAGSVLGCDALIHVVGQE--P 161
EIC30027.1 IVRACLKAHH-HS-----GGGQVVLSLAGGRKTMSADMQWAGSLLGCQALLHVISAD--G 158
WP_125219787.1 IARACLKAHA-EA-----RGGQVVLSLAGGRKTMSADLQWAAGLFGCQALLHVVGKDFKV 162
YP_001740934.1 IYSIVYKVYKTG------NTQDLYLCLSGGRKTMSSDIQQAAYLFGCKAMIHILAEGMVN 160
EPR33305.1 IFRLALMAAAACG------DDGPLFSLTGGRKTMSSDLQEAAGWFGCRALVHVIDNPEYG 165
EIC21510.1 IYRVVLRGREVTR------NGRLYLSLAGGRKTMSAEMQQAGMLFGCDALLHVVDRQLPS 152
WP_096410069.1 IYRVALHARLHTR----RNNSCLYFALAGGRKTMSAELQQAAHLFGADALLHVVDRFAKQ 161
WP_114985529.1 --EVYN------R--HP-ETITDIRKGFDEDLVKVIEMGAEPGRSGILL-------Q--- 185
mfCARF-AD ------------------SSMEEVKEALKTNKIKFVSLGNEKGWTQLKK-------LSFE 181
KDA53021.1 --PQGN------R--EA-STLEEVEQAIATNALRFVRLGPEPGWPQLRL-------LSAP 176
WP_041193240.1 MPAALMARTEAERQAQINLFQHPLPPELAAAITPLVA-GTGTRNELLDIEIDGLR-VDDR 221
WP_119632434.1 LPKLL-------REPTVEFFTRELPAEIVPTVVPLVA-GVGRRSELLDVELDGRGAVSSR 213
EIC30027.1 LPKEF-------REALPSVFTQPLPPEWAALVTPLVA-GQTVRSDLLDIKVDAAEPIKTE 210
WP_125219787.1 MPKAL-------QDAAPELFTRPFDTDCCLAVTPLIT-GKTQRSDLLDLRMDRTEAIATE 214
YP_001740934.1 ----F----------DLETTPENIAYQEINKITPVLYNKDIPAGKLAELMED-------N 199
EPR33305.1 --NAL-------RVLEPPDFAGPLPPETAGGAMPLVTLGRCEPSALLEMDSDGAGAVRAD 216
EIC21510.1 GIERL-------SDQDVGAFCAPLPKPYADIYQPLVTLGRCPPNAALEVHPS----IRSA 201
WP_096410069.1 EREQF-------NSLSFQSLCQPLPKEFADSIRPLVTNGDLPGNEALLNSFQELDCEFEE 214
WP_114985529.1 ---LEALRAA--PHK-------MN---FIRRQLHQVDQLAGHAET-----LQELPFPVLA 225
mfCARF-AD EFPLNEIERK--PRLYLATVADRKLQEKINEVQDYSLRIHYNWGR-----LSELPFPVLA 234
KDA53021.1 SFPLESTLQG--PVHWVRA-SDMRLRQHVEGVLERSRHILAAWEG-----ISELPIPALA 228
WP_041193240.1 SFSVPLPDSGAECA-WPLPQTGPLLQDELQRRLHESSRLMGNFLAHLAESEHHENWRSLY 280
WP_119632434.1 GFPLPLAEPGAPLP-WEHT-EGPSLCDEIRRREAAGSQLLGNYLRALGEAEHHENWRSVY 271
EIC30027.1 DYPLPLPEAGQIAT-FAD--SGASLGQELSKREKASSRLFGNYLQDISRDETHENWRSLY 267
WP_125219787.1 AFPLPLPKSGLPQV-WAA--DTLFLAQELKKREQHGSRLLGNYLSELSRSEHHENWRSLY 271
YP_001740934.1 EVKLPKAE--ITDNIYSYNDEDTSFLDLVEKLQAQQDNLTINYYNKLRYSEPASNFQILH 257
EPR33305.1 QWPLPLAENSVPSEIEEGDLEGPWLSEVLAARRRKAGYHYSQYTVGLIKGDETTNFLALY 276
EIC21510.1 DFPVPPAGGA-------VPPELPLYEE-VRRRLRQADSLIYNFSCQVSQAEAQSNFHGLY 253
WP_096410069.1 RFPLPGFGDQQRFSIESVPAQDDLHAW-VKERQKRAEALLANYRLQVSAHEKLGNFRALY 273

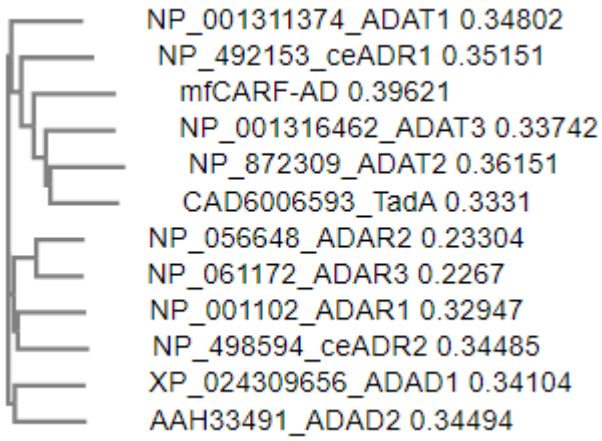

WP_114985529.1 RFDAELLESLRSP------ASAWLKSGLFEQLPKIELHCHHGGFATDGPLLHEVR----- 274
mfCARF-AD RWTPKELDWLNEP------LDPKRDKEWIEKLPKVELHCHLGGFATHGELLAQVR----- 283
KDA53021.1 AWPPSHLRWLHEP------LDPVQDKAWVQALPKVELHCHLGGFATHGELLHKVR----- 277
WP_041193240.1 RLPPAQIQMLRSTVLTA------AHLPALHALPKADLHRHLGGCLGMDAQRVVAQAILAT 334
WP_119632434.1 RLPPRVIETLRTRALNP------DLADWLRTLPKADLHRHIGGCLDLDDQRTAGLAVWEA 325
EIC30027.1 RLPPRRIDQLRNSRLTE------NHRDWLTALPKADLHRHLGGCLDLNAQRHVAQAIWRN 321
WP_125219787.1 RLPPRTIEQLRTMPLDA------KHIDWLNMLPKADLHRHIGGCLDLKAQRRVAEAVWQS 325
YP_001740934.1 LLNPDKINELKKKYIN-GIDNSKQDLRWLYNLPKTDLHCHLGGFADTKGLIKIATANAKY 316
EPR33305.1 SLPPRVVRRLKEDRFGLDPARVGAELAWLRRLPKTDLHCHLGGIADAQGMITVAAVERGR 336
EIC21510.1 MLPPDRVKALRATRLGADSARTQADLGWLAQLPKSDLHCHLGGVLDAAGLIEVATTSAAQ 313
WP_096410069.1 GLSPQTVEKLRQTRLGCDQHKQEQELAWLRRLPKAELHCHLGGILDSAGMIRVAEAMADD 333
WP_114985529.1 ----------AAATWPEELRAL----------A 287
mfCARF-AD ----------SSAEAPSSLPHL----------L 296
KDA53021.1 ----------QEAANPESLPPV----------R 290
WP_041193240.1 YSETDRTRFLHRVAPWLSAAEWPWNWPQSLEAH---------- 367
WP_119632434.1 LNASARGRALDHVAPLLEQTLWPSDWPKRLTAP---------- 358
EIC30027.1 LTEPEQRQALQDCQPLLQAETWANDWPVSLKAN---------- 354
WP_125219787.1 LTRQEQQQARAHCRPLLQSIQWPWHWPESLKNQ---------- 358
YP_001740934.1 LGNSL-----KE----------KELEREITASIQKKNLNELRATWKAINMQENNLR--- 357
EPR33305.1 VEAHR-----SRLDPWL------ADWRRRFEAQ----PVEELRAAFNAKDLRN------A 375
EIC21510.1 VRE-R-----EAVNPEL------ARQLKQVRALAAAEDLDGLRALLGAATAQTPDFRQIR 361
WP_096410069.1 LAA-E-----DRRNREF------AIWRRDMETAIRSGNIPYLKSLLPGGSLSGK--PLRD 379
WP_114985529.1 APDFPAA----------W--P------HPEKPIPLNDYRSLGDATGSKL 318
mfCARF-AD EPPLPEG----------W--P------RPAKPIGLETYMELGRATGTDL 327
KDA53021.1 AIPLPPG----------W--P------IPEEPIGLERYMRLGDNNGSAL 321
WP_041193240.1 --GRDRPALSAT-LLLHASPAQL-QHNLYAVTEPRVALRA-SSQGFAAYERPGELSGSAL 422
WP_119632434.1 --DIPRPHWTAA-LLVHADPDRL-TGNLFDATEPRVALKSRHPLGFAAYERPGELSGSAI 414
EIC30027.1 --G-TRSHRTAA-LLLHSSDQQL-QHNLWYVTEPRIGLKQRHEKGFAAYERPGELTGSAL 409
WP_125219787.1 --G-VRSHNSAA-LLRYADDEQL-QDNLWRCTEPRIALTSRHPQKFAAYERPGELTGSAI 413
YP_001740934.1 ------CLDSSLYLLCFENNENLLEQVIWGDMLQE---ENFIAIGLNRYEEIGDFQGSTL 408
EPR33305.1 VPGVPEPLCVAAFVLLFEEDPDLLDALIFGPFRDE---AAFSGIAFDPYERLGDLQGSGL 432
EIC21510.1 QWDTQEPWGVCGFLMAFTDRPALLDALVFDALRDP---ARYQGIGIAAYEPLGDLQGSAL 418
WP_096410069.1 NFSVTQPLSVAALLYAFRDNPQLLDSLLYGVYQQP---SQFTAVGIESYEELGNLQGSGL 436
WP_114985529.1 LKDPGCLEKHIELLYVHFQKERVVYAEVRCSPNNYALPDKGRTALTVLQHMQQVFEDCMK 378
mfCARF-AD LKDPGCLKEQCRLLYQALQKDRVVYAEIRTSPNNYTNISKGRSAWTVLCEIRDHFQQCMN 387
KDA53021.1 LKDPGCLRAQCRLLYEALLADHVAYAEIRCSPANYASA--SRSPWVVLQEIRNHFQQAME 379
WP_041193240.1 LAHPAALAPYAQSIVEQARREGLAYLELRGTPQRYRPQ----DPTGFLRDLCTALERAGA 478
WP_119632434.1 LSHTAAIEPYADCLVRQARAEGLAYVELRGSPHKYGD------GLLFLEHLAAALQAAAA 468
EIC30027.1 LGYPAAIAPYAQAVVDQAVAEGLAYIELRGSPQKYGD------SLTFLESFHQALNDSLT 463
WP_125219787.1 LSHPAAIEPYAQAIVEQAVQEGLAYIELRGSPQKYGD------GLVFLECFYSALKKALT 467
YP_001740934.1 LQTEEALREICKQLKDDAKKNNLLYKELRCSPCNYTN---NLKATEVVEILYDELKDA-- 463
EPR33305.1 LQSEATIRAACRILLEKAAAENVRYLEVRCSPINYTRG--GVSERRVAEIIADALTGA-- 488
EIC21510.1 LQCEASLRAACAMVRRAARRDRLRYLELRCSPHNYTRG--GLSAQQVVDILLDSLSDE-- 474
WP_096410069.1 LQSEKTLRAAMAELGAICRRERIGYLELRCSPLNYVRG--DLDKDDVVRILVEEAERI-- 492
WP_114985529.1 K----AQDDGSFFVQVNLIIIVTRKRDGDLSDISRHLALAVTAHQPFRQADSPRCRVVGV 434
mfCARF-AD E----NQGAD-TYCHVNLIIIVTRKDKGDLSDISRHLSLAVTAAQHFPPK-ESYCGVVGV 441
KDA53021.1 E----T-PED-RRCHVNLLLTATREEGGDRSRIARHLALAITAAEHWK----NGCRVVGV 429
WP_041193240.1 RVRPVADPAATAWPRVGFVWILNRRH-SD------HAQVVQQ---AVQAHRELGHFLLGL 528
WP_119632434.1 SLP------HGPRPLFRFILIADRRQ-HD-----RLPEVIDL---AVRAKERLPDFVVGL 513
EIC30027.1 SLP------TEKKPCFRFIIIADRRA-KEIE----LADTIRQ---AVIAKRRLPDFIAGL 509
WP_125219787.1 SLP------DGKKPQFRFIVIADRRD-IEGDRAARLQKTIEL---AITAQHQWPDFVVGL 517
YP_001740934.1 ----------ECVFRLIIIGS-RH-KSPKILNEHINLCKE---LKGQNGKISDFICGF 506
EPR33305.1 ----------DGIDTVLIFTAS-RH-AEIEKVRRHVELAGR---LTASGNGFP-ILRGF 531
EIC21510.1 ----------TDCDIRLLFIAS-RH-RRMSEAIQHIELAQD---LRAASADFRRRFVGF 518
WP_096410069.1 ----------EDCDVRLLFIAS-RH-RDPEQTKEHIELALH---WFDNSKSFRERFVGF 536



Supplementary Fig. 1 Multiple sequence alignment of mfCARF-AD with members of the CARF5 clade using ClustalO. Highlighted residues are essential for cOA-degradation (orange), cOA-binding (yellow), zinc-binding (red), substrate-binding (green), and substrate deamination (blue).

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c6ii7A_	Alignment		100.0	25	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**adenosine deaminase; **PDBTitle:** crystal structure of plasmodium falciparum adenosine deaminase2 c27q+l227i mutant co-complexed with zn ion, hypoxanthine and inosine **PDB Entry:** PDBe RCSB PDBj Run Investigator
2	c6n91A_	Alignment		100.0	26	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**adenosine deaminase; **PDBTitle:** crystal structure of adenosine deaminase from vibrio cholerae2 complexed with pentostatin (deoxycoformycin) **PDB Entry:** PDBe RCSB PDBj Run Investigator
3	c3ou8A_	Alignment		100.0	27	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**adenosine deaminase; **PDBTitle:** the crystal structure of adenosine deaminase from pseudomonas2 aeruginosa **PDB Entry:** PDBe RCSB PDBj Run Investigator
4	c3ou8B_	Alignment		100.0	26	**PDB header:**hydrolase **Chain:** B: **PDB Molecule:**adenosine deaminase; **PDBTitle:** the crystal structure of adenosine deaminase from pseudomonas2 aeruginosa **PDB Entry:** PDBe RCSB PDBj Run Investigator
5	d1vfla1	Alignment		100.0	24	**Fold:** TIM beta/alpha-barrel **Superfamily:** Metallo-dependent hydrolases **Family:** Adenosine/AMP deaminase **PDB entry:** PDBe RCSB PDBj Run Investigator
6	d1a4ma_	Alignment		100.0	24	**Fold:** TIM beta/alpha-barrel **Superfamily:** Metallo-dependent hydrolases **Family:** Adenosine/AMP deaminase **PDB entry:** PDBe RCSB PDBj Run Investigator
7	d2amxa1	Alignment		100.0	23	**Fold:** TIM beta/alpha-barrel **Superfamily:** Metallo-dependent hydrolases **Family:** Adenosine/AMP deaminase **PDB entry:** PDBe RCSB PDBj Run Investigator
8	c6jpA_	Alignment		100.0	22	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**adenosine/amp deaminase family protein; **PDBTitle:** the structure of the adal-imp complex **PDB Entry:** PDBe RCSB PDBj Run Investigator
9	c3rysA_	Alignment		100.0	22	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**adenosine deaminase 1; **PDBTitle:** the crystal structure of adenine deaminase (aaur1117) from2 arthrobacter aurescens **PDB Entry:** PDBe RCSB PDBj Run Investigator
10	d2a3la1	Alignment		100.0	22	**Fold:** TIM beta/alpha-barrel **Superfamily:** Metallo-dependent hydrolases **Family:** Adenosine/AMP deaminase **PDB entry:** PDBe RCSB PDBj Run Investigator

Supplementary Fig. 2 Prediction of the mfCARF-AD function using Phyre2. Proteins with the highest structural homology to mfCARF-AD are shown at the top of the list.



Supplementary Fig. 3 Phylogenetic neighbour-joining tree of mfCARF-AD with adenosine deaminases acting on nucleic acids (ADNs) using ClustalO.

```
CLUSTAL O(1.2.4) multiple sequence alignment


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         ------------------------------------------------------------    0
NP_001102_ADAR1         --MNPRQGYSLSGYYTHPFQGYEHR------QLRYQQPGPGSSPSSFLLKQIEFLKGQLP    52
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        MDQNPNYNFGYGQAYGSGTDHTSDSTNNYNWASQWSQPESAASLAT--------------    46
mfCARF-AD               MDSSPQK-------------------------------------------------------    7
NP_001316462_ADAT3      ------------------------------------------------------------    0
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         ------------------------------------------------------------    0
NP_001102_ADAR1         EAPVIGKQTPSLPPSLPGLRPRFPVLLASSTRGRQVDIRGVPRGVHLRSQGLQRGFQHPS    112
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        --------------------HTFPQYVSQQQQQQQQQ-----------------------    63
mfCARF-AD               ------------------------------------------------------------    7
NP_001316462_ADAT3      ------------------------------------------------------------    0
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         ------------------------------------------------------------    0
NP_001102_ADAR1         PRGRSLPQRGVDCLSSHFQELSIYQDQEQRILKFLEELGEGKATTAHDLSGKLGTPKKEI    172
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        ------AQQQAQNTYAAMNPISTFMQQQQRAQTFPQK----------KYGQQGGAPKPSA    107
mfCARF-AD               ------------------------------------------------------------    7
NP_001316462_ADAT3      ------------------------------------------------------------    0
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         ------------------------------------------------------------    0
NP_001102_ADAR1         NRVLYSLAKKGKLQKEAGTPPLWKIAVSTQAWNQHSGVVRPDGHSQGAPN----------    222
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        IRN----NNFGAFGGGHALSQEWVQPMSQNQMGGPQGNRFFNNQKGGPFNQNKPNWRQNK    163
mfCARF-AD               --------NVLLCTLGSTLSKTWLIAVEALY---------FVNHYHQKLD----EVHLLT    46
NP_001316462_ADAT3      ------------------------------------------------------------    0
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0
```

```
NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         ------------------------------------------------------------    0
NP_001102_ADAR1         -SDPSLEPEDRNSTSVS-EDLLEPFIAVSAQAWNQHSGVVRPDSHSQGSPNSDPGLEPED    280
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        PKGPAA-PKKFDSTGKSPAMLLHELFKDVSEEYT------------------------    196
mfCARF-AD               TCNPEA-EECFQNI--------------------------------------------    59
NP_001316462_ADAT3      ------------------------------------------------------------    0
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         ----------------------------------------------------MASVLGSGRG    10
NP_001102_ADAR1         SNSTSALEDPLEFLDMAEIKEKICDY-LFNVSDSSALNLAKNIGLTKARDINAVLIDMER    339
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        --------------EVEGVPKKYCCTLKVNGRTFQME----SVNKKAAKQK-------CS    231
mfCARF-AD               --------------------EQCLKIHFPGLIYTFT----QVDECA--DI-------CS    85
NP_001316462_ADAT3      ------------------------------------------------------------    0
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ------------------------------------------------------------    0
NP_061172_ADAR3         SG------GLSSQLKCKSKRRRRRS-------------------KRK---DKVSILSTF    42
NP_001102_ADAR1         QGDVYRQGTTPPIWHLTDKKRERMQI------------------KRN----TNSVPETA    376
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        EL---VVRDLRPDVHVTPFEEGVAAKAAAPVKKEIDAASGNGQNNKRNLLQADAISNQPT    288
mfCARF-AD               QEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTMSAALLKGASLFGAKYVFDVLAEEPF    145
NP_001316462_ADAT3      ------------------------------------------------------MEPA    4
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0


NP_001311374_ADAT1      ------------------------------------------------------------    0
NP_056648_ADAR2         ----------------------------------------MDIEDEENMSSSS-------    13
NP_061172_ADAR3         LA--P-----------------------FKHLS---PGITNTEDDDTLSTSS-------    66
NP_001102_ADAR1         PAAIPETKRN------------------AEFLT---CNIPTSNASNNMVTTEKVENGQ-    413
NP_498594_ceADR2        ------------------------------------------------------------    0
XP_024309656_ADAD1      ------------------------------------------------------------    0
AAH33491_ADAD2          ------------------------------------------------------------    0
NP_492153_ceADR1        PKKVSAVKK----AKLQLTPV----------------------E-----SALSLLDLMQK    317
mfCARF-AD               PSSMEEVKEALKTNKIKFVSLGNEKGWTQLKKLSFEEFPLNEIERKPRLYLATVADRKLQ    205
NP_001316462_ADAT3      PGLVEQPKC-LEAG----SPEPEPAPWQALPVLSEKQSG--DVEL-VLAYAAPVLDKRQT    56
NP_872309_ADAT2         ------------------------------------------------------------    0
CAD6006593_TadA         ------------------------------------------------------------    0
```

```
NP_001311374_ADAT1  ------------------------------------------------------------  0
NP_056648_ADAR2     --------TDVKENRNLDNVSPKDG-------------------------------------  30
NP_061172_ADAR3     --------AEVKENRNVGNLAARPP-------------------------------------  83
NP_001102_ADAR1     --------EPVIKLENRQEARPEPAR------------------------------------  431
NP_498594_ceADR2    ------------------------------------------------------------  0
XP_024309656_ADAD1  ------------------------------------------------------------  0
AAH33491_ADAD2      ------------------------------------------------------------  0
NP_492153_ceADR1    IIAESAEKYS-----PVFEASEVPKDPEIPEVEVKKEEVDTNGENVANEKKSGWRKNETM  372
mfCARF-AD           EKINEVQDYSLRIHYNWGRLSELPF-PVLARWT--PKELDWLNEPLDPKRDKEWIEKLPK  262
NP_001316462_ADAT3  SRL----------LKEVSALHPLPAQPHLK--------------RVRPSRDAG-------  85
NP_872309_ADAT2     ------------------------------------------------------------  0
CAD6006593_TadA     ------------------------------------------------------------  0

NP_001311374_ADAT1  ------------------------------------------------------------  0
NP_056648_ADAR2     ------------------------------------------------------STPGPGEG  38
NP_061172_ADAR3     ------------------------------------------------------------PS  85
NP_001102_ADAR1     ---------------------------------------------L----KPPVHYNGPSKA  444
NP_498594_ceADR2    ------------------------------------------------------------  0
XP_024309656_ADAD1  ------------------------------------------------------------  0
AAH33491_ADAD2      ------------------------------------------------------------  0
NP_492153_ceADR1    HNVTLKF---VEQN-----------------------------------------------K  384
mfCARF-AD           VELHCHLGGFATHGELLAQVRSSAEAPSSLPHLLEPPLP--EGW------PRPAKPIGLE  314
NP_001316462_ADAT3  ---------SPHALEMLLCLAGPASGPRSLAELLPRPAVDPRGLGQPFLVPVPARPPLTR  136
NP_872309_ADAT2     ------------------------------------------------MEAKAAPKPAAS  12
CAD6006593_TadA     ------------------------------------------------------------  0

NP_001311374_ADAT1  ------------------------------------------------------------  0
NP_056648_ADAR2     SQL--SNGG------------GGGPGRKRPLEEGS--NGHSKYRLKKRRKTPGP--VLP  79
NP_061172_ADAR3     GDR--ARGG------------APGAKRKRPLEEGN--GGHLCKLQLVWKKLSWS--VAP  126
NP_001102_ADAR1     GYVDFENGQWATDDIPDDLNSIRAAPGEFRAIMEMPSFYSHGLPRCSPYKKLTECQLKNP  504
NP_498594_ceADR2    ------------------------------------------------------------  0
XP_024309656_ADAD1  ------------------------------------------------------------  0
AAH33491_ADAD2      ------------------------------------------------------------  0
NP_492153_ceADR1    QYTKMGPSR---GVLKDMVI---------REALRD-------------------------  407
mfCARF-AD           TYMELGRATG-TDLLKDP----G-C---LKEQCRL-------------------------  340
NP_001316462_ADAT3  GQFEEARAHWPTSFHEDKQVTSA-LAGRLFSTQER-------------------------  170
NP_872309_ADAT2     G--------------------A-CSVSAEETEKW-------------------------  25
CAD6006593_TadA     ----------------------MSEVEFSHEYW-------------------------  11

NP_001311374_ADAT1  ------------------------------------------------------------  0
NP_056648_ADAR2     KNALMQLN-EIKPGLQYTLLSQTGPVHAPLFVMSVEVNGQVFEGSGPTKKK-AKLHAAEK  137
NP_061172_ADAR3     KNALVQLH-ELRPGLQYRTVSQTGPVHAPVFAVAVEVNGLTFEGTGPTKKK-AKMRAAEL  184
NP_001102_ADAR1     ISGLLEYAQFASQTCEFNMIEQSGPPHEPRFKFQVVINGREFPPAEAGSKKVAKQDAAMK  564
NP_498594_ceADR2    ------------------------------------------------------------  0
XP_024309656_ADAD1  ------------------------------------------------------------  0
AAH33491_ADAD2      ------------------------------------------------------------  0
NP_492153_ceADR1    ---------------------------------------LFNVSHADITTVARRHASNR  427
mfCARF-AD           ---------------------------------------LYQALQKDRVVYAEI-----  355
NP_001316462_ADAT3  ---------------------------------------AAMQSHMERAVWAAR-----  185
NP_872309_ADAT2     ---------------------------------------MEEAMHMAKEALENT-----  40
CAD6006593_TadA     ---------------------------------------MRHALTLAKRAWDER-----  26
```

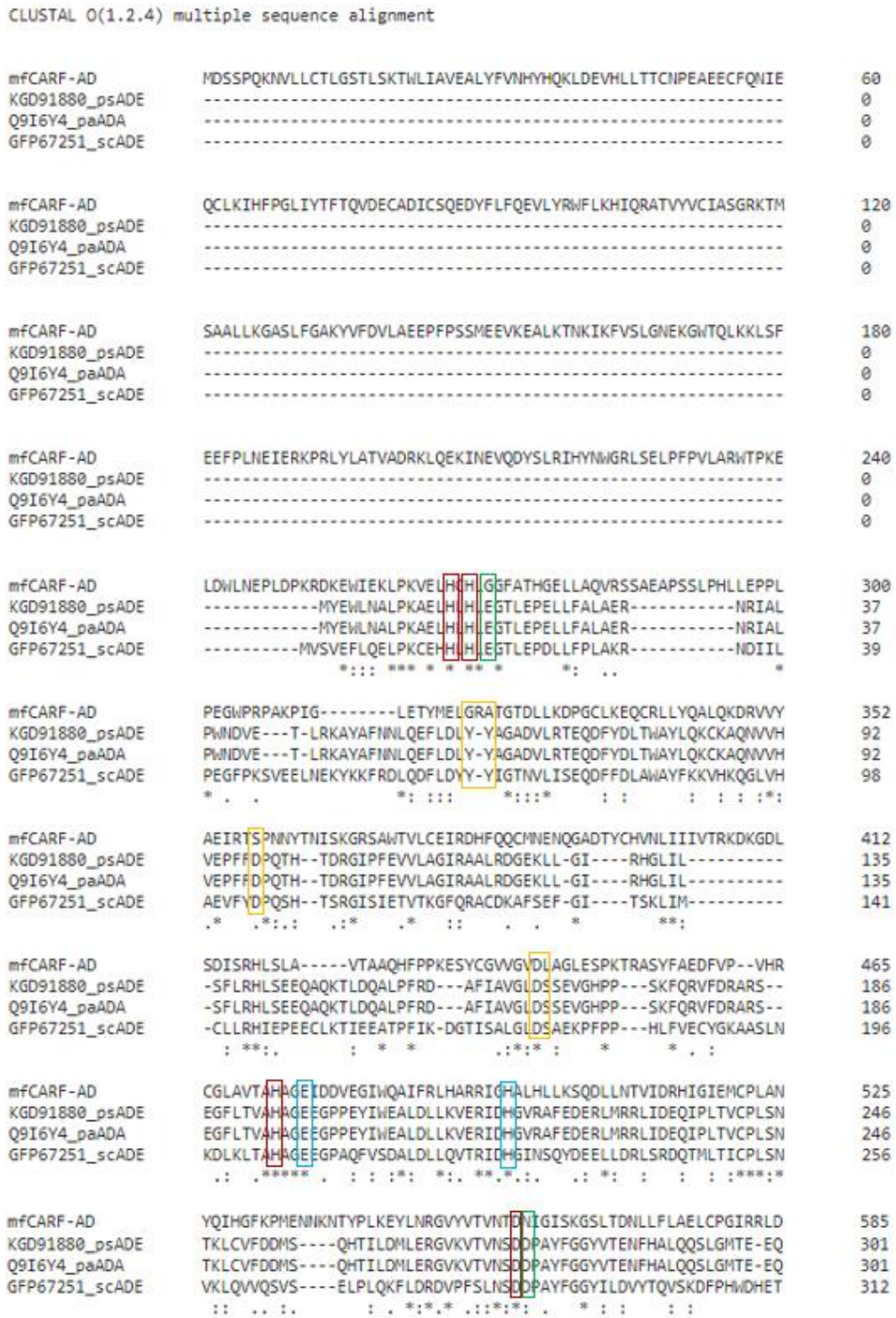
```
NP_001311374_ADAT1      ------------------------------------------------------------  0
NP_056648_ADAR2         A--LRSFVQFPNASEAHLAMGRTLSVNTDFTSDQADFPDTLFNGFETPDKAEPPFYVGSN  195
NP_061172_ADAR3         A--LRSFVQFPNACQAHLAMGGGPGPGTDFTSDQADFPDTLFQEFEPPAPR-PGLAGGRP  241
NP_001102_ADAR1         A--MTILLEEAKAKDSG-----KSEESSHYSTEKE---SEKTAESQTPTPSATSFFSGKS  614
NP_498594_ceADR2        ------------------------------------------------------------  0
XP_024309656_ADAD1      ---------MA-------------------------SNNHWFQSSQVPSFA--QML----  20
AAH33491_ADAD2          ---------MASASQGADDDGSRRKPRLAASLQISPQPRPWRPL---PAQ-----A----  39
NP_492153_ceADR1        LGHDTTILQCLNTICSI----------LNCTLTIECEPAE-----------DRPLGIGRA  466
mfCARF-AD               -----------RT----------------------SPNN-----------YTNISKGRS  370
NP_001316462_ADAT3      -----------RAAARG----------LRAVGAVVVDPAS-----------DRVLATGHD  213
NP_872309_ADAT2         --------------------------EVPVGCLMVY--N-----------NEVVGKGRN  60
CAD6006593_TadA         --------------------------EVPVGAVLVH--N-----------NRVIGEGWN  46

NP_001311374_ADAT1      ------------------------------------------------------------  0
NP_056648_ADAR2         ------------------------------------------------------GDD  198
NP_061172_ADAR3         ------------------------------------------------------GDA  244
NP_001102_ADAR1         PVTTLLECMHKLGNSCEFRLLSKEGPAHEPKFQYCVAV--------------------GAQ  655
NP_498594_ceADR2        ------------------------------------------------------------  0
XP_024309656_ADAD1      --------------------KKNLPVQPATK-TI---T--------------------TPT  37
AAH33491_ADAD2          --------------------QSAMGPAPAPA-TY---R--------------------AEG  56
NP_492153_ceADR1        YFMAKCTIIDHNENDLKFEVK-SSSLASK--------AMAKDWVA---------------  502
mfCARF-AD               AWTVLCEIRDHFQQCMNENQG-ADTYCHV--NLIIIVTRKDKGDLSDISRHLSLAVTAAQ  427
NP_001316462_ADAT3      C---------------S--CA-DNPLLHA--VMVCV-DLVARGQGR---GTYDFRPFPAC  249
NP_872309_ADAT2         E--------------VN--QT-KNATRHA--EMVAI-DQVLDW-----------------  83
CAD6006593_TadA         R--------------PI--GR-HDPTAHA--EIMAL-RQGGL------------------  68

NP_001311374_ADAT1      ------------------------------------------------------------  0
NP_056648_ADAR2         SFSSSG----------------------DLSLSASPVPA----------S--LAQPPL  222
NP_061172_ADAR3         ALLSAAYG---------------------RRRLLCRAL-------------D---LVGPT  267
NP_001102_ADAR1         TFPSVSAPSKKVAKQMAAEEAMKALHGEATNSMASDNQPEGMISESLDNLES--MMPNKV  713
NP_498594_ceADR2        --------------MSVEEGMEVEL----------------KSEK-MDLDD--NIPDFV  26
XP_024309656_ADAD1      GWSSESYGLSKMASKV------TQVTGNFPEPL----L-----SKNLSSISN--PVLPPK  80
AAH33491_ADAD2          GWPQVSVLRDSGP-GA------GAGVGELGAARAWENL-----GEQMGKAPR--VPVPPA  102
NP_492153_ceADR1        ------------------------------------------------QETLKNYF  510
mfCARF-AD               HFPPKESY--------------------------------CGVVGVDLAGLESPKTRASYF  456
NP_001316462_ADAT3      SFAPAAAP--------------------------------QAVRAGAVRKLDADEDGLPYL  278
NP_872309_ADAT2         ------------------------------------------------CRQSGKSPSEV  94
CAD6006593_TadA         ------------------------------------------------VMQNYR  74

NP_001311374_ADAT1      ------------------------------------------------------------  0
NP_056648_ADAR2         P----VLPPFPPPSGKNPVMILNELR-------PGLKYDFLSESGESHAK-S--FVMSVV  268
NP_061172_ADAR3         P----AT--PAAPGERNPVVLLNRLR-------AGLRYVCLAEPAERRAR-S--FVMAVS  311
NP_001102_ADAR1         R----KIGELVRYLNTNPVGGLL-EYAR--SHGFAAEFKLVDQSGPPHEP-K--FVYQAK  763
NP_498594_ceADR2        K-------ETVERSGKNPMSLFSELYVH--MTGNTPVFDFYNRNQPNGSM-K--FICVVI  74
XP_024309656_ADAD1      KIPKEFIMKYKR-GEINPVSALH-QFAQ--MQRVQLDLKETVTTGNVMGP-YFAFC--AV  133
AAH33491_ADAD2          GLS----LPLKDPPASQAVSLLT-EYAA--SLGIFLLFR----EDQPPGP-CFPFSVSAE  150
NP_492153_ceADR1        AIDPSSCVKTDAVSSQGPCALLHAMLNKQTKQKCKIAYEFKDNVPPVAGQATTTFYCECV  570
mfCARF-AD               AE---------DFVPVHRCGLAVTAH--------------AGEIDDVEGIWQA-------  486
NP_001316462_ADAT3      CTG------YDLYVTREPCAMCAMALV-----HARILRVFYGAPSP-DGALGT-------  319
NP_872309_ADAT2         FEH------TVLYVTVEPCIMCAAALR-----LMKIPLVVYGCQNERFGGCGS-------  136
CAD6006593_TadA         LID------ATLYVTLEPCVMCAGAMI-----HSRIGRVVFGARDAKTGAAGS-------  116
```

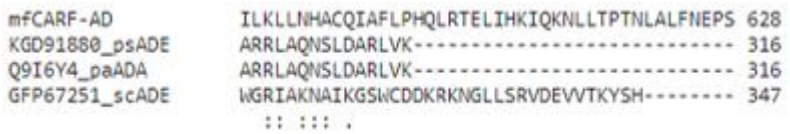
```
NP_001311374_ADAT1      ------------------------------------------------------------   0
NP_056648_ADAR2         VDGQFFEGSGRNKK-LAKARAAQSALAAIFNLH-LDQ-----TPSRQPIPSEGLQL----   317
NP_061172_ADAR3         VDGRTFEGSGRSKK-LARGQAAQAALQELFDIQ-MPG-----H-----APGRARRT----   355
NP_001102_ADAR1         VGGRWFPAVCAHSKKQGKQEAADAALRVLIGEN-EKAERM-GFTEVTPVTGASLRRTMLL   821
NP_498594_ceADR2        LNGERIEGNVKSKKKEAKVSCSLKGLEVV----------------------------   103
XP_024309656_ADAD1      VDGIQYKTGLGQNKKESRSNAAKLALDEL-LQLDEPEPRILETSGPPPFPAEPVVLSEL-   191
AAH33491_ADAD2          LDGVVCPAGTANSKTEAKQQAALSALCYIRSQLENPE--SPQTSSRPPLAP--LSV----   202
NP_492153_ceADR1        IDETDRYIGVGRSKKLAKSEAAMQALKKLFKIDYD--PAG----------NYPLAL----   614
mfCARF-AD               ----------------------------IFRLHA---RRI----------GHALHL----   501
NP_001316462_ADAT3      ----------------------------RFRI--HARPDL----------NHRFQV----   335
NP_872309_ADAT2         ----------------------------VLNIASADLPNT----------GRPFQC----   154
CAD6006593_TadA         ----------------------------LMDVLHH--PGM----------NHRVEI----   132

NP_001311374_ADAT1      ----------------MWTADEIAQLCYEHY---GIRLPKKGKPEPNHEWTLLA-AVVK   39
NP_056648_ADAR2         --------------HLPQVLADAVSRLVLGKF---GDLTD--NFSSPHARRKVLAGVVMT   358
NP_061172_ADAR3         --------------PMPQEFADSISQLVTQKF---REVTT--DLTPMHARHKALAGIVMT   396
NP_001102_ADAR1         LSRSPEA-QPKTLPLTGSTFHDQIAMLSHRCFNTLTNSF----QPSLLG-RKILAAIIMK   875
NP_498594_ceADR2        LKHVSEYVVPAVKFEEKTTFFE---MLREHTYAKFYELCK--NNALIYGFEKVIASVFLK   158
XP_024309656_ADAD1      ------AYVSKVHYEGRHIQYAKISQIVKERF---NQLIS--NRSEYLKYSSSLAAFIIE   240
AAH33491_ADAD2          --------------ENILTHEQRCAALVSAGF---DLLLD--ERSPYWACKGTVAGVILE   243
NP_492153_ceADR1        -TS-RAMTESKVSPLCRH-IAEFCKREYHQM----TEYYQ------IPPS-NLFAAFLLV   660
mfCARF-AD               -LKSQDLLN---TVIDRHIGIEMCPLANYQI----HGFKP------MENNKNTYPL----   543
NP_001316462_ADAT3      -FR--------------GVLEEQCRWL---D----PDT----------------------   351
NP_872309_ADAT2         -IP--------------GYRAEEAVEM---L----KTFYK------QENP----------   176
CAD6006593_TadA         -TE--------------GILADECAAL---L----SDFFR------MRRQ----------   154

NP_001311374_ADAT1      IQSPADKACDTPDKPVQVTKEVVSMGTGTKC------IGQSKMRKNGDILNDSHAEVIAR   93
NP_056648_ADAR2         TGT-----------DVKDAKVISVSTGTKC------INGEYMSDRGLALNDCHAEIISR   400
NP_061172_ADAR3         KGL-----------DARQAQVVALSSGTKC------ISGEHLSDQGLVVNDCHAEVVAR   438
NP_001102_ADAR1         KDS-----------E-DMGVVVSLGTGNRC------VKGDSLSLKGETVNDCHAEIISR   916
NP_498594_ceADR2        IN---------------GNLQIIALSTGNKG------LRGDKIVNDGTALIDCHAEILAR   197
XP_024309656_ADAD1      RAG--------------QHEVVAIGTGEYN------YSQD-IKPDGRVLHDTHAVVTAR   278
AAH33491_ADAD2          REIPRARG------HVKEIYKLVALGTGSSC------CAGW-LEFSGQQLHDCHGLVIAR   290
NP_492153_ceADR1        N-------------AQEEKRVLAMGSSIQYIVEPDTLSGANG----TSLLHLDAIILAR   702
mfCARF-AD               -------------------KEYLNRGV--YVTVNTDNIGISKGSL-TDNLLFLAELCPGI   581
NP_001316462_ADAT3      ------------------------------------------------------------   351
NP_872309_ADAT2         -----------------------------------NAPK--------------------   180
CAD6006593_TadA         -----------------------------------EIKA--------------------   158

NP_001311374_ADAT1      RSFQRYLLHQLQLAAT----LKEDSIFVPGTQK-GVWKLRRDLIFVFFSSHTPCGDASII   148
NP_056648_ADAR2         RSLLRFLYTQLELYLNN-KDDQKRSIFQKSER-G-GFRLKENVQFHLYISTSPCGDARIF   457
NP_061172_ADAR3         RAFLHFLYTQLELHLSKRREDSERSIFVRLKE-G-GYRLRENILFHLYVSTSPCGDARLH   496
NP_001102_ADAR1         RGFIRFLYSELMKYNSQ---TAKDSIFEPAKG-GEKLQIKKTVSFHLYISTAPCGDGALF   972
NP_498594_ceADR2        RGLLRFLYSEVLKFSTE---PPN-SIF--TKG-KNALVLKPGISFHLFINTAPCGVARID   250
XP_024309656_ADAD1      RSLLRYFYRQLLLFYSKNPAMMEKSIFCTEPTS-NLLTLKQNINICLYMNQLPKGSAQIK   337
AAH33491_ADAD2          RALLRFLFRQLLLATQGGPKGKEQSVLAPQPGPGPPFTLKPRVFLHLYISNTPKGAARD-   349
NP_492153_ceADR1        RAMLKAFIHELSTVDS------ECSIFEKKEE--GKAALKPNLRLVLYSNYSPPCIHAVD   754
mfCARF-AD               RR-----LDILKLLNH------ACQIAFLP----------HQLRTE--------LIHKI-   611
NP_001316462_ADAT3      ------------------------------------------------------------   351
NP_872309_ADAT2         ----------SKVRKK------ECQKS-----------------------------------   191
CAD6006593_TadA         ----------QKKAQS------STD-------------------------------------   167
```

```
NP_001311374_ADAT1      PMLEFEDQPC---CPVFRNWAHNSSVEASSNLEAPGNERKCEDPDSPVTKKMRLEPGTAA   205
NP_056648_ADAR2         SPHEPILEGSRSYTQAGVQWCNHGSLQPR----PPG--------------LLSDPSTST   498
NP_061172_ADAR3         SPYEITTD-------------------------------------------------   504
NP_001102_ADAR1         DKSCSDRAMESTE--------------------------------------------   985
NP_498594_ceADR2        KKLKPGTSDD-----------------------------------------------   260
XP_024309656_ADAD1      SQLRLNPHSISAF--------------------------------------------   350
AAH33491_ADAD2          --IYLPPTSEGGL--------------------------------------------   360
NP_492153_ceADR1        DAATKK---L-----------------------------------------------   761
mfCARF-AD               -----Q---K-----------------------------------------------   613
NP_001316462_ADAT3      -------------------------------------------------------   351
NP_872309_ADAT2         -------------------------------------------------------   191
CAD6006593_TadA         -------------------------------------------------------   167

NP_001311374_ADAT1      RE-VTNGAAHHQSFGKQKSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCV   264
NP_056648_ADAR2         FQGAGTTEPADRHPNRKARGQLRTKIES-------------------------GEGTI   531
NP_061172_ADAR3         -------LHSSKHLVRKFRGHLRTKIES-------------------------GEGTV   530
NP_001102_ADAR1         -------SRHYPVFENPKQGKLRTKVEN-------------------------GEGTI   1011
NP_498594_ceADR2        --------------LQNSSRLRFKIDK-------------------------GMGTV   278
XP_024309656_ADAD1      --------EANEE------LCLHVAVE-----------------------------GKI   366
AAH33491_ADAD2          --------PHSPP------MRLQAHVL-----------------------------GQL   376
NP_492153_ceADR1        -------------------SYVTPT-------------------------NLTCV   772
mfCARF-AD               -------------------NLLTPT-------------------------NLALF   624
NP_001316462_ADAT3      -------------------------------------------------------   351
NP_872309_ADAT2         -------------------------------------------------------   191
CAD6006593_TadA         -------------------------------------------------------   167

NP_001311374_ADAT1      PGEAGDSGKPGAAFHQVGLLRVKPGRGDRTRSMSCSDKMARWNVLGCQGALLMHLLEEPI   324
NP_056648_ADAR2         PVRSN---------ASI-QTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFV-EPI   580
NP_061172_ADAR3         PVRGP---------SAV-QTWDGVLLGEQLITMSCTDKIARWNVLGLQGALLSHFV-EPV   579
NP_001102_ADAR1         PVESS---------DIV-PTWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFL-QPI   1060
NP_498594_ceADR2        LGGAS-------EFEAP-QTFDGIMMGERMRTMSCSDKLLRANVLGVQGAILSHFI-DPI   329
XP_024309656_ADAD1      YL--------------T-VY-CPKDGVNRISSMSSSDKLTRWEVLGVQGALLSHFI-QPV   409
AAH33491_ADAD2          KP--------------V-CYVAPSLCDTHVGCLSASDKLARWAVLGLGGALLAHLV-SPL   420
NP_492153_ceADR1        PD--------------DVLTYEQIKETKSLRVHCTADKLFKWNTLGIQGALLSNVL-HPI   817
mfCARF-AD               NE--------------PS----------------------------------------   628
NP_001316462_ADAT3      -------------------------------------------------------   351
NP_872309_ADAT2         -------------------------------------------------------   191
CAD6006593_TadA         -------------------------------------------------------   167

NP_001311374_ADAT1      YLSAVVIGK-CPYSQEAMQRALIGRCQNVS-----ALPKGFGVQELKILQSDLLFEQSRS   378
NP_056648_ADAR2         YFSSIILGS--LYHGDHLSRAMYQRISNIE-----DLPPLYTLNKPLLSGISNAE-----   628
NP_061172_ADAR3         YLQSIVVGS--LHHTGHLARVMSHRMEGVG-----QLPASYRHNRPLLSGVSDAE-----   627
NP_001102_ADAR1         YLKSVTLGY--LFSQGHLTRAICCRVTRDGSAFEDGLRHPFIVNHPKVGRVSIYDS-KR-   1116
NP_498594_ceADR2        YYSSIAVAE--LNNADRLRKAVYSRAAT----F--KPPAPFHVQDVEIGECQVEDT-EQ-   379
XP_024309656_ADAD1      YISSILIGDGNCSDTRGLEIAIKQ---RVDDALTSKLPMFYLVNRPHISLVPSAYP-L--   463
AAH33491_ADAD2          YSTSLILAD-SCHDPPTLSRAIHTRP-CLDSVLGPCLPPPYVRTALHLFAGPPVAP-S--   475
NP_492153_ceADR1        FIDNIFFGSEAPVSDESLSYALQGRLGPNENE------------REIIVESMP------   858
mfCARF-AD               -------------------------------------------------------   628
NP_001316462_ADAT3      -------------------------------------------------------   351
NP_872309_ADAT2         -------------------------------------------------------   191
CAD6006593_TadA         -------------------------------------------------------   167
```

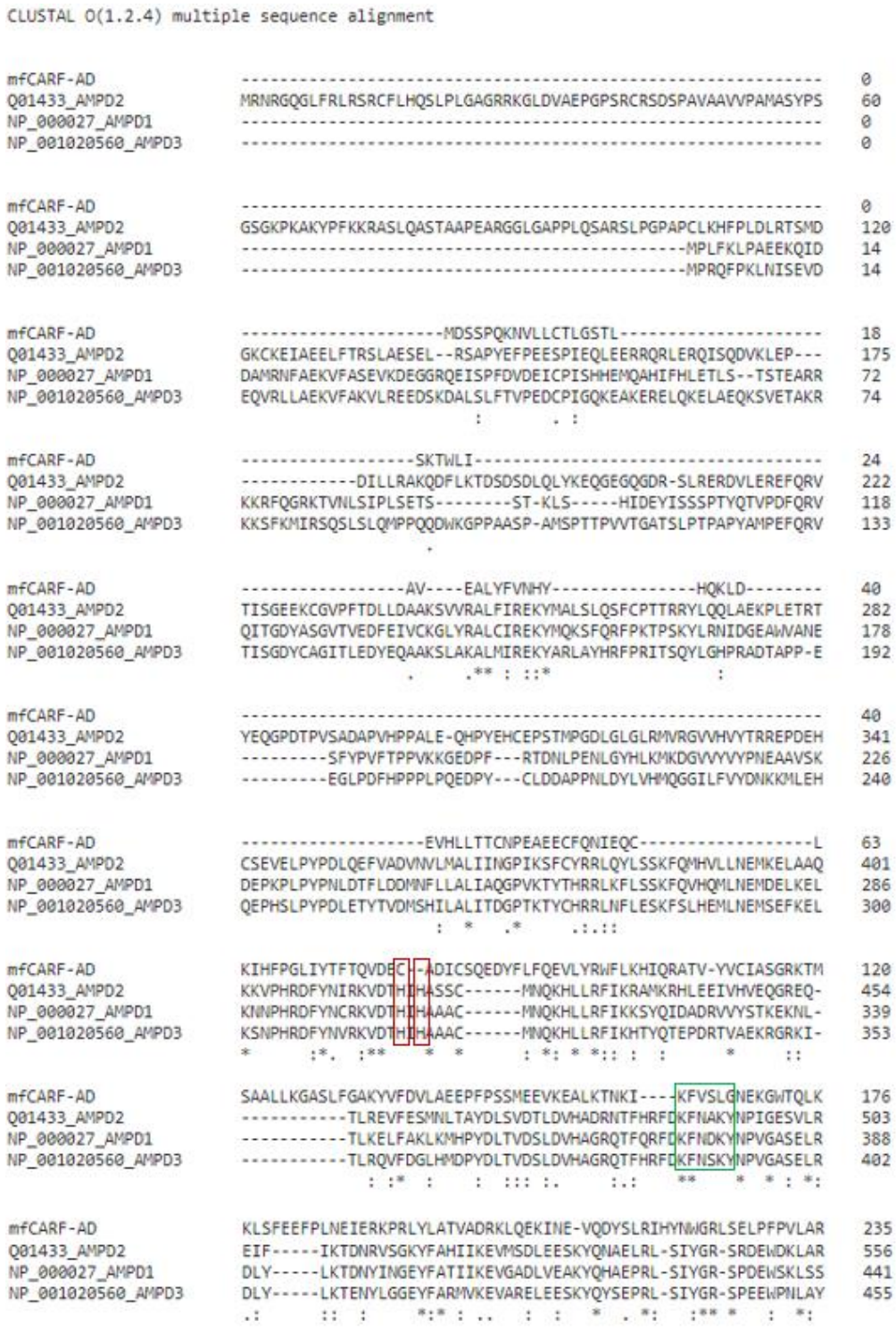
```
NP_001311374_ADAT1      AVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANGFPQGTTK---KTIGSLQARSQISK    435
NP_056648_ADAR2         ------ARQPGKA--PNFSVNWTVGDSAIEVIN-AT--TGK--------DELGRASRLCK    669
NP_061172_ADAR3         ------ARQPGKS--PPFSMNWVVGSADLEIIN-AT--TGR--------RSCGGPSRLCK    668
NP_001102_ADAR1         --------QSGKT--KETSVNWCLADGYDLEIL-DG-TRGTVD------GPRNELSRVSK    1158
NP_498594_ceADR2        -----STSAAARS--TISSMNWNLADGNTEVVR--T-SDGMVHDKDMSGADITTPSRLCK    429
XP_024309656_ADAD1      -------QMNLEY--KFLSLNWAQGDVSLEIVD--G-LSGKITESSPFKSGMSMASRLCK    511
AAH33491_ADAD2          ------EPTPDTC--RGLSLNWSLGDPGIEVVD--V-ATGRVKANA----ALGPPSRLCK    520
NP_492153_ceADR1        ---------VQMRMHMGISHLWHRGVDSVETLD---YNTGR--------TSKGSPSRVCK    898
mfCARF-AD               ------------------------------------------------------------    628
NP_001316462_ADAT3      ------------------------------------------------------------    351
NP_872309_ADAT2         ------------------------------------------------------------    191
CAD6006593_TadA         ------------------------------------------------------------    167

NP_001311374_ADAT1      VELFRSFQKLLSRIARDKWPHSLRVQKLDTYQEYKEAASSYQEAWS----TLRKQVFGSW    491
NP_056648_ADAR2         HALYCRWMRVHGKVPSHLL--RSKITKPNVYHESKLAAKEYQAAKARLFTAFIKAGLGAW    727
NP_061172_ADAR3         HVLSARWARLYGRLSTRT---PSPGDTPSMYCEAKLGAHTYQSVKQQLFKAFQKAGLGTW    725
NP_001102_ADAR1         KNIFLLFKKLCSF-----R--YRRDLLRLSYGEAKKAARDYETAKNYFKKGLKDMGYGNW    1211
NP_498594_ceADR2        KNMAELMITICTL-----T--KTSVDYPISYEELKAGSQEYAAAKKSFITWLRQKDLGIW    482
XP_024309656_ADAD1      AAMLSRFNLLAKEAK------KE-LLEAGTYHAAKCMSASYQEAKCKLKSYLQQHGYGSW    564
AAH33491_ADAD2          ASFLRAFHQAARAVG------KPYLLALKTYEAAK--AGPYQEARRQLSLLLDQQGLGAW    572
NP_492153_ceADR1        AEIFEAYRKLNGV-----------DQAVVNYAKAKEMASEYQYEKKVFYEKLEAAGLGKW    947
mfCARF-AD               ------------------------------------------------------------    628
NP_001316462_ADAT3      ------------------------------------------------------------    351
NP_872309_ADAT2         ------------------------------------------------------------    191
CAD6006593_TadA         ------------------------------------------------------------    167

NP_001311374_ADAT1      IRNPPDYH-QFK-----       502
NP_056648_ADAR2         VEKPTEQD-QFSLTP--       741
NP_061172_ADAR3         VRKPPEQQ-QFLLTL--       739
NP_001102_ADAR1         ISKPQEEK-NFYLCPV-       1226
NP_498594_ceADR2        QRKPREFQ-MFTIN---       495
XP_024309656_ADAD1      IVKSPCIE-QFNM----       576
AAH33491_ADAD2          PSKPLV-G-KFRN----       583
NP_492153_ceADR1        QTKPAELVDSFTLAAFD       964
mfCARF-AD               ----------------        628
NP_001316462_ADAT3      ----------------        351
NP_872309_ADAT2         ----------------        191
CAD6006593_TadA         ----------------        167
```

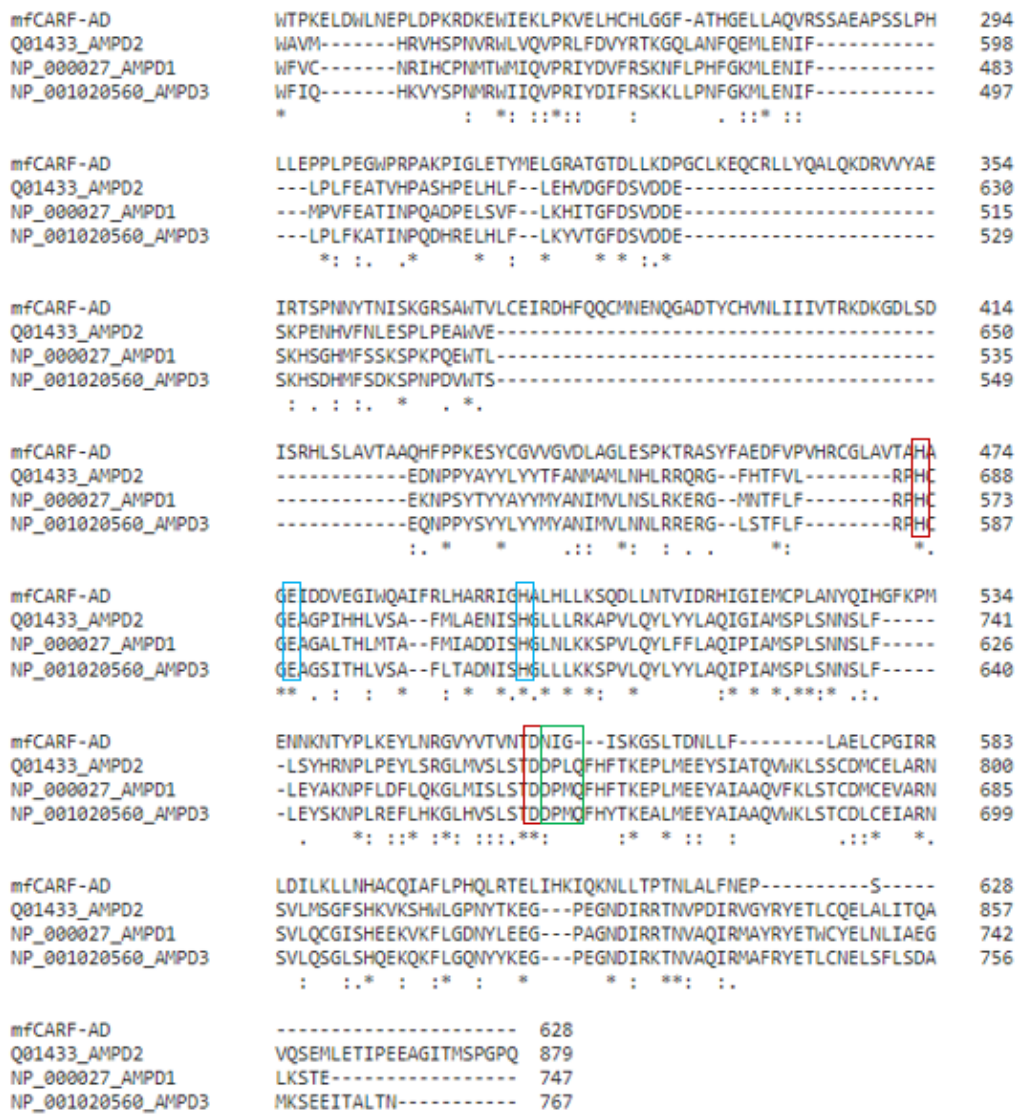
Supplementary Fig. 3 Multiple sequence alignment of mfCARF-AD with adenosine deaminases acting on nucleic acids (ADNs) using ClustalO.


CLUSTAL O(1.2.4) multiple sequence alignment
mfCARF-AD MDSSPQKNVLLCTLGSTLSKTWLIAVEALYFVNHYHQKLDEVHLLTTCNPEAEECFQNIE 60
KGD91880_psADE ---------- 0
Q9I6Y4_paADA ---------- 0
GFP67251_scADE ---------- 0
mfCARF-AD QCLKIHFPGLIYTFTQVDECADICSQEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTM 120
KGD91880_psADE ---------- 0
Q9I6Y4_paADA ---------- 0
GFP67251_scADE ---------- 0
mfCARF-AD SAALLKGASLFGAKYVFDVLAEEPFPSSMEEVKEALKTNKIKFVSLGNEKGWTQLKKLSF 180
KGD91880_psADE ---------- 0
Q9I6Y4_paADA ---------- 0
GFP67251_scADE ---------- 0
mfCARF-AD EEFPLNEIERKPRLYLATVADRKLQEKINEVQDYSLRIHYNWGRLSELPFPVLARWTPKE 240
KGD91880_psADE ---------- 0
Q9I6Y4_paADA ---------- 0
GFP67251_scADE ---------- 0
mfCARF-AD LDWLNEPLDPKRDKEWIEKLPKVELHGHLGGFATHGELLAQVRSSAEAPSSLPHLLEPPL 300
KGD91880_psADE ----------MYEWLNALPKAELHLHLEGTLEPELLFALAER----------NRIAL 37
Q9I6Y4_paADA ----------MYEWLNALPKAELHLHLEGTLEPELLFALAER----------NRIAL 37
GFP67251_scADE ----------MVSVEFLQELPKCEHHLHLEGTLEPDLLFPLAKR----------NDIIL 39
mfCARF-AD PEGWPRPAKPIG--------LETYMELGRATGTDLLKDPGCLKEQCRLLYQALQKDRVVY 352
KGD91880_psADE PWNDVE---T-LRKAYAFNNLQEFLDLY-YAGADVLRTEQDFYDLTWAYLQKCKAQNVVH 92
Q9I6Y4_paADA PWNDVE---T-LRKAYAFNNLQEFLDLY-YAGADVLRTEQDFYDLTWAYLQKCKAQNVVH 92
GFP67251_scADE PEGFPKSVEELNEKYKKFRDLQDFLDYY-YIGTNVLISEQDFFDLAWAYFKKVHKQGLVH 98
mfCARF-AD AEIRTSPNNYTNISKGRSAWTVLCEIRDHFQQCMNENQGADTYCHVNLIIIVTRKDKGDL 412
KGD91880_psADE VEPFFDPQTH--TDRGIPFEVVLAGIRAALRDGEKLL-GI----RHGLIL---------- 135
Q9I6Y4_paADA VEPFFDPQTH--TDRGIPFEVVLAGIRAALRDGEKLL-GI----RHGLIL---------- 135
GFP67251_scADE AEVFYDPQSH--TSRGISIETVTKGFQRACDKAFSEF-GI----TSKLIM---------- 141
mfCARF-AD SDISRHLSLA-----VTAAQHFPPKESYCGVVGVDLAGLESPKTRASYFAEDFVP--VHR 465
KGD91880_psADE -SFLRHLSEEQAQKTLDQALPFRD---AFIAVGLDSSEVGHPP---SKFQRVFDRARS-- 186
Q9I6Y4_paADA -SFLRHLSEEQAQKTLDQALPFRD---AFIAVGLDSSEVGHPP---SKFQRVFDRARS-- 186
GFP67251_scADE -CLLRHIEPEECLKTIEEATPFIK-DGTISALGLDSAEKPFPP---HLFVECYGKAASLN 196
mfCARF-AD CGLAVTAHAGEIDDVEGIWQAIFRLHARRIGHALHLLKSQDLLNTVIDRHIGIEMCPLAN 525
KGD91880_psADE EGFLTVAHAGEEGPPEYIWEALDLLKVERIDHGVRAFEDERLMRRLIDEQIPLTVCPLSN 246
Q9I6Y4_paADA EGFLTVAHAGEEGPPEYIWEALDLLKVERIDHGVRAFEDERLMRRLIDEQIPLTVCPLSN 246
GFP67251_scADE KDLKLTAHAGEEGPAQFVSDALDLLQVTRIDHGINSQYDEELLDRLSRDQTMLTICPLSN 256
mfCARF-AD YQIHGFKPMENNKNTYPLKEYLNRGVYVTVNTDNIGISKGSLTDNLLFLAELCPGIRRLD 585
KGD91880_psADE TKLCVFDDMS----QHTILDMLERGVKVTVNSDDPAYFGGYVTENFHALQQSLGMTE-EQ 301
Q9I6Y4_paADA TKLCVFDDMS----QHTILDMLERGVKVTVNSDDPAYFGGYVTENFHALQQSLGMTE-EQ 301
GFP67251_scADE VKLQVVQSVS----ELPLQKFLDRDVPFSLNSDDPAYFGGYILDVYTQVSKDFPHWDHET 312



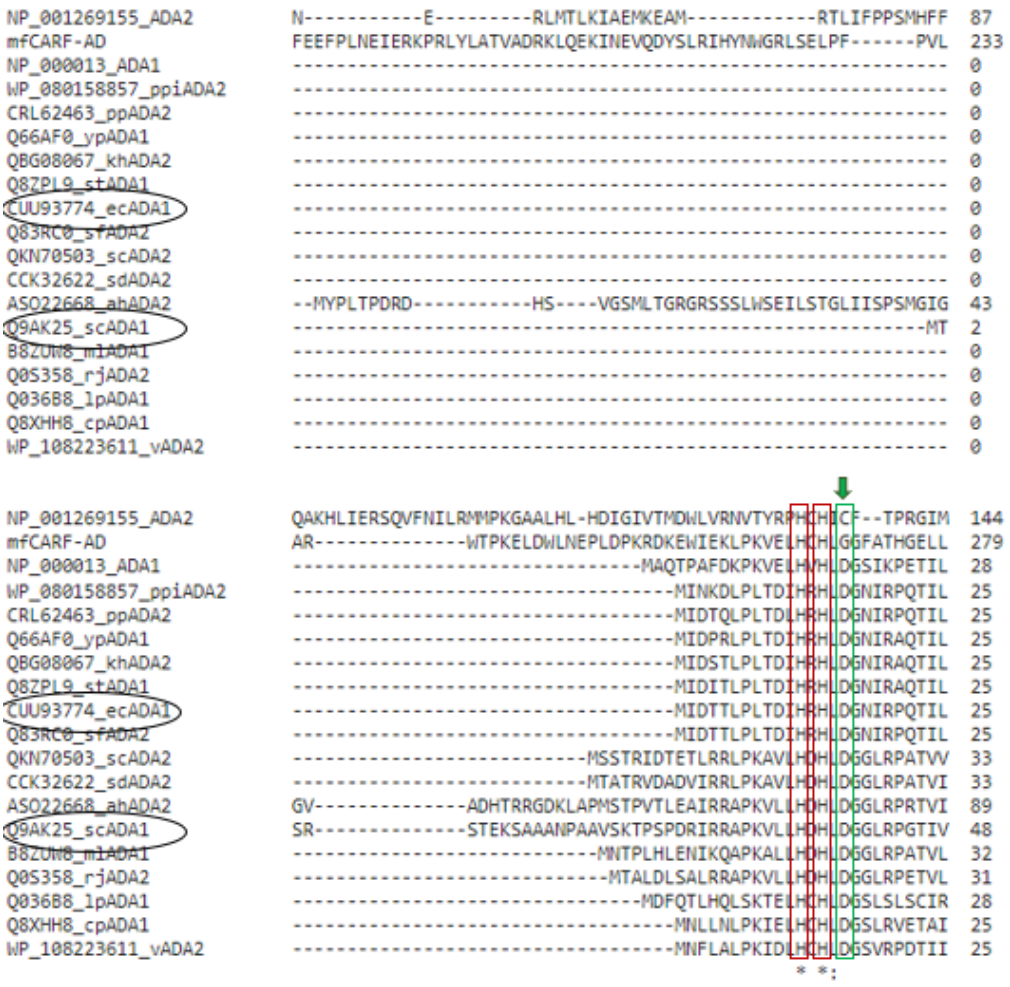
Supplementary Fig. 4 Multiple sequence alignment of mfCARF-AD with adenine deaminases acting on free adenines (ADEs) using ClustalO. Red, green, yellow, and blue boxes indicate residues essential for zinc-binding, substrate-binding, ribose occlusion, and substrate deamination, respectively.

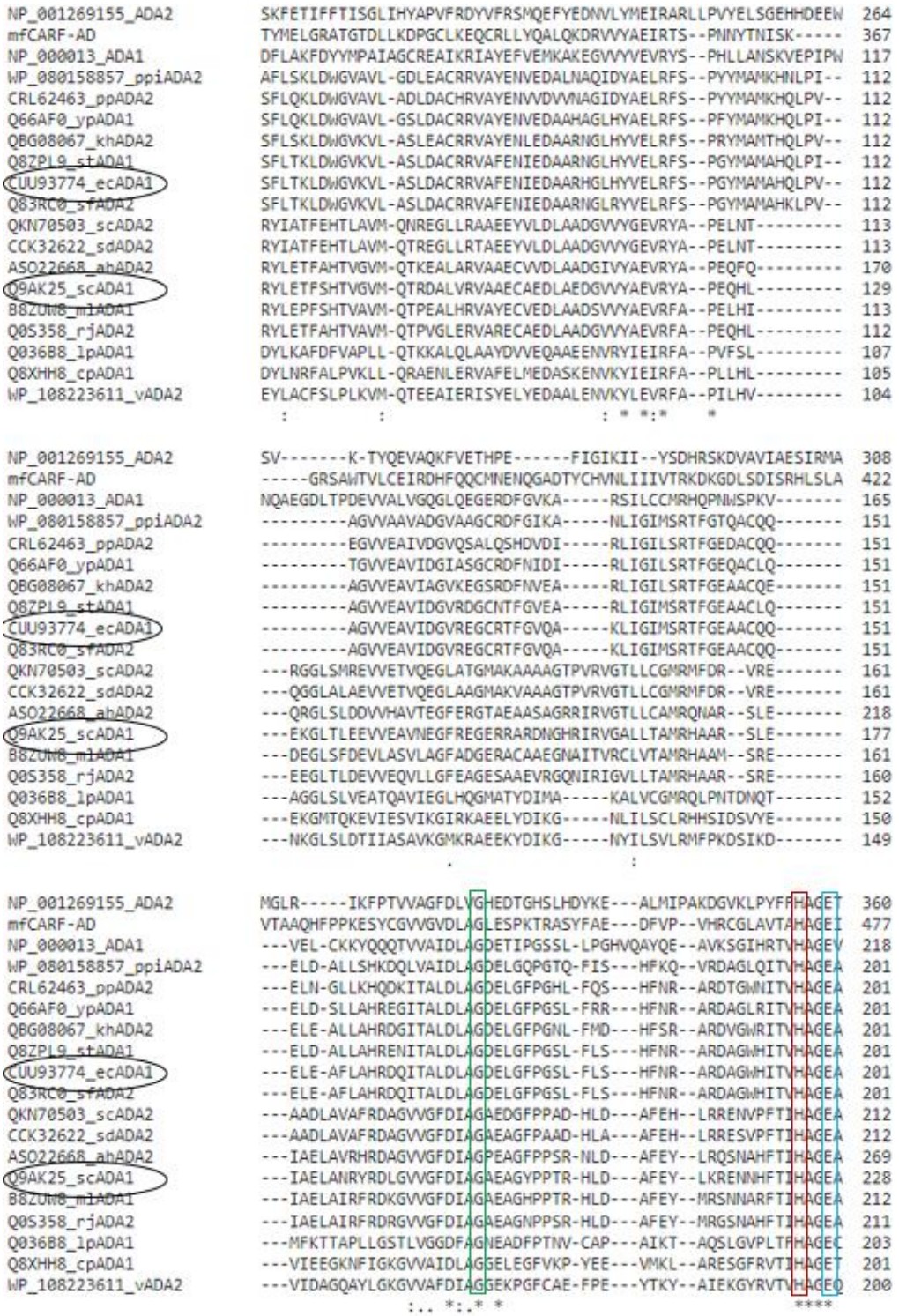

CLUSTAL O(1.2.4) multiple sequence alignment
mfCARF-AD
Q01433_AMPD2
NP_000027_AMPD1
NP_001020560_AMPD3

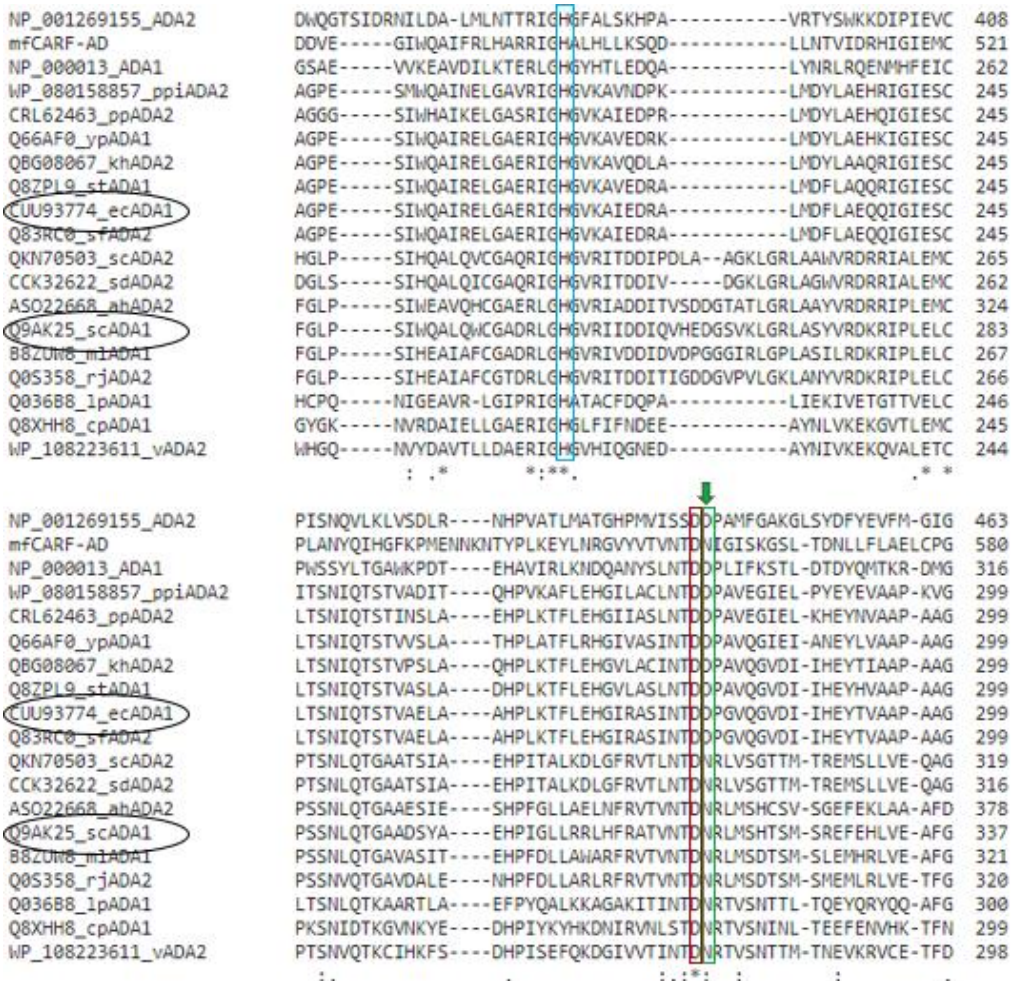


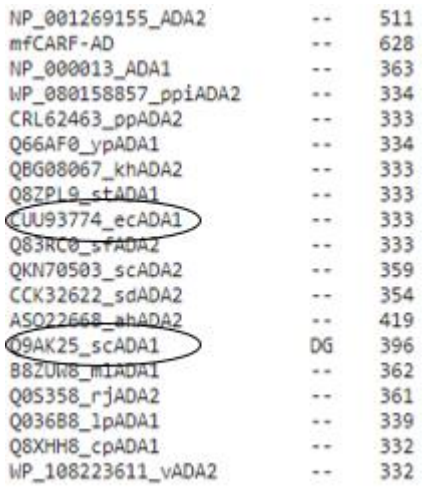
Supplementary Fig. 5 Multiple sequence alignment of mfCARF-AD with AMP deaminases acting on AMP (AMPDs) using ClustalO. Red, green, and blue boxes indicate residues essential for zinc-binding, phosphate binding, and substrate deamination, respectively.

```
CLUSTAL O(1.2.4) multiple sequence alignment

NP_001269155_ADA2          ------------------------------------------------------------  0
mfCARF-AD                  MDSSPQKNVLLCTLGSTLSKTWLIAVEALYFVNHYHQKLDEVHLLTTCNPEAEECFQNIE  60
NP_000013_ADA1             ------------------------------------------------------------  0
WP_080158857_ppiADA2       ------------------------------------------------------------  0
CRL62463_ppADA2            ------------------------------------------------------------  0
Q66AF0_ypADA1              ------------------------------------------------------------  0
QBG08067_khADA2            ------------------------------------------------------------  0
Q8ZPL9_stADA1              ------------------------------------------------------------  0
CUU93774_ecADA1            ------------------------------------------------------------  0
Q83RC0_sfADA2              ------------------------------------------------------------  0
QKN70503_scADA2            ------------------------------------------------------------  0
CCK32622_sdADA2            ------------------------------------------------------------  0
ASQ22668_ahADA2            ------------------------------------------------------------  0
Q9AK25_scADA1              ------------------------------------------------------------  0
B8ZUW8_m1ADA1              ------------------------------------------------------------  0
Q0S358_rjADA2              ------------------------------------------------------------  0
Q036B8_lpADA1              ------------------------------------------------------------  0
Q8XHH8_cpADA1              ------------------------------------------------------------  0
WP_108223611_vADA2         ------------------------------------------------------------  0

NP_001269155_ADA2          ----------------------------------------------MLVDGPSERPAL  12
mfCARF-AD                  QCLKIHFPGLIYTFTQVDECADICSQEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTM  120
NP_000013_ADA1             ------------------------------------------------------------  0
WP_080158857_ppiADA2       ------------------------------------------------------------  0
CRL62463_ppADA2            ------------------------------------------------------------  0
Q66AF0_ypADA1              ------------------------------------------------------------  0
QBG08067_khADA2            ------------------------------------------------------------  0
Q8ZPL9_stADA1              ------------------------------------------------------------  0
CUU93774_ecADA1            ------------------------------------------------------------  0
Q83RC0_sfADA2              ------------------------------------------------------------  0
QKN70503_scADA2            ------------------------------------------------------------  0
CCK32622_sdADA2            ------------------------------------------------------------  0
ASQ22668_ahADA2            ------------------------------------------------------------  0
Q9AK25_scADA1              ------------------------------------------------------------  0
B8ZUW8_m1ADA1              ------------------------------------------------------------  0
Q0S358_rjADA2              ------------------------------------------------------------  0
Q036B8_lpADA1              ------------------------------------------------------------  0
Q8XHH8_cpADA1              ------------------------------------------------------------  0
WP_108223611_vADA2         ------------------------------------------------------------  0

NP_001269155_ADA2          CFLLLAVAMSFFGSALSIDET------------RAHL-LLKEKMMRLGGRLVLNTKEELA  59
mfCARF-AD                  SAA-LLKGASLFGAKYVFDVLAEEPFPSSMEEVKEALKTNKIKFVSLGNEKGWTQLKKLS  179
NP_000013_ADA1             ------------------------------------------------------------  0
WP_080158857_ppiADA2       ------------------------------------------------------------  0
CRL62463_ppADA2            ------------------------------------------------------------  0
Q66AF0_ypADA1              ------------------------------------------------------------  0
QBG08067_khADA2            ------------------------------------------------------------  0
Q8ZPL9_stADA1              ------------------------------------------------------------  0
CUU93774_ecADA1            ------------------------------------------------------------  0
Q83RC0_sfADA2              ------------------------------------------------------------  0
QKN70503_scADA2            ------------------------------------------------------------  0
CCK32622_sdADA2            ------------------------------------------------------------  0
ASQ22668_ahADA2            ------------------------------------------------------------  0
Q9AK25_scADA1              ------------------------------------------------------------  0
B8ZUW8_m1ADA1              ------------------------------------------------------------  0
Q0S358_rjADA2              ------------------------------------------------------------  0
Q036B8_lpADA1              ------------------------------------------------------------  0
Q8XHH8_cpADA1              ------------------------------------------------------------  0
WP_108223611_vADA2         ------------------------------------------------------------  0
```


NP_001269155_ADA2 N----------E---------RLMTLKIAEMKEAM------------RTLIFPPSMHFF 87
mfCARF-AD FEEFPLNEIERKPRLYLATVADRKLQEKINEVQDYSLRIHYNMGRLSELPF------PVL 233
NP_000013_ADA1 ---------- 0
WP_080158857_ppiADA2 ---------- 0
CRL62463_ppADA2 ---------- 0
Q66AF0_ypADA1 ---------- 0
QBG08067_khADA2 ---------- 0
Q8ZPL9_stADA1 ---------- 0
CUU93774_ecADA1 ---------- 0
Q83RC0_sfADA2 ---------- 0
QKN70503_scADA2 ---------- 0
CCK32622_sdADA2 ---------- 0
ASQ22668_ahADA2 --MYPLTPDRD----------HS----VGSMLTGRGRSSSLWSEILSTGLIISPSMGIG 43
Q9AK25_scADA1 ----------MT 2
B8ZUM8_mlADA1 ---------- 0
Q0S358_rjADA2 ---------- 0
Q036B8_lpADA1 ---------- 0
Q8XHH8_cpADA1 ---------- 0
WP_108223611_vADA2 ---------- 0
NP_001269155_ADA2 QAKHLIERSQVFNILRMMPKGAALHL-HDIGIVTMDWLVRNVTYRPHCHHCF--TPRGIM 144
mfCARF-AD AR--------------WTPKELDWLNEPLDPKRDKEWIEKLPKVELHCHLGGFATHGELL 279
NP_000013_ADA1 ----------MAQTPAFDKPKVELHVHLDGSIKPETIL 28
WP_080158857_ppiADA2 ----------MINKDLPLTDIHRHLDGNIRPQTIL 25
CRL62463_ppADA2 ----------MIDTQLPLTDLHRHLDGNIRPQTIL 25
Q66AF0_ypADA1 ----------MIDPRLPLTDIHRHLDGNIRAQTIL 25
QBG08067_khADA2 ----------MIDSTLPLTDIHRHLDGNIRAQTIL 25
Q8ZPL9_stADA1 ----------MIDITLPLTDIHRHLDGNIRAQTIL 25
CUU93774_ecADA1 ----------MIDTTLPLTDIHRHLDGNIRPQTIL 25
Q83RC0_sfADA2 ----------MIDTTLPLTDIHRHLDGNIRPQTIL 25
QKN70503_scADA2 ----------MSSTRIDTETLRRLPKAVLHDHLDGGLRPATVV 33
CCK32622_sdADA2 ----------MTATRVDADVIRRLPKAVLHDHLDGGLRPATVI 33
ASQ22668_ahADA2 GV----------ADHTRRGDKLAPMSTPVTLEAIRRAPKVLLHDHLDGGLRPRTVI 89
Q9AK25_scADA1 SR----------STEKSAAANPAAVSKTPSPDRIRRAPKVLLHDHLDGGLRPGTIV 48
B8ZUM8_mlADA1 ----------MNTPLHLENIKQAPKALLHDHLDGGLRPATVL 32
Q0S358_rjADA2 ----------MTALDLSALRRAPKVLLHDHLDGGLRPETVL 31
Q036B8_lpADA1 ----------MDFQTLHQLSKTELHCHLDGSLSLSCIR 28
Q8XHH8_cpADA1 ----------MNLLNLPKIELHCHLDGSLRVETAI 25
WP_108223611_vADA2 ----------MNFLALPKIDLHIHLDGSVRPDTII 25
NP_001269155_ADA2 QFRFAHPTPRPSEKCSKWILLEDYRKRVQNVTEFDDSLLRNFTLVTQHPEVIYTNQNVVW 204
mfCARF-AD AQVRSSAE----------APSSLPHLLEPPLPEGWPRPAKPI-------GLE 314
NP_000013_ADA1 YYGRRRGI----------ALPANTAEGLLNVIGMD-KPL-------TLP 59
WP_080158857_ppiADA2 ELGQQFNI----------ALPATEIEALRPHVQIVEAEP-------SLV 57
CRL62463_ppADA2 DLAKQHNI----------ALPAYELEALRPHVQITKNEP-------SLV 57
Q66AF0_ypADA1 DLGQQFNL----------NLPANELDTLRPHVQITKTEP-------DLV 57
QBG08067_khADA2 DLGRQFNL----------YLPANTLDTLRPHVQVTSNEP-------DLV 57
Q8ZPL9_stADA1 DLGRQFNI----------ALPAQTLETLIPHVQVTSTEP-------DLV 57
CUU93774_ecADA1 ELGRQYNI----------SLPAQSLETLIPHVQVIANEP-------DLV 57
Q83RC0_sfADA2 ELGRQYNI----------SLPAQSLETLIPHVQVIANEP-------DLV 57
QKN70503_scADA2 ELARSVGH----------TLPTTDPDELAAWYYEAANSG-------DLV 65
CCK32622_sdADA2 ELAAEVGH----------TLPTTDPDELAAWYFEAANSG-------DLV 65
ASQ22668_ahADA2 ELAEELGY----------AGLPSTEEDALGVWFREAADSG-------SLE 122
Q9AK25_scADA1 ELARETGY----------GDLPETDADLLGTWFRQAADSG-------SLE 81
B8ZUM8_mlADA1 DIAGQVGY----------DRLPATDVESLETWFRTASHSG-------SLE 65
Q0S358_rjADA2 ELARDCGY----------DALPADTAPALATWFREAADSG-------SLE 64
Q036B8_lpADA1 QLAKMIDR----------KLPA-TDDELRRLVQAPADSE-------NLG 59
Q8XHH8_cpADA1 ELAKKEGV----------KLDSYEYDKVKELLVIPKECN-------SLE 57
WP_108223611_vADA2 DLAKQYDI----------ELPA-DRDAVVESLTVPEDCK-------NLD 56


NP_001269155_ADA2 SKFETIFFTISGLIHYAPVFRDYVFRSMQEFYEDNVLYMEIRARLLPVYELSGEHHDEEW 264
mfCARF-AD TYMELGRATGTDLLKDPGCLKEQCRLLYQALQKDRVVYAEIRTS--PNNYTNISK----- 367
NP_000013_ADA1 DFLAKFDYYMPAIAGCREAIKRIAYEFVEMKAKEGVVYVEVRYS--PHLLANSKVEPIPW 117
WP_080158857_ppiADA2 AFLSKLDWGVAVL-GDLEACRRVAYENVEDALNAQIDYAELRFS--PYYMAMKHNLPI-- 112
CRL62463_ppADA2 SFLQKLDWGVAVL-ADLDACHRVAYENVVDVVNAGIDYAELRFS--PYYMAMKHQLPV-- 112
Q66AF0_ypADA1 SFLQKLDWGVAVL-GSLDACRRVAYENVEDAAHAGLHYAELRFS--PFYMAMKHQLPI-- 112
QBG08067_khADA2 SFLSKLDWGVKVL-ASLEACRRVAYENLEDAARNGLHYVELRFS--PRYMAMTHQLPV-- 112
Q8ZPL9_stADA1 SFLTKLDWGVKVL-ASLDACRRVAFENIEDAARNGLHYVELRFS--PGYMAMAHQLPI-- 112
CUU93774_ecADA1 SFLTKLDWGVKVL-ASLDACRRVAFENIEDAARHGLHYVELRFS--PGYMAMAHQLPV-- 112
Q83RC0_sfADA2 SFLTKLDWGVKVL-ASLDACRRVAFENIEDAARNGLRYVELRFS--PGYMAMAHKLPV-- 112
QKN70503_scADA2 RYIATFEHTLAVM-QNREGLLRAAEEYVLDLAADGVVYGEVRYA--PELNT--------- 113
CCK32622_sdADA2 RYIATFEHTLAVM-QTREGLLRTAEEYVLDLAADGVVYGEVRYA--PELNT--------- 113
ASO22668_ahADA2 RYLETFAHTVGVM-QTKEALARVAAECVVDLAADGIVYAEVRYA--PEQFQ--------- 170
Q9AK25_scADA1 RYLETFSHTVGVM-QTRDALVRVAAECAEDLAEDGVVYAEVRYA--PEQHL--------- 129
B8ZUM8_mlADA1 RYLEPFSHTVAVM-QTPEALHRVAYECVEDLAADSVVYAEVRFA--PELHI--------- 113
Q0S358_rjADA2 RYLETFAHTVAVM-QTPVGLERVARECAEDLAADGVVYAEVRFA--PEQHL--------- 112
Q036B8_lpADA1 DYLKAFDFVAPLL-QTKKALQLAAYDVVEQAAEENVRYIEIRFA--PVFSL--------- 107
Q8XHH8_cpADA1 DYLNRFALPVKLL-QRAENLERVAFELMEDASKENVKYIEIRFA--PLLHL--------- 105
WP_108223611_vADA2 EYLACFSLPLKVM-QTEEAIERISYELYEDAALENVKYLEVRFA--PILHV--------- 104
SV-------K-TYQEVAQKFVETHPE------FIGIKII--YSDHRSKDVAVIAESIRMA 308
-----GRSAWTVLCEIRDHFQQCMNENQGADTYCHVNLIIIVTRKDKGDLSDISRHLSLA 422
NQAEGDLTPDEVVALVGQGLQEGERDFGVKA-----RSILCCMRHQPNWSPKV------- 165
---------AGVVAAVADGVAAGCRDFGIKA-----NLIGIMSRTFGTQACQQ------- 151
---------EGVVEAIVDGVQSALQSHDVDI-----RLIGILSRTFGEDACQQ------- 151
---------TGVVEAVIDGIASGCRDFNIDI-----RLIGILSRTFGEQACLQ------- 151
---------AGVVEAVIAGVKEGSRDFNVEA-----RLIGILSRTFGEAACQE------- 151
---------AGVVEAVIDGVRDGCNTFGVEA-----RLIGIMSRTFGEAACLQ------- 151
---------AGVVEAVIDGVREGCRTFGVQA-----KLIGIMSRTFGEAACQQ------- 151
---------AGVVEAVIDGVREGCRTFGVQA-----KLIGIMSRTFGEAACQQ------- 151
---RGGLSMREVVETVQEGLATGMAKAAAAGTPVRVGTLLCGMRMFDR--VRE------- 161
---QGGLALAEVVETVQEGLAAGMAKVAAAGTPVRVGTLLCGMRMFDR--VRE------- 161
---QRGLSLDDVVHAVTEGFERGTAEAASAGRRIRVGTLLCAMRQNAR--SLE------- 218
---EKGLTLEEVVEAVNEGFREGERRARDNGHRIRVGALLTAMRHAAR--SLE------- 177
---DEGLSFDEVLASVLAGFADGERACAAEGNAITVRCLVTAMRHAAM--SRE------- 161
---EEGLTLDEVVEQVLLGFEAGESAAEVRGQNIRIGVLLTAMRHAAR--SRE------- 160
---AGGLSLVEATQAVIEGLHQGMATYDIMA-----KALVCGMRQLPNTDNQT------- 152
---EKGMTQKEVIESVIKGIRKAEELYDIKG-----NLILSCLRHHSIDSVYE------- 150
---NKGLSLDTIIASAVKGMKRAEEKYDIKG-----NYILSVLRMFPKDSIKD------- 149
MGLR-----IKFPTVVAGFDLVGHEDTGHSLHDYKE---ALMIPAKDGVKLPYFFHAGET 360
VTAAQHFPPKESYCGVVGVDLAGLESPKTRASYFAE---DFVP--VHRCGLAVTAHAGEI 477
---VEL-CKKYQQQTVVAIDLAGDETIPGSSL-LPGHVQAYQE--AVKSGIHRTVHAGEV 218
---ELD-ALLSHKDQLVAIDLAGDELGQPGTQ-FIS---HFKQ--VRDAGLQITVHAGEA 201
---ELN-GLLKHQDKITALDLAGDELGFPGHL-FQS---HFNR--ARDTGWNITVHAGEA 201
---ELD-SLLAHREGITALDLAGDELGFPGSL-FRR---HFNR--ARDAGLRITVHAGEA 201
---ELE-ALLAHRDGITALDLAGDELGFPGNL-FMD---HFSR--ARDVGWRITVHAGEA 201
---ELD-ALLAHRENITALDLAGDELGFPGSL-FLS---HFNR--ARDAGWHITVHAGEA 201
---ELE-AFLAHRDQITALDLAGDELGFPGSL-FLS---HFNR--ARDAGWHITVHAGEA 201
---ELE-AFLAHRDQITALDLAGDELGFPGSL-FLS---HFNR--ARDAGWHITVHAGEA 201
---AADLAVAFRDAGVVGFDIAGAEDGFPPAD-HLD---AFEH--LRRENVPFTIHAGEA 212
---AADLAVAFRDAGVVGFDIAGAEAGFPAAD-HLA---AFEH--LRRESVPFTIHAGEA 212
---IAELAVRHRDAGVVGFDIAGPEAGFPPSR-NLD---AFEY--LRQSNAHFTIHAGEA 269
---IAELANRYRDLGVVGFDIAGAEAGYPPTR-HLD---AFEY--LKRENNHFTIHAGEA 228
---IAELAIRFRDKGVVGFDIAGAEAGHPPTR-HLD---AFEY--MRSNNARFTIHAGEA 212
---IAELAIRFRDRGVVGFDIAGAEAGNPPSR-HLD---AFEY--MRGSNAHFTIHAGEA 211
---MFKTTAPLLGSTLVGGDFAGNEADFPTNV-CAP---AIKT--AQSLGVPLTFHAGEC 203
---VIEEGKNFIGKGVVAIDLAGGELEGFVKP-YEE---VMKL--ARESGFRVTIHAGET 201
---VIDAGQAYLGKGVVAFDIAGGEKPGFCAE-FPE---YTKY--AIEKGYRVTVHAGEQ 200

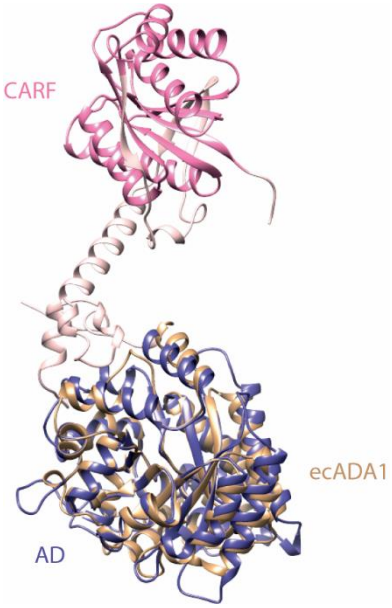

NP_001269155_ADA2 DWQGTSIDRNILDA-LMLNTTRIGHGFALSKHPA----------VRTYSWKKDIPIEVC 408
mfCARF-AD DDVE-----GIWQAIFRLHARRIGHALHLLKSQD----------LLNTVIDRHIGIEMC 521
NP_000013_ADA1 GSAE-----VVKEAVDILKTERLGHGYHTLEDQA----------LYNRLRQENMHFEIC 262
WP_080158857_ppiADA2 AGPE-----SMWQAINELGAVRIGHGVKAVNDPK----------LMDYLAEHRIGIESC 245
CRL62463_ppADA2 AGGG-----SIWHAIKELGASRIGHGVKAIEDPR----------LMDYLAEHQIGIESC 245
Q66AF0_ypADA1 AGPE-----SIWQAIRELGAERIGHGVKAVEDRK----------LMDYLAEHKIGIESC 245
QBG08067_khADA2 AGPE-----SIWQAIRELGAERIGHGVKAVQDLA----------LMDYLAAQRIGIESC 245
Q8ZPL9_stADA1 AGPE-----SIWQAIRELGAERIGHGVKAVEDRA----------LMDFLAQQRIGIESC 245
CUU93774_ecADA1 AGPE-----SIWQAIRELGAERIGHGVKAIEDRA----------LMDFLAEQQIGIESC 245
Q83RC0_sfADA2 AGPE-----SIWQAIRELGAERIGHGVKAIEDRA----------LMDFLAEQQIGIESC 245
QKN70503_scADA2 HGLP-----SIHQALQVCGAQRIGHGVRITDDIPDLA--AGKLGRLAAWVRDRRIALEMC 265
CCK32622_sdADA2 DGLS-----SIHQALQICGAQRIGHGVRITDDIV-----DGKLGRLAGWVRDRRIALEMC 262
ASO22668_ahADA2 FGLP-----SIWEAVQHCGAERLGHGVRIADDITVSDDGTATLGRLAAYVRDRRIPLEMC 324
Q9AK25_scADA1 FGLP-----SIWQALQWCGADRLGHGVRIIDDIQVHEDGSVKLGRLASYVRDKRIPLELC 283
B8ZUM8_mlADA1 FGLP-----SIHEAIAFCGADRLGHGVRIVDDIDVDPGGGIRLGPLASILRDKRIPLELC 267
Q0S358_rjADA2 FGLP-----SIHEAIAFCGTDRLGHGVRITDDITIGDDGVPVLGKLANYVRDKRIPLELC 266
Q036B8_lpADA1 HCPQ-----NIGEAVR-LGIPRIGHATACFDQPA----------LIEKIVETGTTVELC 246
Q8XHH8_cpADA1 GYGK-----NVRDAIELLGAERIGHGLFIFNDEE----------AYNLVKEKGVTLEMC 245
WP_108223611_vADA2 WHGQ-----NVYDAVTLLDAERIGHGVHIQGNED----------AYNIVKEKQVALETC 244
NP_001269155_ADA2 PISNQVLKLVSDLR----NHPVATLMATGHPMVISSDDPAMFGAKGLSYDFYEVFM-GIG 463
mfCARF-AD PLANYQIHGFKPMENNKNTYPLKEYLNRGVYVTVNTDNIGISKGSL-TDNLLFLAELCPG 580
NP_000013_ADA1 PWSSYLTGAWKPDT----EHAVIRLKNDQANYSLNTDDPLIFKSTL-DTDYQMTKR-DMG 316
WP_080158857_ppiADA2 ITSNIQTSTVADIT----QHPVKAFLEHGILACLNTDDPAVEGIEL-PYEYEVAAP-KVG 299
CRL62463_ppADA2 LTSNIQTSTINSLA----EHPLKTFLEHGIIASLNTDDPAVEGIEL-KHEYNVAAP-AAG 299
Q66AF0_ypADA1 LTSNIQTSTVVSLA----THPLATFLRHGIVASINTDDPAVQGIEI-ANEYLVAAP-AAG 299
QBG08067_khADA2 LTSNIQTSTVPSLA----QHPLKTFLEHGVLACINTDDPAVQGVDI-IHEYTIAAP-AAG 299
Q8ZPL9_stADA1 LTSNIQTSTVASLA----DHPLKTFLEHGVLASLNTDDPAVQGVDI-IHEYHVAAP-AAG 299
CUU93774_ecADA1 LTSNIQTSTVAELA----AHPLKTFLEHGIRASINTDDPGVQGVDI-IHEYTVAAP-AAG 299
Q83RC0_sfADA2 LTSNIQTSTVAELA----AHPLKTFLEHGIRASINTDDPGVQGVDI-IHEYTVAAP-AAG 299
QKN70503_scADA2 PTSNLQTGAATSIA----EHPITALKDLGFRVTLNTDNRLVSGTTM-TREMSLLVE-QAG 319
CCK32622_sdADA2 PTSNLQTGAATSIA----EHPITALKDLGFRVTLNTDNRLVSGTTM-TREMSLLVE-QAG 316
ASO22668_ahADA2 PSSNLQTGAAESIE----SHPFGLLAELNFRVTVNTDNRLMSHCSV-SGEFEKLAA-AFD 378
Q9AK25_scADA1 PSSNLQTGAADSYA----EHPIGLLRRLHFRATVNTDNRLMSHTSM-SREFEHLVE-AFG 337
B8ZUM8_mlADA1 PSSNLQTGAVASIT----EHPFDLLAWARFRVTVNTDNRLMSDTSM-SLEMHRLVE-AFG 321
Q0S358_rjADA2 PSSNVQTGAVDALE----NHPFDLLARLRFRVTVNTDNRLMSDTSM-SMEMLRLVE-TFG 320
Q036B8_lpADA1 LTSNLQTKAARTLA----EFPYQALKKAGAKITINTDNRTVSNTTL-TQEYQRYQQ-AFG 300
Q8XHH8_cpADA1 PKSNIDTKGVNKYE----DHPIYKYHKDNIRVNLSTDNRTVSNINL-TEEFENVHK-TFN 299
WP_108223611_vADA2 PTSNVQTKCIHKFS----DHPISEFQKDGIVVTINTDNRTVSNTTM-TNEVKRVCE-TFD 298
NP_001269155_ADA2 GMKADLRTLKQLAMNSIKYSTLLESEKNTFMEIWKK-RWDKFIADVATK---------- 511
mfCARF-AD IRR---LDILKLLNHACQIAFLPHQLRTELIHKIQK-NLLTPTNLALFNEPS-------- 628
NP_000013_ADA1 FTE---EEFKRLNINAAKSSFLPEDEKRELLDLLYK-AYGMPPSASAGQNL--------- 363
WP_080158857_ppiADA2 LSG---AQIRQAQINGLELAFLSATEKQALRDIASK-RG---------------------- 334
CRL62463_ppADA2 LTP---EQIRQTQINGLKMAFISQAEREALIKKASQ-R---------------------- 333
Q66AF0_ypADA1 LTP---HEIRQAQANGLEMAFISEQEKQALRDKVFP-IS---------------------- 334
QBG08067_khADA2 LSR---EQIRQAQVNGLELAFLSPQEKASLIARVEQ-A---------------------- 333
Q8ZPL9_stADA1 LSR---EQIRQAQINGLEIAFLSDSEKRALREKVAE-A---------------------- 333
CUU93774_ecADA1 LSR---EQIRQAQINGLEMAFLSAEEKRALREKVAA-K---------------------- 333
Q83RC0_sfADA2 LSR---EQIRQAQINGLEMAFLSAEEKRALREKVAA-K---------------------- 333
QKN70503_scADA2 WSV---EDLRTVTVNALKSAFVPFDERTALIEDVVLPAYASVL----------------- 359
CCK32622_sdADA2 WTV---EDLRTVTVNALKSAFIPFDERKALIEDVVLPGYAV------------------ 354
ASO22668_ahADA2 YGW---DDLRRFTLNAMKSAFIGYDERLALIDDLIKPGYAALLD---------------- 419
Q9AK25_scADA1 YTL---DDMQWFSVNAMKSAFIPFDERLAMINDVIKPGYAELKSEWLFQQTASTSGSSES 394
B8ZUM8_mlADA1 YGW---GDLERFTINAMKSAFIPFDQRLAIIDEVIKPRFAVLVG---------------- 362
Q0S358_rjADA2 YGW---TDLERYTINAMKSAFIPFDQRLQLIDEVIKPGFAVLVG---------------- 361
Q036B8_lpADA1 TTA---ADFLAFNLNAIDAAFIPDADKKSLRDQLHQ-DYATYC----------------- 339
Q8XHH8_cpADA1 IDF---EDYKKIYLNSVEASFCSEELKEKLKLSIII----------------------- 332
WP_108223611_vADA2 LTK---EDYAQIYKYSVENAFASDEVKQHLMSFADQ-I---------------------- 332



Supplementary Fig. 6 Multiple sequence alignment of mfCARF-AD with adenosine deaminases acting on free adenosines (ADAs) using ClustalO. Red, green, and blue boxes indicate residues essential for zinc-binding, substrate-binding, and substrate deamination, respectively. Arrows show residues of mfCARF-AD that are not conserved to the rest ADAs.

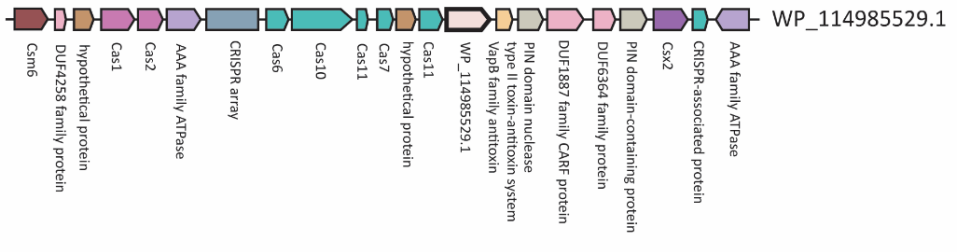

CLUSTAL O(1.2.4) multiple sequence alignment
mfCARF-AD MDSSPQKNVLLCTLGSTLSKTWLIAVEALYFVNHYHQKLDEVHLLTTCNPEAEECFQNIE 60
Q8IJA9_pfADA ---------- 0
Q9KNI7_vcADA ---------- 0
Q9I6Y4_paADA ---------- 0
mfCARF-AD QCLKIHFPGLIYTFTQVDECADICSQEDYFLFQEVLYRWFLKHIQRATVYVCIASGRKTM 120
Q8IJA9_pfADA ---------- 0
Q9KNI7_vcADA ---------- 0
Q9I6Y4_paADA ---------- 0
mfCARF-AD SAALLKGASLFGAKYVFDVLAEEPFPSSMEEVKEALKTNKIKFVSLGNEKGWTQLKKLSF 180
Q8IJA9_pfADA ---------- 0
Q9KNI7_vcADA ---------- 0
Q9I6Y4_paADA ---------- 0
mfCARF-AD EEFPLNEIERKPRLYLATVADRKLQEKINEVQDYSLRIHYNWGRLSELPFPVLARWTPKE 240
Q8IJA9_pfADA ----------MN-CKNMDTSYEIINYLTKDE 20
Q9KNI7_vcADA ---------- 0
Q9I6Y4_paADA ---------- 0
mfCARF-AD LDWLNEPLDPKRDKEWIEKLPKVELHCHLGGFATHGELLAQVRSSAEAP-SSLPHLLEPP 299
Q8IJA9_pfADA LDIDLSCMDKKERYKIWKRLPKCELHCHLDVCFSVDFFLNVIRKYNIQPNMS-DEEIIDY 79
Q9KNI7_vcADA ----------MITSSLPLTDLHRHLDGNIRTQTILELGQKFGVKLPANTLQTLTPY 46
Q9I6Y4_paADA ----------MYEWLNALPKAELHLHLEGTLEPELLFALAERNRIALPWNDVETLRKA 48
mfCARF-AD LPEGWPRPAKPIGLETYMELGRATGTDLLKDPGCLKEQCRLLYQALQKDRVVYAEIRTSP 359
Q8IJA9_pfADA YLFSKPG----KSLDEFVEKA-LRLTDIYIDYTVVEDLAKHAVFNKYKEGVVLMEFRYSP 134
Q9KNI7_vcADA VQIVEAE----PSLVAFLSKL-DWGVAVLGDLDACRRVAYENVEDALNARIDYAELRFSP 101
Q9I6Y4_paADA YAFNN--------LQEFLDLY-YAGADVLRTEQDFYDLTWAYLQKCKA----QNVVHVEP 95
mfCARF-AD NNYTNIS--KGRSAWTVLCEIRDHFQQCMNENQGADTYCHVNLIIIVTRKDKGDLSDISR 417
Q8IJA9_pfADA SFMSF---KHNLDKDLIHEAIVKGLNEAVALL---EYKIQVGLLC----------TGDG 177
Q9KNI7_vcADA YYMAM---KHSLPVTGVVEAVVDGVRAGVRDF-----GIQANLIG----------IMSR 142
Q9I6Y4_paADA FFDPQTHTDRGIPFEVVLAGIRAALRDGEKLL-----GIRHGLIL----------SFLR 139
mfCARF-AD HLSLA---VTAAQHFPPKESYCGVVGVDLAGLESPKTRASYFAEDFVPVHRCGLAVTAHA 474
Q8IJA9_pfADA GLSHERMKEAAEFCIKHKK---DFVGYDHAGHEVDL-K--PFKDIFDNIREEGISLSVHA 231
Q9KNI7_vcADA TFGTDACQQELDAILSQKN---HIVAVDLAGDELGQ-PGDRFIQHFKQVRDAGLHVTVHA 198
Q9I6Y4_paADA HLSEEQAQKTLDQALPFRD---AFIAVGLDSSEVGH-PPSKFQRVFDRARSEGFLTVAHA 195
mfCARF-AD GEIDD---VEGIWQAIFRLHARRIGHALHLLKSQDLLNTVIDRHIGIEMCPLANYQIHGF 531
Q8IJA9_pfADA GEDVSIPNLNSLYTAINLLHVKRIGHGIRVSESQELIDLVKEKDILLEVCPISNVLLNNV 291
Q9KNI7_vcADA GEAAG---PESMWQAIRDLGATRIGHGVKAIHDPKLMDYLAQHRIGIESCLTSNLQTSTV 255
Q9I6Y4_paADA GEEGP---PEYIWEALDLLKVERIDHGVRAFEDERLMRRLIDEQIPLTVCPLSNTKLCVF 252
mfCARF-AD KPMENNKNTYPLKEYLNRGVYVTVNTDNIGISKGSLTDNLLFLAELCPGIRRLDILKLLN 591
Q8IJA9_pfADA KSMD----THPIRMLYDAGVKVSVNSDDPGMFLTNITDNYEELYT-HLNFTLADFMKMNL 346
Q9KNI7_vcADA DSLA----THPLKRFLEHGILACINTDDPAVEGIELPYEYEVAAP-QAGLSQEQIRQAQL 310
Q9I6Y4_paADA DDMS----QHTILDMLERGVKVTVNSDDPAYFGGYVTENFHALQQ-SLGMTEEQARRLAQ 307

Supplementary Fig. 7 Multiple sequence alignment (ClustalO) of mfCARF-AD with proteins that showed the highest structural homology to mfCARF-AD in Phyre2 analysis. Red, green, yellow, and blue boxes indicate residues essential for zinc-binding, substrate-binding, substrate specificity, and substrate deamination, respectively.



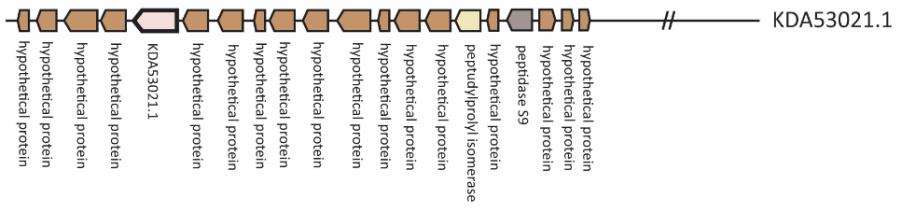
Supplementary Fig. 8 Structural comparison of mfCARF-AD with ecADA1 using AlphaFold and Chimera. Blue and yellow colours indicate the adenosine deaminase domain of mfCARF-AD and ecADA1, respectively, while pink colour shows the CARF domain of mfCARF-AD.

a



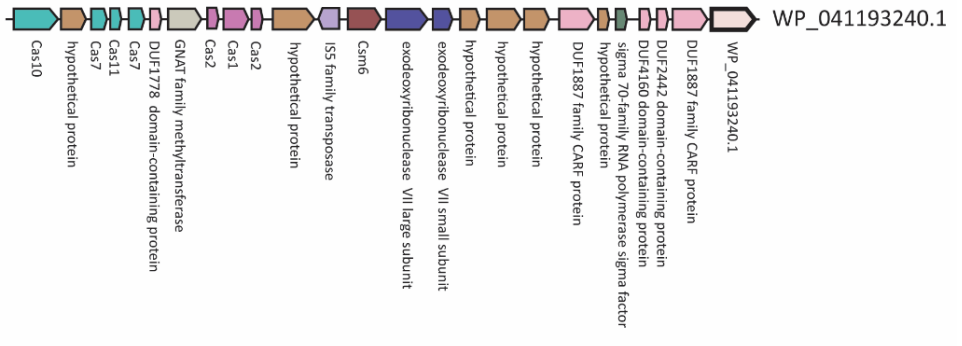
Candidatus Cyclonatronum proteinivorum (Type III-A CRISPR-Cas system)

b



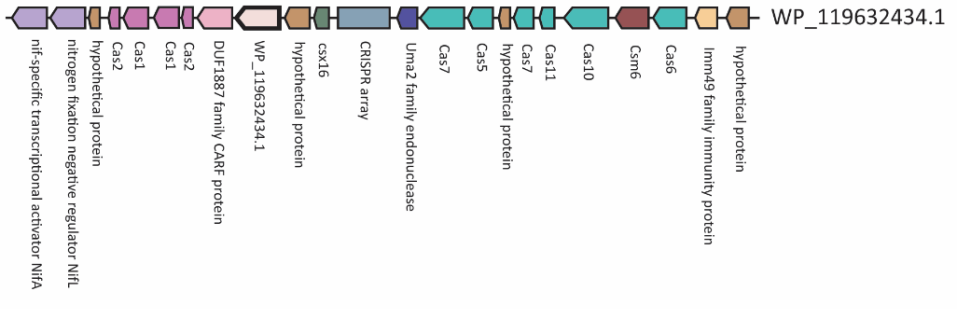
Thermoanaerobaculum aquaticum

c

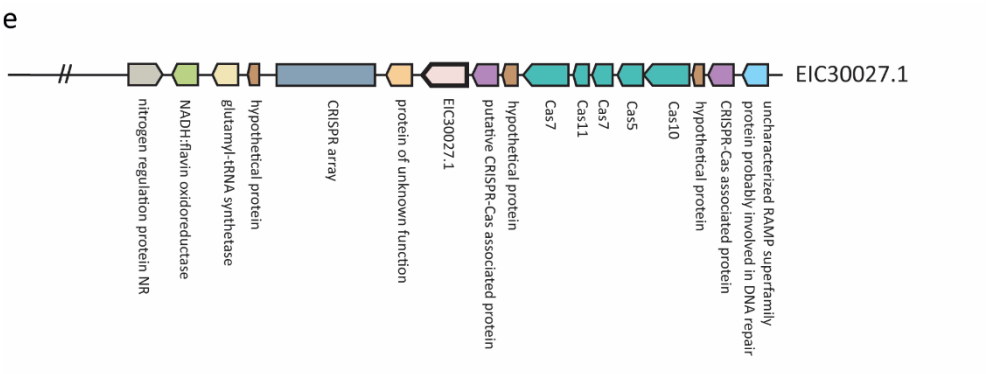


Candidatus Symbiobacter mobilis (Type III-B CRISPR-Cas system)

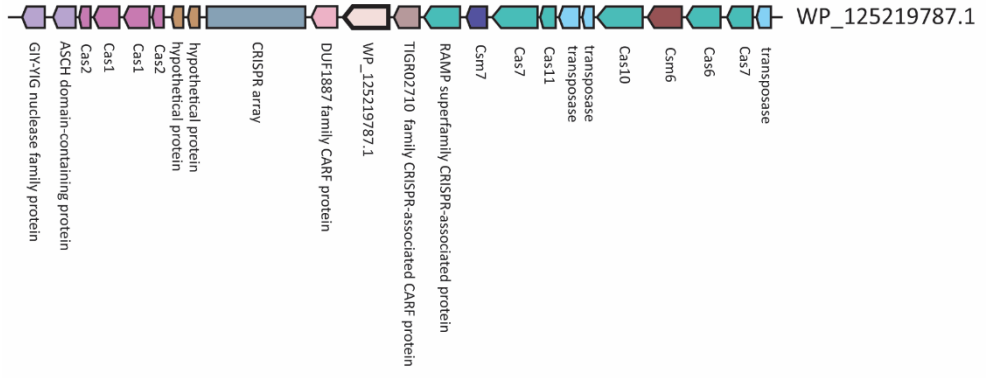
d



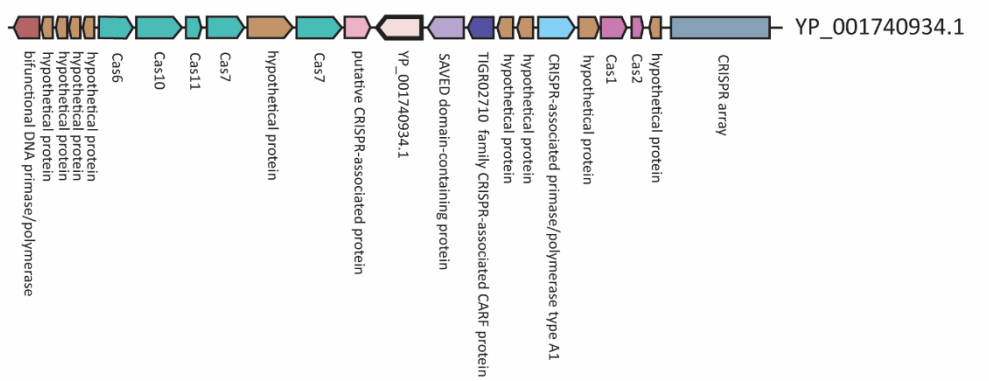
Methylocaldum marinum (Type III-A CRISPR-Cas system)



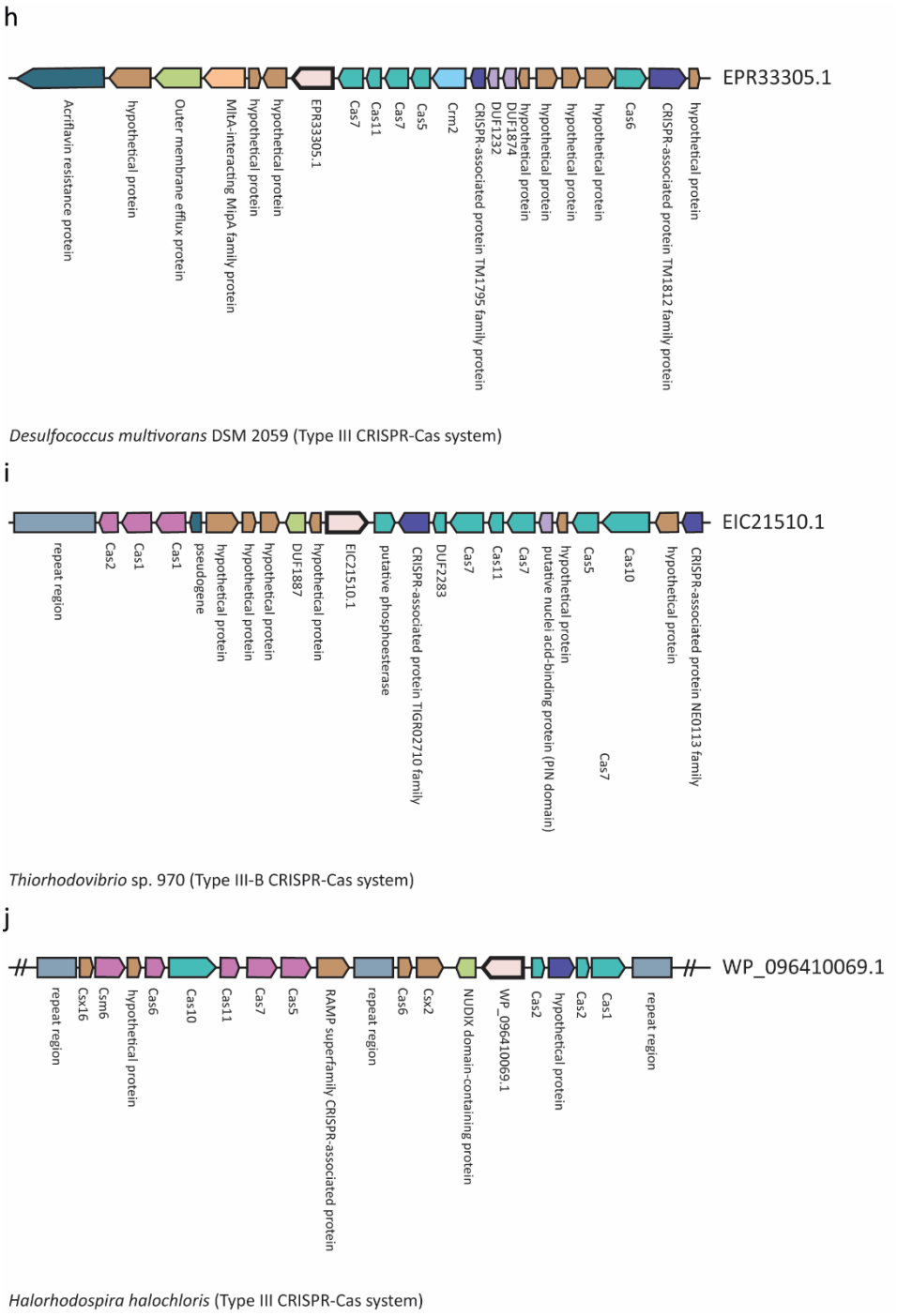
Methylomicrobium album BG8 (Type III-B CRISPR-Cas system)



Methylomicrobium sp. wino1 (Type III-A CRISPR-Cas system)

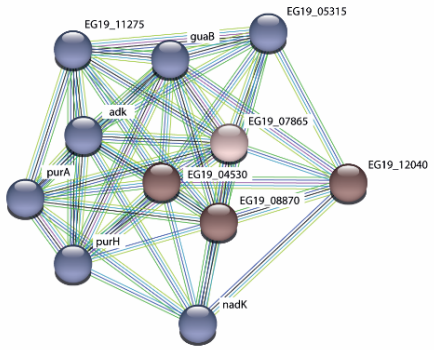


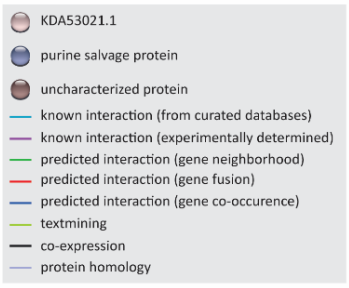
Candidatus Cloacimonas acidaminovorans str. Evry (Type III-A CRISPR-Cas system)



Supplementary Fig. 9 Genomic loci of CARF5-clade members that carry both a CARF and an adenosine deaminase domain, according to Makarova et al., 2020[26].

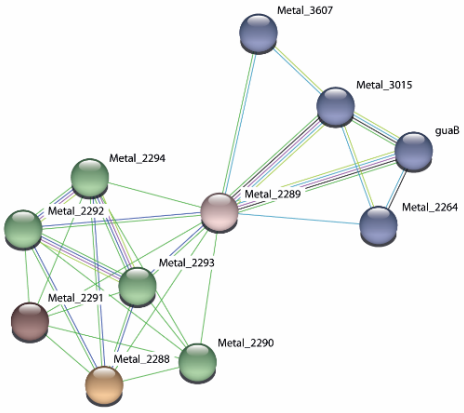
a

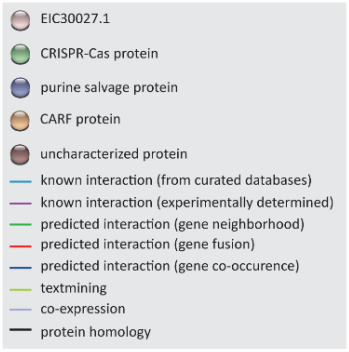




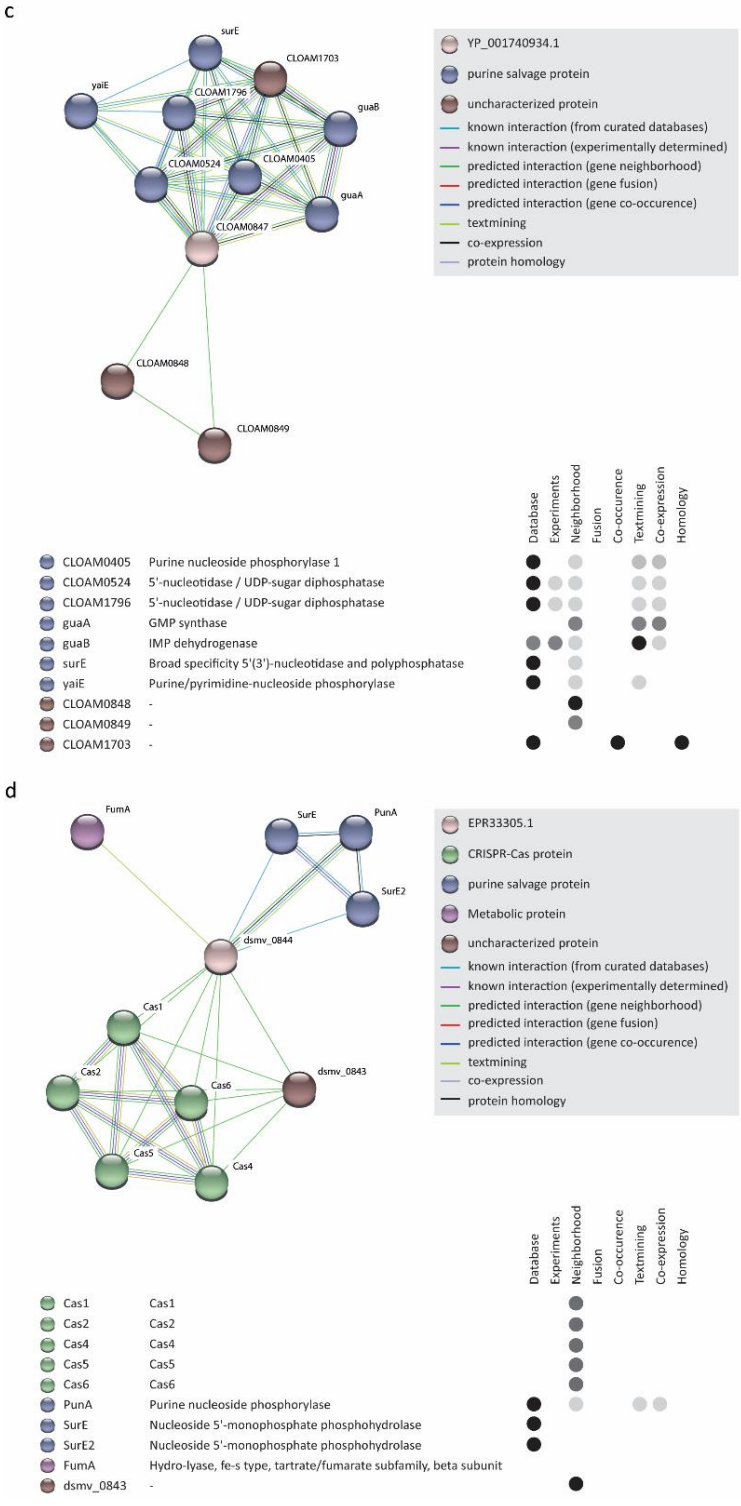
adk	Adenylate kinase
EG19_05315	Purine nucleoside phosphorylase
EG19_11275	Hypoxanthine phosphoribosyltransferase
guaB	IMP dehydrogenase
nadK	NAD kinase
purA	Adenylosuccinate synthetase
purH	Phosphoribosylaminoimidazolecarboxamide formyltransferase
EG19_04530	-
EG19_08870	-
EG19_12040	-

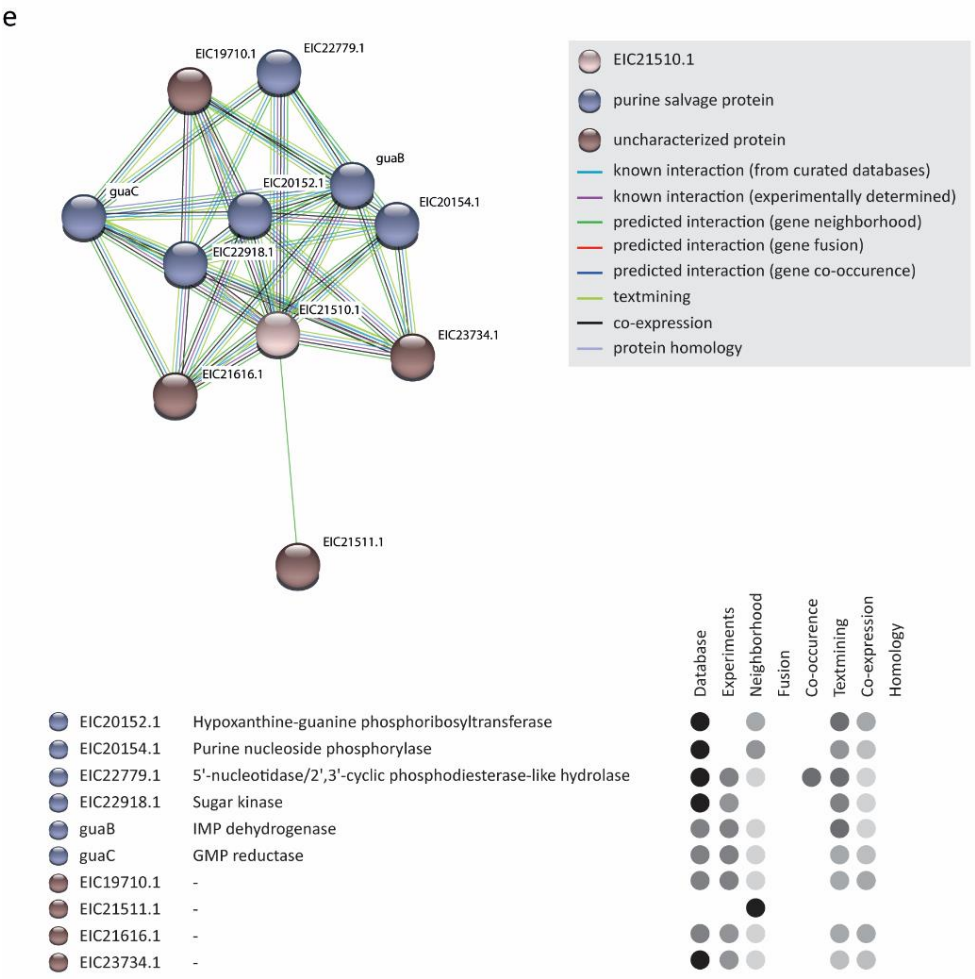
b



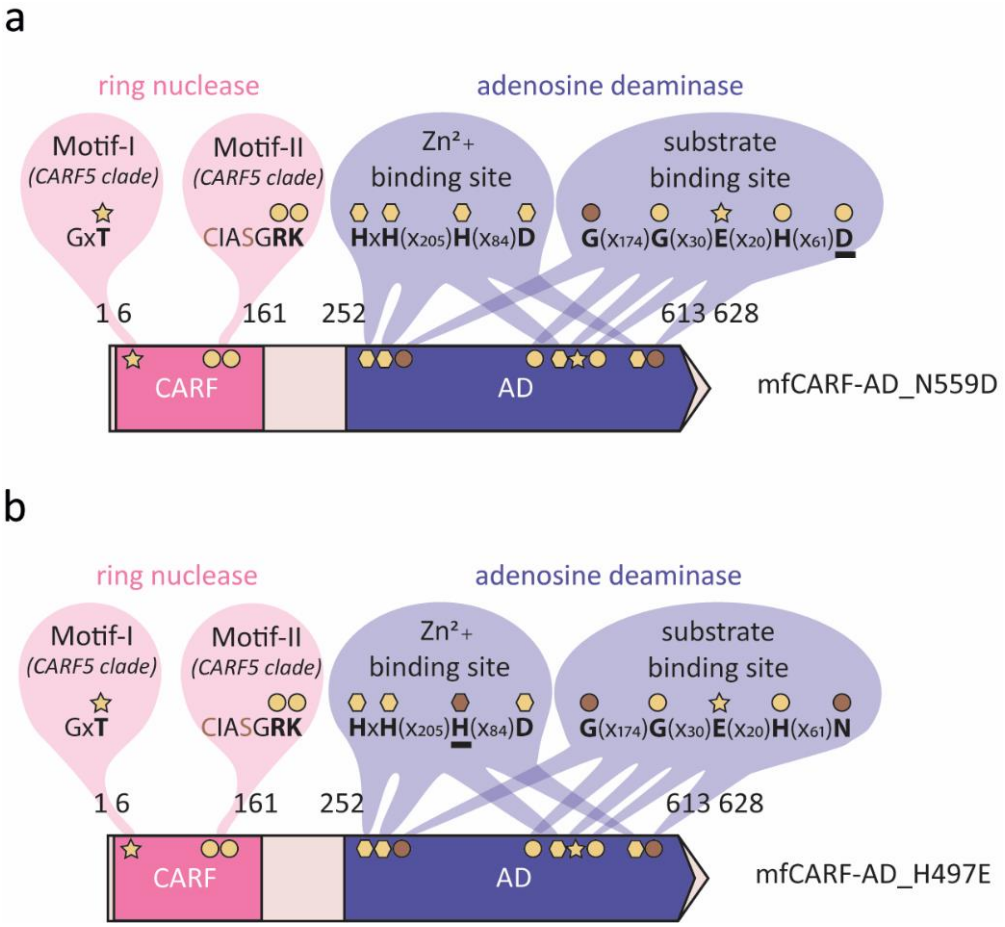


Metal_2290	TIGRFAM: putative CRISPR-associated protein, VVA1548 family
Metal_2292	Cas7
Metal_2293	Cas11
Metal_2294	Cas7
guaB	IMP dehydrogenase
Metal_2264	5'-nucleotidase
Metal_3015	5'-nucleotidase/2',3'-cyclic phosphodiesterase-like hydrolase
Metal_3607	Putative HD superfamily hydrolase
Metal_2288	DUF1887 domain containing-protein
Metal_2291	-

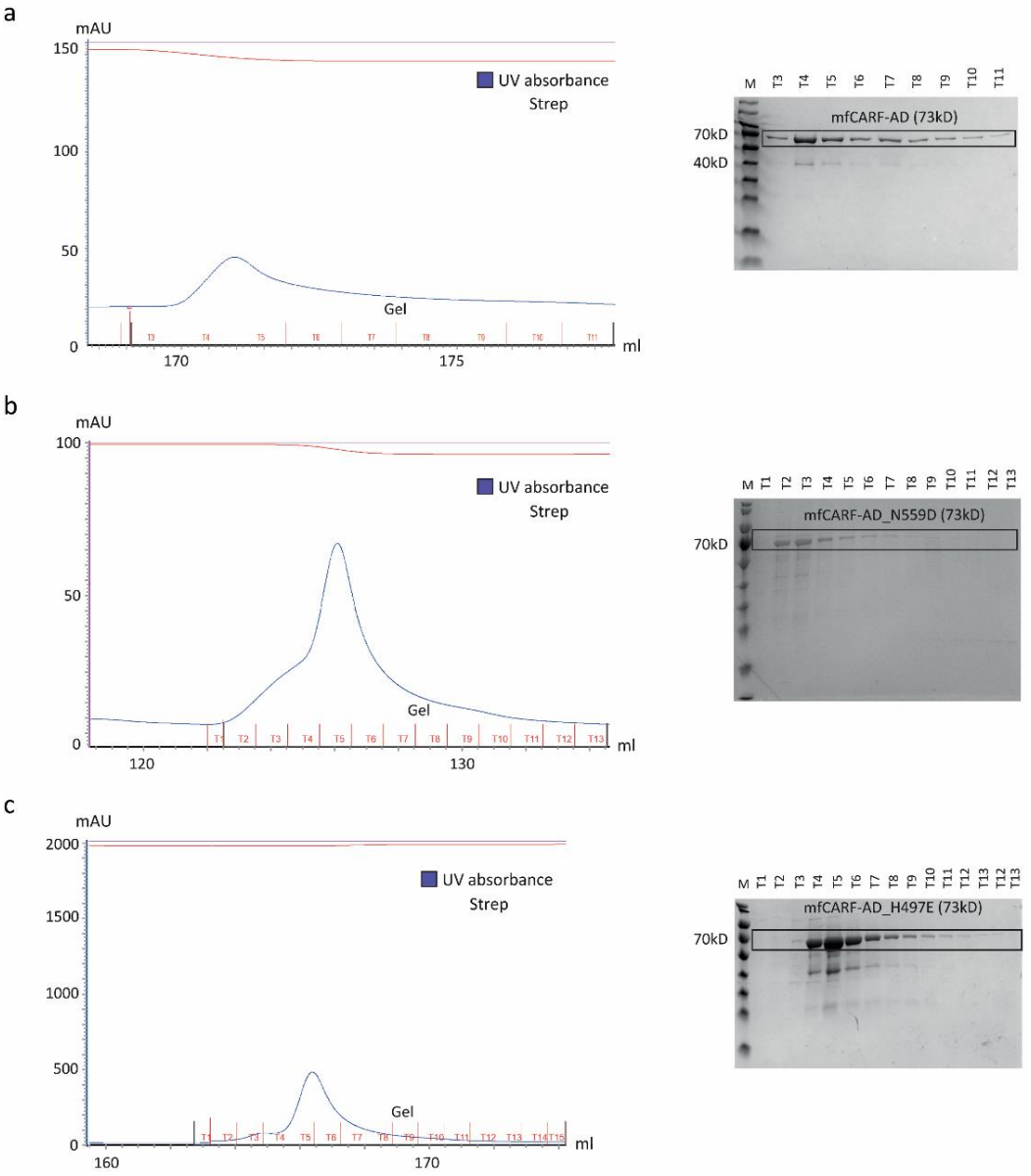

c
surE
CLOAM1703
yaiE
CLOAM1796
guaB
CLOAM0524
CLOAM0405
guaA
CLOAM0847
CLOAM0848
CLOAM0849
YP_001740934.1
purine salvage protein
uncharacterized protein
known interaction (from curated databases)
known interaction (experimentally determined)
predicted interaction (gene neighborhood)
predicted interaction (gene fusion)
predicted interaction (gene co-occurence)
textmining
co-expression
protein homology
Database
Experiments
Neighborhood
Fusion
Co-occurence
Textmining
Co-expression
Homology
CLOAM0405 Purine nucleoside phosphorylase 1
CLOAM0524 5'-nucleotidase / UDP-sugar diphosphatase
CLOAM1796 5'-nucleotidase / UDP-sugar diphosphatase
guaA GMP synthase
guaB IMP dehydrogenase
surE Broad specificity 5'(3')-nucleotidase and polyphosphatase
yaiE Purine/pyrimidine-nucleoside phosphorylase
CLOAM0848 -
CLOAM0849 -
CLOAM1703 -
d
FumA
SurE
PunA
SurE2
dsmv_0844
Cas1
Cas2
Cas6
dsmv_0843
Cas5
Cas4
EPR33305.1
CRISPR-Cas protein
purine salvage protein
Metabolic protein
uncharacterized protein
known interaction (from curated databases)
known interaction (experimentally determined)
predicted interaction (gene neighborhood)
predicted interaction (gene fusion)
predicted interaction (gene co-occurence)
textmining
co-expression
protein homology
Database
Experiments
Neighborhood
Fusion
Co-occurence
Textmining
Co-expression
Homology
Cas1 Cas1
Cas2 Cas2
Cas4 Cas4
Cas5 Cas5
Cas6 Cas6
PunA Purine nucleoside phosphorylase
SurE Nucleoside 5'-monophosphate phosphohydrolase
SurE2 Nucleoside 5'-monophosphate phosphohydrolase
FumA Hydro-lyase, fe-s type, tartrate/fumarate subfamily, beta subunit
dsmv_0843 -



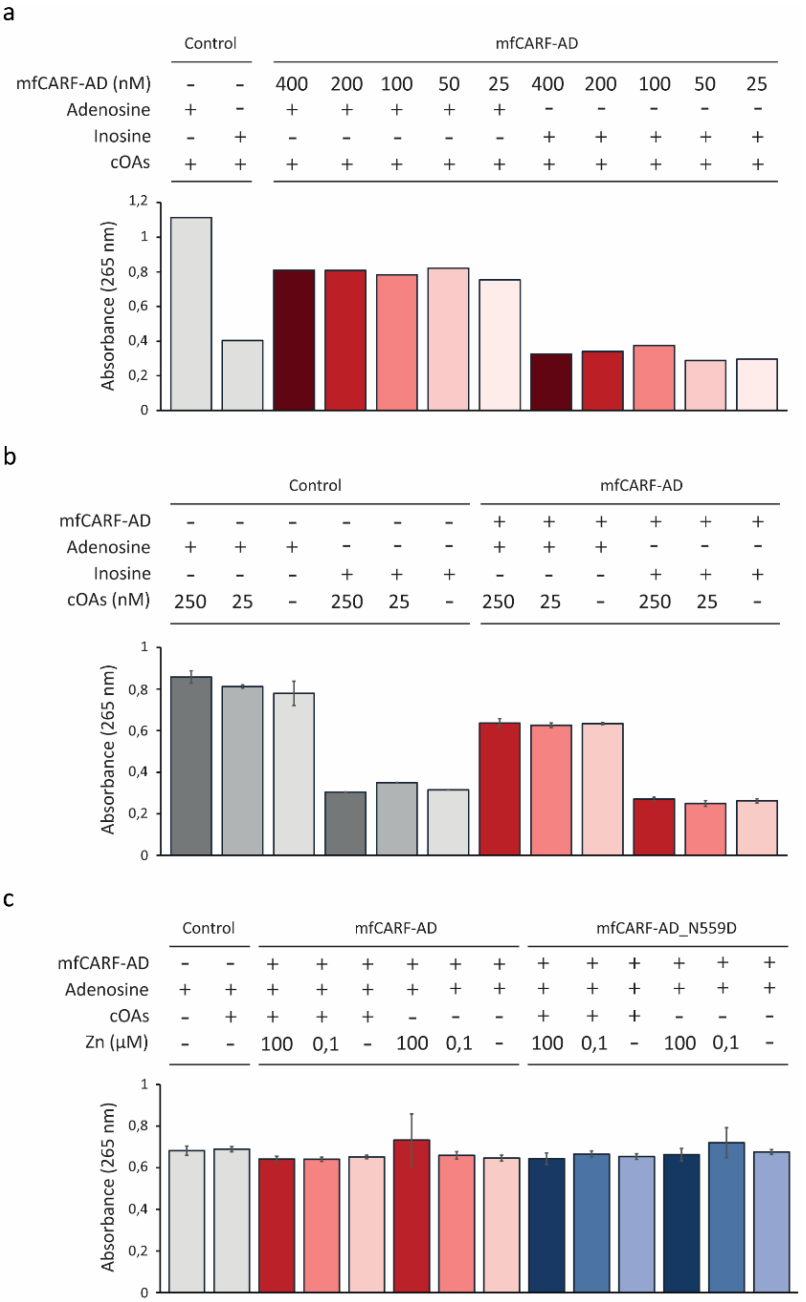
Supplementary Fig. 10 Functional protein association network of CARF5-clade members using STRING analysis. Blue and brown circles indicate purine salvage and uncharacterized proteins, respectively, while coloured lines show known or predicted interactions, textmining, co-expression, and protein homology.



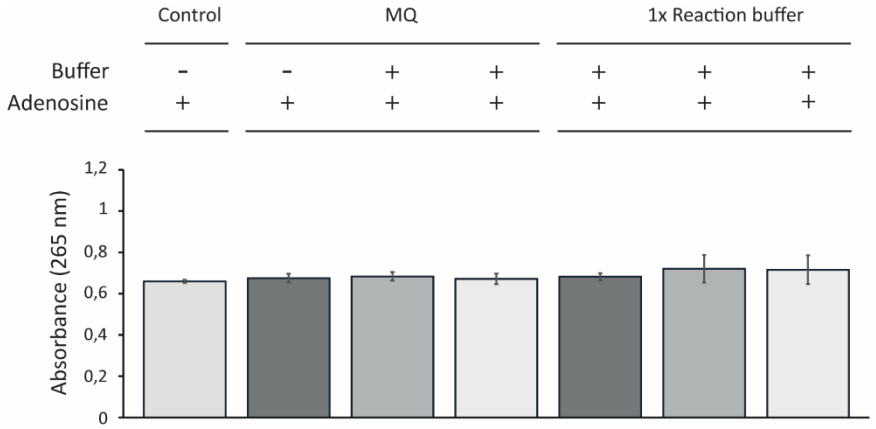
Supplementary Fig. 11 Domain architecture of mutant mfCARF-AD_N559D and mfCARF-AD_H497E variants. The CARF and AD domains are shown as pink and blue boxes, respectively. Yellow stars, circles, and hexagons represent conserved active sites, substrate binding sites, and zinc binding sites, respectively. Brown circles indicate substrate binding residues that are not conserved among other adenosine deaminases.



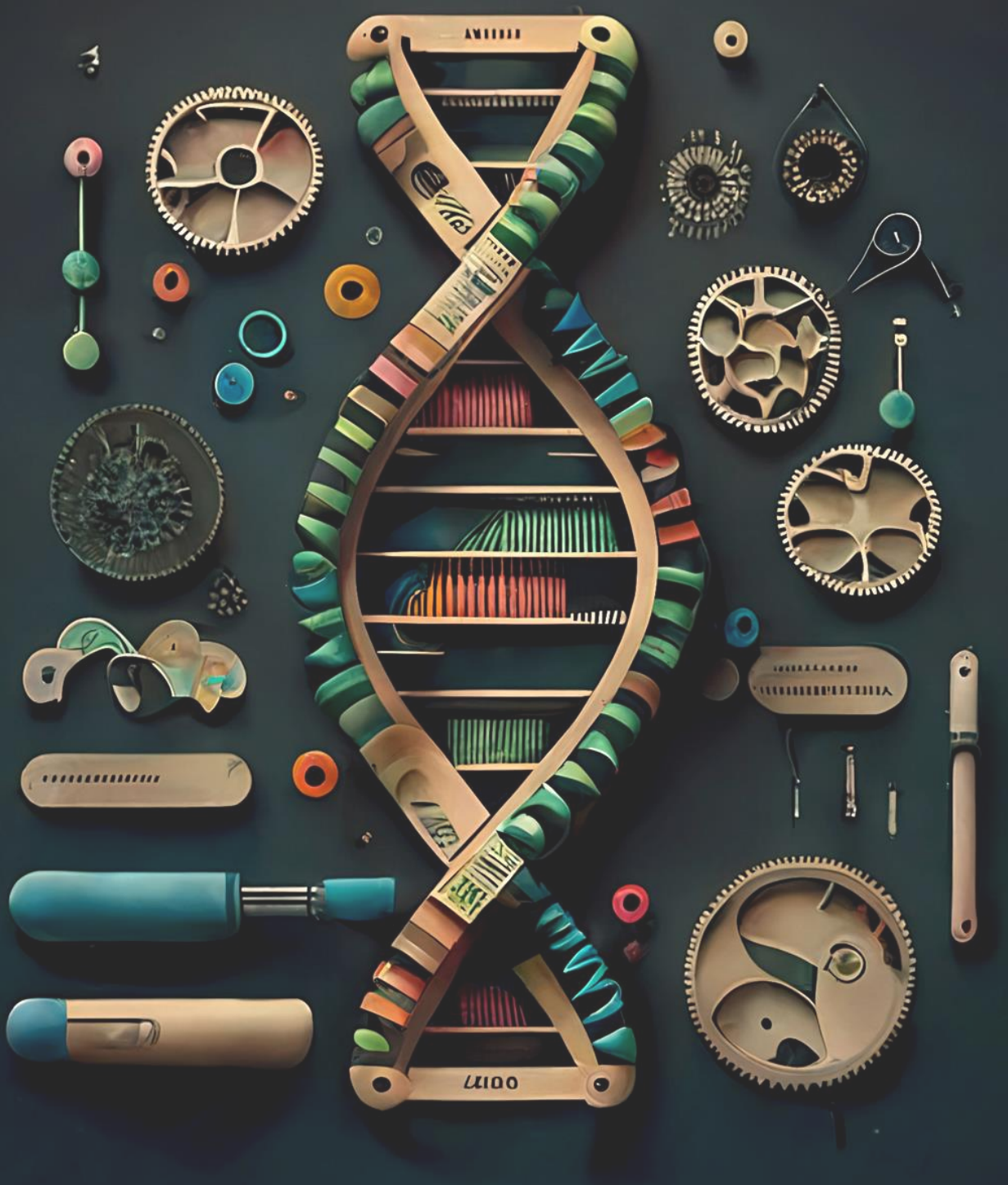
Supplementary Fig. 12 Strep-purification of mfCARF-AD variants. Chromatograph and SDS-PAGE gels from the purification of (**a**) mfCARF-AD, (**b**) mfCARF-AD_N559D, and (**c**) mfCARF-AD_H497E.



Supplementary Fig. 13 *In vitro* assays for deamination of free adenosine. Absorbance measured at 265 nm after 3h of reaction in a mixture of mfCARF-AD or mfCARF-AD_N559D with adenosine or inosine, in the presence or the absence of cOAs and zinc. Increasing concentrations of (**a**) mfCARF-AD, (**b**) cOAs, and (**c**) zinc were examined. Symbols '+' and '-' indicate presence and absence of the corresponding element, respectively. Bar graphs in section b and c were created based on results from three independent replicates.



Supplementary Fig. 14 Mock control *in vitro* assay. Absorbance measured at 265 nm after 3h of reaction in a mixture of adenosine with or without MQ or 1x Reaction buffer. Symbols '+' and '-' indicate presence and absence of the corresponding element, respectively. Bar graphs were created based on results from three independent replicates.



Chapter 4

Keeping CRISPR in check: diverse mechanisms of phage–encoded anti–CRISPRs

Despoina Trasanidou[1], Ana Sousa Geros[2], Prarthana Mohanraju[1], Anna Cornelia Nieuwenweg[1], Franklin L. Nobrega[2,*] & Raymond Staals[1,*]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

[2]Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands.

* Corresponding authors

Adapted from publication:

Trasanidou, D., Geros, A. S., Mohanraju, P., Nieuwenweg, A. C., Nobrega, F. L., & Staals, R. H. (2019). Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS microbiology letters*, 366(9), fnz098.

Abstract

CRISPR-Cas represents the only adaptive immune system of prokaryotes known to date. These immune systems are widespread among bacteria and archaea, and provide protection against invasion of mobile genetic elements, such as bacteriophages and plasmids. As a result of the arms-race between phages and their prokaryotic hosts, phages have evolved inhibitors known as anti-CRISPR (Acr) proteins to evade CRISPR immunity. In the recent years, several Acr proteins have been described in both temperate and virulent phages targeting diverse CRISPR-Cas systems. Here, we describe the strategies of Acr discovery and the multiple molecular mechanisms by which these proteins operate to inhibit CRISPR immunity. We discuss the biological relevance of Acr proteins and speculate on the implications of their activity for the development of improved CRISPR-based research and biotechnological tools.

Keywords

CRISPR-Cas; phage; genome editing; anti-CRISPR

Introduction

Viruses are ubiquitous entities co-existing with cellular life forms, present in almost all explored environments[1]. Viruses that infect bacteria (bacteriophages or phages) are the most abundant biological entities on the planet with population numbers in the order of 10^{31}[1-3]. The ability of phages to easily maneuver between different biomes, operating as vehicles of horizontal gene transfer (HGT), makes them major agents of evolution[4]. Bacteriophages are classified based on their life-cycle into virulent and temperate. Virulent phages rely exclusively on productive infection cycles for propagation, which ultimately kills the host for the release of new viral particles that can engage in another round of infection. Temperate phages have the choice to multiply in their host cells leading to cell lysis or to integrate their phage genome into the bacterial chromosome as a prophage. Prophages are propagated passively by the replication machinery of the bacterial cell[5].

As a response to the constant threat of phage infection, a diverse arsenal of defense mechanisms has evolved in bacterial hosts. Because phages evolve rapidly to counter these immune systems[6], the hosts need to constantly evolve new means of self-protection, leading to a perennial arms-race between hosts and their phages[7]. The defense systems evolved by bacteria provide both innate and adaptive immunity against phage infection. Innate immunity systems interfere at different levels of the phage's infection cycle via receptor masking, superinfection exclusion (Sie), restriction–modification (R-M), bacteriophage exclusion (BREX), toxin–antitoxin (TA) modules, abortive infection (Abi), prokaryotic Argonautes (pAgos), production of antiphage chemicals and defense island system associated with restriction–modification (DISARM) systems[8-13], whereas other remain to be characterized[14]. Adaptive and heritable immunity is provided by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated (Cas) systems[15], which work as a fascinating complementation to the innate defense strategies.

Diverse variants of the CRISPR-Cas defense system are present in most of the sequenced genomes of archaea and half of those of bacteria[10]. A CRISPR-Cas locus typically consists of a CRISPR array and an operon of CRISPR-associated (*cas*) genes. The CRISPR array is composed of a series of short, partially palindromic and direct repetitive sequences (repeats) interspaced by variable sequences (spacers), originating from phage genomes or other invading mobile genetic elements (MGE), such as (conjugative) plasmids[16-19]. The

cas genes encode for the Cas proteins, which are necessary for the generation of new spacers or are involved in the targeting of the MGE, as explained below. Collectively, these two elements of CRISPR-Cas systems mediate sequence-specific immunity against invasive MGEs[20-23].

The continuous arms-race between prokaryotic hosts and their cognate MGEs is speculated to be responsible for the rapid evolution of highly diverse CRISPR-Cas systems. The current CRISPR-Cas classification scheme distinguishes two broad classes based on the protein composition of the effector Cas complex. Class 1 systems (types I, III and IV) use multi-subunit Cas protein complexes for the recognition of targeted nucleic acids, while the less common class 2 systems (types II, V and VI) employ a single multi-domain effector protein complex that performs target recognition and cleavage. These classes are further subdivided into a total of six CRISPR types with 25 subtypes[24].

Despite substantial structural and functional diversity, all CRISPR-Cas systems mediate immunity through three distinct steps: adaptation, expression and interference[25]. During adaptation, short DNA fragments (known as protospacers) are acquired from invading MGEs and subsequently processed and inserted as spacers into the CRISPR locus, typically by the Cas1-Cas2 complex[26]. Next, during expression, the CRISPR array is transcribed as a long precursor CRISPR RNA (pre-crRNA) and the Cas proteins are expressed. The pre-crRNA is then processed within repeat regions to yield mature CRISPR RNAs (crRNAs) by dedicated Cas proteins and/or host factors[20,27-29]. The crRNAs are packaged with one or more Cas proteins into effector Cas complexes that scrutinize the microbial cell for potential invasion. Finally, during interference, the Cas complexes recognize complementary target sequences of invading MGEs by Watson–Crick base pairing. Upon binding to a cognate target sequence, the complex either recruits a nuclease or stimulates its intrinsic nuclease activity to neutralize the invader[20,23,30]. Type I, II and V CRISPR-Cas systems target DNA and rely on a short stretch (2 to 7 nucleotides) of conserved nucleotides adjacent to the protospacer, known as the protospacer adjacent motif (PAM), for spacer selection during adaptation and target identification during interference[17,21]. The PAM allows for self/nonself discrimination, as its absence in the CRISPR array prevents autoimmunity and self-cleavage.

In response to the microbial antiviral defense mechanisms, phages have evolved numerous mechanisms to overcome prokaryotic immunity. Phage evasion from prokaryotic CRISPR-Cas systems was first found to rely on mutational drifts, predominantly occurring in regions that require perfect complementarity between the

crRNA and the protospacer (so-called seed region) for interference, or in the PAM sequences[31-33]. Depending on the location of the mismatch (between the crRNA and the protospacer), a single mutation can be sufficient to abolish CRISPR-Cas immunity[31,34]. Deletion of the protospacer sequence and/or the PAM have also been shown to provide an effective way for phages to escape CRISPR-Cas targeting, despite the potential of imposing a fitness cost[31]. Similar to the evasion strategy from R-M systems, phages can also modify their bases with hydroxymethylcytosine (HMC) and its bulkier glycosylated form to reduce target binding affinity and thereby protect from CRISPR-mediated targeting by both type I and type II systems[35,36], whereas other modifications do not disturb Cas9 recognition[37]. Finally, some phages encode their own CRISPR locus that targets host antiviral genomic regions, such as chromosomal (defense) islands[38].

The first examples of phage-encoded anti-CRISPR (Acr) proteins were found in class 1 type I-F and I-E systems of *Pseudomonas aeruginosa*[39,40]. Acr proteins have distinct sequences (Tables 1, 2), structures[41-43] and mechanisms[33] and they provide phages with a direct and specific means to inhibit targeting by the CRISPR-Cas system. To date, 45 unique families of Acr proteins have been discovered, and categorized into class 1 (Table 1) and class 2 (Table 2) CRISPR-Cas inhibitors. These highly diverse and small (typically 50–330 amino acids) proteins do not share much sequence or protein domain similarity to each other or to any protein of known function[44].

Here, we explore the biological relevance and detail the recent insights into the molecular mechanisms and structures of anti-CRISPR proteins. We also address the development of anti-CRISPRs as ‘off-switches’ for genome editing and discuss the impact of their use in other biotechnological applications.

Table 1. Class 1 anti-CRISPR protein families.

Family	Size (aa)	Origin (characterized homolog)	Method of discovery	Accession number	CRISPR-Cas type inhibited [organism]	Mode-of-action	Structure	PDB code	References
AcrIC1	190	*Moraxella bovoculi* prophage	Self-targeting, Guilt-by-association [*acrIF11*]	AKG19229.1	I-C [Mbo]	-	-	-	44
AcrID1	98	*Sulfolobus islandicus* rudivirus 3	Functional assays	YP_009272954.1	I-D [Sis]	Binds as a dimer to the Cas10d, mimicking DNA (blocks DNA binding)	Compact dimeric αβ-sandwich; each monomer 5-stranded antiparallel β-sheet + 2 α-helices at one side of the β-sheet	6EXP	49
AcrIE1	100	*Pseudomonas aeruginosa* phage JBD5	Functional assays	YP_007392738.1	I-E [Pae]	Binds as a dimer to the Cas3 (blocks DNA cleavage)	Elongated dimeric structure; each monomer 1 antiparallel β-sheet + 3 α-helices	6ARZ, 6AS4	40,66
AcrIE2	84	*P. aeruginosa* phage JBD88a	Functional assays	YP_007392439.1	I-E [Pae]	-	-	-	40
AcrIE3	68	*P. aeruginosa* phage DMS3	Functional assays	YP_950454.1	I-E [Pae]	Probably binds to the Cascade (blocks DNA binding)			40
AcrIE4	52	*P. aeruginosa* phage D3112	Functional assays	NP_938238.1	I-E [Pae]	-	-	-	40
AcrIE5	65	*Pseudomonas otitidis* mobile genetic element	Guilt-by-association [*aca1*]	WP_074973300.1	I-E [Pae]	-	-	-	44
AcrIE6	79	*P. aeruginosa* mobile genetic element	Guilt-by-association [*aca1*]	WP_087937214.1	I-E [Pae]	-	-	-	44
AcrIE7	106	*P. aeruginosa* mobile genetic element	Guilt-by-association [*aca1*]	WP_087937215.1	I-E [Pae]	-	-	-	44
AcrIF1	78	*P. aeruginosa* phage JBD30	Functional assays	YP_007392342.1	I-F [Pae, Pec]	2-3 copies interact with the hexameric Cas7f spine of the Cascade (block DNA binding)	4-stranded antiparallel β-sheet + 2 α-helices at one side of the β-sheet	2LW5, 5UZ9, 6ANV, 6B46	39,62,63
AcrIF2	90	*P. aeruginosa* phage D3112	Functional assays	NP_938237	I-F [Pae, Pec]	Binds to the Cas5f:Cas8f tail of the Cascade, mimicking DNA (blocks DNA binding)	4-stranded antiparallel β-sheet + 2 antiparallel α-helices at either side of the β-sheet	5UZ9, 6B47	39,62,63
AcrIF3	139	*P. aeruginosa* phage JBD88a	Functional assays	YP_007392440.1	I-F [Pae]	Binds as a dimer to the Cas3, preventing its recruitment to the Cascade-dsDNA (blocks DNA binding) or spacer acquisition by the Cas1-2/3 complex (blocks adaptation)	Dimeric structure; each monomer 6 α-helices	5GNF, 5GQH, 5B7I	39,42,65
AcrIF4	100	*P. aeruginosa* phage JBD26	Functional assays	WP_016068584.1	I-F [Pae]	Binds to the Cascade (blocks DNA binding)	-	-	39
AcrIF5	79	*P. aeruginosa* phage JBD5	Functional assays	YP_007392740.1	I-F [Pae]	-	-	-	39
AcrIF6	100	*P. aeruginosa* prophage	Guilt-by-association [*aca1*]	WP_043884810	I-F [Pae, Pec], I-E [Pae]	-	-	-	45
AcrIF7	83	*P. aeruginosa* prophage	Guilt-by-association [*aca1*]	ACD38920.1	I-F [Pae, Pec]	-	-	-	45
AcrIF8	92	*Pectobacterium carotovorum* phage ZF40	Guilt-by-association [*aca2*]	AFC22483.1	I-F [Pae, Pec]	-	-	-	45
AcrIF9	68	*Vibrio parahaemolyticus* mobile genetic element	Guilt-by-association [*aca2*]	WP_031500045.1	I-F [Pae, Pec]	-	-	-	45

Table 1. Class 1 anti-CRISPR protein families. *(continuation)*

Family	Size (aa)	Origin (characterized homolog)	Method of discovery	Accession number	CRISPR-Cas type inhibited [organism]	Mode-of-action	Structure	PDB code	References
AcrIF10	97	*Shewanella xiamenensis* prophage	Guilt-by-association [*aca2*]	KEK29119	I-F [Pae, Pec]	Binds to the Cas7f:Cas8f tail, mimicking DNA (blocks DNA binding)	4-stranded antiparallel β-sheet + 3 antiparallel α-helices at one side of the β-sheet	6ANW, 6B48	45,63
AcrIF11	132	*P. aeruginosa* mobile genetic element	Guilt-by-association [*aca1*]	WP_038819808.1	I-F [Pae]	-	-	-	44
AcrIF12	124	*P. aeruginosa*	Guilt-by-association [*aca4*]	ABR13388.1	I-F [Pae]	-	-	-	44
AcrIF13	115	*Moraxella catarrhalis* prophage	Self-targeting, Guilt-by-association [*acrIF11*]	EGE18854.1	I-F [Mbo]	-	-	-	44
AcrIF14	124	*M. catarrhalis* phage Mcat5	Self-targeting, Guilt-by-association [*acrIF11*]	AKI27193.1	I-F [Mbo]	-	-	-	44
AcrIE4-F7	119	*Pseudomonas citronellolis* mobile genetic element	Guilt-by-association [*aca1*]	WP_064584002.1	I-E [Pae], I-F [Pae]	-	-	-	44

Table 2. Class 2 anti-CRISPR protein families.

Family	Size (aa)	Origin (characterized homolog)	Method of discovery	Accession number	CRISPR-Cas type inhibited [organism]	Mode-of-action	Structure	PDB code	References
AcrIIA1	149	*Listeria monocytogenes* prophage J0161a	Self-targeting	WP_003722518.1	II-A [Lmo]	Recognizes nucleic acids (putative transcriptional regulation)	Dimeric structure with pseudo 2-fold symmetry; each monomer 5 α-helices + 1 3_{10} helix at N-terminus and 3 α-helices + 2 3_{10} helices at C-terminus (all helical 2-domain)	5Y6A	46,82
AcrIIA2	123	*L. monocytogenes* prophage J0161a	Self-targeting	WP_003722517.1	II-A [Lmo, Spy]	Binds to the PAM-interacting, the WED, the HNH, and the REC2 domains (blocks DNA recognition, binding/unwinding, and cleavage, respectively)	Bent 4-stranded antiparallel β-sheet + 2 α-helices at either side of the β-sheet	6MCB, 6MCC, 6IFO	46,77,78
AcrIIA3	125	*L. monocytogenes* prophage SLCC2482	Self-targeting, Guilt-by-association [*acrIIA1*]	WP_014930691.1	II-A [Lmo]	-	-	-	46
AcrIIA4	87	*L. monocytogenes* prophage J0161b	Self-targeting, Guilt-by-association [*acrIIA1*]	WP_003723290.1	II-A [Lmo, Spy]	Binds to the PAM-interacting, the Topo-homology, and the RuvC domains (blocks DNA recognition, binding/unwinding, and cleavage, respectively)	3-stranded antiparallel β-sheet + 3 α-helices at one side of the β-sheet + 1 3_{10} helix	5XN4, 5XBL, 5VW1, 5VZL	46,73,74,76 ,100
AcrIIA5	140	*Streptococcus thermophilus* (virulent) phage D4276	Functional assays	ASD50988.1	II-A [Sth1, Sth3, Spy]	-	-	-	87
AcrIIA6	183	*S. thermophilus* (virulent) phage D1811	Functional assays, Guilt-by-association [*acrIIA5*]	MH000604	II-A [Sth1]	-	Dimeric structure; each monomer 4-stranded antiparallel β-sheet + 8 α-helices	6EYX, 6EYY	87
AcrIIC1	86	*Neisseria meningitidis* mobile genetic element	Guilt-by-association [*aca2*]	WP_049360089.1	II-C [Nme,Cje, Geo, Hpa, Smu]	Binds to the HNH active site (allows DNA binding, blocks DNA cleavage)	5-stranded β-bundle interspaced by 2 α-helices	5VGB	43,79
AcrIIC2	123	*N. meningitidis* prophage	Guilt-by-association [*aca3*]	WP_042743678.1	II-C [Nme, Hpa, Smu]	Binds to the HNH domain (blocks DNA cleavage)	-	-	79
AcrIIC3	116	*N. meningitidis* prophage	Guilt-by-association [*aca3*]	WP_042743676.1	II-C [Nme, Hpa, Smu]	Forces Cas9 dimerization (blocks DNA binding)	-	-	79
AcrIIC4	88	*Haemophilus parainfluenzae* prophage	Guilt-by-association [*aca* not reported]	WP_049372635	II-C [Nme, Hpa, Smu]	Binds to the Cas9 (blocks DNA binding)	-	-	81
AcrIIC5	130	*Simonsiella muelleri* transfer element	Guilt-by-association [*aca* not reported]	WP_002642161.1	II-C [Nme, Hpa, Smu]	Binds to the Cas9 (blocks DNA binding)	-	-	81
AcrVA1	170	*Moraxella bovoculi* prophage	Self-targeting, Guilt-by-association [*acrIF11*]	WP_046701302.1	V-A [Mbo, Asp, Lba, Fno]	-	-	-	44,101

Table 2. Class 2 anti-CRISPR protein families. *(continuation)*

Family	Size (aa)	Origin (characterized homolog)	Method of discovery	Accession number	CRISPR-Cas type inhibited [organism]	Mode-of-action	Structure	PDB code	References
AcrVA2	322	*M. bovoculi* prophage	Self-targeting, Guilt-by-association [*acrIF11*]	AKG19228.1	V-A [Mbo]	-	-	-	44
AcrVA3	168	*M. bovoculi* prophage	Self-targeting, Guilt-by-association [*acrIF11*]	AKG19230.1	V-A [Mbo], I-C [Mbo]	-	-	-	44
AcrVA4	234	*M. bovoculi* prophage	Self-targeting	WP_046699156.1	V-A [Mbo, Lba]	-	-	-	101
AcrVA5	92	*M. bovoculi* prophage	Self-targeting	WP_046699157.1	V-A [Mbo, Lba]	-	-	-	101
Csx27	201	*Bergeyella zoohelcum*	Guilt-by-association [*cas13b*]	WP_034985946.1	VI-B	-	-	-	48

Biological relevance of anti–CRISPR proteins

The emergence of widespread, specialized and highly diverse phage-encoded proteins that thwart CRISPR-Cas immunity, suggests that Acr proteins play an important role in phage biology. The first identified Acr proteins were shown to inactivate the type I-F CRISPR-Cas system of *P. aeruginosa*, halting the host CRISPR machinery upon phage infection[39]. However, finding other Acr proteins by homology searches proved to be a challenging task due to their low sequence similarity. Instead, it was noted that the genomic neighbourhood of *acr* genes had interesting similarities that could be exploited to discover new Acrs. Typically, many *acr* genes co-occur with a group of genes that were collectively called 'anti-CRISPR-associated genes' (*aca's*)[45]. To date, seven *aca* genes have been identified (Table 3). While the function of aca's is not yet understood, these genes often encode for a protein containing a helix-turn-helix (HTH) motif, suggesting they fulfil a regulatory function[45]. Nevertheless, the presence of *aca's* has been instrumental in finding new Acr proteins and vice-versa, a method that is now known as 'guilt-by-association' (Fig. 1a). In addition, the occurrence of a so-called 'self-targeting' spacer (i.e. a spacer that targets the host's own genome) within the CRISPR array is often indicative of a suppressed CRISPR-Cas system due to the presence of a (prophage encoded) *acr* gene[46] (Fig. 1b). Furthermore, novel Acr proteins can be found using (high-throughput) screening and testing assays, including transformation of metagenomic libraries in an Acr-selection strain (Fig. 1c) either combined or not with synthetic genetic circuit-based selection for CRISPR-Cas suppression activity[47].

The high diversity of the Acr proteins, their ability to inhibit different (sub)types of CRISPR-Cas systems (I-C, I-D, I-E, I-F, II-A, II-C, V-A, VI-B)[44,48,49], their widespread presence and their usual coexistence in the same locus, demonstrate the strong evolutionary pressure that CRISPR-Cas systems exert in Acr arsenal diversification, and vice versa, meeting the Red Queen Hypothesis on the continuous shaping of the host-invader dynamics[50,51].

The origin of Acr proteins remains to be understood, but it is hypothesized that these proteins do not share a common ancestor due to their low structural similarity[45]. While it is theorized they represent a product of *de novo* evolution from intergenic regions[52,53], parallel studies show that they might have derived from other bacterial or viral proteins, as specific nuclease inhibitors, regulatory or even phage capsid proteins[53,54]. Due to their function, *acr* genes were classified as accessory, or 'morons', since they are not strictly necessary in a phage lifecycle[55-57]. However, when facing specific CRISPR-active hosts, the

presence of these genes was shown to increase the fitness of Acr-positive phage populations[39,45]. While previous studies show the activity of CRISPR-Cas systems *in vivo* can clear a targeted phage in as little as 2 min[23,58], the presence of Acr proteins seems to decrease or completely abolish bacterial immunity, classifying it as a major CRISPR-counteracting mechanism for successful phage infection and replication. However, the fast-acting nature of CRISPR-Cas limits the potential of a single phage to overcome the host's defense by Acr activity. Recently, it was shown that even though CRISPR-Cas systems are partially affected by the expression of Acr proteins, the latter are not able to confer full protection to their phage associated genome upon a single infection[58,59]. Instead, a critical Acr concentration inside each single cell is necessary for successful host immunosuppression, allowing posterior lytic re-infection or genomic integration of temperate phages[58,59]. It was then demonstrated that a single clonal phage population could inactivate CRISPR-Cas immunity through phage cooperation, where failed infections from 'sacrificial Acr donors' allow accumulation of Acr inhibitors inside a cell, which, upon a certain threshold, leads the 'acceptor' phages to successfully infect and amplify[58,59]. The quantitative demand of Acr proteins for full host immunosuppression postulated that phage concentration has a key role on CRISPR evasion, which is inversely proportional to the strength of each Acr protein[58,59].

The existence of Acr proteins might also explain the incomplete, absent or deficient CRISPR-Cas systems found in bacteria[53]. Prophage integration into the host chromosome and consistent Acr expression might result in CRISPR-Cas inactivating mutations, loss of *cas* genes and even complete loss of CRISPR-Cas systems[53]. Interestingly, although it is evident that Acr proteins are relevant for the phage it originated from, bacteria might also benefit from the stable expression of these proteins (e.g. from prophage regions). For example, inhibition of CRISPR-Cas immunity might enhance HGT in these hosts, which can have a positive contribution to bacterial fitness upon acquisition of beneficial foreign genetic material[39,40,53,58,60,].

Table 3. Anti-CRISPR-associated (*aca*) genes used in the guilt-by-association approach.

Name	Size (aa)	Accession number	References
Aca1	79	YP_007392343	39
Aca2	125	WP_019933869.1	45
Aca3	70	WP_049360086.1	45
Aca4	67	ABR13385.1	44
Aca5	60	WP_039494319.1	44
Aca6	65	WP_035450933.1	44
Aca7	68	WP_064702654.1	44

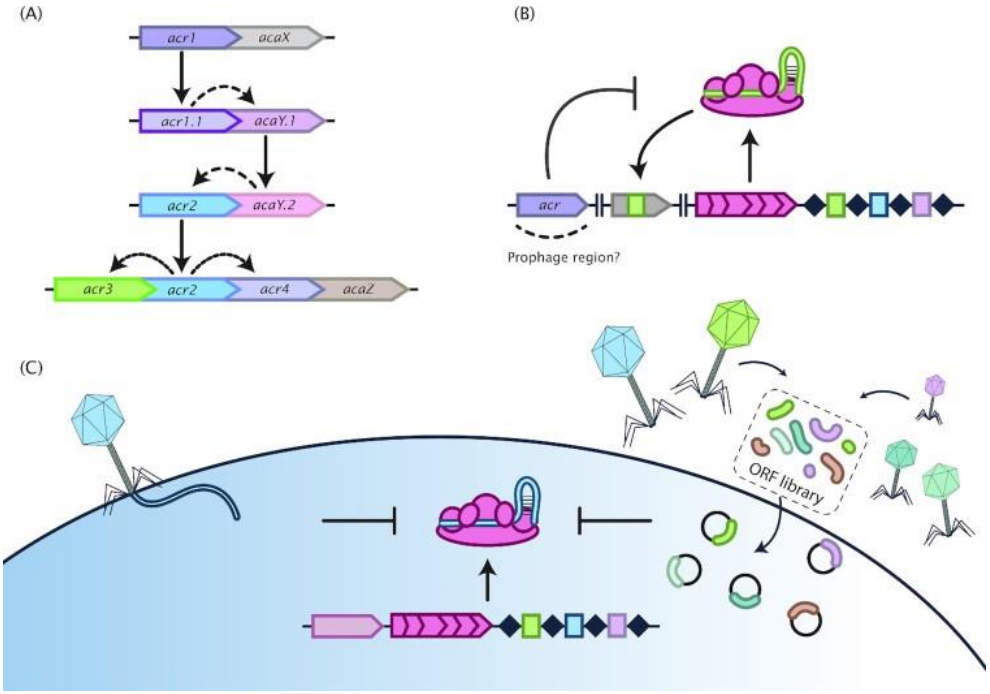


Fig. 1 Different discovery and testing methods of Acrs. (**a**) Guilt–by–association discovery method[45]. This discovery method is based on the strong co–occurrence and clustering of *acr* and *aca* genes through proximity and homology searches. In this example, homology searches using the *acr1* gene yields its homologue acr1.1. Inspection of genes in close proximity yielded acaY.1, which in turn can be used for further iterative rounds of *acr* and/or *aca* gene discovery. Both *acr* and *aca* genes typically appear in clusters leading to the discovery of new *acr* and *aca* genes. (**b**) The self–targeting discovery method[46]. The presence of a self–targeting spacer (in green) within the CRISPR array hints at the presence of a (set of) *acr* gene(s) (in purple) somewhere within the host's genome, often within prophage regions. (**c**) Low– and high–throughput functional assays to identify phage–encoded Acrs. In a low–throughput assay, individual phages are used to screen for anti–CRISPR activity in hosts with a CRISPR–Cas system targeting the phage (left) using (for example) plaque assays. High–throughput screening can be performed by transforming phage ORF libraries that are placed on a plasmid containing a protospacer. Successful transformants can be screened further to pinpoint the gene with the Acr activity within the collection of ORFs.

Mechanisms of anti–CRISPR proteins

Over the last six years, a series of studies interlacing genetic, biochemical and structural analyses have elucidated the mechanism of action of 12 Acr proteins from different families. Although many Acrs remain to be tested for anti-adaptation activity, the vast majority of the currently characterized Acr proteins act at the interference phase by directly blocking target DNA binding or cleavage (Fig. 2). These two general modes-of-action are spread among both class 1 and class 2 Acr proteins, with distinct molecular mechanisms[57].

Class 1 Anti–CRISPR Proteins

The class 1 Acr proteins studied up till now all impede type I (subtype C, D, E or F) CRISPR-Cas systems (Table 1). Two mechanistic routes have been described for Acrs to perturb CRISPR interference: the most common is the direct interaction with the Cascade surveillance complex to prevent DNA binding[33,41,61-64], while the less common involves the direct interaction with the effector nuclease Cas3, which typically gets recruited upon successful target binding by the Cascade, to block DNA cleavage[33,40,42,45,65,66].

A. Preventing DNA Binding via Interaction with the Cascade Complex

Steric occlusion of DNA binding

AcrIF1 from P. aeruginosa phage JBD30 binds along the hexameric Cas7f spine with a stoichiometry of 2–3 molecules per Cascade complex[33]. Several high resolution cryo–electron microscopy and nuclear magnetic resonance (NMR) studies combined with site-directed mutagenesis indicated that AcrIF1 molecules bind tightly at different positions of the *P. aeruginosa* Cascade complex. More specifically, two AcrIF1 monomers sit on the Cas7f.4 and Cas7f.6 thumbs (Tyr6, Tyr20 and Glu31 lying on a single interaction surface of each monomer interact with the conserved Lsy85 on the Cas7f thumb), resulting in a conformational change that sterically blocks access of the crRNA guide to the target DNA, while a possible third monomer binds to a Cas7 region in close proximity to the Cas8f-Cas5f tail, which is crucial for target DNA binding[33,41,62-64,67].

DNA mimicry

While AcrIF1 exploits three different binding modes to disrupt target DNA recognition by the Cascade complex, AcrIF2 from *P. aeruginosa* phage D3112 mediates inhibition by interacting with a single site within the complex. AcrIF2 directly competes with the target DNA for a positively charged binding interface on the Cas5f:Cas8f tail between the Cas7f.6 thumb and the Cas8f hook, a region called the 'lysine-rich vise'[33]. The small acidic AcrIF2 protein behaves as a DNA mimic, as the numerous acidic residues on its surface adopt a pseudo-helical distribution, resembling a double-stranded DNA (dsDNA) molecule. The interaction sites of AcrIF2 and DNA on the Cas5f:Cas8f heterodimer overlap partially, and AcrIF2 binding shoves the Cas8f hook away from the DNA-association pocket, sterically hampering the access of the dsDNA to the Cascade complex. Additional interactions of AcrIF2 with basic residues crucial for DNA binding further ensure obstruction of target DNA binding[61-64]. Given the close proximity of the interaction sites of AcrIF2 and the third AcrIF1 monomer, when cooperating, AcrIF1 exhibits a maximum of two binding modes while AcrIF2 activity remains intact[64].

Similar to AcrIF2, AcrIF10 from *Shewanella xiamenensis* prophage also mimics DNA by occupying a region on the Cas5f:Cas8f heterodimer that closely overlaps with the binding site of AcrIF2 (possibly Cas8f K71 and R78, Cas5f R90 and Cas7f K299). However, instead of wrenching the Cas8f hook away, AcrIF10 triggers a partially closed state of the hook swinging it toward Cas7.6f (similar but smaller than the movement caused by DNA binding), displaying the conformational flexibility of this domain and implying the need of additional interactions for absolute closure of the hook. Intriguingly, AcrIF10 and dsDNA display different charge profiles on the interaction surface; nevertheless, they interact with closely overlapping regions in the Cascade complex to prevent DNA binding[63].

The first archaeal Acr protein identified, AcrID1, was shown to directly interact as a homodimer with two copies of the large subunit (Cas10d) of the type I-D Cascade complex in *Sulfolobus islandicus*. The strong negatively charged surface of this protein suggests that it may behave as a DNA mimic, such as AcrIF2. In addition, conserved residues on the surface of AcrID1, such as Glu21, Lys34, Tyr55, Glu81, Arg92 and Trp91, may have a key role in inter-protein interactions. However, the exact mechanism of AcrID1 still remains to be explored[49].

Unknown mechanisms

AcrIE3 and AcrIF4 from P. aeruginosa phage DMS3 and JBD26, respectively, have been shown to associate with the Cascade complex to hinder DNA binding, though via an unknown mechanism[33,40].

B. Preventing DNA Cleavage via Interaction with the Cas3 Nuclease

Disruption of binding to the Cascade:dsDNA chimera

Cryo-electron microscopy and X-ray crystallography demonstrated that the AcrIF3 protein from *P. aeruginosa* phage JBD5 forms a homodimer that binds to the Cas3 nuclease[42,65]. High binding affinity was observed, since more than half of the AcrIF3 surface interacts with the Cas3 protein, forming several hydrogen bonds and hydrophobic interactions. As a consequence, the interaction sites for both the non-complementary DNA strand and the Cascade complex are blocked. Specifically, one AcrIF3 monomer occupies the helicase domain (HD) and the Linker region of Cas3, while the other monomer relates to the C-terminal domain (CTD) (Tyr97, Trp93, and a large network of hydrogen bonds), which altogether constitute the internal cleft of the Cas3 structure. By covering this cleft, AcrIF3 disrupts association with the target DNA (non-complementary strand) and locks the ATP-dependent Cas3 nuclease in an inactive ADP-bound form[42,65]. Noteworthy, the Cas3 effector nuclease/helicase is fused to the Cas2 protein in type I-F systems and thereby forms an integral part of the type I-F (primed) adaptation machinery, hinting at the functional link between adaptation and interference, as shown recently[68-70]. Interestingly, AcrIF3 dimer binds to the opposite site of Cas2, thus not influencing the assembly of the Cas1-Cas2-Cas3 complex[71]. However, the dimer obstructs the recruitment of the Cascade:dsDNA chimera to Cas3, preventing the generation of precursor protospacer DNA. Consequently, AcrIF3 blocks both primed spacer acquisition and crRNA interference[65,72].

Unknown mechanisms

Akin to AcrIF3, AcrIE1 from *P. aeruginosa* phage JBD5 directly associates with the ATP-dependent Cas3 nuclease, without affecting the binding ability of the Cascade to the target DNA. Due to the structural homology between Cas3 proteins of type I-F and I-E

CRISPR-Cas systems, it is likely that AcrIF3 and AcrIE1 either adopt distinct modes of binding to the same surface or target unrelated regions on the Cas3 protein, hindering target DNA cleavage[40,66].

Class 2 Anti−CRISPR Proteins

Class 2 Acr proteins have been discovered for type II (subtype A and C), type V (subtype A) and type VI (subtype B) CRISPR-Cas systems (Table 2). Almost all type II Acrs characterized to date directly interact with the Cas9 endonuclease, although by distinct mechanisms, as described below.

A. Preventing DNA Binding via Interaction with the Cas9 Protein

DNA mimicry and steric occlusion of DNA binding and cleavage

Both AcrIIA2 and AcrIIA4 from *Listeria monocytogenes* prophage J0161a/b have been demonstrated *in vivo* and *in vitro* to directly interact with single-guide RNA (sgRNA) - oaded SpyCas9, abolishing DNA binding and cleavage[46,73-75]. Biochemical and structural studies have indicated that AcrIIA4 binds to several regions within SpyCas9. First, AcrIIA4 (Asp14, Asp37, Glu40, Asp69 and Glu70) associates with the PAM-interacting domain (Glu1108, Ser1109, Ser1216, Lys1200, Arg1335 and Arg1333) through an acidic surface that mimics a negatively-charged dsDNA molecule, thereby hampering the initial PAM searching stage. Second, AcrIIA4 (Asp14 and Asn36) interacts with the Topo-homology domain (Glu1108, Ser1109 and Ser1136), also known as DNA-melting region, putatively preventing DNA binding or unwinding. Third, AcrIIA4 (Leu19-Gln29) forms numerous surface complementarities with the concave surface of SpyCas9 at the RuvC domain (Asn767, Thr13, Ala764 and Arg976), abrogating the endonuclease activity. In addition, AcrIIA4 binds the linker between RuvC and HNH domains, sterically blocking conformational changes necessary for DNA cleavage[73,74,76]. Similar to AcrIIA4, AcrIIA2 prevents target DNA recognition, binding and cleavage. Specifically, AcrIIA2 (Asp71 and Glu72) associates with the PAM-interacting domain (Arg1335 and Arg1333), the WED domain (Lys1107, Glu1108, Ser1109 and Ser1136 of WED domain interact with His37, Asp38, Glu93 and Asp96 of AcrIIA2), the HNH domain (Gln774 and Arg778 interact with Asn19 of AcrIIA2) and the REC2 domain (Lys268 and Asp269 interact with Gln7, Thr28 and Asp30 of AcrIIA2)[77,78]. Notably, both AcrIIA4 and AcrIIA2 are not able to bind to SpyCas9

in the absence of a preloaded sgRNA, as sgRNA-binding is required for the formation of the Acr-interaction surface of SpyCas9[73,74,76-78].

DNA and sgRNA mimicry

Although AcrIIC2 from Neisseria meningitidis prophage was previously speculated to associate with catalytic residues of the NmeCas9 HNH domain[43,79], a recent biochemical and structural study revealed interaction with the NmeCas9 bridge helix (BH)-REC1 domain (residues 51-241)[80]. Indeed, AcrIIC2 forms an homodimer, forming an acidic groove on top of the dimer. The residues in this groove (E18, N23/D24/E25 and residues 109-124) strongly bind to the arginine-rich α helix of BH (residues 56-79; mainly R62, R63, R70 and R73). As such, the AcrIIC2 dimer significantly impedes sgRNA loading to apoNmeCas9, and abrogates dsDNA binding to the sgRNA-NmeCas9-AcrIIC2 complex. Remarkably, AcrIIC2 requires the apo form for effective inhibition, being the first Acr reported to directly interfere with the sgRNA loading to Cas9[80].

Dimerization of Cas9

AcrIIC3 from *N. meningitidis* prophage hinders DNA binding and induces NmeCas9 dimerization by associating with the HNH domain and the REC lobe, respectively[80]. AcrIIC3 (K532-Y540, R557-H563) interacts with a non-conserved region of the NmeCas9 HNH domain opposite to the active site (L58, N60, R33, V34 and D38 among others)[80], allowing PAM detection but hampering complete R-loop formation[43]. Hence, the binding affinity for target DNA is decreased, while DNA cleavage is abrogated. AcrIIC3 additionally associates with the REC lobe, triggering NmeCas9 dimerization with a 2:2 stoichiometry. Each AcrIIC3 molecule binds to the HNH domain as well as the REC lobe of the same or another NmeCas9 molecule, forcing AcrIIC3-Cas9 dimerization and preventing target DNA loading[80].

Unknown mechanisms

Recently, AcrIIC4 and AcrIIC5 were discovered in a *Haemophilus parainfluenzae* prophage and a *Simonsiella muelleri* transfer element. Both were shown to impede the Cas9:sgRNA complex from binding to the target DNA, following an unknown mode-of-action[81].

B. Preventing DNA Cleavage via Interaction with the Cas9 Protein at the Catalytic Site

Like the type I AcrIF3, AcrIIC1 from N. meningitidis MGE allows binding of the CRISPR interference complex to the target DNA, though hampering DNA cleavage. Biochemical and structural characterization have revealed that AcrIIC1 specifically binds to the active site of the NmeCas9 HNH domain (D587, H588), blocking cleavage of the target strand and preventing conformational changes necessary for the activation of the RuvC domain, which would theoretically catalyze cleavage of the non-target strand. Thus, the sgRNA-loaded Cas9 remains bound to the target DNA, being trapped in a catalytically inactive state. To achieve high stability of the inter-protein interaction, AcrIIC1 additionally associates with five neighboring residues of the HNH domain (K549, K551, D598, K603, N616). Similar to AcrIIC1, AcrIIC2 associates with catalytic residues of the NmeCas9 HNH domain, albeit the exact mode-of-action is still elusive[79,43].

C. Preventing CRISPR–Cas Immunity via Putative Binding to RNA or DNA Molecules

Similar to AcrID1, AcrIIA1 from *L. monocytogenes* prophage J0161a forms a homodimer, though with an unusual two helical domain structure. The N-terminal domain resembles the HTH motif of transcriptional factors, whilst the CTD adopts an architecture of unknown function. It is anticipated that AcrIIA1 recognizes and associates with heterogeneous RNA molecules to abolish CRISPR-Cas immunity. However, no binding to CRISPR RNA (crRNA), trans-activating RNA (tracrRNA) or their duplex has been observed. AcrIIA1 harbors a positively charged surface around the HTH region, resembling nucleic acid binding motifs of many transcriptional factors that are crucial for RNA and dsDNA recognition. Hence, it would be possible that AcrIIA1 binds to the promoter regions of crRNA or tracrRNA to hinder CRISPR-Cas immunity. However, Cas9 expression levels

appeared to be unaffected by AcrIIA1[82]. The unique structure and function of AcrIIA1 reveals a novel mechanism of action yet unknown among Acr proteins, strengthening our understanding about the versatile and sophisticated ways in which these small proteins may hamper CRISPR-Cas systems[46,82].

D. Preventing RNA Binding or Cleavage via Interaction with the Cas13b Protein

Although not associated to a phage genome, through the computational pipeline that guided the discovery of subtype VI-B CRISPR-Cas loci, the accessory protein Csx27 was recently found to repress the interference stages of its associated CRISPR-Cas system[48]. Experimental testing of *Bergeyella zoohelcum* Csx27 (201 aa) has demonstrated an inhibitory effect of the protein when expressed together with different Cas13b proteins, weakening their RNA interference activity. Even though important details on the mechanism of Csx27 remain to be identified, the current findings suggest a broad activity of the protein among type VI-B loci, relating to a possible regulatory mechanism of phage interference[48].

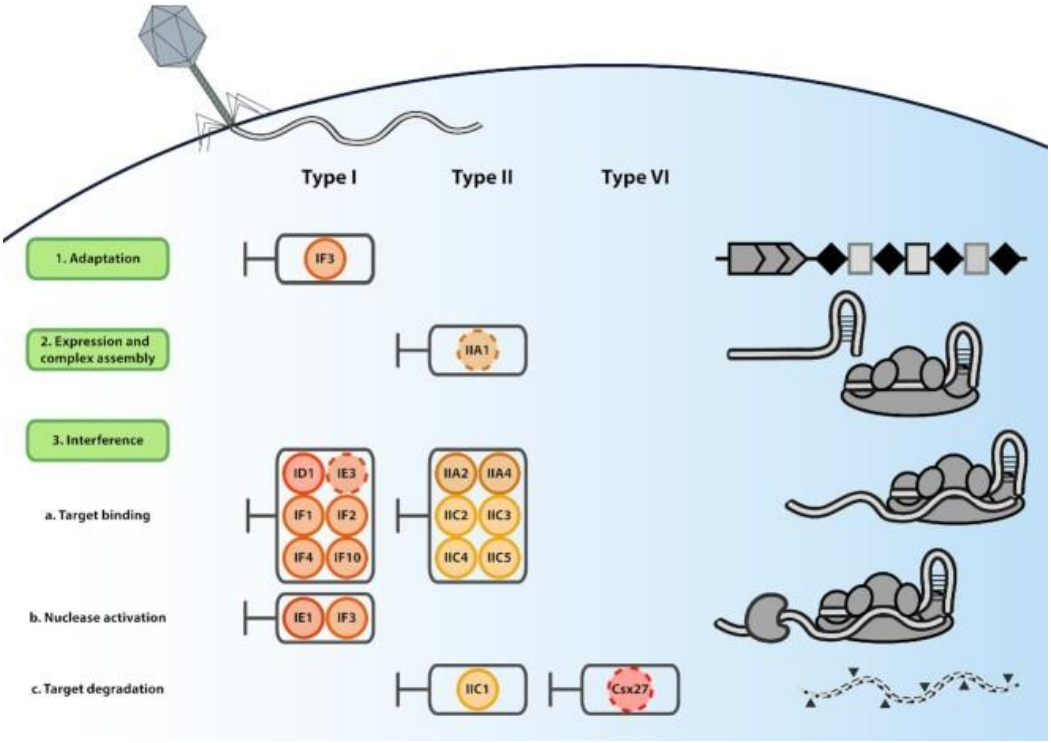


Fig. 2 Schematic overview of the different Acrs and their mechanisms. The green boxes on the left show the different stages of CRISPR-Cas immunity. The columns indicate which CRISPR-Cas type is suppressed by which (group of) Acrs. Acrs are depicted as circles with their abbreviated names (e.g. AcrIF3 is abbreviated to IF3). A dashed line indicates a suggested role for the particular Acr or that the Acr mechanism remains to be elucidated. Note that most Acrs appear to suppress the interference stage, whereas only one Acr (AcrIF3) suppressed different stages.

Applications of anti-CRISPR proteins

CRISPR-Cas systems have recently been scrutinized for their potential in biotechnological applications. Type II CRISPR-Cas9 systems raise special interest among the scientific community, due to their programmability and specific nuclease activity, representing the most promising tool on genome editing and modulation studied to date[83]. The use of catalytically inactive Cas9 ('dead', dCas9) was also previously shown to have powerful biotechnological applications. These include CRISPR interference and activation (CRISPRi and CRISPRa), modification of epigenetic marks and gene expression modulation when fused dCas9 to other metabolic key enzymes[84]. Other applications encompass dynamic genomic imaging, identification of specific genes or genomic loci, monitoring of gene copy and follow-up of chromatin formation and telomere elongation, when combining site-specific sgRNA molecules with fluorescent tagging of dCas9 (EGFT-dCas9)[46,85]. Recently, a Cas9-Assisted Targeting of Chromosome segments (CATCH) was also used for nanopore sequencing of a breast cancer gene[67]. Taken together, the quick emergence of CRISPR-based technologies and the continuous quest for finding new CRISPR-Cas nucleases and variants hereof indicates that more (advanced) applications are expected to be developed in the near future. However, full specificity is crucial in these applications, especially those with therapeutic purposes, as off-target events are still one of the main bottlenecks that limit the efficiency of this technology[86].

Deeper insight in Acr inhibitory mechanisms might soon allow for precise temporal, spatial and conditional control of CRISPR-Cas systems through an 'on-off switch' regulation. Several AcrIIA and AcrIIC proteins were found to work as 'off-switches' of Cas9 activity in human cell lines[46,79]. The potential of combining these natural CRISPR inhibitors with Cas9 editing systems, tuning its activity in cellular environments, could result in full optimization of gene editing processes, substantially decreasing off-target events by allowing Acr proteins to accumulate whenever or wherever editing activity is unwanted[46,76,79,87]. Innovation in the use of Acrs to control CRISPR-Cas editing are expected to quickly emerge, as demonstrated by the combination of an AcrIIA4 hybrid with a LOV2 photosensor for the light-mediated control of genome and epigenome editing by CRISPR-Cas9 effectors in human cells[88]. Acr-mediated inhibition was also proven to be effective towards the activity of dCas9 processes[46,85,89] as well as the control of genomic circuits and gene editing with Cas9[89], once again representing a promising tool for optimization of the activity of these CRISPR-based technologies and future therapeutic and biotechnological applications.

Although no research was performed yet on the possible applications of Csx27, this represents the only protein known to date to repress Cas13b, a type VI CRISPR associated protein. In contrast to most other CRISPR-Cas systems currently classified, prokaryotes carrying type VI CRISPR loci are able to target foreign RNA molecules[48]. The discovery and employment of Acr proteins able to inhibit these systems might allow modulation of future RNA-based biotechnology applications.

Acr proteins might also benefit gene drive technology. Gene drives are powerful tools to eradicate vector-borne diseases, eliminate pests (e.g. agricultural pests and invasive species) and even increase animal welfare. The technology enables the rapid dissemination of genetic mutation(s) through a population by surpassing Mendelian inheritance rules regardless of the fitness-affecting properties of the introduced mutation[90]. This is accomplished by turning heterozygous organisms into homozygotes through the incorporation of the desired mutated gene flanked by a homing endonuclease (such as CRISPR-Cas) and a corresponding guide RNA (targeting the wild-type gene) in the genome. Upon recognition of the target sequence (i.e. the wild-type gene) on the homologous chromosome, the homing endonuclease will introduce a break at the target site. If this event is followed by homologous recombination, a cassette consisting of the mutated gene flanked by the homing nuclease and the guide RNA will replace the wildtype locus. However, one of the main technological hurdles to overcome is the current inability to effectively control the spreading of a gene drive once out in the environment[91,92]. Acrs can be used as a molecular tool to control this spread. These proteins have been shown to reduce the efficiency of the homing nuclease in a tweakable manner[75,93]. Acrs are also envisioned to enable timed drive activation and to aid anti-gene drives in destroying the original gene drive construct (immunization)[75]. The latter can be achieved by introducing an *acr*-encoding gene instead of a mutated gene, as part of a gene drive. This gene drive should make use of a homing endonuclease which is not suppressed by the respective *acr* gene. Though gene drive efficiency can be partially controlled via sgRNA design[93-95], incorporation of Acrs in the molecular design to control spreading is advantageous over regulation via sgRNA design since Acrs work directly against the homing endonuclease whereas sgRNA based molecular principles are case specific.

Finally, Acr proteins might represent a powerful tool to enable phage therapy in CRISPR-active hosts. The emergence of multidrug resistant bacteria represents a rising scientific concern due to the possible implications of antibiotic unresponsive infections. Phage therapy poses an interesting alternative to the control of these bacterial infections[96], and Acr-mediated inhibition of active CRISPR-Cas systems might facilitate the employment of

known phages that would be otherwise targeted by the bacterial immune system. The possibility of using studied phages, instead of the constant search and characterization of new ones not yet targeted by the bacterial CRISPR-Cas system, might represent a therapeutic advantage and lead to faster treatment.

Outlook

A limited number of Acr proteins has been identified so far, but their sequence and structural diversity is already remarkable. The discovery of new Acrs, especially those targeting CRISPR-Cas (sub)types for which Acr proteins have not yet been found, is expected to clarify the number of distinct Acr protein families and how widespread they are. The development of new Acr identification strategies will certainly be required to avoid biases created by current pipelines. Clarification of the mechanisms of multiple Acr proteins, including the characterization of AcrVA1, AcrVA4 and AcrVA5[97,98] while our study was under review, is expected to fuel the Acr-based finetuning of CRISPR-Cas applications, such as gene editing or gene drives.

Because bacteria and phages have co-evolved together for billions of years, it is anticipated that bacteria have developed mechanisms to counteract Acr protein activity. Possible strategies have been hinted, including the accumulation of multiple types of CRISPR-Cas systems in a single cell (e.g. type I-E and I-F systems in *P. aeruginosa*), mutation of *cas* genes[99], or silencing of *acr* gene expression. Proper research on the field will certainly increase our understanding on bacterial evolution, and also expand the CRISPR toolbox for biotechnological applications.

Funding

This work was supported by the Netherlands Organization for Scientific Research (NWO) Veni grant [016.Veni.171.047 to R.H.J.S.], Veni grant [016.Veni.181.092 to F.L.N.] and the Alexander S. Onassis Public Benefit Foundation [F ZM 083–2/2018–2019 to D.T.]. The authors declare that they have no conflicts of interest to declare.

Conflicts of interest

None declared.

References

1. Koonin, E. V., & Dolja, V. V. (2013). A virocentric perspective on the evolution of life. *Current opinion in virology*, *3*(5), 546-557.
2. Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356-361.
3. Cobián Güemes, A. G., Youle, M., Cantú, V. A., Felts, B., Nulton, J., & Rohwer, F. (2016). Viruses as winners in the game of life. *Annual review of virology*, *3*, 197-214.
4. Sano, E., Carlson, S., Wegley, L., & Rohwer, F. (2004). Movement of viruses between biomes. *Applied and Environmental Microbiology*, *70*(10), 5842-5846.
5. Gandon, S. (2016). Why be temperate: lessons from bacteriophage λ. *Trends in microbiology*, *24*(5), 356-365.
6. Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, *148*(4), 1667-1686.
7. Forterre, P., & Prangishvili, D. (2009). The great billion-year war between ribosome-and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Annals of the New York Academy of Sciences*, *1178*(1), 65-77.
8. Chopin, M. C., Chopin, A., & Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology*, *8*(4), 473-479.
9. Makarova, K. S., Wolf, Y. I., Snir, S., & Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of bacteriology*, *193*(21), 6039-6056.
10. Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic acids research*, *41*(8), 4360-4377.
11. Samson, J. E., Magadán, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology*, *11*(10), 675-687.
12. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., ... & Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO journal*, *34*(2), 169-183.
13. Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A. I., Mantel, I., ... & Maxwell, K. L. (2018). A chemical defence against phage infection. *Nature*, *564*(7735), 283-286.
14. Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., ... & Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, *359*(6379), eaar4120.
15. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712.
16. Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551-2561.
17. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*(2), 174-182.
18. Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653-663.
19. Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio*, *8*(5), e01397-17.
20. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., ... & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960-964.
21. Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, *322*(5909), 1843-1845.
22. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., ... & Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, *139*(5), 945-956.

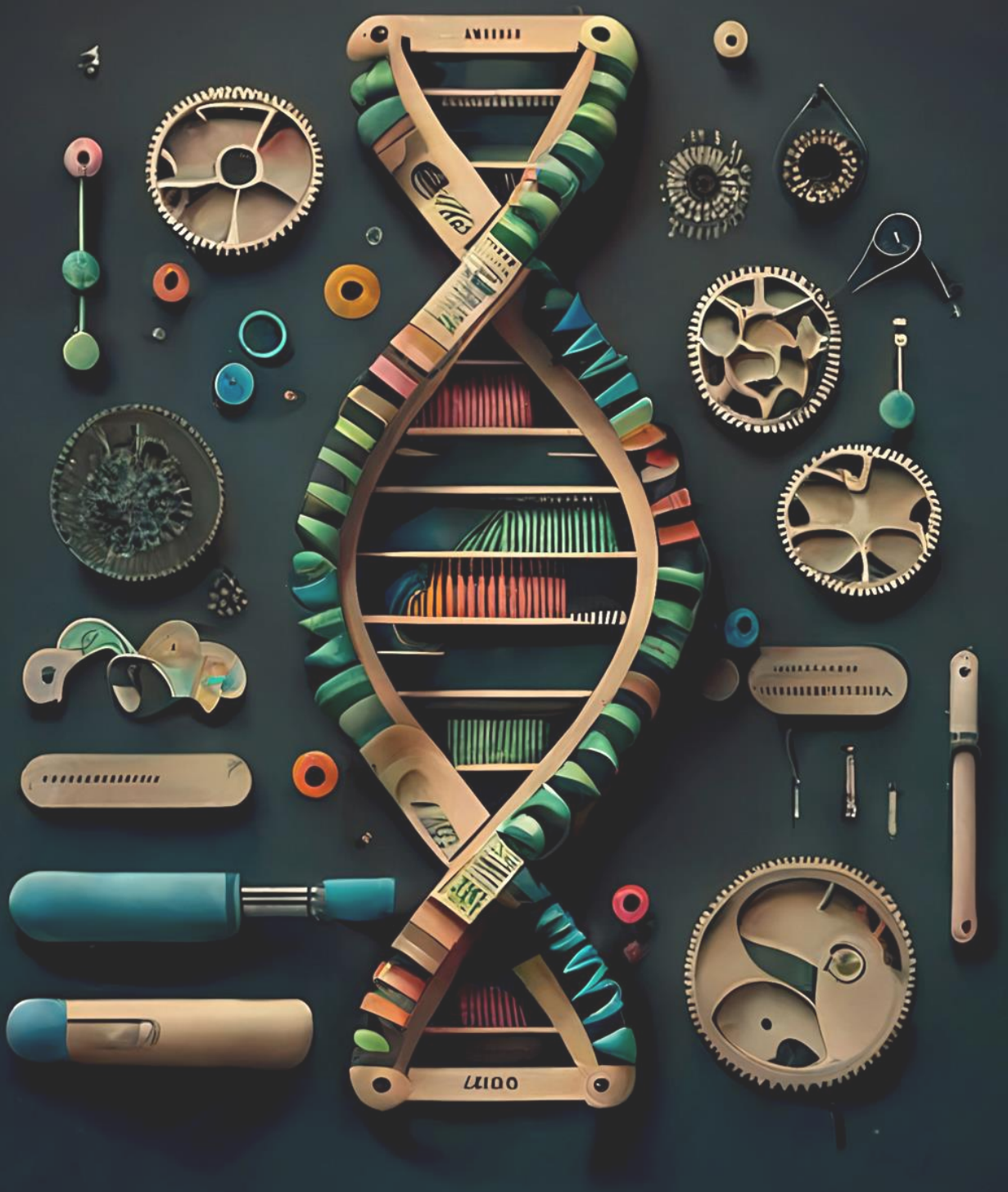
23. Garneau, J. E., Dupuis, M. È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., ... & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67-71.
24. Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology*, *37*, 67-78.
25. Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., & Van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, *353*(6299), aad5147.
26. Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. (2017). CRISPR-Cas: adapting to change. *Science*, *356*(6333), eaal5056.
27. Hale, C., Kleppe, K., Terns, R. M., & Terns, M. P. (2008). Prokaryotic silencing (psi) RNAs in Pyrococcus furiosus. *Rna*, *14*(12), 2572-2579.
28. Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence-and structure-specific RNA processing by a CRISPR endonuclease. *Science*, *329*(5997), 1355-1358.
29. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., ... & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602-607.
30. Westra, E. R., van Erp, P. B., Künne, T., Wong, S. P., Staals, R. H., Seegers, C. L., ... & van der Oost, J. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell*, *46*(5), 595-605.
31. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., ... & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology*, *190*(4), 1390-1400.
32. Sun, C. L., Barrangou, R., Thomas, B. C., Horvath, P., Fremaux, C., & Banfield, J. F. (2013). Phage mutations in response to CRISPR diversification in a bacterial population. *Environmental microbiology*, *15*(2), 463-470.
33. Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M. F., Hidalgo-Reyes, Y., ... & Davidson, A. R. (2015). Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*, *526*(7571), 136-139.
34. Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., ... & Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences*, *108*(25), 10098-10103.
35. Bryson, A. L., Hwang, Y., Sherrill-Mix, S., Wu, G. D., Lewis, J. D., Black, L., ... & Bushman, F. D. (2015). Covalent modification of bacteriophage T4 DNA inhibits CRISPR-Cas9. *MBio*, *6*(3), e00648-15.
36. Vlot, M., Houkes, J., Lochs, S. J. A., Swarts, D. C., Zheng, P., Kunne, T., ... & Brouns, S. J. J. (2018). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR–Cas effector complexes. *Nucleic acids research*, *46*(2), 873-885.
37. Yaung, S. J., Esvelt, K. M., & Church, G. M. (2014). CRISPR/Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. *PloS one*, *9*(6), e98811.
38. Seed, K. D., Lazinski, D. W., Calderwood, S. B., & Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, *494*(7438), 489-491.
39. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, *493*(7432), 429-432.
40. Pawluk, A., Bondy-Denomy, J., Cheung, V. H., Maxwell, K. L., & Davidson, A. R. (2014). A new group of phage anti-CRISPR genes inhibits the type IE CRISPR-Cas system of Pseudomonas aeruginosa. *MBio*, *5*(2), e00896-14.
41. Maxwell, K. L., Garcia, B., Bondy-Denomy, J., Bona, D., Hidalgo-Reyes, Y., & Davidson, A. R. (2016). The solution structure of an anti-CRISPR protein. *Nature communications*, *7*(1), 1-5.
42. Wang, J., Ma, J., Cheng, Z., Meng, X., You, L., Wang, M., ... & Wang, Y. (2016). A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. *Cell research*, *26*(10), 1165-1168.
43. Harrington, L. B., Doxzen, K. W., Ma, E., Liu, J. J., Knott, G. J., Edraki, A., ... & Doudna, J. A. (2017). A broad-spectrum inhibitor of CRISPR-Cas9. *Cell*, *170*(6), 1224-1233.
44. Marino, N. D., Zhang, J. Y., Borges, A. L., Sousa, A. A., Leon, L. M., Rauch, B. J., ... & Bondy-Denomy, J. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, *362*(6411), 240-242.
45. Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., ... & Davidson, A. R. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature microbiology*, *1*(8), 1-6.
46. Rauch, B. J., Silvis, M. R., Hultquist, J. F., Waters, C. S., McGregor, M. J., Krogan, N. J., & Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell*, *168*(1-2), 150-158.
47. Uribe, R. V., van der Helm, E., Misiakou, M. A., Lee, S. W., Kol, S., & Sommer, M. O. (2019). Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. *Cell host & microbe*, *25*(2), 233-241.
48. Smargon, A. A., Cox, D. B., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., ... & Zhang, F. (2017). Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Molecular cell*, *65*(4), 618-630.

49. He, F., Bhoobalan-Chitty, Y., Van, L. B., Kjeldsen, A. L., Dedola, M., Makarova, K. S., ... & Peng, X. (2018). Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype ID immunity. *Nature microbiology*, *3*(4), 461-469.
50. Westra, E. R., van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J. M., Best, A., ... & Buckling, A. (2015). Parasite exposure drives selective evolution of constitutive versus inducible defense. *Current biology*, *25*(8), 1043-1049.
51. van Houte, S., Ekroth, A. K., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., ... & Westra, E. R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature*, *532*(7599), 385-388.
52. Tautz, D. (2014). The discovery of de novo gene evolution. *Perspectives in biology and medicine*, *57*(1), 149-161.
53. Stanley, S. Y., & Maxwell, K. L. (2018). Phage-encoded anti-CRISPR defenses. *Annual review of genetics*, *52*, 445-464.
54. Stone, N. P., Hilbert, B. J., Hidalgo, D., Halloran, K. T., Lee, J., Sontheimer, E. J., & Kelch, B. A. (2018). A hyperthermophilic phage decoration protein suggests common evolutionary origin with herpesvirus triplex proteins and an anti-CRISPR protein. *Structure*, *26*(7), 936-947.
55. Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., & Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *Journal of molecular biology*, *299*(1), 27-51.
56. Brüssow, H., Canchaya, C., & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews*, *68*(3), 560-602.
57. Borges, A. L., Davidson, A. R., & Bondy-Denomy, J. (2017). The discovery, mechanisms, and evolutionary impact of anti-CRISPRs. *Annual review of virology*, *4*(1), 37.
58. Borges, A. L., Zhang, J. Y., Rollins, M. F., Osuna, B. A., Wiedenheft, B., & Bondy-Denomy, J. (2018). Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity. *Cell*, *174*(4), 917-925.
59. Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., ... & van Houte, S. (2018). Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. *Cell*, *174*(4), 908-916.
60. Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R., & Marraffini, L. A. (2013). Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS genetics*, *9*(9), e1003844.
61. Van Erp, P. B., Jackson, R. N., Carter, J., Golden, S. M., Bailey, S., & Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic acids research*, *43*(17), 8381-8391.
62. Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., ... & Wiedenheft, B. (2017). Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell*, *169*(1), 47-57.
63. Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., ... & Subramaniam, S. (2017). Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR-Cas surveillance complex. *Cell*, *171*(2), 414-426.
64. Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., ... & Gao, G. F. (2017). Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. *Cell research*, *27*(7), 853-864.
65. Wang, X., Yao, D., Xu, J. G., Li, A., Xu, J., Fu, P., ... & Zhu, Y. (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nature structural & molecular biology*, *23*(9), 868-870.
66. Pawluk, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T. F., Davidson, A. R., & Maxwell, K. L. (2017). Disabling a type IE CRISPR-Cas nuclease with a bacteriophage-encoded anti-CRISPR protein. *MBio*, *8*(6), e01751-17.
67. Gabrieli, T., Sharim, H., Fridman, D., Arbib, N., Michaeli, Y., & Ebenstein, Y. (2018). Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic acids research*, *46*(14), e87-e87.
68. Künne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I., Miellet, W. R., Klein, M., ... & Brouns, S. J. (2016). Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Molecular cell*, *63*(5), 852-864.
69. Staals, R. H., Jackson, S. A., Biswas, A., Brouns, S. J., Brown, C. M., & Fineran, P. C. (2016). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. *Nature communications*, *7*(1), 1-13.
70. Fagerlund, R. D., Wilkinson, M. E., Klykov, O., Barendregt, A., Pearce, F. G., Kieper, S. N., ... & Fineran, P. C. (2017). Spacer capture and integration by a type IF Cas1–Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences*, *114*(26), E5122-E5128.
71. Rollins, M. F., Chowdhury, S., Carter, J., Golden, S. M., Wilkinson, R. A., Bondy-Denomy, J., ... & Wiedenheft, B. (2017). Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proceedings of the National Academy of Sciences*, *114*(26), E5113-E5121.
72. Vorontsova, D., Datsenko, K. A., Medvedeva, S., Bondy-Denomy, J., Savitskaya, E. E., Pougach, K., ... & Semenova, E. (2015). Foreign DNA acquisition by the IF CRISPR–Cas system requires all components of the interference machinery. *Nucleic acids research*, *43*(22), 10848-10860.
73. Dong, D., Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., ... & Huang, Z. (2017). Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. *Nature*, *546*(7658), 436-439.

74. Yang, H., & Patel, D. J. (2017). Inhibition mechanism of an anti-CRISPR suppressor AcrIIA4 targeting SpyCas9. *Molecular cell*, *67*(1), 117-127.
75. Basgall, E. M., Goetting, S. C., Goeckel, M. E., Giersch, R. M., Roggenkamp, E., Schrock, M. N., ... & Finnigan, G. C. (2018). Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology*, *164*(4), 464.
76. Shin, J., Jiang, F., Liu, J. J., Bray, N. L., Rauch, B. J., Baik, S. H., ... & Doudna, J. A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. *Science advances*, *3*(7), e1701620.
77. Jiang, F., Liu, J. J., Osuna, B. A., Xu, M., Berry, J. D., Rauch, B. J., ... & Doudna, J. A. (2019). Temperature-responsive competitive inhibition of CRISPR-Cas9. *Molecular cell*, *73*(3), 601-610.
78. Liu, L., Yin, M., Wang, M., & Wang, Y. (2019). Phage AcrIIA2 DNA mimicry: structural basis of the CRISPR and anti-CRISPR arms race. *Molecular cell*, *73*(3), 611-620.
79. Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., ... & Davidson, A. R. (2016). Naturally occurring off-switches for CRISPR-Cas9. *Cell*, *167*(7), 1829-1838.
80. Zhu, Y., Gao, A., Zhan, Q., Wang, Y., Feng, H., Liu, S., ... & Gao, P. (2019). Diverse mechanisms of CRISPR-Cas9 inhibition by type IIC anti-CRISPR proteins. *Molecular cell*, *74*(2), 296-309.
81. Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H. E., ... & Sontheimer, E. J. (2018). Potent Cas9 inhibition in bacterial and human cells by AcrIIC4 and AcrIIC5 anti-CRISPR proteins. *MBio*, *9*(6), e02321-18.
82. Ka, D., Jang, D. M., Han, B. W., & Bae, E. (2018). Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. *Nucleic acids research*, *46*(18), 9805-9815.
83. Komor, A. C., Badran, A. H., & Liu, D. R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, *168*(1-2), 20-36.
84. Hilton, I. B., D'ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology*, *33*(5), 510-517.
85. Liu, X. S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., ... & Jaenisch, R. (2018). Rescue of fragile X syndrome neurons by DNA methylation editing of the FMR1 gene. *Cell*, *172*(5), 979-992.
86. Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids*, *4*, e264.
87. Hynes, A. P., Rousseau, G. M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., ... & Moineau, S. (2018). Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. *Nature communications*, *9*(1), 1-10.
88. Bubeck, F., Hoffmann, M. D., Harteveld, Z., Aschenbrenner, S., Bietz, A., Waldhauer, M. C., ... & Niopek, D. (2018). Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9. *Nature methods*, *15*(11), 924-927.
89. Nakamura, M., Srinivasan, P., Chavez, M., Carter, M. A., Dominguez, A. A., La Russa, M., ... & Qi, L. S. (2019). Anti-CRISPR-mediated control of gene editing and synthetic circuits in eukaryotic cells. *Nature communications*, *10*(1), 1-11.
90. Burt, A. (2003). Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proceedings of the Royal Society of London. *Series B: Biological Sciences*, *270*(1518), 921-928.
91. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature biotechnology*, *33*(12), 1250-1255.
92. Webber, B. L., Raghu, S., & Edwards, O. R. (2015). Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat?. *Proceedings of the National Academy of Sciences*, *112*(34), 10565-10567.
93. Roggenkamp, E., Giersch, R. M., Schrock, M. N., Turnquist, E., Halloran, M., & Finnigan, G. C. (2018). Tuning CRISPR-Cas9 gene drives in Saccharomyces cerevisiae. *G3: Genes, Genomes, Genetics*, *8*(3), 999-1018.
94. Noble, C., Olejarz, J., Esvelt, K. M., Church, G. M., & Nowak, M. A. (2017). Evolutionary dynamics of CRISPR gene drives. *Science advances*, *3*(4), e1601964.
95. Yan, Y., & Finnigan, G. C. (2018). Development of a multi-locus CRISPR gene drive system in budding yeast. *Scientific reports*, *8*(1), 1-12.
96. Nobrega, F. L., Costa, A. R., Kluskens, L. D., & Azeredo, J. (2015). Revisiting phage therapy: new applications for old resources. *Trends in microbiology*, *23*(4), 185-191.
97. Dong, L., Guan, X., Li, N., Zhang, F., Zhu, Y., Ren, K., ... & Huang, Z. (2019). An anti-CRISPR protein disables type V Cas12a by acetylation. *Nature Structural & Molecular Biology*, *26*(4), 308-314.
98. Knott, G. J., Thornton, B. W., Lobba, M. J., Liu, J. J., Al-Shayeb, B., Watters, K. E., & Doudna, J. A. (2019). Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nature structural & molecular biology*, *26*(4), 315-321.
99. Pausch, P., Müller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural variation of type IF CRISPR RNA guided DNA surveillance. *Molecular cell*, *67*(4), 622-632.

100. Kim, I., Jeong, M., Ka, D., Han, M., Kim, N. K., Bae, E., & Suh, J. Y. (2018). Solution structure and dynamics of anti-CRISPR AcrIIA4, the Cas9 inhibitor. *Scientific reports*, *8*(1), 1-9.

101. Watters, K. E., Fellmann, C., Bai, H. B., Ren, S. M., & Doudna, J. A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science*, *362*(6411), 236-239.



Chapter 5

In vivo characterization of the AcrIIC1 anti–CRISPR protein for Cas9–based genome engineering

Despoina Trasanidou[1], Ana Potocnik[1*], Patrick Barendse[1*], Prarthana Mohanraju[1], Evgenios Bouzetos[1], Efthymios Karpouzis[1], Amber Desmet[1], Richard van Kranenburg[1,2], John van der Oost[1], Raymond H.J. Staals[1**] & Ioannis Mougiakos[1,3**]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

[2]Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands.

[3]SNIPR Biome, Lersø Parkallé 44, 2100 Copenhagen, Denmark.

*Equal contributions

**Corresponding authors

Manuscript under revision (Nature Communications)

Abstract

Anti-CRISPR proteins (Acrs) block the activity of CRISPR-associated (Cas) proteins, either by inhibiting DNA interference or by preventing crRNA loading and complex formation. Although the main use of Acrs in genome engineering applications is to lower the cleavage activity of Cas proteins, they can also be instrumental for various other CRISPR-based applications. Here, we explore the genome editing potential of the thermoactive type II-C Cas9 variants from *Geobacillus thermodenitrificans* T12 (ThermoCas9) and *Geobacillus stearothermophilus* (GeoCas9) in *Escherichia coli*. We then demonstrate that the anti-CRISPR protein AcrIIC1 from *Neisseria meningitidis* robustly inhibits their DNA cleavage activity, but not their DNA binding capacity. Finally, we exploit these AcrIIC1:Cas9 complexes for gene silencing and base-editing, developing the first reported Acr base-editing tools. With these tools we pave the way for future engineering applications in mesophilic and thermophilic bacteria combining the activities of Acr and CRISPR-Cas proteins.

Keywords

ThermoCas9, GeoCas9, thermostable, genome editing, base editing

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins are part of prokaryotic adaptive immune systems. Several Cas nucleases have been repurposed as useful tools for genome editing and transcriptional control[1,2]. CRISPR-Cas systems are divided into 2 classes, 6 types, and >30 subtypes based on their signature Cas genes. Class 2 systems have been widely used as genetic engineering tools, due to their streamlined architecture of a single effector protein (i.e. Cas9, Cas12 and Cas13), rather than a multi-protein effector complex, as is the case for Class 1 systems[3]. Despite the numerous applications of Cas12 (type V)[4] and Cas13 (type VI)[5] systems, the most extensive toolbox is based on Cas9 (type II) proteins. The type II-A Cas9 from *Streptococcus pyogenes* strain SF370 (SpyCas9)[6] is the best characterized CRISPR-Cas variant to date.

The Cas9-based genome editing technology relies on a Cas9 endonuclease and a target-specific single-guide RNA (sgRNA)[6]. The sgRNA allows the Cas9 nuclease to bind to a target DNA sequence (protospacer). After recognition of a matching target sequence, a double-stranded DNA break (DSDB) will be introduced. The protospacer is complementary to the 5'-end of the sgRNA (spacer) and flanked downstream by a short conserved motif (protospacer adjacent motif, PAM)[7,8]. Cas9-mediated DSDBs are lethal for most prokaryotes, due to the scarcity of non-templated DNA repair mechanisms, including Non-Homologous End Joining and Alternative End Joining[9]. Hence, an exogenous DNA template is often required for surviving a DSDB, using either the host's homologous recombination (HR) machinery, or heterologously-expressed (phage) recombinases (e.g. lambda-red)[10-12]. The typical low efficiency of native HR-mediated repair mechanisms from the host is the reason why Cas-mediated DSDBs are predominantly used as a counter-selection system, i.e. to kill the unedited cells after phage-recombinase-assisted HR[13,14]. A practical drawback is that this strategy often requires multiple plasmids and/or strictly controlled promoters[15-18].

To overcome this limitation, CRISPR-mediated base-editing has been developed by fusing Cas9 to a deaminase that modifies a single nucleotide base at the target site. Base-editing is used for introducing site-specific mutations, or for gene disruption by generation of premature stop codons. As base-editors generally introduce single strand nicks (not DSDBs) or no nicks at all, there is no need for efficient recombination machineries, repair DNA templates, and other exogenous factors[19]. The first developed CRISPR-based base-

editing systems are "Target-AID" (activation-induced cytidine deaminase)[20] and "BE" (base editor)[21], which have also been applied in eukaryotes. They comprise a fusion between SpyCas9 (either the catalytically dead variant (dSpyCas9) or a nickase (nSpyCas9)) and a cytidine deaminase enzyme (converting a C•G into a T•A base pair). The "Target-AID" system employs the *Petromyzon marinus* cytidine deaminase (PmCDA1) or its human orthologue (human AID), while the "BE" system applies the rat APOBEC1 (rAPOBEC1). Currently reported base-editors mostly rely on the type II-A SpyCas9 variant, which has pros and cons. First, SpyCas9 enables editing at a very narrow window located at the PAM-distal end of a targeted protospacer, which is good for precise editing, but not in case of gene disruption. Second, the protospacer must be flanked immediately downstream by an 5'-NGG-3' PAM, limiting the targetable sites[20,21]. Third, SpyCas9 is only active *in vivo* at temperatures below 42°C[22], restricting base-editing exclusively to mesophiles. Since these restrictions may narrow the flexibility and applicability of base-editors, systems based on alternative CRISPR-Cas variants have recently attracted special attention[23-26].

The relatively underexplored type II-C Cas9 proteins have gained substantial interest because of their small size, high fidelity, variable PAM preferences, activity even at harsh conditions (human plasma, high temperatures/salt concentrations), high flexibility with regards to interacting with different sgRNAs, and off-switch control by anti-CRISPR proteins (Acrs) with unique inhibition mechanisms[27,28]. The more compact REC lobe of II-C Cas9 proteins is probably responsible for their weaker dsDNA unwinding activity, their reduced dsDNA binding affinity and stability, as well as their lower dsDNA cleavage activity[28]. Although these properties may decrease the on-target efficiency, these variants exhibit limited off-targeting compared to SpyCas9. This appealing feature is additionally supported by the natural ability of the II-C Cas9 proteins to recognize longer target sequences and PAMs that minimize editing at undesired sites, preventing toxic off-target effects[29,30]. Despite these attractive properties of II-C CRISPR nucleases, their use for genome editing applications has remained largely unexplored.

Anti-CRISPR proteins (Acrs) are small, phage-encoded proteins that evolved to inhibit CRISPR-Cas systems during the arms-race between phages and their prokaryotic hosts[31]. Acrs are being discovered and characterized at a rapidly increasing rate since their initial discovery, and a great arsenal of Acr mechanisms has already been established for various CRISPR-effectors including II-C Cas9 orthologue[32]. Nonetheless, Acrs have rarely been exploited in genome engineering applications for a function other than their ability to obstruct the DNA binding activities of CRISPR-Cas nucleases[31,33]. It was recently

demonstrated that AcrIIC1 from *Neisseria meningitidis* binds the HNH domain of many II-C Cas9s, blocking target DNA cleavage (both *in vitro* and *in vivo*) while still allowing the binding to the DNA target (shown only *in vitro*)[34,35]. These findings provide opportunities to use this family of Acrs in combination with Cas9 variants not only for inhibition of (excessive) Cas9 cleavage activity during genome editing, as reported for most Acrs, but also as an alternative to catalytically inactive Cas9 variants in gene silencing and base-editing applications.

Here, we selected two II-C Cas9 orthologues with a wide-temperature range (20 to 70 °C) from *Geobacillus thermodenitrificans* T12 (ThermoCas9)[36] and *Geobacillus stearothermophilus* (GeoCas9)[37] to study their DNA cleaving and binding activities, as well as their potential for genome editing, transcriptional silencing and base-editing in *E. coli*. These Cas9 variants were selected for their small size and their high stability in harsh conditions, making them suitable candidates for diverse *in vitro* and *in vivo* applications. In addition, we characterized the ability of AcrIIC1 to inhibit the DNA cleavage activity of ThermoCas9 and GeoCas9 *in vivo*, providing an 'off-switch' for genome editing applications. Moreover, we examined the effect of AcrIIC1 on the *in vivo* DNA binding stability of these Cas9 nucleases, and we developed the first Class 2 CRISPR-Acr tools for gene silencing and base-editing. Altogether, we expand the genetic engineering toolbox.

Results

AcrIIC1 inhibits *in vivo* DNA cleavage by ThermoCas9 and GeoCas9 in *E. coli*

Due to the thermophilic nature of the selected Cas nucleases, we first evaluated the *in vivo* DNA cleavage activity of ThermoCas9 and GeoCas9 in *E. coli*, at 37°C. For this purpose, we employed an *E. coli* DH10b strain with a genome-integrated and constitutively expressed *gfp* gene (*E. coli_gfp*; Supplementary Table 1). We constructed targeting ThermoCas9 and GeoCas9 plasmids, namely pTCas9 and pGCas9 respectively, with expression of the *cas9* genes under the control of a synthetic, IPTG-inducible promoter, while opting for constitutive expression of the sgRNA module (Fig. 1a; Supplementary Table 2). We used both nucleases to target the same three protospacers: two in the promoter region (gfp1 and gfp2) and one in the coding region of the *gfp* gene (gfp3) (Supplementary Fig. 1; Supplementary Table 3). Protospacers gfp1 and gfp2 had a 5'-

N4CACA-3' PAM, while protospacer gfp3 had a 5'-N4CCAA-3' PAM. Both PAM sequences have been shown to allow ThermoCas9 cleavage *in vitro*[36], with varying degrees of preference at 37°C. A previous study[37] *in silico* predicted and experimentally validated the 5'-N4CRAA-3' motif as the GeoCas9 PAM. Nonetheless, to gain more insight in the PAM preference of GeoCas9, we tested the nuclease activity on protospacers gfp1, gfp2 and gfp3. We transformed all the targeting plasmids in the *E. coli_gfp* strain. In most cases, the relative transformation efficiency of the *E. coli_gfp* strain was remarkably reduced at high Cas9 induction conditions (Fig. 1b, c). ThermoCas9 could efficiently target the protospacer with the 5'-N4CCAA-3' PAM (gfp3), and to a lesser degree one of the protospacers with the 5'-N4CACA-3' PAM (gfp1) (Fig. 1b). In contrast, GeoCas9 could efficiently target both protospacers with the 5'-N4CACA-3' PAM (gfp1 and gfp2), and more weakly the protospacer gfp3 (Fig. 1c). Overall, this strongly suggests that ThermoCas9 and GeoCas9 induce lethal DSDBs at 37°C and they have distinct PAM- and spacer-dependent cleavage activities regardless of their high identity (87% on amino acid level).

Recent *in vitro* studies on the anti-CRISPR protein AcrIIC1 from *Neisseria meningitidis* reported that AcrIIC1 binds to the HNH domains of various II-C and some II-A Cas9 endonucleases, blocking their DNA cleavage activity but not affecting their DNA binding ability[35,38-40]. Prompted by the potential to use this mode of Acr inhibition for a versatile regulatory tool for CRISPR-based applications, we studied the potential of AcrIIC1 to inhibit the *in vivo* DNA cleavage activities of ThermoCas9 and GeoCas9. We set the plasmid-based expression of the *AcrIIC1* gene under the control of the L-rhamnose-inducible promoter (P_{rha}) and transformed the resulting pAcr construct into the *E. coli_gfp* strain (Supplementary Fig. 2a; Supplementary Table 2). We subsequently transformed the resulting *E. coli_gfp*: pAcr strain with the pTCas9 and pGCas9 plasmids targeting the three *gfp* protospacers (Supplementary Fig. 2a; Supplementary Tables 2, 3). In this two-plasmid approach, AcrIIC1 robustly inhibited the cleavage activity of both nucleases (Supplementary Fig. 2b, c). Even uninduced/leaky AcrIIC1 expression was enough to restore the transformation efficiencies for the most efficiently targeted protospacers (gfp1 and gfp3 for ThermoCas9, gfp1 and gfp2 for GeoCas9), due to the high copy number of pAcr and the medium copy number of pTCas9 or pGCas9. Therefore, in an alternative approach, we introduced the AcrIIC1 expressing module in the pTCas9 and pGCas9 targeting plasmids, constructing the pAcrTCas9 and pAcrGCas9 series of plasmids (Fig. 1d; Supplementary Table 2). We repeated the killing assays with the *E. coli_gfp* strain and, also with this approach, the cleavage activities of the nucleases were generally obstructed (Fig. 1e, f). The reduced transformation efficiencies for the most efficient spacers were

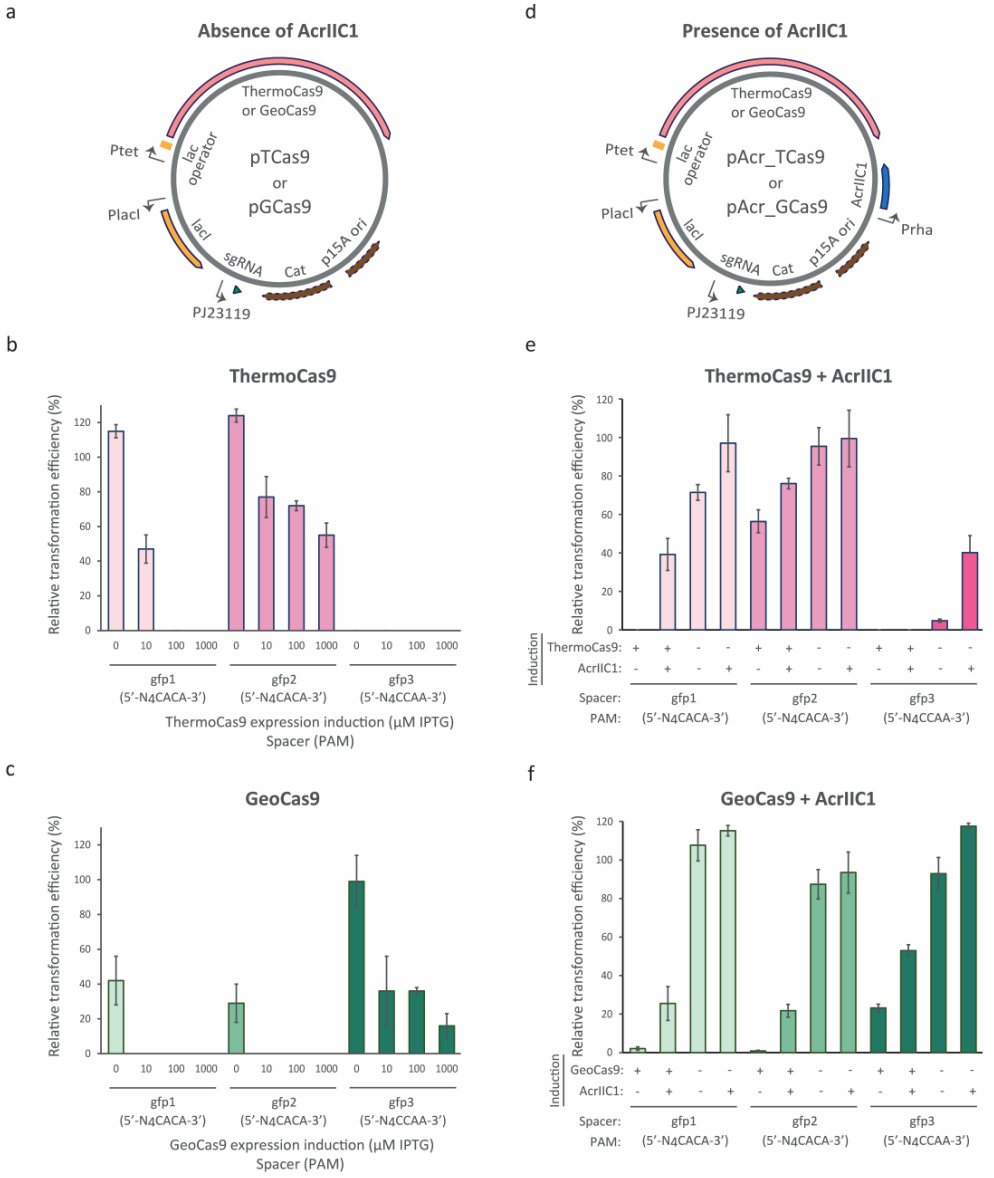


Fig. 1 Cleavage activity of ThermoCas9 and GeoCas9 *in vivo* (37°C) and inhibition by AcrIIC1. a Schematic illustration of the construct transformed into the *E. coli_gfp* strain for killing assays. **b**, **c** Relative to the non-targeting spacer transformation efficiencies for the killing assays with ThermoCas9 and GeoCas9. **d** Schematic illustration of the construct transformed for killing-inhibition assays. **e**, **f** Relative to the non-targeting spacer transformation efficiencies for the killing-inhibition assays with ThermoCas9 and GeoCas9. In the killing assays, the expression of the Cas9 proteins was induced using increasing IPTG concentrations (0, 10, 100, 1000 μM), while in the killing-inhibition assays either null (symbol '-') or maximum IPTG concentration (1000 μM; symbol '+') was applied. The symbols '+' and '-' also represent induction (0.2% L-rhamnose) and absence of induction of AcrIIC1 expression, respectively. Bar graphs were created based on results from three independent biological replicates.

alleviated when we induced the AcrIIC1 expression while keeping the expression of the nucleases uninduced (Fig. 1e, f). Overall, we developed AcrIIC1:Cas9 expression systems in which AcrIIC1 blocks, either constitutively or inducibly, the *in vivo* DNA cleavage activities of ThermoCas9 and GeoCas9 in *E. coli*.

AcrIIC1 blocks ThermoCas9- and GeoCas9-based genome editing in *E. coli*

After establishing that ThermoCas9 and GeoCas9 are capable of providing strong counter-selective pressure in *E. coli* (Fig. 1b, c), we continued by evaluating the potential of these nucleases for homologous recombination (HR)-based genome editing, by combining λ-Red recombineering[41] with Cas9-mediated counter-selection. As a proof of principle, we set out to delete the *gfp* gene from the genome of *E. coli_gfp.* We selected six *gfp* protospacers per nuclease, with varying PAM preferences[36,37], and we incorporated the expression cassettes of the corresponding spacers into the pTCas9 and pGCas9 plasmids (Supplementary Table 3). In the same plasmids, we cloned the genomic regions upstream and downstream of *gfp* as the donor template for HR, resulting in the pHR_TCas9 and pHR_GCas9 series of editing plasmids (Supplementary Fig. 3a; Supplementary Table 2). We transformed the *E. coli_gfp* strain with a plasmid encoding the λ-Red recombineering machinery (pKD46; Supplementary Tables 1, 2) and subsequently with the pHR_TCas9 or pHR_GCas9 plasmids. Finally we screened single colonies by colony PCR and Sanger sequencing for the desired deletion. This ThermoCas9- and GeoCas9-based counter-selection approach improved the genome editing efficiency of λ-Red recombineering in *E. coli* from less than 10%[42] up-to 100% (Fig. 2a, b; Supplementary Fig. 3b, c). We obtained predominantly clean mutants (colonies with solely knock-out genotype) when we targeted protospacers with preferred PAMs[36,37], while poor editing was observed in the case of protospacers with less preferred PAMs (Fig. 2a, b; Supplementary Fig. 3b, c), in line with the results from the killing assays (Fig. 1b, c). Overall, the developed HR-ThermoCas9 or -GeoCas9 counter-selection systems are as efficient as other previously reported Cas9-based editing tools in *E. coli*[43-49].

We subsequently examined the effect of the AcrIIC1 expression on the HR-ThermoCas9 and GeoCas9 counter-selection editing systems in order to examine the potential of AcrIIC1 as an off-switch in genome editing applications. We introduced the AcrIIC1 expressing module into the editing constructs, generating the pHR_AcrTCas9 and

pHR_AcrGCas9 constructs (Supplementary Fig. 4a), and repeated the editing experiments. AcrIIC1 greatly blocked ThermoCas9- and GeoCas9-based counter-selection, resulting in lower editing and transformation efficiencies (Fig. 2c, d; Supplementary Fig. 4b, c, 5). Mixed mutants (colonies with both knock-out and wild-type phenotypes) were mainly detected for protospacers with preferred PAMs[36,37], while basal or no editing was observed in the case of protospacers with less preferred PAMs (Fig. 2c, d; Supplementary Fig. 4b, c). Overall, we demonstrate that AcrIIC1 reduces or completely blocks editing when combining recombineering with ThermoCas9- or GeoCas9-based counter-selection.

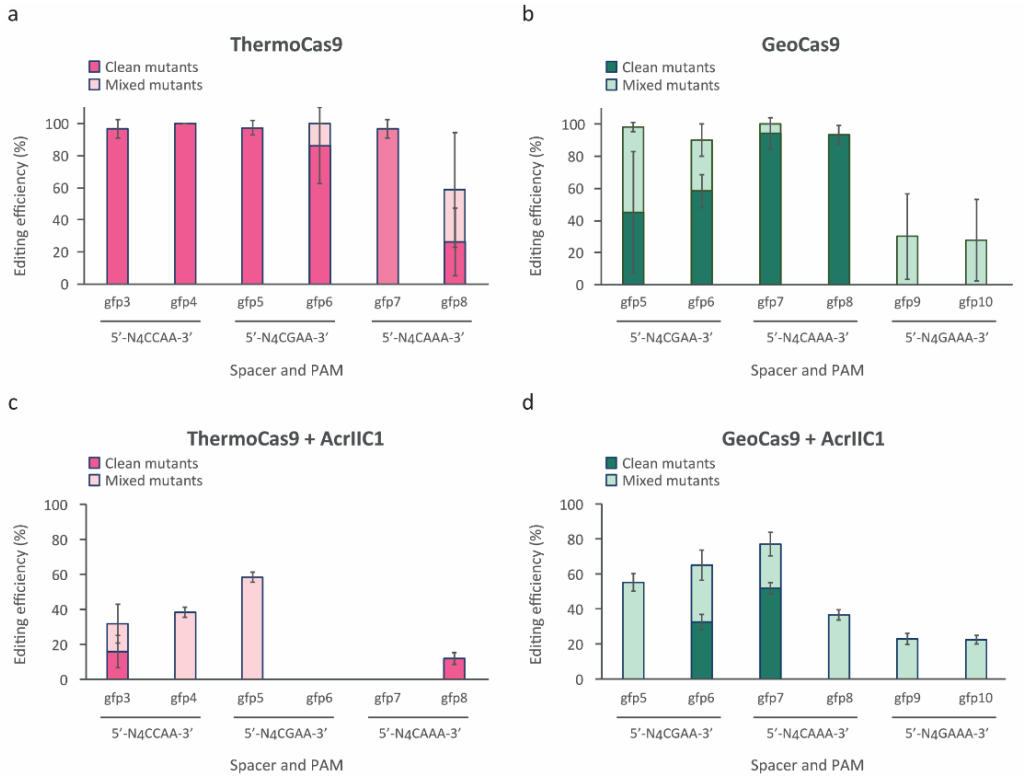


Fig. 2 AcrIIC1-mediated inhibition of ThermoCas9- and GeoCas9-based genome editing in *E. coli*. Editing efficiencies (%) of ThermoCas9 (**a**, **c**) and GeoCas9 (**b**, **d**) in *E. coli_gfp*, in the absence (**a**, **b**) or presence (**c**, **d**) of AcrIIC1. The editing efficiency (%) represents the % of edited versus screened single colonies, while the clean-mutant efficiency (%) represents the % of clean-mutant versus screened colonies. Dark pink and green bars indicate clean mutant editing efficiencies, while light pink and green bars represent mixed mutant editing efficiencies. Cas9 expression was fully induced in all cases. AcrIIC1 expression in (**c**) and (**d**) was fully induced. Bar graphs were created based on results from at least three independent biological replicates.

AcrIIC1:Cas9 complexes mediate successful transcriptional regulation in *E. coli*

It was previously demonstrated *in vitro* that AcrIIC1 does not impede the binding of the II-C Cas9 nuclease from *Neisseria meningitidis* to its DNA target[35]. To investigate whether AcrIIC1-mediated inhibition of Cas9 could be harnessed for transcriptional regulation, we examined the impact of AcrIIC1:ThermoCas9 and AcrIIC1:GeoCas9 complex formation on DNA binding efficiencies *in vivo* and compared them to the DNA binding efficiency of the nuclease deficient ("dead") ThermoCas9 and GeoCas9 variants (dThermoCas9 and dGeoCas9, respectively). For this purpose, we introduced mutations that disrupt the active sites of the RuvC and the HNH domains (ThermoCas9^{D8A,H582A}; GeoCas9^{D8A,H582A}), generating the pdTCas9 and pdGCas9 plasmids (Supplementary Table 2). We then compared the GFP expression of the *E. coli_gfp*: pdTCas9/pdGCas9 strains to the corresponding six *E. coli_gfp*: pAcr + pTCas9/pGCas9 strains we constructed for the killing-inhibition assays (Supplementary Fig. 2a). We found that dThermoCas9, dGeoCas9, AcrIIC1:ThermoCas9 and AcrIIC1:GeoCas9 can efficiently downregulate GFP expression (Supplementary Fig. 6). Although the silencing capabilities of the AcrIIC1:ThermoCas9 and AcrIIC1:GeoCas9 complexes were somewhat lower, these results demonstrate that AcrIIC1 can be used to uncouple the DNA binding and DNA cleaving activity of Cas9 nucleases for silencing purposes. The lower silencing capabilities of the AcrIIC1:Cas9 complexes could be attributed to the negative effect of AcrIIC1 on the expression or the stability of the nucleases it targets[39]. In addition, we cannot rule out differences between the DNA binding capacities of the active and the deactivated versions of the two nucleases (see below).

AcrIIC1 does not hinder the *in vivo* DNA binding of ThermoCas9 and GeoCas9 to their targets

Next we examined the DNA binding strength of the active nucleases (in the absence of AcrIIC1), and we compared them to their deactivated variants and AcrIIC1:Cas9 complexes. We started by studying the DNA cleavage activities of ThermoCas9 and GeoCas9 in the presence of spacer-protospacer mismatches, aiming to identify the number of mismatches that abrogates DNA cleavage. We constructed libraries of targeting plasmids with increasing numbers of 5'-end mutations for each of the aforementioned targeting spacers (Supplementary Fig. 7; Supplementary Tables 2, 3) and

we repeated the transformation-based killing assays. The assays revealed that both ThermoCas9 and GeoCas9 require at least 19 bp long spacer-protospacer complementarity for measurable cleavage of some of the most efficiently targeted protospacers (gfp3 for ThermoCas9, gfp1 and gfp2 for GeoCas9) (Fig. 3a). For the less efficiently targeted protospacers (gfp2 for ThermoCas9 and gfp3 for GeoCas9) full spacer-protospacer complementarity (23 bp) was required for weak cleavage (Fig. 3a). Taken together, both ThermoCas9 and GeoCas9 require at least 19 bp spacer-protospacer complementarity to cleave DNA.

Afterwards, we compared the DNA binding capabilities of the active nucleases with spacer mismatches that prevent DNA cleavage to their deactivated variants and AcrIIC1:Cas9 complexes. We employed the *E. coli_gfp* and *E. coli_gfp*: pAcr strains respectively, and we used silencing of the expression of the *gfp* gene as a readout for DNA binding. We extended the previously constructed spacer-protospacer mismatch plasmid libraries and we kept the numbers of tested spacer-protospacer matches below 19 bp to avoid DNA cleavage when using the active nucleases (Supplementary Fig. 7; Supplementary Tables 2, 3). We subsequently transformed the *E. coli* strains with the plasmid libraries and we measured the effect of each Cas9-spacer combination on the GFP expression. For all the tested strains, the reduction in GFP expression was negligible for spacer-protospacer complementarities of 8 bp or less (Fig. 3b, c). These results are in agreement with the speculated seed region for these nucleases (PAM-proximal region of the protospacer), showing that DNA binding is strongly reduced when introducing spacer-protospacer mismatches that extend into this region[50]. Interestingly, dThermoCas9 blocked the GFP expression more efficiently than active ThermoCas9 for all tested targets and mismatch combinations (Fig. 3b). Although this observation could be attributed to differences between the expression of ThermoCas9 and dThermoCas9, Western-blot ruled out this explanation (Supplementary Fig. 8). The reverse trend, albeit far less pronounced, was observed for GeoCas9 and dGeoCas9 (Fig. 3c). These results indicate that the binding activities of ThermoCas9 and GeoCas9 cannot be linearly associated with the binding activities of their nuclease deficient counterparts. It is also noteworthy that dThermoCas9 did silence the GFP expression when guided to protospacer gfp2 (Supplementary Fig. 6a), whereas ThermoCas9 showed low DNA cleavage activity for this protospacer (Fig. 1b). Hence, the DNA cleavage activities of these nucleases do not always correlate with their binding activities. Furthermore, dThermoCas9 and dGeoCas9 showed similar binding activities for protospacer gfp2 (Supplementary Fig. 6), whereas ThermoCas9 and GeoCas9 presented different DNA cleavage activities for this protospacer (Fig. 1b, c). This suggests

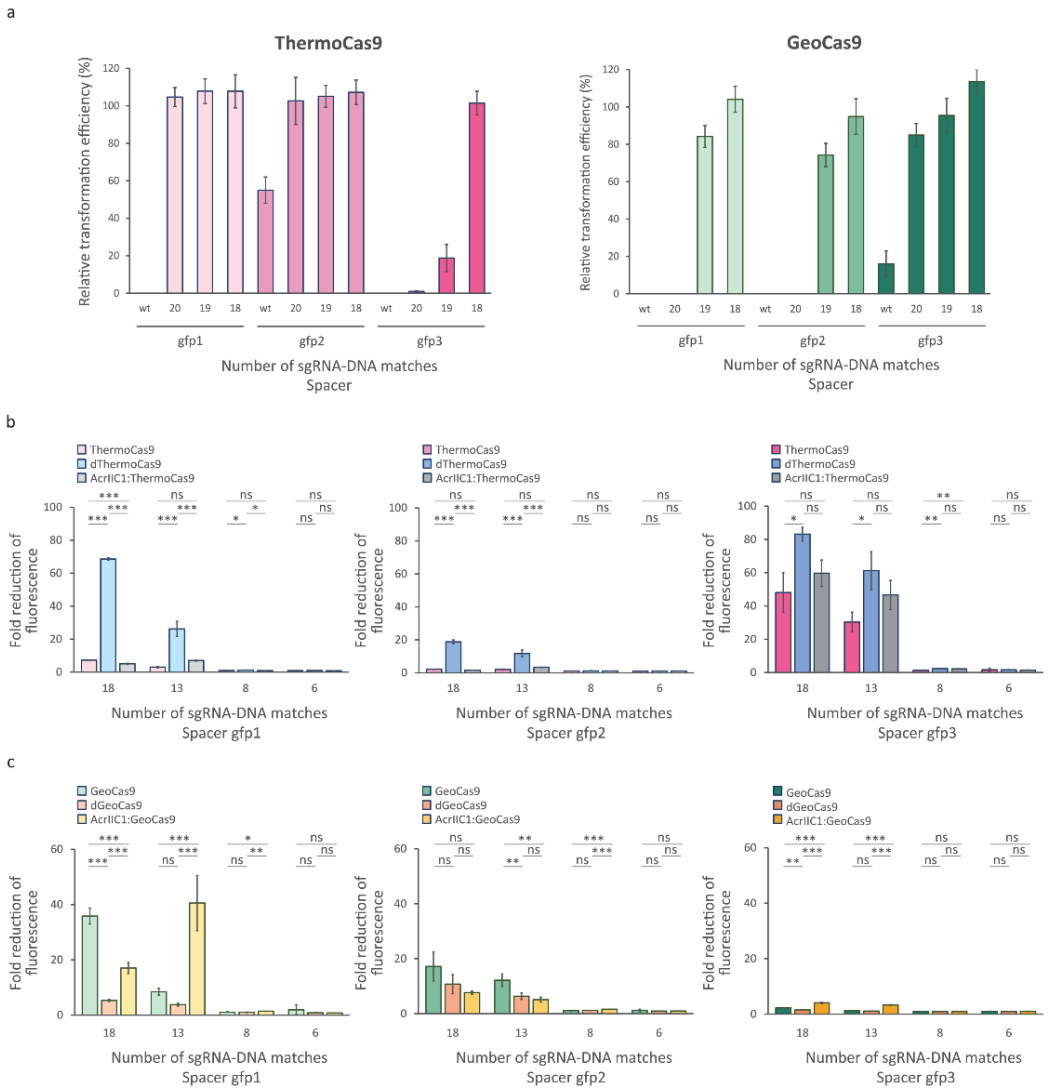


Fig. 3 *In vivo* cleavage and binding specificity of ThermoCas9 and GeoCas9, and comparison to their deactivated variants and AcrIIC1:Cas9 complexes. a Transformation efficiencies of *E. coli_gfp* cells with constructs that express ThermoCas9 (left) or GeoCas9 (right) and sgRNAs with decreasing number of matches to the targeted protospacers (wt = 23 nt). **b, c** Fluorescence reduction assays of *E. coli_gfp* cells that express sgRNAs with decreasing number of matches to the targeted protospacers and (**b**) ThermoCas9 (pink), dThermoCas9 (blue) or AcrIIC1:ThermoCas9 (grey); (**c**) GeoCas9 (green), dGeoCas9 (orange) or AcrIIC1:GeoCas9 (yellow). Bar graphs were created based on results from three independent biological replicates. Statistical significance was calculated with Pairwise Welch's t-tests and the Benjamini & Hochberg p-value adjustment. $p > 0.05$ is shown as ns, $p < 0.05$ is shown as *, $p < 0.01$ is shown as ** and $p < 0.001$ is shown as ***.

differences between the binding and the cleaving requirements of these highly similar Cas9 orthologues. Finally, taking into account the previously reported negative effect of AcrIIC1 on the expression or the stability of II-C Cas9 nucleases[39], we expected that the AcrIIC1:Cas9 complexes would show lower DNA binding efficiencies compared to the sole Cas9s. Nonetheless, the DNA binding activities of the AcrIIC1:ThermoCas9 and AcrIIC1:GeoCas9 complexes were generally similar to the sole ThermoCas9 and GeoCas9 (Fig. 3b, c). Hence, our results demonstrate that AcrIIC1 does not hinder the *in vivo* binding of ThermoCas9 and GeoCas9 to their DNA targets. Moreover, differences in DNA binding abilities of the AcrIIC1:Cas9 complexes and the dCas9 proteins originate from innate differences in DNA binding capacities of the active and deactivated variants.

AcrIIC1:Cas9 complexes as alternatives of dThermocas9 and dGeocas9 for base-editing in *E. coli*

We continued by developing dThermoCas9- and dGeoCas9-associated base-editors, and by studying their editing outcomes. For this purpose, we fused the *Petromyzon marinus* cytidine deaminase (PmCDA1)[20] gene and the phage PBS2 uracil DNA glycosylase inhibitor (UGI)[51] gene to the C-terminus of the *dthermocas9* and *dgeocas9* genes (Supplementary Fig. 9; Supplementary Table 2). We cloned six spacers into the resulting dThermoTarget-AID and dGeoTarget-AID plasmids, designed to target protospacers within the *gfp* gene of the *E. coli_gfp* strain (Supplementary Table 2, 3). Four of the ThermoCas9 spacers (gfp4, gfp5, gfp7, gfp8) and three of the GeoCas9 spacers (gfp5, gfp7, gfp8) were the same spacers we previously used for the corresponding HR-Cas9 counter-selection editing experiments (Fig. 2a, b; Supplementary Table 2, 3). These protospacers were selected to, collectively, contain cytidines at almost all the positions (Supplementary Table 3), allowing a fair assessment of the editing windows of the base-editors. We transformed the dThermoTarget-AID and dGeoTarget-AID plasmids in the *E. coli_gfp* strain and screened single colonies for base-editing activity by sequencing PCR amplicons spanning the targeted regions in the *gfp* gene. We detected C•G to T•A conversions at multiple positions for all the protospacers we targeted with the dThermoTarget-AID system and for four out of the six protospacers we targeted with the dGeoTarget-AID system. However, most of the edited colonies had mixed genotypes and only a few had a clean conversion at one or more positions (Fig. 4a, b; Supplementary Fig. 10). Up to two and four simultaneous clean mutations were generated by the dThermoTarget-AID and dGeoTarget-AID systems, respectively (Supplementary Fig. 10). Both base-editing systems

preferentially edited cytidines at the PAM-distal end of the protospacer (Supplementary Fig. 10), similar to the commonly used SpyCas9 base-editors[52]. The observed editing windows in many occasions were extended outside of the protospacer region and were up to 28 bp for dThermoTarget-AID (from -6 to -33 positions relative to the PAM) and up to 15 bp for dGeoTarget-AID (from -10 to -24 positions relative to the PAM) (Supplementary Fig. 10). These editing windows are larger than the 5 bp (from -16 to -20 positions relative to the PAM) activity window reported for the SpyCas9 base-editors[52,53]. Interestingly, protospacer gfp8 was efficiently edited by both base-editors (Fig. 4a, b), whereas the HR-based genome editing of gfp8 by the ThermoCas9 counter-selection system was rather inefficient (Fig. 2a, b). A similar case is protospacer gfp2, for which high binding strength of dThermoCas9 and low cleavage activity of its active nuclease were observed (Fig. 1b; Supplementary Fig. 6a). In other words, a protospacer that is efficiently cleaved is not necessarily a good target for efficient base-editing. Vice versa, the targeting of the gfp5 and gfp7 protospacers resulted in low base-editing but high HR-editing outcomes (Fig. 2a, b; Fig. 4a, b). These results show that the steady binding of the base-editing complexes to their targets does not ensure high editing efficiencies, suggesting that the conversion process is governed by additional rules. Streaking and incubation of randomly selected colonies resulted in the editing of previously unedited protospacer positions, and the further extension of the editing windows (up to 30 bp for dThermoTarget-AID and 25 bp for dGeoTarget-AID), predominantly towards the PAM proximal region (Fig. 4e). Moreover, the number of simultaneous clean C•G to T•A conversions was increased to up to eight for the dThermoTarget-AID system and up to six for the dGeoTarget-AID system (Supplementary Fig. 10). Hence, the base-editing efficiency and window of activity of these dThermoCas9/dGeoCas9-associated editors can be further increased by prolonging the editing conditions.

Furthermore, we reasoned to combine the DNA binding activities of the AcrIIC1:Cas9 complexes with the base-editing activity of the PmCDA1 enzyme. We expected that the resulting systems can not only be an alternative to the “dThermoTarget-AID” and “dGeoTarget-AID” systems, but also have the additional benefit of being able to select against unedited loci by relieving the Acr-mediated inhibition of DNA cleavage activity of Cas9. In addition, this could enrich for colonies with clean mutations. We introduced the AcrIIC1 expressing module in the Target-AID plasmids and we substituted the *dcas9* genes with their catalytically active counterparts, developing the “AcrThermoTarget-AID” and “AcrGeoTarget-AID” base-editing systems (Supplementary Fig. 9; Supplementary Tables

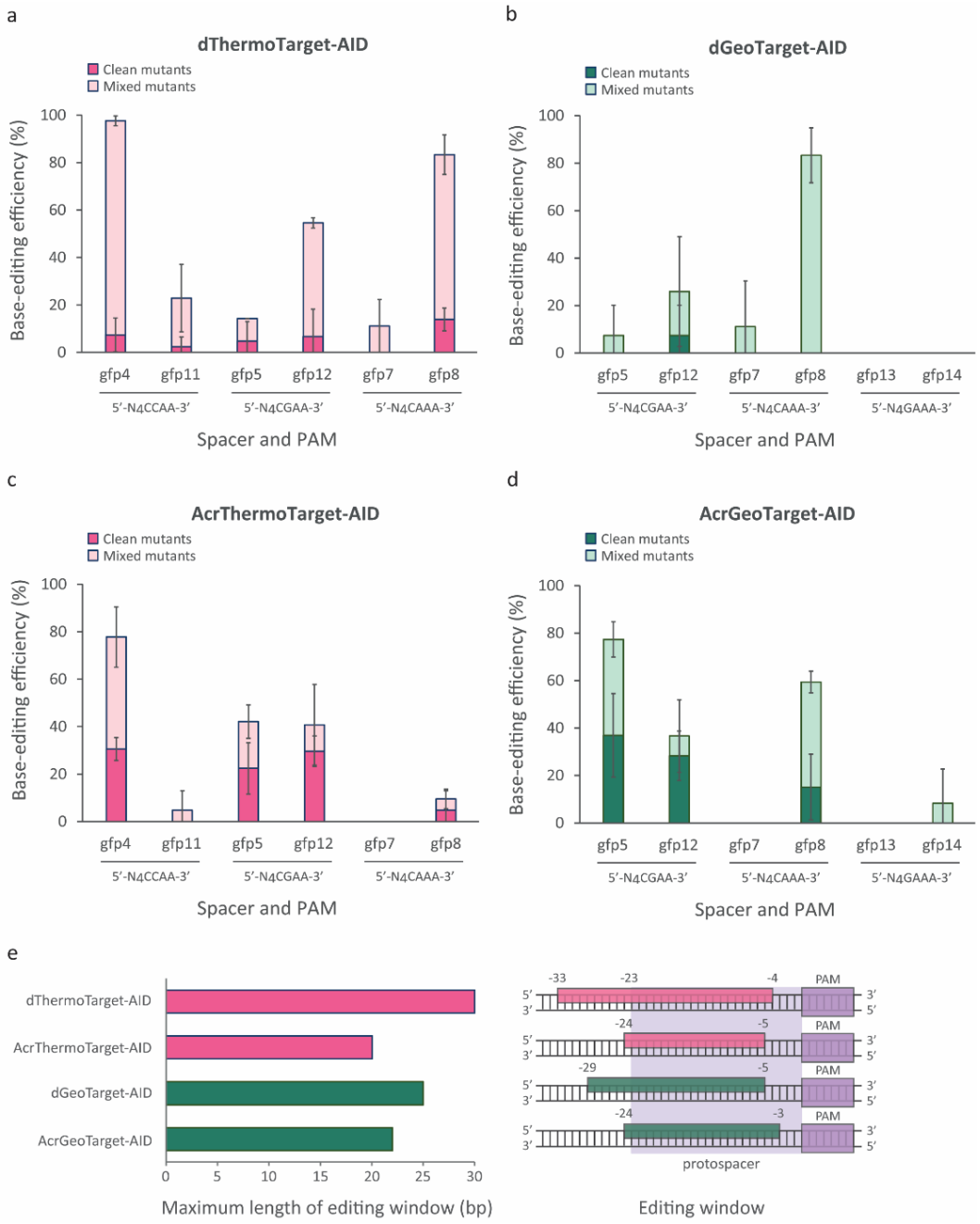


Fig. 4 Type II-C cytidine base-editors and their AcrIIC1:Cas9 complexes that enable base-editing coupled to counter-selection in *E. coli*. Base-editing efficiency (%) of the dThermoCas9 (**a**), dGeoCas9 (**b**), AcrIIC1:ThermoCas9 (**c**), and AcrIIC1:GeoCas9 (**d**) base-editors in *E. coli_gfp*. The base-editing efficiency (%) represents the % of the number of edited versus screened single colonies, while the clean-mutant efficiency

(%) represents the % of clean-mutant versus screened single colonies. Dark pink and green bars indicate clean mutant editing efficiencies, while light pink and green bars represent mixed mutant editing efficiencies. Bar graphs were created based on results from three independent biological replicates. **e** The maximum length (left) and spectrum (right) of the editing window of each base-editing system. The dark purple boxes indicate the PAM region, while the light purple area depicts the protospacer region. Pink and green boxes represent the overall activity window of the base-editing systems, after analysis of all screened colonies either before or after streaking.

2). We transformed the constructed plasmids in the *E. coli_gfp* strain, performed *gfp*-specific PCRs on single colonies and sequenced the products. As expected, the number of clean point mutations in protospacers that were moderately or highly edited was substantially higher compared to the corresponding number from the "dThermoTarget-AID" and "dGeoTarget-AID" systems (Fig. 4a-d; Supplementary Fig. 10). Moreover, we detected an increased preference for editing at the PAM-proximal ends when compared to the dCas9-based editors (Supplementary Fig. 10). This could be attributed to the tolerance of these active Cas9s in mismatches at the PAM-distal end, which results in counter-selection of cells with this mutant genotype (Fig. 3a). In contrast, mismatches at the PAM-proximal end are not tolerated, blocking counter-selection and allowing for cell survival (Fig. 3a). The observed editing windows were up to 15 bp for AcrThermoTarget-AID (from -9 to -23 positions relative to the PAM) and up to 22 bp for AcrGeoTarget-AID (from -3 to -24 positions relative to the PAM), while the number of simultaneous clean mutations was up to four and three, respectively (Supplementary Fig. 10). Streaking and incubation of randomly selected colonies resulted in the editing of previously unedited protospacer positions, the extension of the activity window (in the case of AcrThermoTarget-AID) and in colonies with predominantly clean mutations at all the edited positions (Fig. 4e; Supplementary Fig. 10).

Overall, these type II-C editors mediate efficient base-editing in wide activity windows and multiple positions simultaneously, facilitating gene inactivation in bacteria in the absence of donor DNA. The Acr-mediated control over Cas9 cleavage activity additionally enables counter-selection of non-edited loci, further increasing the number of clean mutants and the preference for editing at the PAM-proximal protospacer region.

Discussion

In this study, we initially demonstrate that the thermostable type II-C ThermoCas9 and GeoCas9 endonucleases are highly active *in vivo* at 37°C and introduce lethal DSDBs in the *E. coli* genome. Moreover, we reveal additional PAM recognition sequences of GeoCas9, for which very few PAMs have been reported earlier[37]. In agreement with our previous *in vitro* study[36], we show that ThermoCas9 and GeoCas9 (under these conditions) exhibit lower *in vivo* mismatch tolerance compared to previous data for SpyCas9[54]. The enhanced fidelity of these II-C Cas9 variants can be attributed not only to their longer PAM (4 bp) and spacer (23 bp), but also to their slower dsDNA cleavage rates (at 37°C, lower k_{cat})[28,36,55]. These characteristics are similar to the II-C Cas9 variants from *N. meningitidis* (NmeCas9), and *Campylobacter jejuni* (CjeCas9), which are also more specific than SpyCas9[29,30]. Leveraging their functionality at 37°C, we apply ThermoCas9 and GeoCas9 as counter-selection tools in combination with λ-Red recombineering in *E. coli*, leading to high editing efficiencies and predominantly clean mutants. In addition, we show that the previously *in vitro* characterized anti-CRISPR protein from *N. meningitidis* (AcrIIC1)[34,35] robustly inhibits the DNA cleavage activity of both ThermoCas9 and GeoCas9 *in vivo*. To the best of our knowledge, this is the first time that an Acr protein is shown to interact with ThermoCas9, enabling applications that require 'off-switch' control of its cleavage activity. In line with its reported cleavage inhibition ability, AcrIIC1 successfully blocks the Cas9-mediated counter-selection, severely dropping the editing efficiencies and eliminating the presence of clean mutants. Overall, ThermoCas9 and GeoCas9 can be used as efficient genome targeting or editing tools for *in vivo* applications at 37°C with the ability of 'off-switch' control by AcrIIC1.

Next, we uncovered the ability of AcrIIC1 to trap these Cas9 nucleases in a DNA-bound and cleavage-inactive state *in vivo*, resulting in silencing of the targeted *gfp* gene in the genome of *E. coli*. The silencing efficiency of the AcrIIC1:Cas9 complexes was lower than that of their dCas9 variants, probably because of inherent differences in the DNA binding ability between the active and inactive form of each nuclease that we reveal in our study. We also show that AcrIIC1 does not destabilize the *in vivo* DNA binding strength of ThermoCas9 and GeoCas9, and we exploit this unique feature to couple base-editing applications (by allowing AcrIIC1 expression) to subsequent counter-selection (by interrupting AcrIIC1 expression). In this context, we develop AcrIIC1:Cas9 and dCas9 base-editors of these II-C variants and we observe up to six times larger editing windows than the previously reported dSpyCas9 base-editor[53], probably due to differences in the size

and the structure of the II-C and II-A variants. The observed editing windows in many occasions were expanded upstream of the protospacer region. The relatively large editing window is explained partly by the fact that these II-C orthologues recognize longer sequences (23 nt) compared to the II-A variants (20 nt), forming a slightly larger R-loop with extended single-stranded DNA region available for deamination[36,37,53]. Similar to the dSpyCas9 editors[52], the dThermoCas9 and dGeoCas9 editors mainly edit the PAM-distal end of the protospacer, which makes sense as the presence of mutations within the seed region hampers the DNA binding ability of all these Cas9s. In contrast, the Acr-editors associated to the active Cas9 variants show preference for editing at the PAM-proximal end of the protospacer, because mutations at the PAM-distal end are tolerated by ThermoCas9 and GeoCas9, resulting in cleavage leading to cell death (counter-selection); hence this explains the enrichment of the PAM-proximal edits. All in all, these are the first Acr tools for base-editing applications reported to date. The extended base-editing windows of all II-C editor designs described here may increase the freedom for gene inactivation (generation of premature stop codons, mutation of start codon), for gene mutagenesis (introduction of multiple nucleotide substitutions), and, theoretically, for abrogation of regulatory sequences (regions coding for small RNAs or promoter regions) and splice sites.

Methods

Bacterial strains and growth conditions

Bacterial strains used in this study are listed in Supplementary Table 1. All *E. coli* strains were cultured in Luria-Bertani broth (LB) or on LB agar plates, supplemented when necessary with chloramphenicol (15 µg/ml), and/or ampicillin (100 µg/ml). L-rhamnose (0.2% w/v), and/or IPTG (0 - 1000 µM) were additionally used for inducing the expression of AcrIIC1 and (d)ThermoCas9/ (d)GeoCas9/ (d)ThermoCas9-PmCDA1-UGI-LVA/ (d)GeoCas9-PmCDA1-UGI-LVA, respectively. L-arabinose (0.15% w/v) was also applied in the genome editing assays to trigger the expression of the λ-Red recombineering proteins (Exo, Beta, Gamma). Moreover, M9TG (11.28 g 1X M9 salts, 10 g tryptone, 5 g glycerol) medium was used instead of LB in the fluorescence loss assays. All strains were grown at 37°C (220 rpm when liquid culture), except for *E. coli_gfp*: pKD46 (Supplementary Table 1) which was grown at 30°C.

Construction of plasmids

Plasmids, primers or oligonucleotides, *cas9* or *acriic1* gene and protein sequences used in this study are presented in Supplementary Tables 2, 4, 5, and 6, respectively. Thanks to the high (94%) nucleotide identity between the ThermoCas9 and GeoCas9 sgRNA modules (with the few differences predicted to be part of the sgRNA loops), we used the ThermoCas9 sgRNA module for all created constructs. The bacterial plasmids were constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB). The fragments for assembling the plasmids were obtained through PCR with Q5® High-Fidelity 2X Master Mix (NEB). The PCR products were run on a 0.8% agarose gel and were subsequently purified using Zymogen gel DNA recovery kit (Zymo Research). The assembled plasmids were transformed to chemically competent *E. coli* DH5α cells[56] (Supplementary Table 1) and plated on LB agar containing chloramphenicol (15 μg/ml) or ampicillin (100 μg/ml) and incubated overnight at 37°C. The next day, single colonies were inoculated in LB medium, grown overnight at 37°C (220 rpm) and the plasmids were isolated using the GeneJet plasmid Miniprep kit (ThermoFisher Scientific). All the constructs were verified using Sanger sequencing (Macrogen). Description of the assembled fragments used for the construction of each plasmid is detailed in Supplementary Table 2. For annealing of oligos to create dsDNA used in the plasmid assembly, 4 μl oligonucleotide pairs (100 μM each) were mixed in Milli-Q water to a final volume of 100 μl, heated at 95°C for 5 min, and slowly cooled to room temperature.

Killing assays

To target bacterial DNA, chemically competent *E. coli_gfp* cells were transformed[56] with equal amounts (3 ng) of *gfp*-targeting plasmid (pTCas9_gfp1-gfp3; pGCas9_gfp1-gfp3) (Fig. 1a; Supplementary Tables 1-3). We also constructed libraries of mismatched targeting plasmids, namely pTCas9_x.y and pGCas9_x.y (where x = the employed spacer gfp1, gfp2, or gfp3; y = the number of consecutive spacer-protospacer matches counting from the PAM-proximal end of the protospacer) (Fig. 1a; Supplementary Fig. 7; Supplementary Tables 2, 3). As control, 3 ng of a non-targeting plasmid (pTCas9_scr; pGCas9_scr) were used (Fig. 1a; Supplementary Fig. 7; Supplementary Tables 2, 3). Transformed cells were cultured on LB agar supplemented with chloramphenicol (15 μg/ml) and different IPTG concentrations (0 μM, 10 μM, 100 μM, 1000 μM IPTG) for 17 h at 37°C. Colony forming units (CFUs) were counted after plating 100 μl of undiluted

biological triplicates (from 500 µl recovery) and used for calculating the relative transformation efficiencies.

Killing-inhibition assays

To inhibit targeting of the bacterial genomic DNA using a two-plasmid approach, *E. coli_gfp* cells harboring an AcrIIC1-expressing plasmid (pAcr) were transformed[56] with equal amounts (5 ng) of *gfp*-targeting plasmid (pTCas9_gfp1-gfp3; pGCas9_gfp1-gfp3) (Supplementary Fig. 2a; Supplementary Tables 1-3). As control, 5 ng of a non-targeting plasmid (pTCas9_scr; pGCas9_scr) were used (Supplementary Fig. 2a; Supplementary Tables 2, 3). Transformed cells were grown on LB agar supplemented with chloramphenicol (15 µg/ml), ampicillin (100 µg/ml), IPTG (0 µM or 1000 µM), and L-rhamnose (0% or 0.2% w/v) for 17 h at 37°C. Colony forming units (CFUs) were counted after plating 100 µl undiluted biological triplicates and used for calculating the relative transformation efficiencies.

Alternatively following a single-plasmid approach, *E. coli_gfp* cells were transformed with equal amounts (5 ng) of targeting inhibition plasmid additionally carrying the *acriic1* gene (pAcr_TCas9_gfp1-gfp3; pAcr_GCas9_gfp1-gfp3) (Fig. 1d; Supplementary Tables 2, 3). As control, 5 ng of a non-targeting plasmid (pAcr_TCas9_scr; pAcr_GCas9_scr) were used (Fig. 1d; Supplementary Tables 2, 3). Transformed cells were grown on LB agar supplemented with chloramphenicol (15 µg/ml), IPTG (0 µM or 1000 µM), and L-rhamnose (0% or 0.2% w/v) for 17 h at 37°C. CFUs were counted after plating 100 µl undiluted biological triplicates and used for calculating the relative transformation efficiencies.

Genome editing assays

For the deletion of the genomic *gfp* gene and its promoter (P_{lacUV5}), chemically competent *E. coli_gfp* cells harboring the λ-Red operon (*exo*, *beta*, and *gam* genes)[41] were transformed[56] with 5 ng of genome editing plasmid (pHR_TCas9_gfp3-gfp8; pHR_GCas9_gfp5-gfp10) (Supplementary Fig. 3a; Supplementary Tables 1-3). The λ-Red operon was transcribed from the arabinose-inducible promoter (P_{araB})[57] on a thermo-sensitive, low copy number plasmid (pKD46)[58]. During recovery, cells grew for 2 and half hours at 30°C (220 rpm) and the expression of both the λ-Red recombineering system and

the Cas9 nuclease was induced (0.15% w/v L-arabinose and 1 mM IPTG, respectively) for maximum HR efficiency. The induction of the Cas9 expression was prolonged for counter-selection of the non-edited cells by plating 100 μl of undiluted biological triplicates on LB agar supplemented with chloramphenicol (15 μg/ml) and IPTG (1000 μM). After 17 h of incubation at 37°C, several colonies were screened for genome editing through OneTaq® 2X Master Mix with Standard Buffer (NEB) PCR amplification of the targeted region with genome specific primers (Supplementary Table 4), and the size of the PCR products was verified by 1.2% agarose gel electrophoresis. The results were analyzed with GelAnalyzer 19.1 and colonies were divided in 3 distinct categories: (a) wild-type, (b) mixed wild-type and knockout, and (c) clean knock-out. Sanger sequencing was indicatively performed to verify successful gene deletion in the resulting mutants. Bacterial cells either lacking (*E. coli* DH10b cells) or containing the targeted fragment (*E. coli_gfp*) (Supplementary Table 1) were used as positive and negative genome editing control, respectively.

Counter-selection inhibition assays

To inhibit the Cas9-mediated counter-selection of unedited cells after HR, *E. coli_gfp*: pKD46 cells were transformed with equal amounts (5 ng) of counter-selection inhibition plasmid additionally carrying the *acriic1* gene (pHR_AcrTCas9_gfp3-gfp8; pHR_AcrGCas9_gfp5-gfp10) (Supplementary Fig. 4a; Supplementary Tables 1-3). Maintaining maximum expression of the λ-Red recombineering system (0.15% w/v L-arabinose) during recovery, we fully expressed AcrIIC1 (0.2% w/v L-rhamnose) and Cas9 (1000 μM IPTG) and plated 100 μl of undiluted biological triplicates on LB agar supplemented with chloramphenicol (15 μg/ml) and L-rhamnose (0.2% w/v). Screening and analysis of the editing events were performed as described above.

Binding assays

To quantify the fluorescence loss, *E. coli* DH10B_*gfp* cells were transformed[56] with equal amounts (3 ng) of *gfp*-silencing plasmid (pdTCas9_gfp1-gfp3; pdGCas9_gfp1-gfp3 and their derivatives with partial spacer-protospacer complementarity) (Supplementary Fig. 7; Supplementary Tables 1-3). As positive fluorescence control, 3 ng of empty vector (pACYC184) or non-targeting plasmid (pdTCas9_scr; pdGCas9_scr) (Supplementary Tables 2, 3) were used, while as a negative fluorescence control *E. coli* DH10b cells transformed

with 3 ng of the empty vector were employed (Supplementary Table 1). Post-transformation, 2 µl of the recovered cells were cultured in 198 µl M9TG containing chloramphenicol (15 µg/ml) in a Masterblock® 96-well deep microplate (Greiner Bio-One) for 22 h at 37°C with vigorous shaking (900 rpm). The second day, 2 µl of overnight cultures were diluted in 198 µl M9TG containing the same antibiotic, and 2 µl of these were re-diluted in 198 µl M9TG with antibiotic and IPTG (0 µM, 10 µM, 50 µM) in a Masterblock® 96-well deep micro-plate (Greiner Bio-One), and incubated for 22 h at 37°C with vigorous shaking (900 rpm). The third day, 2 µl of overnight cultures were diluted in 998 µl 1X PBS (8 g NaCl, 0.2 g KCl, 1.44 g Na2HPO4 2H2O, 0.24 g KH2PO4; pH=6.8) in a Masterblock® 96-well deep microplate (Greiner Bio-One). The fluorescence signal and the presence of subpopulations were examined using the Attune NxT Flow Cytometer device (Thermo Fisher Scientific) (GFP intensity 405-512/25 of at least 30,000 single cells per sample). All assays were performed in three biological replicates.

Similarly, for the AcrIIC1:Cas9 complexes, *E. coli_gfp* cells harboring the AcrIIC1-expressing plasmid (pAcr) were transformed[56] with equal amounts (5 ng) of *gfp*-targeting plasmid (pTCas9_gfp1-gfp3; pGCas9_gfp1-gfp3 and their derivatives with partial spacer-protospacer complementarity) (Supplementary Fig. 2a, 7; Supplementary Tables 1-3). As positive fluorescence controls, 5 ng of the empty vectors (pACYC184; pUC19) or the non-targeting plasmids (pTCas9_scr; pGCas9_scr) were used (Supplementary Tables 2, 3), while as negative fluorescence control 5 ng of *E. coli* DH10B cells electroporated with 5 ng of each empty vector (pACYC184; pUC19) were applied (Supplementary Table 1). The same fluorescence loss assay protocol as above was followed, with the only difference that ampicillin (100 µg/ml) and L-rhamnose (0.2% w/v) were also applied.

Statistical significance was calculated with Pairwise Welch's t-tests and the Benjamini & Hochberg p-value adjustment.

Western blot assays

To compare the expression of ThermoCas9 and dThermoCas9, *E. coli_gfp* cells were transformed with equal amounts (3 ng) of a His6-(d)ThermoCas9-encoding, non-targeting plasmid (pHis-TCas9_scr; pHis-dTCas9_scr) (Supplementary Tables 2, 3). As negative control, 3 ng of a non-targeting plasmid lacking a *his6-(d)thermocas9* gene (pscr) were used (Supplementary Tables 2, 3). Transformed cells were grown overnight at 37°C on LB agar supplemented with chloramphenicol (15 µg/ml). Single colonies were inoculated in

5 ml LB containing chloramphenicol (15 µg/ml), and cell cultures were incubated at 37°C under shaking (220 rpm) until reaching OD600 = 0.4 - 0.6. After 1 h incubation on ice, the His6-(d)ThermoCas9 expression was induced with 1 mM IPTG, and the cultures were incubated overnight at 20°C for maximum protein yields. The equivalents of 1 mL of cells at OD600 = 0.8 were centrifuged at maximum speed for 1 min. The cell pellets were resuspended in 100 µl of 1x SDS-PAGE buffer, heated at 98°C for 10 min, centrifuged at maximum speed for 5 min, and loaded in a 10% Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad). 10 µl of PageRuler™ Prestained Protein Ladder and 10 µl of blot positive control marker were loaded in the gel, which was run in a Mini–PROTEAN® Tetra Vertical Electrophoresis Cell (Bio-Rad) containing 800 ml 1x SDS running buffer at constant 20 mA for 1 h (PowerPac™ Basic Power Supply; Bio-Rad). After 4 h staining in PageBlue™ Protein Staining Solution (Thermo Scientific) under gentle shaking, washing and visualization with UV light (G:BOX Chemi XX6), the gel was transferred to the iBlot2 dry blotting machine (Invitrogen) and then to the Transfer buffer (3.03 g Trizma-base, 14.4 g Glycine, 200 ml methanol, dH2O up to 1 l) in a big agar plate. After blocking in PBST (137 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.1% Tween-20; pH 7.4) + 3% BSA for 1 h at room temperature, the membrane was incubated overnight at 4°C under slow rotation with 6x-His Tag Monoclonal Antibody (HIS.H8; Invitrogen) diluted 1:5,000 in the same solution. The membrane was washed with PBST and incubated for 1 h at room temperature with Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody Alexa Fluor Plus 488 (Invitrogen) diluted 1:20,000 in the same solution. After final washing of the membrane with PBST, the blot was developed using SuperSignal™ West Pico PLUS Chemiluminescent Substrate kit (Thermo Scientific) and visualized according to the instructions.

Base-editing assays

To introduce single nucleotide substitution(s) in the genomic *gfp* gene, *E. coli_gfp* cells were transformed[56] with 3 ng of base-editing plasmid (pdThermoTarget-AID_gfp4/gfp5/gfp7/gfp8/gfp11/gfp12; pAcrThermoTarget-AID_gfp4/gfp5/gfp7/ gfp8/gfp11/gfp12; pdGeoTarget-AID_gfp5/gfp7/gfp8/gfp12/gfp13/gfp14; pAcrGeo Target-AID_gfp5/gfp7/gfp8/gfp12/gfp13/gfp14) (Supplementary Fig. 9; Supplementary Tables 1-3). The expression of the dCas9-editors was induced with the addition of 50 µM IPTG during both recovery and plating, while in the case of the Acr-editors base-editing (50 µM IPTG and 0.2% w/v L-rhamnose) and counter-selection (1 mM IPTG) conditions

were separately applied during recovery and plating, respectively. After 17 h of incubation at 37°C in the presence of chloramphenicol (15 μg/ml), several colonies were streaked on "master" selection plates with no inducers and were simultaneously screened for base-editing through PCR amplification (Q5® High-Fidelity 2X Master Mix; NEB) of the targeted region with genome-specific primers. The amplified fragments were purified (DNA clean and concentrator kit; Zymo Research) and sequenced (Sanger), followed by high-throughput *in silico* analysis of the results employing a variation of the on-line tool "EditR"[59], previously developed in our lab. Each base-editing experiment was performed in three biological replicates. In addition, the streaks on the "master" plates from single colonies were streaked on plates supplemented with 1 mM IPTG to screen for enhanced base-editing efficiency and/or purity, as described above.

References

1. Mougiakos, I., Bosma, E. F., Ganguly, J., van der Oost, J., & van Kranenburg, R. (2018). Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects. *Current opinion in biotechnology*, *50*, 146-157.
2. Fokum, E., Zabed, H. M., Guo, Q., Yun, J., Yang, M., Pang, H., An, Y., Li, W., & Qi, X. (2019). Metabolic engineering of bacterial strains using CRISPR/Cas9 systems for biosynthesis of value-added products. *Food bioscience*, *28*, 125-132.
3. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., ... & Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67-83.
4. Meliawati, M., Schilling, C., & Schmid, J. (2021). Recent advances of Cas12a applications in bacteria. *Applied Microbiology and Biotechnology*, 1-10.
5. Kordyś, M., Sen, R., & Warkocki, Z. (2021). Applications of the versatile CRISPR-Cas13 RNA targeting system. *Wiley Interdisciplinary Reviews: RNA*, e1694.
6. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821.
7. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology*, *190*(4), 1390-1400.
8. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733-740.
9. Finger-Bou, M., Orsi, E., van der Oost, J., & Staals, R. H. (2020). CRISPR with a happy ending: Non-templated DNA repair for prokaryotic genome engineering. *Biotechnology journal*, *15*(7), 1900404.
10. Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., & Blaschek, H. P. (2015). Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. *Journal of biotechnology*, *200*, 1-5.
11. Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., & Gu, Y. (2016). CRISPR/Cas9-based efficient genome editing in Clostridium ljungdahlii, an autotrophic gas-fermenting bacterium. *ACS synthetic biology*, *5*(12), 1355-1361.
12. Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., & Yang, S. (2016). CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology journal*, *11*(7), 961-972.
13. Court, D. L., Sawitzke, J. A., & Thomason, L. C. (2002). Genetic engineering using homologous recombination. *Annual review of genetics*, *36*(1), 361-388.
14. Pines, G., Freed, E. F., Winkler, J. D., & Gill, R. T. (2015). Bacterial recombineering: genome engineering via phage-based homologous recombination. *ACS synthetic biology*, *4*(11), 1176-1185.

15. Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. J., Chen, T., & Zhao, X. (2015). Metabolic engineering of Escherichia coli using CRISPR–Cas9 meditated genome editing. *Metabolic engineering*, *31*, 13-21.
16. Tong, Y., Charusanti, P., Zhang, L., Weber, T., & Lee, S. Y. (2015). CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS synthetic biology*, *4*(9), 1020-1029.
17. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., & Yang, S. (2015). Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Applied and environmental microbiology*, *81*(7), 2506-2514.
18. Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., & Blaschek, H. P. (2016). Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in Clostridium beijerinckii as an example. *ACS synthetic biology*, *5*(7), 721-732.
19. Wang, Y., Liu, Y., Zheng, P., Sun, J., & Wang, M. (2020). Microbial base editing: a powerful emerging technology for microbial genome engineering. *Trends in Biotechnology*, 39(2), 165-180.
20. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., Shimatani, Z., & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, *353*(6305), aaf8729.
21. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424.
22. Mougiakos, I., Bosma, E. F., Weenink, K., Vossen, E., Goijvaerts, K., van der Oost, J., & van Kranenburg, R. (2017). Efficient genome editing of a facultative thermophile using mesophilic spCas9. *ACS synthetic biology*, *6*(5), 849-861.
23. Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, *368*(6488), 290-296.
24. Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., ... & Li, Z. (2020). Precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. *BMC biology*, *18*(1), 1-14.
25. Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., ... & Chen, J. (2018). Base editing with a Cpf1–cytidine deaminase fusion. *Nature biotechnology*, *36*(4), 324-327.
26. Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., ... & Liu, D. R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C: G-to-T: A base editors with higher efficiency and product purity. *Science advances*, *3*(8), eaao4774.
27. Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, *42*(10), 6091-6105.
28. Mir, A., Edraki, A., Lee, J., & Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS chemical biology*, *13*(2), 357-365.
29. Amrani, N., Gao, X. D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K. E., Wu, T., Donohoue, P. D., Settle, A. H., Lied, A. M., McGovern, K., Fuller, C. K., Cameron, P., Fazzio, T. G., Zhu, L. J., Wolfe, S. A., & Sontheimer, E. J. (2018). NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome biology*, *19*(1), 1-25.
30. Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., Song, D. W., Lee., K. J., Jung, M. H., Kim, S., Kim, J. H., Kim, J. H., & Kim, J. S. (2017). *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nature communications*, *8*(1), 1-12.
31. Trasanidou, D., Gerós, A. S., Mohanraju, P., Nieuwenweg, A. C., Nobrega, F. L., & Staals, R. H. (2019). Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS microbiology letters*, *366*(9), fnz098.
32. Davidson, A. R., Lu, W. T., Stanley, S. Y., Wang, J., Mejdani, M., Trost, C. N., ... & Sontheimer, E. J. (2020). Anti-CRISPRs: protein inhibitors of CRISPR-Cas systems. *Annual review of biochemistry*, *89*, 309-332.
33. Marino, N. D., Pinilla-Redondo, R., Csörgő, B., & Bondy-Denomy, J. (2020). Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature methods*, *17*(5), 471-479.
34. Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E. J., Maxwell, K .L., & Davidson, A. R. (2016). Naturally occurring off-switches for CRISPR-Cas9. *Cell*, *167*(7), 1829-1838.
35. Harrington, L. B., Doxzen, K. W., Ma, E., Liu, J. J., Knott, G. J., Edraki, A., Garcia, B., Amrani, N., Chen, J. S., Cofsky, J. C., Kranzusch, P .J., Sontheimer, E. J., Davidson, A. R., Maxwell, K. L., & Doudna, J. A. (2017). A broad-spectrum inhibitor of CRISPR-Cas9. *Cell*, *170*(6), 1224-1233.
36. Mougiakos, I., Mohanraju, P., Bosma, E. F., Vrouwe, V., Bou, M. F., Naduthodi, M. I., Gussak, A., Brinkman, R. B. L., Kranenburg, R., & Van Der Oost, J. (2017). Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications*, *8*(1), 1-11.
37. Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C., & Doudna, J. A. (2017). A thermostable Cas9 with increased lifetime in human plasma. *Nature communications*, *8*(1), 1-8.
38. Song, G., Zhang, F., Zhang, X., Gao, X., Zhu, X., Fan, D., & Tian, Y. (2019). AcrIIA5 inhibits a broad range of Cas9 orthologs by preventing DNA target cleavage. *Cell reports*, *29*(9), 2579-2589.

39. Mathony, J., Harteveld, Z., Schmelas, C., Upmeier zu Belzen, J., Aschenbrenner, S., Sun, W., ... & Niopek, D. (2020). Computational design of anti-CRISPR proteins with improved inhibition potency. *Nature Chemical Biology*, *16*(7), 725-730.
40. Garcia, B., Lee, J., Edraki, A., Hidalgo-Reyes, Y., Erwood, S., Mir, A., ... & Davidson, A. R. (2019). Anti-CRISPR AcrIIA5 potently inhibits all Cas9 homologs used for genome editing. *Cell reports*, *29*(7), 1739-1746.
41. Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., & Copeland, N. G. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences*, *97*(11), 5978-5983.
42. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., & Court, D. L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nature protocols*, *4*(2), 206.
43. Arroyo-Olarte, R. D., Bravo Rodríguez, R., & Morales-Ríos, E. (2021). Genome editing in bacteria: CRISPR-Cas and beyond. *Microorganisms*, *9*(4), 844.
44. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, *31*(3), 233-239.
45. Jakočiūnas, T., Jensen, M. K., & Keasling, J. D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. *Metabolic engineering*, *34*, 44-59.
46. Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2015). Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in Escherichia coli. *Applied and environmental microbiology*, *81*(15), 5103-5114.
47. Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., & Yang, S. (2015). Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Applied and environmental microbiology*, *81*(7), 2506-2514.
48. Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. J., ... & Zhao, X. (2015). Metabolic engineering of Escherichia coli using CRISPR–Cas9 meditated genome editing. *Metabolic engineering*, *31*, 13-21.
49. Reisch, C. R., & Prather, K. L. (2015). The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Scientific reports*, *5*(1), 1-12.
50. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173-1183.
51. Zhigang, W., Smith, D. G., & Mosbaugh, D. W. (1991). Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene*, *99*(1), 31-37.
52. Porto, E. M., Komor, A. C., Slaymaker, I. M., & Yeo, G. W. (2020). Base editing: advances and therapeutic opportunities. *Nature Reviews Drug Discovery*, *19*(12), 839-859.
53. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., & Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. *Nature microbiology*, *3*(4), 423-429.
54. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., & Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research*, *41*(15), 7429-7437.
55. Bisaria, N., Jarmoskaite, I., & Herschlag, D. (2017). Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell systems*, *4*(1), 21-29.
56. Green, R., & Rogers, E. J. (2013). Transformation of chemically competent E. coli. *Methods Enzymol*, *529*, 329-336.
57. Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS microbiology reviews*, *34*(5), 779-796.
58. Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences*, *97*(12), 6640-6645.
59. Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: a method to quantify base editing from Sanger sequencing. *The CRISPR journal*, *1*(3), 239-250.

Acknowledgements

We would like to thank Dr. Thijs Nieuwkoop for his technical assistance with the flow cytometer, Dr. Sjoerd Creutzburg for offering the *E. coli_gfp* strain, and Dr. Asimenia Gavriilidou for assisting with the statistical analysis. D.T. is financially supported by the Alexander S. Onassis Foundation grant (F ZM 083-2/2018-2019). J.v.d.O. is supported by

the Dutch Research Council (NWO) through a TOP grant (714.015.001) and a Spinoza award (SPI 93-537), and by the European Research Council (ERC) by an Advanced Grant (ARGO AdG-834279). R.H.J.S. is supported by the Dutch Technology Foundation STW, which is part of The Netherlands Organization for Scientific Research (NWO) and which is partly funded by the Ministry of Economic Affairs.

Author contributions

D.T., R.H.J.S., and I.M. conceived this study and design of experiments. D.T., A.P., P.B., E.B., E.K., and A.D. conducted the experiments. P.M., R.v.K., J.v.d.O, R.H.J.S., and I.M. supervised this project. D.T., J.v.d.O, R.H.J.S., and I.M. wrote the manuscript.

Competing interests

J.v.d.O. is cofounder of NTrans Technologies, and member of the Scientific Advisory Board of NTrans Technologies and Hudson River Biotechnology. J.O. and R.S. are members of the Scientific Advisory Board of Scope Biosciences. R.v.K. and I.M. are employed by the commercial companies Corbion (Gorinchem, The Netherlands) and SNIPR Biome (Copenhagen, Denmark), respectively. D.T., J.O., R.S. and I.M. are inventors on CRISPR-Cas related patents/patent applications.

Supplementary tables & figures

Supplementary Table 1. Strains used in this study.

Strain	Description	Plasmid	Antibiotic resistance	Reference
E. coli DH5α	-	-	-	Lab stock
E. coli DH10b	-	-	-	NEB (#C3019H)
*E. coli*_gfp	Promoter lacUV5 and *gfp* gene integrated in the genome of *E. coli* DH10b	-	-	Lab stock
*E. coli*_gfp:pKD46	Promoter lacUV5 and *gfp* gene integrated in the genome of *E. coli* DH10b	pKD46	Ampicillin	This study
*E. coli*_gfp:pAcr	Promoter lacUV5 and *gfp* gene integrated in the genome of *E. coli* DH10b	pAcr	Ampicillin	This study

Supplementary Table 2. Plasmids used in this study.

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pACYC184	-	-	Negative control in killing, binding, and killing-inhibition assays. Positive control in binding assays. PCR template for pscr.	Lab stock
pKD46	-	-	Lambda Red recombineering genes (*exo*, *bet*, *gam*) for genome editing assays.	58
pUC19	-	-	Negative control in binding, and killing-inhibition assays. PCR template for pAcr.	Lab stock
pET_MBP_Geo_st	-	-	PCR template for pGCas9_gfp1/gfp2/gfp3/scr.	37
pPtet_Acr_Prha_ TCas9	-	-	PCR template for pGCas9_gfp1/gfp2/gfp3/scr, and pTCas9_gfp1/gfp2/gfp3/scr.	Lab stock
pPrha_TCas9 _gfp1	-	-	PCR template for pGCas9_gfp1.	Lab stock
pPrha_TCas9 _gfp2	-	-	PCR template for pGCas9_gfp2.	Lab stock
pPrha_TCas9 _gfp3	-	-	PCR template for pGCas9_gfp3.	Lab stock
pPrha_TCas9 _scr	-	-	PCR template for pGCas9_scr.	Lab stock
pPrha_GCas9 _gfp1	-	-	PCR template for pAcr.	Lab stock
pMK-RQ-_Acr_ coK12_coT12	-	-	PCR template for pAcr.	Lab stock
pThermoCas9i	-	-	PCR template for pdTCas9_gfp1/gfp2/gfp3/scr.	36
pScI_dCas9_CDA1_ UGI_LVA_noRNA	-	-	PCR template for all pdThermoTarget-AID, pdGeoTarget-AID, pAcrThermoTarget-AID, and pAcrGeoTarget-AID plasmids.	53
pscr	Gibson assembly	Fragment 1: PJ23119, sgRNA scr, and terminator from pAcrTCas9_scr amplified with primers BG15197 and BG15198	SgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in Western blot assays (negative control).	This study
		Fragment 2: pACYC184 backbone from pACYC184 amplified with primers BG15199 and BG15200		
pTCas9_gfp1	Gibson assembly	Fragment 1: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, pACYC184 backbone, and terminator from pGCas9_gfp1 amplified with primers BG13459 and BG13460	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 in killing, binding, and killing-inhibition assays. PCR template for pdTCas9_gfp1, and pAcrTCas9_gfp1.	This study
		Fragment 2: *thermocas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462		
pTCas9_gfp1.20	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.20 from pTCas9_scr amplified with primers BG17337 and BG17340	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 3 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 1.20, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17355 and BG17368		
pTCas9_gfp1.19	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.19 from pTCas9_scr amplified with primers BG17337 and BG17341	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 4 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 1.19, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17356 and BG17368		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pTCas9_gfp1.18	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.18 from pTCas9_scr amplified with primers BG17337 and BG17342	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in killing and binding assays.	This study
		Fragment 2: sgRNA 1.18, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17357 and BG17368		
pTCas9_gfp1.13	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.13 from pTCas9_scr amplified with primers BG17337 and BG18458	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.13, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18464 and BG17368		
pTCas9_gfp1.8	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.8 from pTCas9_scr amplified with primers BG17337 and BG18983	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.8, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18992 and BG17368		
pTCas9_gfp1.6	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.6 from pTCas9_scr amplified with primers BG17337 and BG18984	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.6, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18993 and BG17368		
pTCas9_gfp2	Gibson assembly	Fragment 1: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, pACYC184 backbone, and terminator from pGCas9_gfp2 amplified with primers BG13459 and BG13460	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 in killing, binding, and killing-inhibition assays. PCR template for pdTCas9_gfp2, and pAcrTCas9_gfp2.	This study
		Fragment 2: *thermocas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462		
pTCas9_gfp2.20	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.20 from pTCas9_scr amplified with primers BG17337 and BG17345	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 3 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 2.20, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17360 and BG17368		
pTCas9_gfp2.19	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.19 from pTCas9_scr amplified with primers BG17337 and BG17346	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 4 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 2.19, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17361 and BG17368		
pTCas9_gfp2.18	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.18 from pTCas9_scr amplified with primers BG17337 and BG17347	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in killing and binding assays.	This study
		Fragment 2: sgRNA 2.18, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17362 and BG17368		
pTCas9_gfp2.13	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.13 from pTCas9_scr amplified with primers BG17337 and BG18460	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: sgRNA 2.13, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18466 and BG17368		
pTCas9_gfp2.8	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.8 from pTCas9_scr amplified with primers BG17337 and BG18986	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.8, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18995 and BG17368		
pTCas9_gfp2.6	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.6 from pTCas9_scr amplified with primers BG17337 and BG18987	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.6, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18996 and BG17368		
pTCas9_gfp3	Gibson assembly	Fragment 1: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, pACYC184 backbone, and terminator from pGCas9_gfp3 amplified with primers BG13459 and BG13460	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated PlacUV5 in killing, binding, killing-inhibition assays. PCR template for pdTCas9_gfp3, pAcrTCas9_gfp3, and pHR_TCas9_gfp3.	This study
		Fragment 2: *thermocas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462		
pTCas9_gfp3.20	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.20 from pTCas9_scr amplified with primers BG17337 and BG17350	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 3 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 3.20, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17365 and BG17368		
pTCas9_gfp3.19	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.19 from pTCas9_scr amplified with primers BG17337 and BG17351	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 4 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 3.19, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17366 and BG17368		
pTCas9_gfp3.18	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.18 from pTCas9_scr amplified with primers BG17337 and BG17352	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in killing and binding assays.	This study
		Fragment 2: sgRNA 3.18, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG17367 and BG17368		
pTCas9_gfp3.13	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.13 from pTCas9_scr amplified with primers BG17337 and BG19690	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.13, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18468 and BG17368		
pTCas9_gfp3.8	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.8 from pTCas9_scr amplified with primers BG17337 and BG19693	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.8, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18998 and BG17368		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pTCas9_gfp3.6	Gibson assembly	Fragment 1: part of the terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.6 from pTCas9_scr amplified with primers BG17337 and BG19694 Fragment 2: sgRNA 3.6, terminator, pACYC184 backbone, and part of the terminator from pTCas9_scr amplified with primers BG18999 and BG17368	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pTCas9_scr	Gibson assembly	Fragment 1: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, pACYC184 backbone, and terminator from pGCas9_scr amplified with primers BG13459 and BG13460 Fragment 2: *thermocas9* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13461 and BG13462	ThermoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in killing, binding, and killing-inhibition assays. PCR template for pTCas9_gfp1.20/gfp1.19/gfp1.18/gfp1.13/gfp1.8/gfp1.6, pTCas_gfp2.20/gfp2.19/gfp2.18/gfp2.13/gfp2.8/gfp2.6, pTCas_gfp3.20/gfp3.19/gfp3.18/gfp3.13/gfp3.8/gfp3.6, pHis-TCas9_scr, pdTCas9_scr, pAcrTCas9_scr, and pHR_TCas9_scr.	This study
pHis-TCas9_scr	Gibson assembly	Fragment 1: His6 tag, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, partly *lacI* gene from pTCas9_scr amplified with primers BG24374 and BG24375 Fragment 2: partly of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, pACYC184 backbone, and His6 tag from pTCas9_scr amplified with primers BG24376 and BG24377	His6-ThermoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in Western blot assays.	This study
pGCas9_gfp1	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310 Fragment 2: 5' UTR, lac operator, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, PJ23119, sgRNA 1, terminator, pACYC184 backbone, and terminator from pPrha_TCas9 _gfp1 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 in killing, binding, and killing-inhibition assays. Moreover, PCR template for pTCas9_gfp1, pdGCas9_gfp1, and pAcrGCas9_gfp1.	This study
pGCas9_gfp1.20	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.20 from pGCas9_scr amplified with primers BG17337 and BG17340 Fragment 2: sgRNA 1.20, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17355 and BG17368	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 3 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
pGCas9_gfp1.19	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.19 from pGCas9_scr amplified with primers BG17337 and BG17341 Fragment 2: sgRNA 1.19, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17356 and BG17368	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 4 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
pGCas9_gfp1.18	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.18 from pGCas9_scr amplified with primers BG17337 and BG17342 Fragment 2: sgRNA 1.18, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17357 and BG17368	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in killing and binding assays.	This study
pGCas9_gfp1.13	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.13 from pGCas9_scr amplified with primers BG17337 and BG18458 Fragment 2: sgRNA 1.13, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18464 and BG17368	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pGCas9_gfp1.8	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.8 from pGCas9_scr amplified with primers BG17337 and BG18983 Fragment 2: sgRNA 1.8, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18992 and BG17368	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pGCas9_gfp1.6	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.6 from pGCas9_scr amplified with primers BG17337 and BG18984 Fragment 2: sgRNA 1.6, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18993 and BG17368	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pGCas9_gfp2	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310 Fragment 2: 5' UTR, lac operator, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312 Fragment 3: terminator, PJ23119, sgRNA 2, terminator, pACYC184 backbone, and terminator from pPrha_TCas9 _gfp2 amplified with primers BG13313 and BG13309	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 in killing, binding, and killing-inhibition assays. Moreover, PCR template for pTCas9_gfp2, pdGCas9_gfp2, and pAcrGCas9_gfp2.	This study
pGCas9_gfp2.20	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.20 from pGCas9_scr amplified with primers BG17337 and BG17345 Fragment 2: sgRNA 2.20, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17360 and BG17368	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 3 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
pGCas9_gfp2.19	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.19 from pGCas9_scr amplified with primers BG17337 and BG17346 Fragment 2: sgRNA 2.19, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17361 and BG17368	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 4 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
pGCas9_gfp2.18	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.18 from pGCas9_scr amplified with primers BG17337 and BG17347 Fragment 2: sgRNA 2.18, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17362 and BG17368	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in killing and binding assays.	This study
pGCas9_gfp2.13	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.13 from pGCas9_scr amplified with primers BG17337 and BG18460 Fragment 2: sgRNA 2.13, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18466 and BG17368	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pGCas9_gfp2.8	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.8 from pGCas9_scr amplified with primers BG17337 and BG18986 Fragment 2: sgRNA 2.8, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18995 and BG17368	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pGCas9_gfp2.6	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.6 from pGCas9_scr amplified with primers BG17337 and BG18987	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.6, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18996 and BG17368		
pGCas9_gfp3	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene in killing, binding, and killing-inhibition assays. Moreover, PCR template for pTCas9_gfp3, pdGCas9_gfp3, pAcrGCas9_gfp3, and pHR_GCas9_gfp9/gfp10.	This study
		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312		
		Fragment 3: terminator, PJ23119, sgRNA 3, terminator, pACYC184 backbone, and terminator from pPrha_TCas9 _gfp3 amplified with primers BG13313 and BG13309		
pGCas9_gfp3.20	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.20 from pGCas9_scr amplified with primers BG17337 and BG17350	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 3 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 3.20, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17365 and BG17368		
pGCas9_gfp3.19	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.19 from pGCas9_scr amplified with primers BG17337 and BG17351	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 4 consecutive mismatches at the 5' end of the spacer. Used in killing assays.	This study
		Fragment 2: sgRNA 3.19, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17366 and BG17368		
pGCas9_gfp3.18	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.18 from pGCas9_scr amplified with primers BG17337 and BG17352	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in killing and binding assays.	This study
		Fragment 2: sgRNA 3.18, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG17367 and BG17368		
pGCas9_gfp3.13	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.13 from pGCas9_scr amplified with primers BG17337 and BG19690	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.13, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18468 and BG17368		
pGCas9_gfp3.8	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.8 from pGCas9_scr amplified with primers BG17337 and BG19693	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.8, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18998 and BG17368		
pGCas9_gfp3.6	Gibson assembly	Fragment 1: part of the terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.6 from pGCas9_scr amplified with primers BG17337 and BG19694	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.6, terminator, pACYC184 backbone, and part of the terminator from pGCas9_scr amplified with primers BG18999 and BG17368		
pGCas9_scr	Gibson assembly	Fragment 1: *geocas9* gene from pET_MBP_Geo_st amplified with primers BG13307 and BG13310	GeoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in killing, binding, and killing-inhibition assays. Moreover, PCR template	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, and *lacI* gene from pPtet_Acr_Prha_ TCas9 amplified with primers BG13311 and BG13312	for pTCas9_scr, pGCas9_gfp1.20/gfp1.19/gfp1.18/gfp1.13/gfp1.8/ gfp1.6, pGCas9_gfp2.20/gfp2.19/gfp2.18/gfp2.13/gfp2.8/gfp2.6, pGCas9_gfp3.20/gfp3.19/gfp3.18/gfp3.13/gfp3.8/gfp3.6, pdGCas9_scr, pAcrGCas9_scr, and pHR_GCas9_scr.	
		Fragment 3: terminator, PJ23119, sgRNA scr, terminator, pACYC184 backbone, and terminator from pPrha_TCas9 _scr amplified with primers BG13313 and BG13309		
pdTCas9_gfp1	Gibson assembly	Fragment 1: *dthermocas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684	dThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 in binding assays.	This study
		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, pACYC184 backbone, and terminator from pTCas9_gfp1 amplified with primers BG13311 and BG13685		
pdTCas9_gfp1.18	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.18 from pdTCas9_scr amplified with primers BG17337 and BG17342	dThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.18, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG17357 and BG17368		
pdTCas9_gfp1.13	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.13 from pdTCas9_scr amplified with primers BG17337 and BG18458	dThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.13, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18464 and BG17368		
pdTCas9_gfp1.8	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.8 from pdTCas9_scr amplified with primers BG17337 and BG18983	dThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.8, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18992 and BG17368		
pdTCas9_gfp1.6	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.6 from pdTCas9_scr amplified with primers BG17337 and BG18984	dThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.6, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18993 and BG17368		
pdTCas9_gfp2	Gibson assembly	Fragment 1: *dthermocas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684	dThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 in binding assays.	This study
		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, pACYC184 backbone, and terminator from pTCas9_gfp2 amplified with primers BG13311 and BG13685		
pdTCas9_gfp2.18	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.18 from pdTCas9_scr amplified with primers BG17337 and BG17347	dThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.18, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG17362 and BG17368		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pdTCas9_gfp2.13	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.13 from pdTCas9_scr amplified with primers BG17337 and BG18460 Fragment 2: sgRNA 2.13, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18466 and BG17368	dThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_gfp2.8	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.8 from pdTCas9_scr amplified with primers BG17337 and BG18986 Fragment 2: sgRNA 2.8, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18995 and BG17368	dThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_gfp2.6	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.6 from pdTCas9_scr amplified with primers BG17337 and BG18987 Fragment 2: sgRNA 2.6, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18996 and BG17368	dThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_gfp3	Gibson assembly	Fragment 1: *dthermocas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684 Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, pACYC184 backbone, and terminator from pTCas9_gfp3 amplified with primers BG13311 and BG13685	dThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated PlacUV5 in binding assays. PCR template for pdThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7/scr.	This study
pdTCas9_gfp3.18	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.18 from pdTCas9_scr amplified with primers BG17337 and BG17352 Fragment 2: sgRNA 3.18, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG17367 and BG17368	dThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_gfp3.13	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.13 from pdTCas9_scr amplified with primers BG17337 and BG19690 Fragment 2: sgRNA 3.13, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18468 and BG17368	dThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_gfp3.8	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.8 from pdTCas9_scr amplified with primers BG17337 and BG19693 Fragment 2: sgRNA 3.8, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18998 and BG17368	dThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_gfp3.6	Gibson assembly	Fragment 1: part of the terminator, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.6 from pdTCas9_scr amplified with primers BG17337 and BG19694 Fragment 2: sgRNA 3.6, terminator, pACYC184 backbone, and part of the terminator from pdTCas9_scr amplified with primers BG18999 and BG17368	dThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
pdTCas9_scr	Gibson assembly	Fragment 1: *dthermocas9* gene from pThermoCas9i amplified with primers BG12625 and BG13684	dThermoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in binding assays. PCR template for	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, pACYC184 backbone, and terminator from pTCas9_scr amplified with primers BG13311 and BG13685	gfp2.8/gfp2.6, pdTCas9_gfp3.18/gfp3.13/gfp3.8/gfp3.6, pHis-dCas9_scr, pAcrTCas9_scr, and pHR_TCas9_scr.	
pHis-dTCas9_scr	Gibson assembly	Fragment 1: His6 tag, *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, partly *lacI* gene from pdTCas9_scr amplified with primers BG24374 and BG24375	His6-dThermoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in Western blot assays.	This study
		Fragment 2: part of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, pACYC184 backbone, and His6 tag from pdTCas9_scr amplified with primers BG24376 and BG24377		
pdGCas9_gfp1	Gibson assembly	Fragment 1: part of the *dgeocas9* gene from pGCas9_gfp1 amplified with primers BG13680 and BG13681	dGeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 in binding assays.	This study
		Fragment 2: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, pACYC184 backbone, and terminator from pGCas9_gfp1 amplified with primers BG13682 and BG13683		
pdGCas9_gfp1.18	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.18 from pdGCas9_scr amplified with primers BG17337 and BG17342	dGeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.18, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG17357 and BG17368		
pdGCas9_gfp1.13	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.13 from pdGCas9_scr amplified with primers BG17337 and BG18458	dGeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.13, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18464 and BG17368		
pdGCas9_gfp1.8	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.8 from pdGCas9_scr amplified with primers BG17337 and BG18983	dGeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.8, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18992 and BG17368		
pdGCas9_gfp1.6	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 1.6 from pdGCas9_scr amplified with primers BG17337 and BG18984	dGeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 1.6, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18993 and BG17368		
pdGCas9_gfp2	Gibson assembly	Fragment 1: part of the *dgeocas9* gene from pGCas9_gfp2 amplified with primers BG13680 and BG13681	dGeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 in binding assays.	This study
		Fragment 2: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, pACYC184 backbone, and terminator from pGCas9_gfp2 amplified with primers BG13682 and BG13683		
pdGCas9_gfp2.18	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.18 from pdGCas9_scr amplified with primers BG17337 and BG17347	dGeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.18, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG17362 and BG17368		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pdGCas9_gfp2.13	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.13 from pdGCas9_scr amplified with primers BG17337 and BG18460	dGeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.13, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18466 and BG17368		
pdGCas9_gfp2.8	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.8 from pdGCas9_scr amplified with primers BG17337 and BG18986	dGeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.8, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18995 and BG17368		
pdGCas9_gfp2.6	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 2.6 from pdGCas9_scr amplified with primers BG17337 and BG18987	dGeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 2.6, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18996 and BG17368		
pdGCas9_gfp3	Gibson assembly	Fragment 1: part of the *dgeocas9* gene from pGCas9_gfp3 amplified with primers BG13680 and BG13681	dGeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated PlacUV5 in binding assays. PCR template for pdGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14/scr.	This study
		Fragment 2: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, pACYC184 backbone, and terminator from pGCas9_gfp3 amplified with primers BG13682 and BG13683		
pdGCas9_gfp3.18	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.18 from pdGCas9_scr amplified with primers BG17337 and BG17352	dGeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 5 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.18, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG17367 and BG17368		
pdGCas9_gfp3.13	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.13 from pdGCas9_scr amplified with primers BG17337 and BG19690	dGeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 10 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.13, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18468 and BG17368		
pdGCas9_gfp3.8	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.8 from pdGCas9_scr amplified with primers BG17337 and BG19693	dGeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 15 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.8, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18998 and BG17368		
pdGCas9_gfp3.6	Gibson assembly	Fragment 1: part of the terminator, *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and part of the sgRNA 3.6 from pdGCas9_scr amplified with primers BG17337 and BG19694	dGeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and harboring 17 consecutive mismatches at the 5' end of the spacer. Used in binding assays.	This study
		Fragment 2: sgRNA 3.6, terminator, pACYC184 backbone, and part of the terminator from pdGCas9_scr amplified with primers BG18999 and BG17368		
pdGCas9_scr	Gibson assembly	Fragment 1: part of the *dgeocas9* gene from pGCas9_scr amplified with primers BG13680 and BG13681	dGeoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in binding assays. PCR template for	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, pACYC184 backbone, and terminator from pGCas9_scr amplified with primers BG13682 and BG13683	pdGCas9_gfp1.18/gfp1.13/gfp1.8/gfp1.6, pdGCas9_gfp2.18/gfp2.13/ gfp2.8/gfp2.6, and pdGCas9_gfp3.18/gfp3.13/gfp3.8/gfp3.6.	
pHR_TCas9_gfp3	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp3) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. PCR template for pHR_TCas9_gfp4/gfp5/gfp6/gfp7/gfp8, and pHR_AcrTCas9_gfp3/gfp4/gfp5/gfp6/gfp7/gfp8.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, and part of the *lacI* gene from pAcrTCas9_gfp3 amplified with primers BG15040 and BG14002		
		Fragment 4: part of the *lacI* gene, terminator, PJ23119, sgRNA gfp3, terminator, and pACYC184 backbone from pAcrTCas9_gfp3 amplified with primers BG12722 and BG14003		
pHR_TCas9_gfp4	Gibson assembly	Fragment 1: sgRNA gfp4, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_TCas9_gfp3 amplified with primers BG14530 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp4) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the thermocas9 gene from pHR_TCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp4 from pHR_TCas9_gfp3 amplified with primers BG16902 and BG14529		
pHR_TCas9_gfp5	Gibson assembly	Fragment 1: sgRNA gfp5, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_TCas9_gfp3 amplified with primers BG14523 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp5) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the *thermocas9* gene from pHR_TCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp5 from pHR_TCas9_gfp3 amplified with primers BG16902 and BG14522		
pHR_TCas9_gfp6	Gibson assembly	Fragment 1: sgRNA gfp6, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_TCas9_gfp3 amplified with primers BG16898 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp6) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the *thermocas9* gene from pHR_TCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp6 from pHR_TCas9_gfp3 amplified with primers BG16902 and BG16903		
pHR_TCas9_gfp7	Gibson assembly	Fragment 1: sgRNA gfp7, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_TCas9_gfp3 amplified with primers BG14692 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp7) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene,		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		terminator, and part of the *thermocas9* gene from pHR_TCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp7 from pHR_TCas9_gfp3 amplified with primers BG16902 and BG14691		
pHR_TCas9_gfp8	Gibson assembly	Fragment 1: sgRNA gfp8, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_TCas9_gfp3 amplified with primers BG16899 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp8) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the thermocas9 gene from pHR_TCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp8 from pHR_TCas9_gfp3 amplified with primers BG16902 and BG16904		
pHR_TCas9_scr	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in genome editing assays (negative control).	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, and part of the *lacI* gene from pAcrTCas9_scr amplified with primers BG15040 and BG14002		
		Fragment 4: part of the *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, and pACYC184 backbone from pAcrTCas9_scr amplified with primers BG12722 and BG14003		
pHR_GCas9_gfp5	Gibson assembly	Fragment 1: sgRNA gfp5, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_GCas9_scr amplified with primers BG14523 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer gfp5) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the *geocas9* gene from pHR_GCas9_scr amplified with primers BG13019 and BG16905		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp5 from pHR_GCas9_scr amplified with primers BG16906 and BG14522		
pHR_GCas9_gfp6	Gibson assembly	Fragment 1: sgRNA gfp6, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_GCas9_scr amplified with primers BG16898 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer gfp6) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the *geocas9* gene from pHR_GCas9_scr amplified with primers BG13019 and BG16905		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp6 from pHR_GCas9_scr amplified with primers BG16906 and BG16903		
pHR_GCas9_gfp7	Gibson assembly	Fragment 1: sgRNA gfp7, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_GCas9_scr amplified with primers BG14692 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5).	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the *geocas9* gene from pHR_GCas9_scr amplified with primers BG13019 and BG16905	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer gfp7) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp7 from pHR_GCas9_scr amplified with primers BG16906 and BG14691		
pHR_GCas9_gfp8	Gibson assembly	Fragment 1: sgRNA gfp8, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_GCas9_scr amplified with primers BG16899 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer gfp8) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, terminator, and part of the *geocas9* gene from pHR_GCas9_scr amplified with primers BG13019 and BG16905		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp8 from pHR_GCas9_scr amplified with primers BG16906 and BG16904		
pHR_GCas9_gfp9	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer gfp9) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from E. coli_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp9 from pGCas9_gfp3 amplified with primers BG15040 and BG15041		
		Fragment 4: sgRNA gfp9, terminator, and pACYC184 backbone from pGCas9_gfp3 amplified with primers BG15042 and BG14003		
pHR_GCas9_gfp10	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer gfp10) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp10 from pGCas9_gfp3 amplified with primers BG15040 and BG15043		
		Fragment 4: sgRNA gfp10, terminator, and pACYC184 backbone from pGCas9_gfp3 amplified with primers BG15044 and BG14003		
pHR_GCas9_scr	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). GeoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in genome editing assays (negative control). PCR template for pHR_GCas9_gfp5/gfp6/gfp7/gfp8.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, and part of the *lacI* gene from pGCas9_scr amplified with primers BG15040 and BG14002		
		Fragment 4: part of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and pACYC184 backbone from pGCas9_scr amplified with primers BG12722 and BG14003		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pHR_AcrTCas9_gfp3	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp3) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells. PCR template for pHR_AcrTCas9_gfp4/gfp5/gfp6/gfp7/gfp8.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: Prha, *acriic1* gene, terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, and part of the *lacI* gene from pAcrTCas9_gfp3 amplified with primers BG15039 and BG14002		
		Fragment 4: part of the *lacI* gene, terminator, PJ23119, sgRNA gfp3, terminator, and pACYC184 backbone from pAcrTCas9_gfp3 amplified with primers BG12722 and BG14003		
pHR_AcrTCas9_gfp4	Gibson assembly	Fragment 1: sgRNA gfp4, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrTCas9_gfp3 amplified with primers BG14530 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp4) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the *thermocas9* gene from pHR_AcrTCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp4 from pHR_AcrTCas9_gfp3 amplified with primers BG16902 and BG14529		
pHR_AcrTCas9_gfp5	Gibson assembly	Fragment 1: sgRNA gfp5, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrTCas9_gfp3 amplified with primers BG14523 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp5) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the *thermocas9* gene from pHR_AcrTCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp5 from pHR_AcrTCas9_gfp3 amplified with primers BG16902 and BG14522		
pHR_AcrTCas9_gfp6	Gibson assembly	Fragment 1: sgRNA gfp6, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrTCas9_gfp3 amplified with primers BG16898 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp6) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the *thermocas9* gene from pHR_AcrTCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp6 from pHR_AcrTCas9_gfp3 amplified with primers BG16902 and BG16903		
pHR_AcrTCas9_gfp7	Gibson assembly	Fragment 1: sgRNA gfp7, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrTCas9_gfp3 amplified with primers BG14692 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp7) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene,		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Prha, *acriic1* gene, terminator, and part of the *thermocas9* gene from pHR_AcrTCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp7 from pHR_AcrTCas9_gfp3 amplified with primers BG16902 and BG14691		
pHR_AcrTCas9_gfp8	Gibson assembly	Fragment 1: sgRNA gfp8, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrTCas9_gfp3 amplified with primers BG16899 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp8) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the *thermocas9* gene from pHR_AcrTCas9_gfp3 amplified with primers BG13019 and BG16901		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp8 from pHR_AcrTCas9_gfp3 amplified with primers BG16902 and BG16904		
pHR_AcrTCas9_scr	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer scr) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: Prha, *acriic1* gene, terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, and part of the *lacI* gene from pAcrTCas9_scr amplified with primers BG15039 and BG14002		
		Fragment 4: part of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and pACYC184 backbone from pAcrTCas9_scr amplified with primers BG12722 and BG14003		
pHR_AcrGCas9_gfp5	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp5) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells. PCR template for pHR_AcrGCas9_gfp6/gfp8.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: Prha, *acriic1* gene, terminator, geocas9 gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and sgRNA gfp5 from pAcrGCas9_gfp3 amplified with primers BG15039 and BG14522		
		Fragment 4: sgRNA gfp5, terminator, and pACYC184 backbone from pAcrGCas9_gfp3 amplified with primers BG14523 and BG14003		
pHR_AcrGCas9_gfp6	Gibson assembly	Fragment 1: sgRNA gfp6, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrGCas9_gfp5 amplified with primers BG16898 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp6) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells. PCR template for pHR_AcrGCas9_gfp7.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the *geocas9* gene from pHR_AcrGCas9_gfp5 amplified with primers BG13019 and BG16905		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp6 from pHR_AcrGCas9_gfp5 amplified with primers BG16906 and BG16903		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pHR_AcrGCas9_gfp7	Gibson assembly	Fragment 1: sgRNA gfp7, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrGCas9_gfp6 amplified with primers BG14692 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp7) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the *geocas9* gene from pHR_AcrGCas9_gfp6 amplified with primers BG13019 and BG16905		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp7 from pHR_AcrGCas9_gfp6 amplified with primers BG16906 and BG14691		
pHR_AcrGCas9_gfp8	Gibson assembly	Fragment 1: sgRNA gfp8, terminator, pACYC184 backbone, and part of the homologous recombination arm upstream of the genome-integrated PlacUV5 from pHR_AcrGCas9_gfp5 amplified with primers BG16899 and BG16900	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp8) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: part of the homologous recombination arm upstream of the genome-integrated PlacUV5, homologous recombination arm downstream of the genome-integrated *gfp* gene, Prha, *acriic1* gene, terminator, and part of the geocas9 gene from pHR_AcrGCas9_gfp5 amplified with primers BG13019 and BG16905		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of sgRNA gfp8 from pHR_AcrGCas9_gfp5 amplified with primers BG16906 and BG16904		
pHR_AcrGCas9_gfp9	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp9) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: Prha, *acriic1* gene, terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and sgRNA gfp9 from pAcrGCas9_gfp3 amplified with primers BG15039 and BG15041		
		Fragment 4: sgRNA gfp9, terminator, and pACYC184 backbone from pAcrTCas9_scr amplified with primers BG15042 and BG14003		
pHR_AcrGCas9_gfp10	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer gfp10) targeting the *gfp* gene for counter-selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		
		Fragment 3: Prha, *acriic1* gene, terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, terminator, PJ23119, and sgRNA gfp10, from pAcrTCas9_scr amplified with primers BG15039 and BG15043		
		Fragment 4: sgRNA gfp10, terminator, and pACYC184 backbone from pAcrTCas9_scr amplified with primers BG15044 and BG14003		
pHR_AcrGCas9_scr	Gibson assembly	Fragment 1: homologous recombination arm upstream of the genome-integrated PlacUV5 from *E. coli*_gfp amplified with primers BG15035 and BG15036	Upstream and downstream homologous recombination arms for the deletion of the genome-integrated *gfp* gene and its promoter (PlacUV5). ThermoCas9 and sgRNA (spacer scr) targeting the *gfp* gene for counter-	This study
		Fragment 2: homologous recombination arm downstream of the genome-integrated *gfp* gene from *E. coli*_gfp amplified with primers BG15037 and BG15038		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 3: Prha, *acriic1* gene, terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, and part of the *lacI* gene from pAcrTCas9_scr amplified with primers BG15039 and BG14002	selection of the non-edited cells in genome editing assays. AcrIIC1Nme inhibiting the ThermoCas9-mediated counter-selection of unedited cells.	
		Fragment 4: part of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and pACYC184 backbone from pAcrTCas9_scr amplified with primers BG12722 and BG14003		
pdThermoTarget-AID_gfp4	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp4) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp4 from pdTCas9_gfp3 amplified with primers BG14521 and BG14529		
		Fragment 4: sgRNA gfp4, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG14530 and BG14524		
pdThermoTarget-AID_gfp11	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp11) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp11 from pdTCas9_gfp3 amplified with primers BG14521 and BG17241		
		Fragment 4: sgRNA gfp11, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG17242 and BG14524		
pdThermoTarget-AID_gfp5	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp5) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp5 from pdTCas9_gfp3 amplified with primers BG14521 and BG14522		
		Fragment 4: sgRNA gfp5, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG14523 and BG14524		
pdThermoTarget-AID_gfp12	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp12) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp12 from pdTCas9_gfp3 amplified with primers BG14521 and BG14525		
		Fragment 4: sgRNA gfp12, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG14526 and BG14524		
pdThermoTarget-AID_gfp8	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp8) targeting the genome-integrated *gfp* gene in base-editing assays.	This study

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp8 from pdTCas9_gfp3 amplified with primers BG14521 and BG16904		
		Fragment 4: sgRNA gfp8, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG16899 and BG14524		
pdThermoTarget-AID_gfp7	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp7) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp7 from pdTCas9_gfp3 amplified with primers BG14521 and BG14691		
		Fragment 4: sgRNA gfp7, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG14692 and BG14524		
pdThermoTarget-AID_scr	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	dThermoCas9-PmCDA1-UGI-LVA and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in base-editing assays.	This study
		Fragment 2: part of the *dthermocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *dthermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA scr from pdTCas9_gfp3 amplified with primers BG14521 and BG14543		
		Fragment 4: sgRNA scr, terminator, pACYC184 backbone, terminator, and stop codon from pdTCas9_gfp3 amplified with primers BG14544 and BG14524		
pdGeoTarget-AID_gfp5	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp5) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp5 from pdGCas9_gfp3 amplified with primers BG14538 and BG14522		
		Fragment 4: sgRNA gfp5, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG14542 and BG14524		
pdGeoTarget-AID_gfp12	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp12) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp12 from pdGCas9_gfp3 amplified with primers BG14538 and BG14525		
		Fragment 4: sgRNA gfp12, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG14526 and BG14524		
pdGeoTarget-AID_gfp7	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp7) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp7 from pdGCas9_gfp3 amplified with primers BG14538 and BG14691		
		Fragment 4: sgRNA gfp7, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG14692 and BG14524		
pdGeoTarget-AID_gfp8	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp8) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp8 from pdGCas9_gfp3 amplified with primers BG14538 and BG16904		
		Fragment 4: sgRNA gfp8, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG16899 and BG14524		
pdGeoTarget-AID_gfp13	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp13) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp13 from pdGCas9_gfp3 amplified with primers BG14538 and BG14541		
		Fragment 4: sgRNA gfp13, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG14542 and BG14524		
pdGeoTarget-AID_gfp14	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVA and sgRNA (spacer gfp14) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp14 from pdGCas9_gfp3 amplified with primers BG14538 and BG17551		
		Fragment 4: sgRNA gfp14, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG17552 and BG14524		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pdGeoTarget-AID_scr	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	dGeoCas9-PmCDA1-UGI-LVAand sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in base-editing assays.	This study
		Fragment 2: part of the *dgeocas9* gene lacking stop codon from pdTCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *dgeocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA scr from pdGCas9_gfp3 amplified with primers BG14538 and BG14543		
		Fragment 4: sgRNA scr, terminator, pACYC184 backbone, terminator, and stop codon from pdGCas9_gfp3 amplified with primers BG14544 and BG14524		
pAcr	Gibson assembly	Fragment 1: Prha, and 5' UTR from pPrha_GCas9_gfp1 amplified with primers BG13427 and BG13428	AcrIIC1Nme inhibiting ThermoCas9 or GeoCas9 in killing-inhibition (two-plasmid approach) and binding assays. PCR template for pAcrTCas9_gfp1, pAcrTCas9_gfp2, pAcrTCas9_gfp3, pAcrTCas9_scr, pAcrGCas9_gfp1, pAcrGCas9_gfp2, pAcrGCas9_gfp3, and pAcrGCas9_scr.	This study
		Fragment 2: *acriic1* gene from pMK-RQ-_Acr_coK12_coT12 amplified with primers BG13429 and BG13430		
		Fragment 3: T7 terminator from annealed oligonucleotides BG13431 and BG13432		
		Fragment 4: pUC19 backbone amplified from pUC19 with primers BG13433 and BG13434		
pAcrTCas9_gfp1	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	ThermoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and AcrIIC1Nme inhibiting ThermoCas9, in killing-inhibition assays (single-plasmid approach).	This study
		Fragment 2: terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pTCas9_gfp1 amplified with primers BG12705 and BG14002		
		Fragment 3: Part of the *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, and pACYC184 backbone from pTCas9_gfp1 amplified with primers BG12722 and BG14003		
pAcrTCas9_gfp2	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	ThermoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and AcrIIC1Nme inhibiting ThermoCas9, in killing-inhibition assays (single-plasmid approach).	This study
		Fragment 2: terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pTCas9_gfp2 amplified with primers BG12705 and BG14002		
		Fragment 3: Part of the *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, and pACYC184 backbone from pTCas9_gfp2 amplified with primers BG12722 and BG14003		
pAcrTCas9_gfp3	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	ThermoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and AcrIIC1Nme inhibiting ThermoCas9, in killing-inhibition assays (single-plasmid approach). PCR template for pTCas9_gfp3, and pAcrThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7/scr.	This study
		Fragment 2: terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pTCas9_gfp3 amplified with primers BG12705 and BG14002		
		Fragment 3: Part of the *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, and pACYC184 backbone from pTCas9_gfp3 amplified with primers BG12722 and BG14003		
pAcrTCas9_scr	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	ThermoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp and AcrIIC1Nme inhibiting ThermoCas9, in killing-inhibition assays (single-plasmid approach). PCR template for pTCas9_scr and pscr.	This study
		Fragment 2: terminator, *thermocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pTCas9_scr amplified with primers BG12705 and BG14002		
		Fragment 3: Part of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and pACYC184 backbone from pTCas9_scr amplified with primers BG12722 and BG14003		
pAcrGCas9_gfp1	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	GeoCas9 and sgRNA (spacer gfp1) targeting the genome-integrated PlacUV5 and AcrIIC1Nme inhibiting GeoCas9, in killing-inhibition assays (single-plasmid approach).	This study
		Fragment 2: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pGCas9_gfp1 amplified with primers BG12705 and BG14002		
		Fragment 3: Part of the *lacI* gene, terminator, PJ23119, sgRNA 1, terminator, and pACYC184 backbone from pGCas9_gfp1 amplified with primers BG12722 and BG14003		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pAcrGCas9_gfp2	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	GeoCas9 and sgRNA (spacer gfp2) targeting the genome-integrated PlacUV5 and AcrIIC1Nme inhibiting GeoCas9, in killing-inhibition assays (single-plasmid approach).	This study
		Fragment 2: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pGCas9_gfp2 amplified with primers BG12705 and BG14002		
		Fragment 3: Part of the *lacI* gene, terminator, PJ23119, sgRNA 2, terminator, and pACYC184 backbone from pGCas9_gfp2 amplified with primers BG12722 and BG14003		
pAcrGCas9_gfp3	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	GeoCas9 and sgRNA (spacer gfp3) targeting the genome-integrated *gfp* gene and AcrIIC1Nme inhibiting GeoCas9, in killing-inhibition assays (single-plasmid approach). PCR template for pGCas9, pHR_AcrGCas9_gfp5/gfp6/gfp7/gfp8/gfp9/gfp10.and pAcrGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14/scr.	This study
		Fragment 2: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pGCas9_gfp3 amplified with primers BG12705 and BG14002		
		Fragment 3: part of the *lacI* gene, terminator, PJ23119, sgRNA 3, terminator, and pACYC184 backbone from pGCas9_gfp3 amplified with primers BG12722 and BG14003		
pAcrGCas9_scr	Gibson assembly	Fragment 1: Prha, and *acriic1* gene from pAcr amplified with primers BG14000 and BG14001	GeoCas9 and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp and AcrIIC1Nme inhibiting GeoCas9, in killing-inhibition assays (single-plasmid approach). PCR template for pHR_AcrGCas9_scr.	This study
		Fragment 2: terminator, *geocas9* gene, 5' UTR, lac operator, Ptet, pACYC184 backbone, PlacI, and part of the *lacI* gene from pGCas9_scr amplified with primers BG12705 and BG14002		
		Fragment 3: part of the *lacI* gene, terminator, PJ23119, sgRNA scr, terminator, and pACYC184 backbone from pGCas9_scr amplified with primers BG12722 and BG14003		
pAcrThermoTarget-AID_gfp4	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp4) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp4 from pAcrTCas9_gfp3 amplified with primers BG14521 and BG14529		
		Fragment 4: sgRNA gfp4, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG14530 and BG14524		
pAcrThermoTarget-AID_gfp11	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp11) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp11 from pAcrTCas9_gfp3 amplified with primers BG14521 and BG17241		
		Fragment 4: sgRNA gfp11, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG17242 and BG14524		
pAcrThermoTarget-AID_gfp5	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp5) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp5 from pAcrTCas9_gfp3 amplified with primers BG14521 and BG14522		
		Fragment 4: sgRNA gfp5, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG14523 and BG14524		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pAcrThermoTarget-AID_gfp12	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp12) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp12 from pAcrTCas9_gfp3 amplified with primers BG14521 and BG14525		
		Fragment 4: sgRNA gfp12, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG14526 and BG14524		
pAcrThermoTarget-AID_gfp8	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp8) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp8 from pAcrTCas9_gfp3 amplified with primers BG14521 and BG16904		
		Fragment 4: sgRNA gfp8, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG16899 and BG14524		
pAcrThermoTarget-AID_gfp7	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp7) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp7 from pAcrTCas9_gfp3 amplified with primers BG14521 and BG14691		
		Fragment 4: sgRNA gfp7, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG14692 and BG14524		
pAcrThermoTarget-AID_scr	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14518	AcrIIC1:ThermoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in base-editing assays.	This study
		Fragment 2: part of the *thermocas9* gene lacking stop codon from pAcrTCas9_gfp3 amplified with primers BG14519 and BG14520		
		Fragment 3: part of the *thermocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA scr from pAcrTCas9_gfp3 amplified with primers BG14521 and BG14543		
		Fragment 4: sgRNA scr, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrTCas9_gfp3 amplified with primers BG14544 and BG14524		
pAcrGeoTarget-AID_gfp5	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp5) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp5 from pAcrGCas9_gfp3 amplified with primers BG14538 and BG14522		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 4: sgRNA gfp5, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG14523 and BG14524		
pAcrGeoTarget-AID_gfp12	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp12) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp12 from pAcrGCas9_gfp3 amplified with primers BG14538 and BG14525		
		Fragment 4: sgRNA gfp12, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG14526 and BG14524		
pAcrGeoTarget-AID_gfp7	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp7) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp7 from pAcrGCas9_gfp3 amplified with primers BG14538 and BG14691		
		Fragment 4: sgRNA gfp7, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG14692 and BG14524		
pAcrGeoTarget-AID_gfp8	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp8) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp8 from pAcrGCas9_gfp3 amplified with primers BG14538 and BG16904		
		Fragment 4: sgRNA gfp8, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG16899 and BG14524		
pAcrGeoTarget-AID_gfp13	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp13) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp13 from pAcrGCas9_gfp3 amplified with primers BG14538 and BG14541		
		Fragment 4: sgRNA gfp13, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG14542 and BG14524		
pAcrGeoTarget-AID_gfp14	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer gfp14) targeting the genome-integrated *gfp* gene in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA gfp14 from pAcrGCas9_gfp3 amplified with primers BG14538 and BG17551		
		Fragment 4: sgRNA gfp14, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG17552 and BG14524		
pAcrGeoTarget-AID_scr	Gibson assembly	Fragment 1: terminator, stop codon, LVA tag, linker, *ugi* gene, linker, *pmcda1* gene, and linker from pScI_dCas9_CDA1_UGI_LVA_noRNA amplified with primers BG15946 and BG14545	AcrIIC1:GeoCas9-PmCDA1-UGI-LVA complex and sgRNA (spacer scr) non-targeting the genome of *E. coli*_gfp in base-editing assays.	This study
		Fragment 2: part of the *geocas9* gene lacking stop codon from pAcrGCas9_gfp3 amplified with primers BG14536 and BG14537		
		Fragment 3: part of the *geocas9* gene, 5' UTR, lac operator, Ptet, PlacI, *lacI* gene, PJ23119, and part of the sgRNA scr from pAcrGCas9_gfp3 amplified with primers BG14538 and BG14543		
		Fragment 4: sgRNA scr, terminator, pACYC184 backbone, Prha, *acriic1* gene, terminator, and stop codon from pAcrGCas9_gfp3 amplified with primers BG14544 and BG14524		

Supplementary Table 3. Spacers used in this study.

Spacer ID	Target strand	Targeting location in the genome (5' -> 3')	3' PAM (5' -> 3')	Sequence of spacer (23 nt) (5' -> 3')	Protein	Application
gfp1	+	PlacUV5 (16 - 38)	NNNNCACA	TCCGGCTCGTATAATGTGTGGGG	(d)ThermoCas9, (d)GeoCas9	Killing, binding, and killing-inhibition assays
gfp1.20	+	PlacUV5 (16 - 38)	NNNNCACA	AGGGGCTCGTATAATGTGTGGGG	ThermoCas9, GeoCas9	Killing assays
gfp1.19	+	PlacUV5 (16 - 38)	NNNNCACA	AGGCGCTCGTATAATGTGTGGGG	ThermoCas9, GeoCas9	Killing assays
gfp1.18	+	PlacUV5 (16 - 38)	NNNNCACA	AGGCCCTCGTATAATGTGTGGGG	(d)ThermoCas9, (d)GeoCas9	Killing and binding assays
gfp1.13	+	PlacUV5 (16 - 38)	NNNNCACA	AGGCCGAGCAATAATGTGTGGGG	dThermoCas9, dGeoCas9	Binding assays
gfp1.8	+	PlacUV5 (16 - 38)	NNNNCACA	AGGCCGAGCATATTAGTGTGGGG	dThermoCas9, dGeoCas9	Binding assays
gfp1.6	+	PlacUV5 (16 - 38)	NNNNCACA	AGGCCGAGCATATTACAGTGGGG	dThermoCas9, dGeoCas9	Binding assays
gfp2	-	PlacUV5 (16 - 38)	NNNNCACA	TAGAGGGAAACCGTTGTGGTCTC	(d)ThermoCas9, (d)GeoCas9	Killing, binding, and killing-inhibition assays
gfp2.20	-	PlacUV5 (16 - 38)	NNNNCACA	ATCAGGGAAACCGTTGTGGTCTC	ThermoCas9, GeoCas9	Killing assays
gfp2.19	-	PlacUV5 (16 - 38)	NNNNCACA	ATCTGGGAAACCGTTGTGGTCTC	ThermoCas9, GeoCas9	Killing assays
gfp2.18	-	PlacUV5 (16 - 38)	NNNNCACA	ATCTCGGAAACCGTTGTGGTCTC	(d)ThermoCas9, (d)GeoCas9	Killing and binding assays
gfp2.13	-	PlacUV5 (16 - 38)	NNNNCACA	ATCTCCCTTTCCGTTGTGGTCTC	dThermoCas9, dGeoCas9	Binding assays
gfp2.8	-	PlacUV5 (16 - 38)	NNNNCACA	ATCTCCCTTTGGCAAGTGGTCTC	dThermoCas9, dGeoCas9	Binding assays
gfp2.6	-	PlacUV5 (16 - 38)	NNNNCACA	ATCTCCCTTTGGCAACAGGTCTC	dThermoCas9, dGeoCas9	Binding assays
gfp3	-	*gfp* gene (82 - 60)	NNNNCCAA	AGAATTTATGCCCATTCACATCC	(d)ThermoCas9, (d)GeoCas9	Killing, binding, killing-inhibition, genome editing assays
gfp3.20	-	*gfp* gene (82 - 60)	NNNNCCAA	TCTATTTATGCCCATTCACATCC	ThermoCas9, GeoCas9	Killing assays
gfp3.19	-	*gfp* gene (82 - 60)	NNNNCCAA	TCTTTTTATGCCCATTCACATCC	ThermoCas9, GeoCas9	Killing assays
gfp3.18	-	*gfp* gene (82 - 60)	NNNNCCAA	TCTTATTATGCCCATTCACATCC	(d)ThermoCas9, (d)GeoCas9	Killing and binding assays
gfp3.13	-	*gfp* gene (82 - 60)	NNNNCCAA	TCTTAAATACCCCATTCACATCC	dThermoCas9, dGeoCas9	Binding assays
gfp3.8	-	*gfp* gene (82 - 60)	NNNNCCAA	TCTTAAATACGGGTATCACATCC	dThermoCas9, dGeoCas9	Binding assays
gfp3.6	-	*gfp* gene (82 - 60)	NNNNCCAA	TCTTAAATACGGGTAAGACATCC	dThermoCas9, dGeoCas9	Binding assays
gfp4	+	*gfp* gene (605 - 627)	NNNNCCAA	CGACCCAGTCCGCTCTGTCGAAA	ThermoCas9	Genome- and base-editing assays
gfp5	+	*gfp* gene (315 - 337)	NNNNCGAA	TTATAAGACCCGAGCAGAAGTTA	ThermoCas9	Genome- and base-editing assays
gfp6	+	*gfp* gene (240 - 262)	NNNNCGAA	GCACGACTTCTTTAAGTCGGCTA	ThermoCas9	Genome editing assays
gfp7	-	*gfp* gene (170 - 148)	NNNNCAAA	CACGGCACCGGTAGTTTACCAGT	ThermoCas9	Genome- and base-editing assays
gfp8	+	*gfp* gene (438 - 460)	NNNNCAAA	TAGCCATAATGTATATATTACCG	ThermoCas9	Genome- and base-editing assays
gfp9	-	*gfp* gene (116 - 94)	NNNNGAAA	TAGGTAGCATCCCCTTCCCCCTC	GeoCas9	Genome editing assays
gfp10	+	*gfp* gene (207 - 229)	NNNNGAAA	GTGTTTCAGCCGTTATCCGGATC	GeoCas9	Genome editing assays
gfp11	-	*gfp* gene (593 - 571)	NNNNCCAA	TTGTCCGGCAACAGAACCGGCCC	(d)ThermoCas9-PmCDA1-UGI-LVA	Base-editing assays
gfp12	+	*gfp* gene (596 - 618)	NNNNCGAA	ACTATCTATCGACCCAGTCCGCT	(d)ThermoCas9-PmCDA1-UGI-LVA	Base-editing assays
gfp13	+	*gfp* gene (610 - 632)	NNNNGAAA	CAGTCCGCTCTGTCGAAAGATCC	(d)GeoCas9-PmCDA1-UGI-LVA	Base-editing assays
gfp14	+	*gfp* gene (597 - 619)	NNNNGAAA	CTATCTATCGACCCAGTCCGCTC	(d)GeoCas9-PmCDA1-UGI-LVA	Base-editing assays
scr	X	X	X	CTAGATCCGCAGTAACCCCATGG	(d)ThermoCas9, (d)GeoCas9	All bacterial assays

***Red color**: spacer-protospacer mismatch (nucleotide substitution in the spacer sequence)*
***Blue color**: spacer-protospacer match (identical sequences of spacer and protospacer)*

Supplementary Table 4. Primers/Oligonucleotides used in this study.

	Primer/ Oligo ID	Sequence (5' -> 3')	Function
Killing, binding, and killing-inhibition (two-plasmid approach) assays	BG13459	tatcaagaccgatttttatacttcatACACGGTGCCTGACTG	Construction of pTCas9_gfp1/gfp2/gfp3/scr.
	BG13460	cggggaaactatccgtccgttataaATCAGACAAAATGGCCTGCTTATG	
	BG13461	TTATAACGGACGGATAGTTTCCC	
	BG13462	ATGAAGTATAAAATCGGTCTTGATATCGG	
	BG13307	tcataagcaggccattttgtctgatTTAGTCACGAGTAGATTGCAGTG	Construction of pGCas9_gfp1/gfp2/gfp3/scr.
	BG13310	ttgctaacgcagtcaggcaccgtgtATGCGTTATAAGATTGGCCTGG	
	BG13311	ACACGGTGCCTGACTG	
	BG13312	tttttggtccatcagtcgattctgaTCACTGCCCGCTTTCC	
	BG13313	TCAGAATCGACTGATGGACC	
	BG13309	ATCAGACAAAATGGCCTGC	
Killing and binding assays	BG17337	CTTCATAAGCAGGCCATTTTGTC	Construction of pTCas9_gfp1.20/gfp1.19/gfp1.18/gfp1.13/gfp1.8/gfp1.6, pGCas9_gfp1.20/gfp1.19/gfp1.18/gfp1.13/gfp1.8/gfp1.6, pdTCas9_gfp1.18/gfp1.13/gfp1.8/gfp1.6, and pdGCas9_gfp1.18/gfp1.13/gfp1.8/gfp1.6.
	BG17340	ccccacacattatacgagcccctGCTAGCATTATACCTAGGACTGAG	
	BG17341	ccccacacattatacgagcgcctGCTAGCATTATACCTAGGACTGAG	
	BG17342	ccccacacattatacgagggcctGCTAGCATTATACCTAGGACTGAG	
	BG18458	ccccacacattattgctcggcctGCTAGCATTATACCTAGGACTGAG	
	BG18983	ccccacacatatatgctcggcctGCTAGCATTATACCTAGGACTGAG	
	BG18984	ccccactgatatatgctcggcctGCTAGCATTATACCTAGGACTGAG	
	BG17355	aggggctcgtataatgtgtggggGTCATAGTTCCCCTGAGATTATCG	
	BG17356	aggcgctcgtataatgtgtggggGTCATAGTTCCCCTGAGATTATCG	
	BG17357	aggccctcgtataatgtgtggggGTCATAGTTCCCCTGAGATTATCG	
	BG18464	aggccgagcaataatgtgtggggGTCATAGTTCCCCTGAGATTATCG	
	BG18992	aggccgagcatattagtgtggggGTCATAGTTCCCCTGAGATTATCG	
	BG18993	aggccgagcatattacagtggggGTCATAGTTCCCCTGAGATTATCG	
	BG17368	ATCAGACAAAATGGCCTGCTTATG	
Killing and binding assays	BG17337	CTTCATAAGCAGGCCATTTTGTC	Construction of pTCas9_gfp3.20/gfp3.19/gfp3.18/gfp3.13/gfp3.8/gfp3.6, pGCas9_gfp3.20/gfp3.19/gfp3.18/gfp3.13/gfp3.8/gfp3.6, pdTCas9_gfp3.18/gfp3.13/gfp3.8/gfp3.6, and pdGCas9_gfp3.18/gfp3.13/gfp3.8/gfp3.6.
	BG17345	gagaccacaacggtttcccctgatGCTAGCATTATACCTAGGACTGAG	
	BG17346	gagaccacaacggtttcccagatGCTAGCATTATACCTAGGACTGAG	
	BG17347	gagaccacaacggtttccgagatGCTAGCATTATACCTAGGACTGAG	
	BG18460	gagaccacaacggaaagggagatGCTAGCATTATACCTAGGACTGAG	
	BG18986	gagaccacttgccaaagggagatGCTAGCATTATACCTAGGACTGAG	
	BG18987	gagacctgttgccaaagggagatGCTAGCATTATACCTAGGACTGAG	
	BG17360	atcagggaaaccgttgtggtctcGTCATAGTTCCCCTGAGATTATCG	
	BG17361	atctgggaaaccgttgtggtctcGTCATAGTTCCCCTGAGATTATCG	
	BG17362	atctcggaaaccgttgtggtctcGTCATAGTTCCCCTGAGATTATCG	
	BG18466	atctccccttccgttgtggtctcGTCATAGTTCCCCTGAGATTATCG	
	BG18995	atctccccttggcaagtggtctcGTCATAGTTCCCCTGAGATTATCG	
	BG18996	atctccccttggcaacaggtctcGTCATAGTTCCCCTGAGATTATCG	
	BG17368	ATCAGACAAAATGGCCTGCTTATG	
	BG17337	CTTCATAAGCAGGCCATTTTGTC	Construction of pTCas9_gfp3.20/gfp3.19/gfp3.18/gfp3.13/gfp3.8/gfp3.6,
	BG17350	ggatgtgaatgggcataaataagaGCTAGCATTATACCTAGGACTGAG	
	BG17351	ggatgtgaatgggcataaaaagaGCTAGCATTATACCTAGGACTGAG	

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

Killing-inhibition (two-plasmid approach), and binding (two-plasmid	BG17352	ggatgtggaatggcatataagaGCTAGCATTATACCTAGGACTGAG	pGCas9_gfp3.20/gfp3.19/gfp
	BG19690	ggatgtgaatgggtatttaagaGCTAGCATTATACCTAGGACTGAG	3.18/gfp3.13/gfp3.8/gfp3.6,
	BG19693	ggatgtcttaccgtatttaagaGCTAGCATTATACCTAGGACTGAG	pdTCas9_gfp3.18/gfp3.13/gf
	BG19694	ggatgtcttaccgtatttaagaGCTAGCATTATACCTAGGACTGAG	p3.8/gfp3.6, and
	BG17365	tctatttatgcccattcacatccGTCATAGTTCCCCTGAGATTATCG	pdGCas9_gfp3.18/gfp3.13/gf
	BG17366	tcttttatgcccattcacatccGTCATAGTTCCCCTGAGATTATCG	p3.8/gfp3.6.
	BG17367	tcttattatgcccattcacatccGTCATAGTTCCCCTGAGATTATCG	
	BG18468	tcttaaatacccattcacatccGTCATAGTTCCCCTGAGATTATCG	
	BG18998	tcttaaatacgggtatcacatccGTCATAGTTCCCCTGAGATTATCG	
	BG18999	tcttaaatacgggtaagacatccGTCATAGTTCCCCTGAGATTATCG	
	BG17368	ATCAGACAAAATGGCCTGCTTATG	
	BG13427	aagcgggcagtgagcgcaacgcaatCACCACAATTCAGCAAATTGTG	Construction of pAcr.
	BG13428	ATGTTATATCTCCTTCTTATAGTTAAACAAAATTATTTCTAG	
	BG13429	ttaactataaagaaggagatatacatATGAATAAAACATATAAAATCGGAAAAAAATGCGG	
	BG13430	gctcagcggtggcggccgctctaggTCATAGTTCAACAAACTCCCAACATG	
Killing-inhibition assays (single-plasmid binding)	BG14000	atcttattaatcagataaaatatttCACCACAATTCAGCAAATTGTG	Construction of
	BG14001	gccatttttgtttaatcctgtcgacTCATAGTTCAACAAACTCCCAACATG	pAcrTCas9_gfp1/gfp2/gfp3/s
	BG12705	GTCGACAGGATTAAACAAAAATGGGCC	cr, and
	BG14002	CAATCAGCAACGACTGTTTGC	pAcrGCas9_gfp1/gfp2/gfp3/s
	BG12722	GCAAACAGTCGTTGCTGATTG	cr.
	BG14003	AAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCG	
Binding assays	**BG12625**	atcttattaatcagataaaatatttGTCGACAGGATTAAACAAAAATGGC	Construction of
	BG13684	ttgctaacgcagtcaggcaccgtgtATGAAGTATAAAATCGGTCTTGCTATCG	pdTCas9_gfp1/gfp2/gfp3/scr
	BG13311	ACACGGTGCCTGACTG	.
	BG13685	AAAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCG	
	BG13680	GATCACAGCATCTACTTCAACGTAAC	Construction of
	BG13681	GCCTGGCCATCGGTATTAC	pdGCas9_gfp1/gfp2/gfp3/scr
	BG13682	GTAATACCGATGGCCAGGC	.
	BG13683	GTTACGTTGAAGTAGATGCTGTGATC	
Western blot assays	BG15197	TTGACAGCTAGCTCAGTCC	Construction of pscr.
	BG15198	AAAAAAGACCTTGACGTTTTCCG	
	BG15199	ggcggaaaacgtcaaggttcttttttATGGAAGCCGGCGGCAC	
	BG15200	tacctaggactgagctagctgctgtcaaAAATATTTTATCTGATTAATAAGATGATCTCTTGAGATCGTTTTGGTC	
	BG24374	taatggtgatggtgatggtgTAACGGACGGATAGTTTCCC	Construction of pHis-TCas9_scr, and pHis-dTCas9_scr.
	BG24375	GTTGATCGGCGCGAGATTTAATC	
	BG24376	GATTAAATCTCGCGCCGATCAAC	
	BG24377	caccatcaccatcaccatTAAATCAGACAAAATGGCCTGCTTATG	
	BG15035	atcttattaatcagataaaatatttCGGGTTGACGTCGTGAC	Construction of
	BG15036	aatgataacatcattatactagtCTTGAAAAAACCCCGCCG	pHR_TCas9_gfp3/gfp4/gfp5/

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

Genome editing assays	BG15037	ACTAGTATAATGATGTGTTATCATTG	gfp6/gfp7/gfp8/scr, and pHR_AcrTCas9_gfp3/gfp4/gfp5/gfp6/gfp7/gfp8/scr.
	BG15038	GTAGTTGTTAACTTACCCG	
	BG15040	tatcagcgggtaagttaacaactacGTCGACAGGATTAAACAAAAATGGC	
	BG14002	CAATCAGCAACGACTGTTTGC	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG14003	AAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCG	
	BG14530	cgacccagtccgctctgtcgaaaGTCATAGTTCCCCTGAGATTATC	
	BG16900	GCAATTATCAAGCGTACCAACAAG	
	BG13019	CTTGTTGGTACGCTTGATAATTGC	
	BG16901	GCAAAGTGGTCAATGCCATTATC	
	BG16902	GATAATGGCATTGACCACTTTGC	
	BG14529	tttcgacagagcggactgggtcgGCTAGCATTATACCTAGGACTG	
	BG14523	ttataagacccgagcagaagttaGTCATAGTTCCCCTGAGATTATC	
	BG14522	taacttctgctcgggtcttataaGCTAGCATTATACCTAGGACTG	
	BG16898	gcacgacttctttaagtcggctaGTCATAGTTCCCCTGAGATTATCG	
	BG16903	tagccgacttaaagaagtcgtgcGCTAGCATTATACCTAGGACTGAG	
	BG14692	cacggcaccggtagtttaccagtGTCATAGTTCCCCTGAGATTATC	
	BG14691	actggtaaactaccggtgccgtgGCTAGCATTATACCTAGGACTG	
	BG16899	tagccataatgtatatattaccgGTCATAGTTCCCCTGAGATTATCG	
	BG16904	cggtaatatatacattatggctaGCTAGCATTATACCTAGGACTGAG	
	BG14523	ttataagacccgagcagaagttaGTCATAGTTCCCCTGAGATTATC	Construction of pHR_GCas9_gfp5/gfp6/gfp7/gfp8/gfp9/gfp10/scr, and pHR_AcrGCas9_gfp5/gfp6/gfp7/gfp8/gfp9/gfp10/scr.
	BG16900	GCAATTATCAAGCGTACCAACAAG	
	BG13019	CTTGTTGGTACGCTTGATAATTGC	
	BG16905	CCTGGAACTGGATGCATACC	
	BG16906	GGTATGCATCCAGTTCCAGG	
	BG14522	taacttctgctcgggtcttataaGCTAGCATTATACCTAGGACTG	
	BG16898	gcacgacttctttaagtcggctaGTCATAGTTCCCCTGAGATTATCG	
	BG16903	tagccgacttaaagaagtcgtgcGCTAGCATTATACCTAGGACTGAG	
	BG14692	cacggcaccggtagtttaccagtGTCATAGTTCCCCTGAGATTATC	
	BG14691	actggtaaactaccggtgccgtgGCTAGCATTATACCTAGGACTG	
	BG16899	tagccataatgtatatattaccgGTCATAGTTCCCCTGAGATTATCG	
	BG16904	cggtaatatatacattatggctaGCTAGCATTATACCTAGGACTGAG	
	BG15035	atcttattaatcagataaaatatttCGGGTTGACGTCGTGAC	
	BG15036	aatgataacacatcattatactagtCTTGAAAAAACCCCGCCG	
	BG15037	ACTAGTATAATGATGTGTTATCATTG	
	BG15038	GTAGTTGTTAACTTACCCG	
	BG15039	tatcagcgggtaagttaacaactacCACCACAATTCAGCAAATTGTG	
	BG15040	tatcagcgggtaagttaacaactacGTCGACAGGATTAAACAAAAATGGC	
	BG15041	gaggggaaggggatgctacctaGCTAGCATTATACCTAGGACTG	
	BG15042	taggtagcatccccttccccctcGTCATAGTTCCCCTGAGATTATC	
	BG14003	AAATATTTTATCTGATTAATAAGATGATCTTCTTGAGATCG	
	BG15043	gatccggataacggctgaaacacGCTAGCATTATACCTAGGACTG	
	BG15044	gtgtttcagccgttatccggatcGTCATAGTTCCCCTGAGATTATC	

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

	BG14002	CAATCAGCAACGACTGTTTGC	
	BG12722	GCAAACAGTCGTTGCTGATTG	
Base-editing assays	BG15946	taagcaggccattttgtctgatTTATGCAACCAGTCCTAGCATC	Construction of pdThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7/scr, and pAcrThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7/scr.
	BG14518	agccggggaaactatccgtccgttaGGTGGAGGAGGTTCTGG	
	BG14519	TAACGGACGGATAGTTTCCC	
	BG14520	CATTCCAGTCCTTCACCGTC	
	BG14521	GACGGTGAAGGACTGGAATG	
	BG14529	tttcgacagagcggactgggtcgGCTAGCATTATACCTAGGACTG	
	BG14530	cgacccagtccgctctgtcgaaaGTCATAGTTCCCCTGAGATTATC	
	BG14524	TAAATCAGACAAAATGGCCTGC	
	BG17241	gggccggttctgttgccggacaaGCTAGCATTATACCTAGGACTGAG	
	BG17242	ttgtccggcaacagaaccggcccGTCATAGTTCCCCTGAGATTATC	
	BG14522	taacttctgctcgggtcttataaGCTAGCATTATACCTAGGACTG	
	BG14523	ttataagacccgagcagaagttaGTCATAGTTCCCCTGAGATTATC	
	BG14525	agcggactgggtcgatagatagtGCTAGCATTATACCTAGGACTG	
	BG14526	actatctatcgacccagtccgctGTCATAGTTCCCCTGAGATTATC	
	BG16904	cggtaatatatacattatggctaGCTAGCATTATACCTAGGACTGAG	
	BG16899	tagccataatgtatatattaccgGTCATAGTTCCCCTGAGATTATCG	
	BG14691	actggtaaactaccggtgccgtgGCTAGCATTATACCTAGGACTG	
	BG14692	cacggcaccggtagtttaccagtGTCATAGTTCCCCTGAGATTATC	
	BG16902	CATCAAAGGATTGTGATAGTTCCCG	
	BG21970	TGATTCAGGACTCTAATGGAGAG	
	BG21906	TTCTCTCCATTAGAGTCCTGAATCA	
	BG16901	CGGGAACTATCACAATCCTTTGATG	
	BG14543	ccatggggttactgcggatctagGCTAGCATTATACCTAGGACTG	
	BG14544	ctagatccgcagtaaccccatggGTCATAGTTCCCCTGAGATTATC	
	BG22757	gagcgcaccgggcgaccgtactgGCTAGCATTATACCTAGGACTGAG	
	BG22758	cagtacggtcgcccggtgcgctcGTCATAGTTCCCCTGAGATTATC	
	BG15946	taagcaggccattttgtctgatTTATGCAACCAGTCCTAGCATC	Construction of pdGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14/scr, and pAcrGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14/scr.
	BG14545	accattcgtccactgcaatctactcGTGACGGTGGAGGAGG	
	BG14536	GAGTAGATTGCAGTGGACGAATG	
	BG14537	GCCACCTATACCTTCCAGTC	
	BG14538	GACTGGAAGGTATAGGTGGC	
	BG14522	taacttctgctcgggtcttataaGCTAGCATTATACCTAGGACTG	
	BG14523	ttataagacccgagcagaagttaGTCATAGTTCCCCTGAGATTATC	
	BG14524	TAAATCAGACAAAATGGCCTGC	
	BG14525	agcggactgggtcgatagatagtGCTAGCATTATACCTAGGACTG	
	BG14526	actatctatcgacccagtccgctGTCATAGTTCCCCTGAGATTATC	
	BG14691	actggtaaactaccggtgccgtgGCTAGCATTATACCTAGGACTG	
	BG14692	cacggcaccggtagtttaccagtGTCATAGTTCCCCTGAGATTATC	
	BG16904	cggtaatatatacattatggctaGCTAGCATTATACCTAGGACTGAG	
	BG16899	tagccataatgtatatattaccgGTCATAGTTCCCCTGAGATTATCG	
	BG14541	ggatctttcgacagagcggactgGCTAGCATTATACCTAGGACTG	

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

	BG14542	cagtccgctctgtcgaaagatccGTCATAGTTCCCCTGAG	
	BG17551	gagcggactgggtcgatagatagGCTAGCATTATACCTAGGACTGAG	
	BG17552	ctatctatcgacccagtccgctcGTCATAGTTCCCCTGAGATTATCG	
	BG21876	GCCAGTTCGATATGGATGGAAAC	
	BG21972	TGATTCAGGACTCTAATGGAGAG	
	BG21906	TTCTCTCCATTAGAGTCCTGAATCA	
	BG21879	CGGTTTCCATCCATATCGAACT	
	BG14543	ccatggggttactgcggatctagGCTAGCATTATACCTAGGACTG	
	BG14544	ctagatccgcagtaaccccatggGTCATAGTTCCCCTGAGATTATC	
Sequencing reactions	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pTCas9_gfp1/gfp2/gfp3/scr, pTCas9_gfp1.20/gfp1.19/gfp 1.18/gfp1.13/gfp1.8/gfp1.6, pTCas9_gfp2.20/gfp2.19/gfp 2.18/gfp2.13/gfp2.8/gfp2.6, pTCas9_gfp3.20/gfp3.19/gfp 3.18/gfp3.13/gfp3.8/gfp3.6, pdTCas9_gfp1/gfp2/gfp3/scr, pdTCas9_gfp1.18/gfp1.13/gf p1.8/gfp1.6, pdTCas9_gfp2.18/gfp2.13/gf p2.8/gfp2.6, pdTCas9_gfp3.18/gfp3.13/gf p3.8/gfp3.6, *pAcr_TCas9_gfp1/gfp2/gfp 3/scr.
	BG12707	GTCGTATCGATGATTTTGATTGTTCG	
	BG12708	GCGTGAGATATAACGGGTATC	
	BG12709	GTAGACTTCGCCTTGTTCC	
	BG12710	GTTCCCATATTCGCGCTG	
	BG12712	ACGCGATGGATATGTTCTG	
	BG12713	CGAAGATTTAGGTGTCCGC	
	BG12714	CATACACATTCCAGTCCTTCACC	
	BG12715	CACCGGTCTCCATCCATATC	
	BG12716	CAACGCCGAGCGATATCG	
	BG12717	CCGTACTCTGAGTGGAAGG	
	BG12718	CAGTCGCGTACCGTCTTC	
	BG12719	GTGCATGCAAGGAGATGG	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	*BG14004	CAGTCGGGAAACCTGTCG	
	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pGCas9_gfp1/gfp2/gfp3/scr, pGCas9_gfp1.20/gfp1.19/gfp 1.18/gfp1.13/gfp1.8/gfp1.6, pGCas9_gfp2.20/gfp2.19/gfp 2.18/gfp2.13/gfp2.8/gfp2.6, pGCas9_gfp3.20/gfp3.19/gfp 3.18gfp3.13/gfp3.8/gfp3.6, pdGCas9_gfp1/gfp2/gfp3/scr , pdGCas9_gfp1.18/gfp1.13/gf p1.8/gfp1.6, pdGCas9_gfp2.18/gfp2.13/gf p2.8/gfp2.6, pdGCas9_gfp3.18/gfp3.13/gf p3.8/gfp3.6,
	BG12712	ACGCGATGGATATGTTCTG	
	BG13157	GCGAATTCGGCAACATGTC	
	BG13158	CAGGGTGAAGTCTACTCCTC	
	BG13159	GCTACATCAGCCGCTTCTTC	
	BG13160	GTCTGCTGGAGCATAACAAC	
	BG13161	GCACCCTGAAACGTTTCG	
	BG13163	CGAAACGTGTAATCTTCGGTC	
	BG13164	GGCCAGTTCCTTGTTCTG	
	BG13165	CGGTTTTCATCCTGTTCC	
	BG13166	CTGGAATGCTTGTTCATACAGC	
	BG13167	CGCTCCAGACGGTGTTTAC	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	BG12718	CAGTCGCGTACCGTCTTC	

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

	BG12719	GTGCATGCAAGGAGATGG	*pAcr_GCas9_gfp1/gfp2/gfp3/scr.
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pHR_TCas9_gfp3/gfp4/gfp5/gfp6/gfp7/gfp8/scr, and pHR_AcrTCas9_gfp3/gfp4/gfp5/gfp6/gfp7/gfp8/scr.
	BG12707	GTCGTATCGATGATTTTGATTGTTCG	
	BG12708	GCGTGAGATATAACGGGTATC	
	BG12709	GTAGACTTCGCCTTGTTCC	
	BG12710	GTTCCCATATTCGCGCTG	
	BG12712	ACGCGATGGATATGTTCTG	
	BG12713	CGAAGATTTAGGTGTCCGC	
	BG12714	CATACACATTCCAGTCCTTCACC	
	BG12715	CACCGGTCTCCATCCATATC	
	BG12716	CAACGCCGAGCGATATCG	
	BG12717	CCGTACTCTGAGTGGAAGG	
	BG12718	CAGTCGCGTACCGTCTTC	
	BG12719	GTGCATGCAAGGAGATGG	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	BG14368	ATCAGACAAAATGGCCTGCTTATG	
	BG14786	GTTATCATTGATGCGAGGCG	
	BG15038	GTAGTTGTTAACTTACCCG	
	BG14789	CGATCAACGTCCTTCTGATC	
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	*BG14004	CAGTCGGGAAACCTGTCG	
	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pHR_GCas9_gfp5/gfp6/gfp7/gfp8/gfp9/gfp10/scr, and pHR_AcrGCas9_gfp5/gfp6/gfp7/gfp8/gfp9/gfp10/scr.
	BG12712	ACGCGATGGATATGTTCTG	
	BG13157	GCGAATTCGGCAACATGTC	
	BG13158	CAGGGTGAAGTCTACTCCTC	
	BG13159	GCTACATCAGCCGCTTCTTC	
	BG13160	GTCTGCTGGAGCATAACAAC	
	BG13161	GCACCCTGAAACGTTTCG	
	BG13163	CGAAACGTGTAATCTTCGGTC	
	BG13164	GGCCAGTTCCTTGTTCTG	
	BG13165	CGGTTTTCATCCTGTTCC	
	BG13166	CTGGAATGCTTGTTCATACAGC	
	BG13167	CGCTCCAGACGGTGTTTAC	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	BG12718	CAGTCGCGTACCGTCTTC	
	BG12719	GTGCATGCAAGGAGATGG	
	BG14368	ATCAGACAAAATGGCCTGCTTATG	
	BG14786	GTTATCATTGATGCGAGGCG	

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

	BG15038	GTAGTTGTTAACTTACCCG	
	BG14789	CGATCAACGTCCTTCTGATC	
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pdThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7, and *pAcrThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7.
	BG12707	GTCGTATCGATGATTTTGATTGTTCG	
	BG12708	GCGTGAGATATAACGGGTATC	
	BG12709	GTAGACTTCGCCTTGTTCC	
	BG12710	GTTCCCATATTCGCGCTG	
	BG12712	ACGCGATGGATATGTTCTG	
	BG12713	CGAAGATTTAGGTGTCCGC	
	BG12714	CATACACATTCCAGTCCTTCACC	
	BG12715	CACCGGTCTCCATCCATATC	
	BG12716	CAACGCCGAGCGATATCG	
	BG12717	CCGTACTCTGAGTGGAAGG	
	BG12718	CAGTCGCGTACCGTCTTC	
	BG12719	GTGCATGCAAGGAGATGG	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	BG14513	GTAATAGAGTTTGCAAGCCCAG	
	BG14695	ATGACCGACGCTGAGTAC	
	BG14516	CGGACAATTCACGATAAATTGG	
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	*BG14004	CAGTCGGGAAACCTGTCG	
	BG12706	GGTAGCTCAGAGAACCTTCG	Sequencing of pdGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14, and *pAcrGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14.
	BG12712	ACGCGATGGATATGTTCTG	
	BG13157	GCGAATTCGGCAACATGTC	
	BG13158	CAGGGTGAAGTCTACTCCTC	
	BG13159	GCTACATCAGCCGCTTCTTC	
	BG13160	GTCTGCTGGAGCATAACAAC	
	BG13161	GCACCCTGAAACGTTTCG	
	BG13163	CGAAACGTGTAATCTTCGGTC	
	BG13164	GGCCAGTTCCTTGTTCTG	
	BG13165	CGGTTTTCATCCTGTTCC	
	BG13166	CTGGAATGCTTGTTCATACAGC	
	BG13167	CGCTCCAGACGGTGTTTAC	
	BG12721	CTGCGTTAGCAATTTAACTGTG	
	BG12722	GCAAACAGTCGTTGCTGATTG	
	BG12723	GTTAACCACCATCAAACAGG	
	BG12718	CAGTCGCGTACCGTCTTC	
	BG12719	GTGCATGCAAGGAGATGG	
	BG14513	GTAATAGAGTTTGCAAGCCCAG	
	BG14695	ATGACCGACGCTGAGTAC	

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

	BG14516	CGGACAATTCACGATAAATTGG	
	*BG12705	GTCGACAGGATTAAACAAAAATGGCC	
	BG13437	CAATACGCAAACCGCCTC	Sequencing of pAcr.
	BG13438	GAGACGGTCACAGCTTGTC	
Screening of colonies for genome/ base-editing	BG13019	CTTGTTGGTACGCTTGATAATTGC	Colony PCR (and sequencing) for screening of possibly genome-edited (deletion of *gfp* gene and its PlacUV5) colonies from assays that pHR_ThermoCas9_gfp3/gfp4/gfp5/gfp6/gfp7/gfp8, and pHR_GeoCas9_gfp5/gfp6/gfp7/gfp8/gfp9/gfp10 were applied.
	BG15769	GCGCGATGGCGATAAATTG	
	BG13019	CTTGTTGGTACGCTTGATAATTGC	PCR amplification for screening of possibly base-edited (nucleotide substitution in *gfp* gene) colonies from assays that pdThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7, pAcrThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8/gfp7, pdGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14, and pAcrGeoTarget-AID_gfp5/gfp12/gfp7/gfp8/gfp13/gfp14 were applied.
	BG15769	GCGCGATGGCGATAAATTG	
	BG14615	GGATGCTACCTATGGTAAACTG	Sequencing for screening of possibly base-edited (nucleotide substitution in gfp gene) colonies from assays that pdThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8, pAcrThermoTarget-AID_gfp4/gfp11/gfp5/gfp12/gfp8, pdGeoTarget-AID_gfp5/gfp12/gfp8/gfp13/gfp14, and pAcrGeoTarget-AID_gfp5/gfp12/gfp8/gfp13/gfp14 were applied.
	BG15431	GTTGAGCTACAGGCGGTC	Sequencing for screening of possibly base-edited

Supplementary Table 4. Primers/Oligonucleotides used in this study. *(continuation)*

			(nucleotide substitution in *gfp* gene) colonies from assays that pdThermoTarget-AID_gfp7, pAcrThermoTarget-AID_gfp7, pdGeoTarget-AID_gfp7, and pAcrGeoTarget-AID_gfp7 were applied.

***Lowercase**: overhang sequence for Gibson assembly*
***Uppercase**: sequence annealing to the PCR template DNA*
***Red colour**: spacer-protospacer mismatch (nucleotide substitution in the spacer sequence)*
***Blue colour**: spacer-protospacer match (identical sequences of spacer and protospacer)*

Supplementary Table 5. DNA sequences of (fusion) genes used in this study.

Gene/ Fusion gene	Sequence (5' -> 3')
thermocas9 (wild-type)	atgaagtataaaatcggtcttgatatcggcattacgtctatcggttgggctgtcattaatttggacattcctcgcatcgaagatttaggtgtccgcatttttgacagagcggaaaacccgaaaaccggggagtcactagctcttccacgtcgcctcgcccgctccgcccgacgtcgtctgc ggcgtcgcaaacatcgactggagcgcattcgccgcctgttcgtccgcgaaggaattttaacgaaggaagagctgaacaagctgtttgaaaaaaagcacgaaatcgacgtctggcagcttcgtgttgaagcactggatcgaaaactaaataacgatgaattagcccgcatccttcttc atctggctaaacggcgtggatttagatccaaccgcaagagtgagcgcaccaacaaagaaaacagtacgatgctcaaacatatttgaagaaaccaatccattcttttcaagttaccgaacggttgcagaaatggttgtcaaggatccgaaatttttccctgcacaagcgtaataaagagg ataattacacccaacactgttgcccgcgacgatcttgaacgggaaatcaaactgattttcgccaaacagcgcgaatatgggaacatcgtttgcacagaagcatttgaacacgagtatatttccatttgggcatcgcaacgccctttgcttctaaggatgatatcgagaaaaaagtcggtt tctgtacgtttgagcctaaagaaaaacgcgcgccaaaagcaacatacacattccagtccttcaccgtctgggaacatattaacaaacttcgtcttgtctcccccgggaggcatccgggcactaaccgatgatgaacgtcgtcttatatacaagcaagcatttcataaaaataaaatcacc ttccatgatgttcgaacattgcttaacttgcctgacgacacccgttttaaaggtcttttatatgaccgaaacaccacgctgaaggaaaatgagaaagttcgcttccttgaactcggcgcctatcataaaatacggaaagcgatcgacagcgtctatggcaaaggagcagcaaaatcatt tcgtccgattgattttgatacatttggctacgcattaacgatgtttaaagacgacaccgacattcgcagttacttgcgaaacgaatacgaacaaaatggaaaacgaatggaaaatctagcggataaagtctatgatgaagaattgattgaagaacttttaaacttatcgttttctaagttt ggtcatctatcccttaaagcgcttcgcaacatccttccatatatggaacaaggcgaagtctactcaaccgcttgtgaacgagcaggatatacatttacagggccaaagaaaaaacagaaaacggtattgctgccgaacattccgccgatcgccaatccggtcgtcatgcgcgcactga cacaggcacgcaaagtggtcaatgccattatcaaaaagtacggctcaccggtctccatccatatcgaactggcccgggaactatcacaatcctttgatgaacgacgtaaaatgcagaaagaacaggaaggaaaccgaaagaaaaacgaaactgccattcgccaacttgttgaatat gggctgacgctcaatccaactgggcttgacattgtgaaattcaaactatggagcgaacaaaacggaaaatgtgcctattcactccaaccgatcgaaatcgagcggttgctcgaaccaggctatacagaagtcgaccatgtgattccatacagccgaagcttggacgatagctatacc aataaagttcttgtgttgacaaaggagaaccgtgaaaaaggaaaccgcaccccagctgaatatttaggattaggctcagaacgttggcaacagttcgagacgtttgtcttgacaaataagcagttttcgaaaaagaagcgggatcgactccttcggcttcattacgatgaaaacgaa gaaaatgagtttaaaaatcgtaatctaaatgataccgttatatctcacgcttcttggctaacttttattcgcgaacatctcaaattcgccgacagcgatgacaaacaaaaagtatacacggtcaacggccgtattaccgcccatttacgcagccgttggaatttaacaaaaccggga agaatcgaatttgcatcatgccgtcgatgctgccatcgtcgcctgcacaacgccgagcgatatcgcccgagtcaccgccttctatcaacggcgcgaacaaaacaaagaactgtccaaaaagacggatccgcagtttccgcagccttggccgcactttgctgatgaactgcaggcgct ttatcaaaaatccaaaggagagtataaaagctctcaatcttggaaattatgataacgagaaactcgaatcgttgcagccggttttttgtctcccgaatgccgaagcggagcataacaggagcggctcatcaagaaacattgcggcgttatatcggcatcgacgaacggagcggaaa aatacagacggtcgtcaaaaagaaactatccgagatccaactggataaaacaggtcatttcccaatgtacgggaaagaaagcgatccaaggacatatgaagccattcgccaacggttgcttgaacataacaatgacccaaaaaaggcgtttcaagagcctctgtataaacccgaag aagaacggagaactaggtcctatcatccgaacaatcaaaatcatcgatacgacaaatcaagttattccgctcaacgatggcaaaacagtcgcctacaacagcaacatcgtgcgggtcgacgtctttgagaaagatggcaaaatattattgtgtccctatctatacaatagatatgatga aagggatcttgccaaacaaggcgatcgagccgaacaaaccgtactctgagtggaaggaaatgacggaggactatacattccgattcagtctatacccaaatgatcttatccgtatcgaatttcccgagaaaaaacaataaagactgctgtgggggaagaaatcaaaattaaggat ctgttcgccttattatcaaaccatcgactcctccaatggagggttaagtttggttagccatgataacaacttttcgctccgcagcatcggttcaagaaccctcaaacgattcgagaaataccaagtagatgtgctaggcaacatctacaaagtgagagggggaaaagagagttgggtgg cgtcatcttctcattcgaaagccggggaaactatccgtccgttataa
geocas9 (wild-type)	atgcgttataagattggcctggacatcggtattacctctgttggttgggcagtcatgaacctggatatccctcgtatcgaagatctgggcgtgcgcatttttgaccgtgcggagaacccgcagaccggtgaatctctggctctgccgcgtcgtctggcacgtagcgcacgccgccgcctgc gtcgtcgtaaacaccgtctggagcgtattcgtcgccctggttattcgtgaaggcatcctgacgaaagaagaactggataaactgttcgaagaaaaacacgagatcgacgtatggcagctgcgtgtagaagcccttgaccgtaagctgaacaacgacgaactggcgcgtgtcctgctgc atctggcaaagcgtcgtggcttcaaatctaaccgtaaatctgaacgctccaataaagagaactccactatgctgaaacatattgaggagaaccgtgcaattctgtctagctaccgtaccgtgggcgaaatgattgttaaagacccgaaattcgcactgcataagcgtaacaaaggcga aaactacaccaacaccattgcacgcgatgacctggaacgtgaaatccgtctgattttctccaaacagcgcgaattcggcaacatgtcttgcaccgaagaattcgaaaacgaatatattaccatttgggcatctcagcgtccggtggcgtctaaagatgatatcgaaaaaaaagtaggc ttttgtactttcgaaccgaaggaaaaacgtgcgccgaaagccacctatacctccagtctttttatcgcgtgggaacatatcaacaaactgcgtctgatttctccgtctggcgcccgcggcctgaccgacgaagaacgtcgtctgctgtatgaacaagcattccagaaaaacaaaattacc taccacgatattcgtacccctgctgcatctgccggacgacacctacttcaagggcatcgtttacgatcgcggtgaatctcgtaagcagaacgaaaacattcgtttcctggaactggatgcatacaccagatccgtaaagctgtagataaagtttacggcgaagggtaaatccagcagctt cctgccgatcgacttttgatacctcggttacgcgctgaccctgtttatagagacgatgcggatatccactcttacctgcgcaacgagtacgaacagaacggcaaacgtatgcctaacctggctaacaaagtttacgataacgagctgattgaagaactgctgaacctgtccttcactaaatt cggtcacctgtctctgaaagctctgcgttccatcctgccgtatatggaacagggtgaagtctactcctccgcttgtgaacgtgcaggctacacctttcaccggtccgaaaaagaagcaaaaaactatgctgctgccgaacatcccgccgattgcgaaccctgtagtaatgcgtgcactga cccaggcgcgcaaagtagtcaacgcgatcatcaaaaagtacggcagcccggtttccatccatatcgaactggcgcgcgacctgagccagacttttgacgagcgtcgtaaaactaaaaaggaacaggatgaaaaccgtaaaaaaaacgaaaccgcgatccgccagctgatggaat acggtctgactctgaaccctactggtcacgatattgtgaagttcaagctgtggtctgaacagaacggtcgctgtgcttactctctgcagccgatcgagatcgaacgtctgctggagccaggttacgttgaagtagatcatgtgatcccgtactcccgctctctggatgattcttataccaac aaagttctggttctgactcgcgaaaaccgtgagaaaggcaaccgcatccgagctgaatatctgggtgttggcactgagcgttggcaacagttcgaaaccttcgtcctgaccaataaacagttctctaaaaagaaacgtgaccgtctgctgcgtctgcactacgatgaaaacgaagaga ctgaattcaaaaaccgtaacctgaacgatactcgctacatcagcgcgttcttcgcaaacttcattcgtgaacacctgaaatttgcggaatccgacgataaacagaaagttttaccgtaaacggccgtgttaccgccccacctgcgttctcgctgggagttcaacaagaaccgtgaggaa agcgatctgcaccacgctgttgacgccgttattgtggcgtgcaccaccccaagcgatatcgctaaggtgaccgcattctaccagcgtcgtgagcagaacaaggaactggccaaaaaaaccgaaccgcattttccgcagccgtggccgcacttcgcggacgaactgcgtgctcgtctgt ccaaacatcctaaagaaagcatcaaagctctgaacctgggtaactacgatgaccaaaaactggaatctctgcagccggtgtttgtcagccgtatgccgaaacgttctgttactggcgctgcgcaccaggaaacgctgcgccgttacgtgggcatcgacgaacgctccggtaaaatcca gaccgtagtaaaaaccaaactgtccgagattaaactggatgcatccggccacttcccaatgtacggtaaagaatccgatccacgcacttatgaagccatccgccagcgtctgctggagcataacaacgacccgaaaaaggcattccaggagcctctgtacaaaccgaaaaaaaacg gcgaaccgggcccggtaatccgtactgtaaaaattatcgacacgaaaaaccaggtgatccctctgaacgatggtaaaaccgtggcctacaattccaacatcgttcgctggagcgtgttcgaaaaagatggtaaatactactgtgtgtaccggtgtatacctggacatcatgaaaggcat tctgccgaacaaagcgattgaaccgaacaagccgtactctgaatggaaagaaatgaccgaagattacacgtttcgtttcagcctgtatccgaacgacctgatccgcatcgaactgccgcgtgaaaaaaccgttaaaaccgctgcaggcgaagaaattaacgtgaaagacgtgttcgt ttactatataaaacgatcgactccgcaaacggcggcctggaactgatttctcacgaccaccgtttctctctgcgtggcgttggctctcgcaccctgaaacgtttcgagaaatatcaagttgatgttctgggtaacatctataaagtgcgtggcgagaaacgtgtcggtctggcgtcctccgca cacagcaaacctggcaaaaccattcgtccactgcaatctactcgtgactaa
dthermocas9 (D8A, H582A mutant)	atgaagtataaaatcggtcttgCtatcggcattacgtctatcggttgggctgtcattaatttggacattcctcgcatcgaagatttaggtgtccgcatttttgacagagcggaaaacccgaaaaccggggagtcactagctcttccacgtcgcctcgcccgctccgcccgacgtcgtctgc ggcgtcgcaaacatcgactggagcgcattcgccgcctgttcgtccgcgaaggaattttaacgaaggaagagctgaacaagctgtttgaaaaaaagcacgaaatcgacgtctggcagcttcgtgttgaagcactggatcgaaaactaaataacgatgaattagcccgcatccttcttc atctggctaaacggcgtggatttagatccaaccgcaagagtgagcgcaccaacaaagaaaacagtacgatgctcaaacatatttgaagaaaccaatccattcttttcaagttaccgaacggttgcagaaatggttgtcaaggatccgaaatttttccctgcacaagcgtaataaagagg ataattacacccaacactgttgcccgcgacgatcttgaacgggaaatcaaactgattttcgccaaacagcgcgaatatgggaacatcgtttgcacagaagcatttgaacacgagtatatttccatttgggcatcgcaacgccctttgcttctaaggatgatatcgagaaaaaagtcggtt

Supplementary Table 5. DNA sequences of (fusion) genes used in this study. *(continuation)*

	tctgtacgtttgagcctaaagaaaaacgcgcgccaaaagcaacatacacattccagtccttcaccgtctgggaacatattaacaaacttcgtcttgtctccccgggaggcatccgggcactaaccgatgatgaacgtcgtcttatatacaagcaagcatttcataaaaataaaatcacc ttccatgatgttcgaacattgcttaacttgcctgacgacacccgttttaaaggtcttttatatgaccgaaacaccacgctgaaggaaaatgagaaagttcgcttccttgaactcggcgcctatcataaaatacggaaagcgatcgacagcgtctatggcaaaggagcagcaaaatcatt tcgtccgattgattttgatacatttggctacgcattaacgatgtttaaagacgacaccgacattcgcagttacttgcgaaacgaatacgaacaaaatggaaaacgaatggaaaatctagcggataaagtctatgatgaagaattgattgaagaacttttaaacttatcgttttctaagttt ggtcatctatcccttaaagcgcttcgcaacatccttccatatatggaacaaggcgaagtctactcaaccgcttgtgaacgagcaggatatacatttacagggccaaagaaaaaacagaaaacggttattgctgccgaacattccgccgatccgccaatccggtcgtcatgcgcgcactga cacaggcacgcaaagtggtcaatgccattatcaaaaagtacggctcaccggtctccatccatatcgaactggcccgggaactatcacaatcctttgatgaacgacgtaaaatgcagaaagaacaggaaggaaaccgaaagaaaaacgaaactgccattcgccaacttgttgaatat gggctgacgctcaatccaactgggcttgacattgtgaaattcaaactatggagcgaacaaaacggaaaatgtgcctattcactccaaccgatcgaaatcgagcggttgctcgaaccaggctatacagaagtcgacGCtgtgattccatacagccgaagcttggacgatagctatacc aataaagttcttgtgttgacaaaggagaaccgtgaaaaaggaaaccgcaccccagctgaatatttaggattaggctcagaacgttggcaacagttcgagacgtttgtcttgacaaataagcagttttcgaaaaagaagcgggatcgactccttcggcttcattacgatgaaaacgaa gaaaatgagtttaaaaatcgtaatctaaatgataccgttatatctcacgcttcttggctaacttttattcgcgaacatctcaaattcgccgacagcgatgacaaacaaaaagtatacacggtcaacggccgtattaccgcccatttacgcagccgttggaatttaacaaaaaccgggga agaatcgaatttgcatcatgccgtcgatgctgccatcgtcgcctgcacaacgccgagcgatatcgcccgagtcaccgccttctatcaacggcgcgaacaaaacaaagaaactgtccaaaaagacggatccgcagtttccgcagccttggccgcactttgctgatgaactgcaggcgcgt ttatcaaaaaatccaaaggagagtataaaagctctcaatcttggaaattatgataacgagaaactcgaatcgttgcagccggttttttgtctcccgaatgccgaagcggagcataacaggagcggctcatcaagaaaacattgcggcgttatatcggcatcgacgaacggagcggaaa aatacagacggtcgtcaaaaagaaactatccgagatccaactggataaaacaggtcatttcccaatgtacgggaaagaaagcgatccaaggacatatgaagccattcgccaacggttgcttgaacataacaatgacccaaaaaaaggcgtttcaagagcctctgtataaaccgaag aagaacggagaactaggtcctatcatccgaacaatcaaaatcatcgatacgacaaatcaagttattccgctcaacgatggcaaaacagtcgcctacaacagcaacatcgtgcgggtcgacgtctttgagaaagatggcaaatattattgtgtccctatctatacaatagatatgatga aagggatcttgccaaacaaggcgatcgagccgaacaaaccgtactctgagtggaaggaaatgacggaggactatacattccgattcagtctatacccaaatgatcttatccgtatcgaatttccccgagaaaaaacaataaagactgctgtgggggaagaaatcaaaattaaggat ctgttcgcctattatcaaaccatcgactcctccaatggagggttaagtttggttagccatgataacaacttttcgctccgcagcatcggttcaagaaccctcaaacgattcgagaaataccaagtagatgtgctaggcaacatctacaaagtgagaggggaaaagagagttgggtgg cgtcatcttctcattcgaaagccgggggaaactatccgtccgttataa
dgeocas9 (D8A, H582A mutant) | atgcgttataagattggcctggCcatcggtattacctctgttggttgggcagtcatgaacctggatatccctcgtatcgaagatctgggcgtgcgcatttttgaccgtgcggagaacccgcagaccggtgaatctctggctctgccgcgtcgtctggcacgtagcgcacgccgccgcctgc gtcgtcgtaaacaccgtctggagcgtattcgtcgccctggttattcgtgaaggcatcctgacgaaagaagaactggataaactgttcgaagaaaaacacgagatcgacgtatggcagctgcgtgtagaagccctggaccgtaagctgaacaacgacgaactggcgcgtgtcctgctgc atctggcaaagcgtcgtggcttcaaatctaaccgtaaatctgaacgctccaataaagagaactccactatgctgaaacatattgaggagaacccgtgcaattctgtctagctaccgtaccgtgggcgaaatgattgttaaagacccgaaatttgcgactgcataagcgtaacaaaggcga aaactacaccaacaccattgcacgcgatgacctggaacgtgaaatccgtctgattttctccaaacagcgcgaattcggcaacatgtcttgcaccgaagaattcgaaaacgaatatattaccatttgggcatctcagcgtccggtggcgtctaaagatgatatcgaaaaaaaagtaggc ttttgtactttcgaaccgaaggaaaaacgtgcgccgaaagccacctatacctttccagtcttttatcgcgtgggaacatatcaacaaactgcgtctgatttctccgtctggcgcccgcggcctgaccgacgaagaacgtcgtctgctgtatgaacaagcattccagaaaaacaaaattacc taccacgatattcgtacccgctgcatctgccggacgacacctacttcaagggcatcgtttacgatcgcggtgaatctgtaagcagaacgaaaacattcgtttcctggaactggatgcatacaccacgatccgtaaagctgtagataaagtttacggcaagggtaaatccagcagctt cctgccgatcgactttgatacctttcggttacgcgctgaccctgtttaaagacgatgcggatatccactcttacctgcgcaacgagtacgaacagaacggcaaacgtatgcctaacctggctaacaaagtttacgataacgagctgattgaagaactgctgaacctgtccttcactaaatt cggtcacctgtctctgaaagctctgcgttccatcctgccgtatatggaacagggtgaagtctactcctccgcttgtgaacgtcaggctacacctttcaccggtccgaaaaagaagcaaaaaactatgctgctgccgaacatcccgccgattgcgaaccctgtagtaatgcgtgcactga cccaggcgcgcaaagtagtcaacgcgatcatcaaaaagtacggcagcccggtttccatccatatcgaactggcgcgcgacctgagccagacttttgacgagcgtcgtaaaactaaaaaggaacaggatgaaaaccgtaaaaaaaagcaaaccgcgatccgccagctgatggaat acggtctgactctgaacccctactggtcacgatattgtgaagttcaagctgtggtctgaacagaacggtcgctgtgcttactctctgcagccgatcgagatcgaacgtctgctggagccaggttacgttgaagtagatGCtgtgatcccgtactcccgctctctggatgattcttatacccaa caaagttctggttctgactcgcgaaaaccgtgagaaaggcaaccgcatcccagctgaatatctgggtgttggcactgagcgttggcaacagttcgaaaccttcgtcctgaccaataaacagttctctaaaaagaaacgtgaccgtctgctgcgtctgcactacgatgaaaacgaagag actgaattcaaaaaccgtaacctgaacgatactcgctacatcagccgcttctttcgcaaacttcattcgtgaacacctgaaatttgcggaatccgacgataaacagaaagtttatacctaaacggccgttgttaccgcccccactgcgttctcgctgggagttcaacaagaacccgtgagga aagcgatctgcaccacgctgttgacgccgttattgtggcgtgcaccaccccaagcgatatcgctaaggtgaccgcattctaccagcgtcgtgagcagaacaaggaactggccaaaaaaacgaaccgcatttttccgcagccgtggccgcacttcgcggacgaactgcgtgctcgtctg tccaaacatcctaaagaaagcatcaaagctctgaacctgggtaactacgatgaccaaaaactggaatctctgcagccggtgtttgtcagccgtatgccgaaacgttctgttactggcgctgcgcaccaggaaacgctgcgccgttacgtgggcatcgacgaacgctccggtaaaatcc agaccgtagtaaaaaccaaactgtccgagattaaactggatgcatccggccacttcccaatgtacggtaaagaatccgatccacgcacttatgaagccatccgccagcgtctgctggagcataacaacgacccgaaaaaggcattccaggagcctctgtacaaaccgaaaaaaaac ggcgaaccgggcccggtaatccgctactgtaaaaattatcgacacgaaaaaccaggtgatccctctgaacgatggtaaaaccgtggcctacaattccaacatcgttcgcgtggacgtgttcgaaaaagatggtaaatactactgtgtaccggtgtatacatgacatcatgaaaggc attctgccgaacaaagcgattgaaccgaacaagccgtactctgaatggaaagaaatgaccgaagattacacgtttcgtttcagcctgtatccgaacgacctgatccgcatcgaactgccgcgtgaaaaaaccgttaaaaccgctgcaggcgaagaaattaacgtgaaagacgtgttc gtttactataaaacgatcgactccgcaaacggcggcctggaactgatttctcacgaccaccgtttctctctgcgtggcgttggctctcgcaccctgaaacgtttcgagaaatatcaagttgatgttctgggtaacatctataaagtgcgtggcgagaaacgtgtcggtctggcgtcctccg cacacagcaaacctggcaaaacattcgtccactgcaatctactcgtgactaa
thermocas9-**linker**-*pmcda1*-**linker**-*ugi*-**linker**-LVA | atgaagtataaaatcggtcttgatatcggcattacgtctatcggttgggctgtcattaatttggacattcctcgcatcgaagatttaggtgtccgcatttttgacagagcggaaaacccgaaaacccggggagtcactagctcttccacgtcgcctcgcccgctccgcccgacgtcgtctgc ggcgtcgcaaacatcgactggagcgcattcgccgcctgttcgtccgcgaaggaatttttaacgaaggaagagctgaacaagctgtttgaaaaaaagcacgaaatcgacgtctggcagcttcgtgttgaagcactggatcgaaaactaaataacgatgaattagcccgcatccttcttc atctggctaaacggcgtggatttagatccaacccgcaagagtgagcgcaccaacaaagaaaacagtacgatgctcaaacatattgaagaaaaccaatccattcttttcaagttaccgaacggttgcagaaatggttgtcaaggatccgaaatttttccctgcacaagcgtaataaagagg ataattacaccaacactgttgcccgcgacgatcttgaacgggaaatcaaactgattttcgccaaacagcgcgaatatgggaacatcgtttgcacagaagcatttgaacacgagtatatttccatttgggcatcgcaacgccctttgcttctcaaggatgatatcgagaaaaaagtcggtt tctgtacgtttgagcctaaagaaaaacgcgcgccaaaagcaacatacacattccagtccttcaccgtctgggaacatattaacaaacttcgtcttgtctccccgggaggcatccgggcactaaccgatgatgaacgtcgtcttatatacaagcaagcatttcataaaaataaaatcacc ttccatgatgttcgaacattgcttaacttgcctgacgacacccgttttaaaggtcttttatatgaccgaaacaccacgctgaaggaaaatgagaaagttcgcttccttgaactcggcgcctatcataaaatacggaaagcgatcgacagcgtctatggcaaaggagcagcaaaatcatt tcgtccgattgattttgatacatttggctacgcattaacgatgtttaaagacgacaccgacattcgcagttacttgcgaaacgaatacgaacaaaatggaaaacgaatggaaaatctagcggataaagtctatgatgaagaattgattgaagaacttttaaacttatcgttttctaagttt ggtcatctatcccttaaagcgcttcgcaacatccttccatatatggaacaaggcgaagtctactcaaccgcttgtgaacgagcaggatatacatttacagggccaaagaaaaaacagaaaacggttattgctgccgaacattccgccgatccgccaatccggtcgtcatgcgcgcactga cacaggcacgcaaagtggtcaatgccattatcaaaaagtacggctcaccggtctccatccatatcgaactggcccgggaactatcacaatcctttgatgaacgacgtaaaatgcagaaagaacaggaaggaaaccgaaagaaaaacgaaactgccattcgccaacttgttgaatat gggctgacgctcaatccaactgggcttgacattgtgaaattcaaactatggagcgaacaaaacggaaaatgtgcctattcactccaaccgatcgaaatcgagcggttgctcgaaccaggctatacagaagtcgaccatgtgattccatacagccgaagcttggacgatagctatacc aataaagttcttgtgttgacaaaggagaaccgtgaaaaaggaaaccgcaccccagctgaatatttaggattaggctcagaacgttggcaacagttcgagacgtttgtcttgacaaataagcagttttcgaaaaagaagcgggatcgactccttcggcttcattacgatgaaaacgaa gaaaatgagtttaaaaatcgtaatctaaatgataccgttatatctcacgcttcttggctaacttttattcgcgaacatctcaaattcgccgacagcgatgacaaacaaaaagtatacacggtcaacggccgtattaccgcccatttacgcagccgttggaatttaacaaaaaccgggga

Supplementary Table 5. DNA sequences of (fusion) genes used in this study. *(continuation)*

	agaatcgaatttgcatcatgccgtcgatgctgccatcgtcgcctgcacaacgccgagcgatatcgcccgagtcaccgccttctatcaacggcgcgaacaaaacaaagaactgtccaaaaagacggatccgcagtttccgcagccttggccgcactttgctgatgaactgcaggcgcgt ttatcaaaaaatccaaaggagagtataaaagctctcaatcttggaaattatgataacgagaaactcgaatcgttgcagccggtttttgtctcccgaatgccgaagcggagcataacaggagcggctcatcaagaaacattgcggcgttatatcggcatcgacgaacggagcggaaa aatacagacggtcgtcaaaaagaaactatccgagatccaactggataaaacaggtcatttcccaatgtacgggaaagaaagcgatccaaggacatatgaagccattcgccaacggttgcttgaacataacaatgacccaaaaaaggcgtttcaagagccctctgtataaaccgaag aagaacggagaactaggtcctatcatccgaacaatcaaaatcatcgatacgacaaatcaagttattccgctcaacgatggcaaaacagtcgcctacaacagcaacatcgtgcgggtcgacgtctttgagaaagatggcaaatattattgtgtccctatctatacaatagatatgatga aagggatcttgccaaacaaggcgatcgagccgaacaaaccgtactctgagtggaaggaaatgacggaggactatacattccgattcagtctatacccaaatgatcttatccgtatcgaatttcccgagaaaaaacaataaagactgctgtgggggaagaaatcaaaattaaggat ctgttcgcctattatcaaaccatcgactcctccaatggagggttaagtttggttagccatgataacaactttcgctccgcagcatcggttcaagaaccctcaaacgatttcgagaaataccaagtagatgtgctaggcaacatctacaaagtgagaggggaaaagagagttggggtgg cgtcatcttctcattcgaaagccggggaaactatccgtccgttaggtgcgaggaggttctgcgagtatgtgcgagccctctttgactttaatgggaatgatgaagaggatcttccctttaagaaaggagacatcctgagaatccgggataagcctgaggagc agtggtggaatgcagaggacagcgaaggaaagaggggatgattcctgtcccttacgtggagaagtattccggagactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggctccgagtccggagactataaggacca cgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggatgaccgacgctgagtacgtgagaatccatgagaagttggacatctacacgtttaagaaacagttttcaacaacaaaaaatccgtgtcgcatagatgctacgttctctttgaatttaaa acgacggggtgaacgtagagcgtgttttttggggctatgctgtgaataaaccacagagcgggacagaacgtggcattcacgccgaaatctttagcattagaaaagtcgaagaatacctgcgcgacaaccccggacaattcacgataaattggtactcatcctggagtccttgtgcagat tgcgctgaaaagatcttagaatggtatcaccaggagctgcgggggaacggccacacttgaaaatctgggcttgcaaactctattacgagaaaaatgcgaggaatcaaattgggctgtggaacctcagagataacgggggttgggttgaatgtaatggtaagtgaacactaccaatgt tgcaggaaaatattcatccaatcgtcgcacaatcaattgaatgagaatagatggcttgagaagactttgaagcgagctgaaaaacgacggagcgagttgtccattatgattcaggtaaaaatactccacaccactaagagtcctgctgtttctagaatgaccaacctttccgacatcat agagaaggaaacaggcaaacagttggtcatccaagagtcgatactcatgcttcctgaagaagttgaggaggtcattgggaataagccggaaagtgacattctcgtacacactgcgtatgatgagagcaccgatgagaacgtgatgctgctcacgtcagatgccccagagtacaaacc cctgggctctggtgattcaggactctaatggagagaacaagatcaagatgctaggactggttgcataa
geocas9-**linker**-*pmcda1*-**linker**-*ugi*-**linker**-LVA	atgcgttataagattggcctggacatcggtattacctctgttggttgggcagtcatgaacctggatatccctcgtatcgaagatctgggcgtgcgcatttttgaccgtgcggagaacccgcagaccggtgaatctctggctctgccgcgtcgtctggcacgtagcgcacgccgccgcctgc gtcgtcgtaaacaccgtctggagcgtattcgtcgcctggttattcgtgaaggcatcctgacgaaagaagaactggataaactgttcgaagaaaaacacgagatcgacgtatggcagctgcgtgtagaagccctggaccgtaagctgaacaacgacgaactggcgcgtgtcctgctgc atctggcaaagcgtcgtggcttcaaatctaaccgtaaatctgaacgctccaataaagagaactccactatgctgaaacatattgaggagaaccgtgcaattctgtctagctaccgtaccgtgggcgaaatgattgttaaagacccgaaattcgcactgcataagcgtaacaaaggcga aaactacaccaacaccattgcacgcgatgacctggaacgtgaaatccgtctgatttttctccaaacagcgcgaattcggcaacatgtcttgcaccgaagaattcgaaaacgaatatattaccatttgggcatctcagcgtccggtggcgtctaaagatgatatcgaaaaaaaagtaggc ttttgtactttcgaaccgaaggaaaaacgtgcgccgaaagccacctatacctccagtcttttatcgcgtgggaacatatcaacaaactgcgtctgatttctccgtctggcgcccgcggcctgaccgacgaagaacgtcgtctgctgtatgaacaagcattccagaaaaacaaaattacc taccacgatattcgtaccctgctgcatctgccggacgacacctacttcaagggcatcgtttacgatcgcggtgaatctcgtaagcagaacgaaaacattcgtttcctggaactggatgcataccaccagatccgtaaagctgtagataaagtttacggcaagggtaaatccagcagctt cctgccgatcgactttgatacttcggttacgcgctgaccctgtttaaagacgatgcggatatccactcttacctgcgcaacgagtacgaacagaacggcaaacgtatgcctaacctggctaacaaagtttacgataacgagctgattgaagaactgctgaacctgtccttcactaaatt cggtcacctgtctctgaaagctctgcgttccatcctgccgtatatggaacagggtgaagtctactcctccgcttgtgaacgtgcaggctacacctttcaccggtccgaaaaagaagcaaaaaactatgctgctgccgaacatcccgccgattgcgaaccctgtagtaatgcgtgcactga cccaggcgcgcaaagtagtcaacgcgatcatcaaaaagtacggcagccccggtttccatccatatcgaactggcgcgcgacctgagccagactttgacgagcgtcgtaaaactaaaaaggaacaggatgaaaaccgtaaaaaaaacgaaaccgcgatccgccagctgatggaat acggtctgactctgaaccctactggtcacgatattgtgaagttcaagctgtggtctgaacagaacggtcgctgtgcttactctctgcagccgatcgagatcgaacgtctgctggagccaggttacgttgaagtagatcatgtgatcccgctctctggatgattcttataccaac aaagttctggttctgactcgcgaaaaccgtgagaaaggcaaccgcatcccagctgaatatctgggtgttggcactgagcgttggcaacagttcgaaaccttcgtcctgaccaataaacagttctctaaaaagaaacgtgaccgtctgctgcgtctgcactacgatgaaaacgaagaga ctgaattcaaaaaccgtaacctgaacgatactcgctacatcagccgcttcttcgcaaacttcattcgtgaacacctgaaatttgcggaatccgacgataaacagaaagtttataccgtaaacggccgtgttaccgcccacctgcgttctcgctgggagttcaacaagaaccgtgaggaa agcgatctgcaccacgctgttgacgccgttattgtggcgtgcaccaccccaagcgatatcgctaaggtgaccgcattctaccagcgtcgtgagcagaacaaggaactggccaaaaaaaccgaaccgcatttccgcagccgtggccgcacttcgcggacgaactgcgtgctcgtctgt ccaaacatcctaaagaaagcatcaaagctctgaacctgggtaactacgatgaccaaaaactggaatctctgcagccggtgtttgtcagccgtatgccgaaacgttctgttactggcgctgccgcaccaggaaacgctgcgccgttacgtgggcatcgacgaacgctccggtaaaatcca gaccgtagtaaaaaccaaactgtccgagattaaactggatgcatccggccacttcccaatgtacggtaaagaatccgatccacgcacttatgaagccatccgccagcgtctgctggagcataacaacgaccgaaaaggcattccaggagcctctgtacaaaccgaaaaaaaacg gcgaaccgggcccggtaatccgctactgtaaaaattatcgacacgaaaaaccaggtgatccctctgaacgatggtaaaaccgtggcctacaattccaacatcgttcgcgtggacgtgttcgaaaaagatggtaaatactactgtgtaccggtgtatacccatggacatcatgaaaggcat tctgccgaacaaagcgattgaaccgaacaagccgtactctgaatggaaagaaatgaccgaagattacacgtttcgtttcagcctgtatccgaacgacctgatccgcatcgaactgccgcgtgaaaaaaccgttaaaacccgctgcaggcgaagaaattaacgtgaaagacgtgttcgt ttactataaaacgatcgactccgcaaacggcggcctggaactgatttctcacgaccaccgtttctctctgcgtggcgttggctctcgcaccctgaaacgtttcgagaaatatcaagttgatgttctgggtaacatctataaagtgcgtggcgaagaaacgtgtcggtctggcgtcctccgca cacagcaaacctggcaaaaccattcgtccactgcaatctactcgtgacggtggaggaggttctgaggtggaggttctgctgagtatgtgcgagccctctttgactttaatgggaatgatgaagaggatcttccctttaagaaaggagacatcctgagaatccgggataagcctga ggagcagtggtggaatgcagaggacagcgaaggaaagaggggatgattcctgtcccttacgtggagaagtattccggagactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggatgaccgacgctgagtacgtgagaatccatgagaagttggacatctacacgtttaagaaacagttttcaacaacaaaaaatccgtgtcgcatagatgctacgttctctttgaa ttaaaacgacggggtgaacgtagagcgtgttttttggggctatgctgtgaataaaccacagagcgggacagaacgtggcattcacgccgaaatctttagcattagaaaagtcgaagaatacctgcgcgacaaccccggacaattcacgataaattggtactcatcctggagtccttgtg cagattgcgctgaaaagatcttagaatggtataaccaggagctgcgggggaacggccacacttgaaaatctgggcttgcaaactctattacgagaaaaatgcgaggaatcaaattgggctgtggaacctcagagataacgggggttgggttgaatgtaatggtaagtgaacactacc aatgttgcaggaaaatattcatccaatcgtcgcacaatcaattgaatgagaatagatggcttgagaagactttgaagcgagctgaaaaacgacggagcgagttgtccattatgattcaggtaaaaatactccacaccactaagagtcctgctgtttctagaatgaccaacctttccga catcatagagaaggaaacaggcaaacagttggtcatccaagagtcgatactcatgcttcctgaagaagttgaggaggtcattgggaataagccggaaagtgacattctcgtacacactgcgtatgatgagagcaccgatgagaacgtgatgctgctcacgtcagatgccccagagt acaaaccctgggctctggtgattcaggactctaatggagagaacaagatcaagatgctaggactggttgcataa
dthermocas9-**linker**-*pmcda1*-**linker**-*ugi*-**linker**-LVA	atgaagtataaaatcggtcttgCtatcggcattacgtctatcggttgggctgtcattaatttggacattcctcgcatcgaagatttaggtgtccgcatttttgacagagcggaaaacccgaaaaccggggagtcactagctcttccacgtcgcctcgcccgctccgcccgacgtcgtctgc ggcgtcgcaaacatcgactggagcgcattcgccgcctgttcgtccgcgaaggaattttaacgaaggaagagctgaacaagctgtttgaaaaaaagcacgaaatcgacgtctggcagcttcgtgttgaagcactggatcgaaaactaaataacgatgaattagcccgcatccttcttc atctggctaaacggcgtggatttagatccaaccgcaagagtgagcgcaccaacaaagaaaacagtacgatgctcaaacatattgaagaaaccaatccattctttcaagttaccgaacggttgcagaaatggttgtcaaggatccgaaatttttcctgcacaagcgtaataaagagg ataattacaccaacactgttgcccgcgacgatcttgaacgggaaatcaaactgattttcgccaaacagcgcgaatatgggaacatcgtttgcacagaagcatttgaacacgagtatatttccatttgggcatcgcaacgccctttttgcttctaaggatgatatcgagaaaaaagtcggtt tctgtacgtttgagcctaaagaaaaacgcgcgccaaaagcaacatacacattccagtccttcaccgtctgggaacatattaacaaacttcgtcttgtctccccgggaggcatccgggcactaaccgatgatgaacgtcgtcttatatacaagcaagcatttcataaaaataaaatcacc ttccatgatgttcgaacattgcttaacttgcctgacgacacccgttttaaaggtcttttatatgaccgaaacaccacgctgaaggaaaatgagaaagttcgcttccttgaactcggcgcctatcataaaatacggaaagcgatcgacagcgtctatggcaaaggagcagcaaaatcatt

Supplementary Table 5. DNA sequences of (fusion) genes used in this study. *(continuation)*

	tcgtccgattgattttgatacatttggctacgcattaacgatgtttaaagacgacaccgacattcgcagttacttgcgaaacgaatacgaacaaaatggaaaacgaatggaaaatctagcggataaagtctatgatgaagaattgattgaagaacttttaaacttatcgttttctaagtttggtcatctatcccttaaagcgcttcgcaacatccttccatatatggaacaaggcgaagtctactcaaccgcttgtgaacgagcaggatatacatttacagggccaaagaaaaaacagaaaacggtattgctgccgaacattccgccgatcgccaatccggtcgtcatgcgcgcactgacacaggcacgcaaagtggtcaatgccattatcaaaaagtacggctcaccggtctccatccatatcgaactggcccgggaactatcacaatcctttgatgaacgacgtaaaatgcagaaagaacaggaaggaaaccgaaagaaaaacgaaactgccattcgccaacttgttgaatatgggctgacgctcaatccaactgggcttgacattgtgaaattcaaactatggagcgaacaaaacggaaaatgtgcctattcactccaaccgatcgaaatcgagcggttgctcgaaccaggctatacagaagtcgacGCtgtgattccatacagccgaagcttggacgatagctatacccaataaagttcttgtgttgacaaaggagaaccgtgaaaaaggaaaccgcaccccagctgaatatttaggattaggctcagaacgttggcaacagttcgagacgtttgtcttgacaaataagcagttttcgaaaaagaagcgggatcgactccttcggcttcattacgatgaaaacgaagaaatgagtttaaaaatcgtaatctaaatgataccgttatatctcacgctttctggctaactttattcgcgaacatctcaaattcgccgacagcgatgacaaacaaaaagtatacacggtcaacggccgtattaccgcccatttacgcagccgttggaatttttaacaaaaaccgggaagaatcgaatttgcatcatgccgtcgatgctgccatcgtcgcctgcacaacgccgagcgatatcgcccgagtcaccgccttctatcaacggcgcgaacaaaacaaagaaactgtccaaaaagacggatccgcagttttccgcagccttggccgcactttgctgatgaactgcaggcgcgttttatcaaaaaatccaaaggagagatataaaagctctcaatcttggaaattatgataacgagaaactcgaatcgttgcagccggttttttgtctcccgaatgccgaagcggagcataacaggagcggctcatcaagaaacattgcggcgttatatcggcatcgacgaacggagcggaaaaatacagacggtcgtcaaaaagaaactatccgagatccaactggataaaacaggtcatttcccaatgtacgggaaagaaagcgatccaaggacatatgaagccattcgccaacggttgcttgaacataacaatgacccaaaaaaggcgtttcaagagcctctgtataaaccgaagaagaacggagaactaggtcctatcatccgaacaatcaaaatcatcgatacgacaaatcaagttattccgctcaacgatggcaaaacagtcgcctacaacagcaacatcgtgcgggtcgacgtctttgagaaagatggcaaatattattgtgtccctatctatacaatagatatgatgaaagggatcttgccaaacaaggcgatcgagccgaacaaaccgtactctgagtggaaggaaatgacggaggactatacattccgattcagtctatacccaaatgatcttatccgtatcgaatttccccgagaaaaaacaataaagactgctgtgggggaagaaatcaaaattaaggatctgttcgcctattatcaaaccatcgactcctccaatggaggggttaagtttggttagccatgataacaactttttcgctccgcagcatcggttcaagaaccctcaaacgattcgagaaatacccaagtagatgtgctaggcaacatctacaaagtgagaggggaaaagagagttggggtggcgtcatcttctcattcgaaagccggggaaactatccgtccgttaggtggaggaggttctggaggtggaggttctgctgagtatgtgcgagccctctttgactttaatgggaatgatgaagaggatcttcccttttaagaaaggagacatcctgagaatccgggataagcctgaggagcagtggtgaatgcagaggacagcgaaggaaagaggggatgattcctgtcccttacgtggagaagtattccggagactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggctcgagtccggagactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggatgaccgacgctgagtacgtgagaatccatgagaagttggacatctacacgtttaagaaacagttttcaacaacaaaaaatccgtgtcgcatagatgctacgttctctttgaatttaaacgacggggtgaacgtagagcgtgttttttggggctatgctgtgaataaacccacagagcgggacagaacgtggcattcacgccgaaatctttagcattagaaaagtcgaagaatacctgcgcgacaacccccggacaattcacgataaattggtactcatcctggagtccttgtgcagattgcgctgaaaagatcttagaatggtataaccaggagctgcgggggaacggccacacttgaaaatctgggcttgcaaactctattacgagaaaaatgcgaggaatcaaattgggctgtggaacctcagagataacgggggttgggttgaatgtaatggtaagtgaacactaccaatgttgcaggaaaatattccatccaatcgtcgcacaatcaattgaatgagaatagatggcttgagaagacttgaagcgagctgaaaaacgacggagcgagttgtccattatgattcaggtaaaaatactccacaccactaagagtcctgctgtttctagaatgaccaacctttccgacatcatagagaaggaaacaggcaaacagttggtcatccaagagtcgatactcatgcttcctgaagaagttgaggaggtcattgggaataagccggaaagtgacattctcgtacacactgcgtatgatgagagcaccgatgagaacgtgatgctgctcacgtcagatgccccagagtacaaacctgggctctggtgattcaggactctaatggagagaacaagatcaagatgctaggactggttgcataa
dgeocas9-linker-*pmcda1*-linker-*ugi*-linker-LVA	atgcgttataagattggcctgGCatcggtattacctctgttggttgggcagtcatgaacctggatatccctcgtatcgaagatctgggcgtgcgcatttttgaccgtgcggagaacccgcagaccggtgaatctctggctctgccgcgtcgtctggcacgtagcgcacgccgccgcctgcgtcgtcgtaaacaccgtctggagcgtattcgtcgcctggttattcgtgaaggcatcctgacgaaagaagaactggataaactgttcgaagaaaaacacgagatcgacgtatggcagctgcgtgtagaagccctggaccgtaagctgaacaacgacgaactggcgcgtgtcctgctgcatctggcaaagcgtcgtggcttcaaatctaaccgtaaatctgaacgctccaataaagagaaactccactatgctgaaacatattgaggagaaccgtgcaattctgtctagctaccgtaccgtgggcgaaatgattgttaaagacccgaaattcgcactgcataaggcgtaacaaaggcgaaaactacaccaacatcattgcacgcgatgacctggaacgtgaaatccgtctgattttctccaaacagcgcgaattcggcaacatgtcttgcaccgaagaattcgaaaacgaatatattaccatttgggcatctcagcgtccggtggcgtctaaagatgatatcgaaaaaaagtaggcttttgtactttcgaaccgaaggaaaaacgtgcgccgaaagccacctatacctttccagtctttttatcgcgtgggaacatatcaacaaactgcgtctgatttctccgtctggcgcccgcggcctgaccgacgaagaacgtcgtctgctgtatgaacaagcattccagaaaaacaaaattaccctaccacgatattcgtacccgtcgcatctgccggacgacacctacttcaagggcatcgtttacgatcgcggtgaatctcgtaagcagaacgaaacattcgtttcctggaactggatgcatacccaccagatccgtaaagctgtagataaagtttacggcaagggtaaatccagcagctttcctgccgatcgacttttgatacctcggttacgcgctgaccctgtttaagagacgatgcggatatccactcttactgcgcaacgagtacgaacagaacggcaaacgtatgcctaacctggctacaaagtttacgataacagcgtcgattcgaagaactgcttccttcactaaattcggtcacctgtctctgaaagctctgcgttccatcctgccgtatatggaacagggtgaagtctactcctccccgcttgtgaacgtgcaggctacaccttcaccggtccgaaaaagaagcaaaaaactatgctgctgccgaacatcccgccgattgcgaacccttgtagtaatgcgtgcactgacccaggcgcgcaaagtagtcaacgcgatcatcaaaaagtacggcagcccggtttccatccatatcgaactggcgcgcgacctgagccagacttttgacgagcgtcgtaaaactaaaaaggaacaggatgaaaaccgtaaaaaaaacgaaaccgcgatccgccagctgatggaatacggtctgactctgaacccctactggtcacgatattgtgaagttcaagctgtggtctgaacagaacggtcgctgtgcttactctctgcagccgatcgagatcgaacgtctgctggagccaggttacgttgaagtagatGCtgtgatcccgtactcccgctctctggatgatttcttatacccaaagttctggtttctgactcgcgaaaacctgtagaaaggcaacccgcatcccagctgaatatctgggtgttggcactgagcgttggcaacagttcgaaaccttcgtcctgaccaataaacagttctctaaaaagaaacgtgaccgtctgctgcgtctgcatcactacgatgaaaacgaagagactgaattcaaaaaccgtaacctgaacgatactcgctacatcagccgcttctttcgcaaacttcattcgtgaacacctgaaatttgcggaatccgacgatcaaacagaaagtttatacccgtaaacggccgtgttaccgcccacctgcgttctcgctgggagttcaacaagaaccgtgaggaaagccgatctgcaccacgctgttgacgccgttattgtggcgtgcaccacccccaagcgatatcgctaaggtgaccgcattctaccagcgtcgtgagcagaacaaggaactggccaaaaaaaccgaaccgcatttttccgcagccgtggccgcacttcgcggacgaactgcgtgctcgtctgtccaaacatcctaaagaaagcatcaaagctctgaacctgggtaactacgatgaccaaaaactggaatctctgcagccggtgtttgtcagccgtatgccgaaacgtttctgttactggcgctgcgcaccaggaaacgctgcgccgttacgtgggcatcgacgaacgctccggtaaaatccagaccgtagtaaaaaccaaactgtccgagattaaactggatgcatccggccactttcccaatgtacggtaaagaatccgatccacgcacttatgaagccatccgccagcgtctgctggagcataacaacgacccgaaaaaggcattccaggagcctctgtacaaaccgaaaaaaaacggcgaaccgggcccggtaatccgtactgtaaaaattatcgacacgaaaaaccaggtgatccctctgaacgatggtaaaaccgtggcctacaattccaacatcgttcgcgtggacgtgttcgaaaaagatggtaaatactactgtgtaccggtgtatacctggacatcatgaaaggcatttctgccgaacaaagcgattgaaccgaacaagccgtactctgaatggaaagaaatgaccgaagattacacgtttcgttttcagcctgtatccgaacgacctgatccgcatcgaactgccgcgtgaaaaaccgttaaaaccgctgcaggcgaagaaattaacgtgaaagacgtgttcgtttactatataaaacgatcgactccgcaaacggcggcctggaactgatttctcacgaccaccgtttctctctgcgtggcgttggctctcgcaccctgaaacgtttcgagaaatatcaagttgatgttctgggtaacatctataaagtgcgtggcgagaaacgtgtcggtctggcgtcctccgcacacagcaaacctggcaaaaccattcgtccactgcaatctactcgtgacggtggaggaggttctggaggtggaggttctgctgagtatgtgcgagccctctttgactttaatgggaatgatgaagaggatcttcccttttaagaaaggagacatcctgagaatccgggataagcctgaggagcagtggtgaatgcagaggacagcgaaggaaagaggggatgattcctgtcccttacgtggagaagtattccggagactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggctcgagtccggagactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataagtctaggatgaccgacgctgagtacgtgagaatccatgagaagttggacatctacacgtttaagaaacagttttcaacaacaaaaaatccgtgtcgcatagatgctacgttctctttggaattaaaacgacggggtgaacgtagagcgtgttttttggggctatgctgtgaataaacccacagagcgggacagaacgtggcattcacgccgaaatctttagcattagaaaagtcgaagaatacctgcgcgacaacccccggacaattcacgataaattggtactcatcctggagtccttgtgcagattgcgctgaaaagatcttagaatggtataaccaggagctgcgggggaacggccacacttgaaaatctgggcttgcaaactctattacgagaaaaatgcgaggaatcaaattgggctgtggaacctcagagataacgggggttgggttgaatgtaatggtaagtgaacactaccaatgttgcaggaaaatattccatccaatcgtcgcacaatcaattgaatgagaatagatggcttgagaagacttgaagcgagctgaaaaacgacggagcgagttgtccattatgattcaggtaaaaatactccacaccactaagagtcctgctgtttctagaatgaccaacctttccgacatcatagagaaggaaacaggcaaacagttggtcatccaagagtcgatactcatgcttcctgaagaagttgaggaggtcattgggaataagccggaaagtgacattctcgtacacactgcgtatgatgagagcaccgatgagaacgtgatgctgctcacgtcagatgccccagagtacaaacctgggctctggtgattcaggactctaatggagagaacaagatcaagatgctaggactggttgcataa

Supplementary Table 5. DNA sequences of (fusion) genes used in this study. *(continuation)*

acriic1	atgaataaaacatataaaatcggaaaaaatgcgggatatgacggctgcgggttatgtttagccgcgatctcagaaaatgaagctattaaagttaagtatctgcgcgatatctgtccagactatgacggcgacgacaaagctgaggactggctgagatggggaacggatagccgcgt gaaagctgcggctttagaaatggagcaatatgcttatacgtcggttgggatggcgtcatgttgggagtttgttgaactatga

***Red color and underlined**: nucleotide substitution*

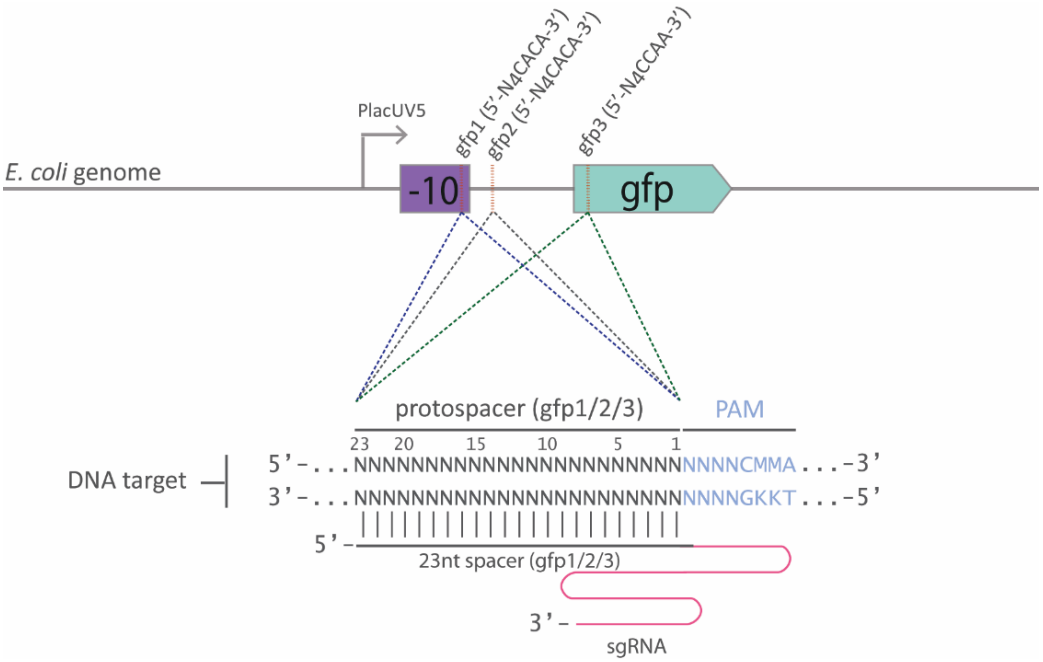
Supplementary Table 6. Amino acid sequences of (fusion) proteins used in this study.

Protein/ Fusion protein	Sequence (5' -> 3')
ThermoCas9 (wild-type)	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSIL SSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQ KTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKG NRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDP QFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIID TKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEIVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIY KVRGEKRVGLASSAHSKPGKTIRPLQSTRD*
GeoCas9 (wild-type)	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRA ILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLL HLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQ KTMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNKVLVLTRENREK GNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKELAKKTE PHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIID TKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEIKVDFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIY KVRGEKRVGLASSAHSKPGKTIRPLQSTRD*
dThermoCas9 (D8A, H582A mutant)	MKYKIGLAIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSIL SSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQ KTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDAVIPYSRSLDDSYTNKVLVLTKENREKG NRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFFANFIREHLKFADSDDKQKVYTVNGRVTAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDP QFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIID TTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTIDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGE KRVGVASSSHSKAGETIRPL*
dGeoCas9 (D8A, H582A mutant)	MRYKIGLAIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRA ILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLL HLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQ KTMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDAVIPYSRSLDDSYTNKVLVLTRENREK GNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKELAKKTE PHFPQPWPHFADELQARLSKHPKESIKALNLGNYDDQKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIID TKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEIKVDFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIYKVRGE KRVGVASSSHSKAGETIRPL*
ThermoCas9-**linker**-PmCDA1-**linker**-UGI-**linker**-LVA fusion	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSIL SSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQ KTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKG NRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFFANFIREHLKFADSDDKQKVYTVNGRVTAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDP QFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTN QVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGE KRVGVASSSHSKAGETIRPL**GGGGSGGGGSAEYVRALFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYSGDYKDHDGDYKDHDIDYKDDDDKSR**LESGDYKDHDGDYKDHDIDYKDDDDKSR MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVG

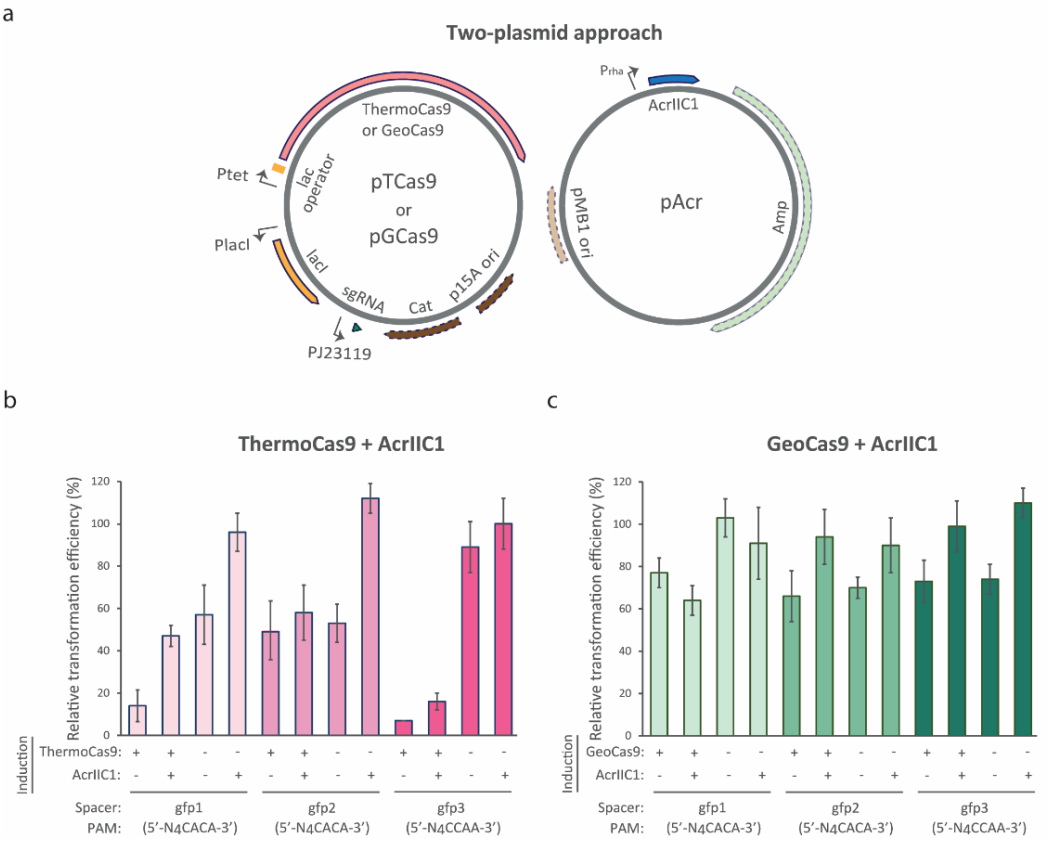
Supplementary Table 6. Amino acid sequences of (fusion) proteins used in this study. *(continuation)*

	LNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV**SR**MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML**G**LVA*
GeoCas9-**linker**-PmCDA1-**linker**-UGI-**linker**-LVA	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRA ILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLL HLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQ KTMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNKVLVLTRENREK GNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKELAKKTE PHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIID TKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIY KVRGEKRVGLASSAHSKPGKTIRPLQSTRD**GGGGSGGGGSAEYVRALFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYSGDYKDHDGDYKDHDIDYKDDDDKSR**LESGDYKDHDGDYKDHDID YKDDDDKSRMTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGL WNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV**SR**MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKML**G**LVA*
dThermoCas9-**linker**-PmCDA1-**linker**-UGI-**linker**-LVA	MKYKIGLAIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSIL SSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQ KTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDAVIPYSRSLDDSYTNKVLVLTKENREKG NRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDP QFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTN QVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGE KRVGVASSSHSKAGETIRPL**GGGGSGGGGSAEYVRALFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYSGDYKDHDGDYKDHDIDYKDDDDKSR**LESGDYKDHDGDYKDHDIDYKDDDDKSR MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVG LNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV**SR**MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML**G**LVA*
dGeoCas9-**linker**-PmCDA1-**linker**-UGI-**linker**-LVA	MRYKIGLAIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDRKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRA ILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLL HLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQ KTMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDAVIPYSRSLDDSYTNKVLVLTRENREK GNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKELAKKTE PHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIRTVKIID TKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIY KVRGEKRVGLASSAHSKPGKTIRPLQSTRD**GGGGSGGGGSAEYVRALFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSEGKRGMIPVPYVEKYSGDYKDHDGDYKDHDIDYKDDDDKSR**LESGDYKDHDGDYKDHDID YKDDDDKSRMTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGL WNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV**SR**MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG ENKIKML**G**LVA*
AcrIIC1	MNKTYKIGKNAGYDGCGLCLAAISENEAIKVKYLRDICPDYDGDDKAEDWLRWGTDSRVKAAALEMEQYAYTSVGMASCWEFVEL*

***Red color and underlined**: amino acid substitution*

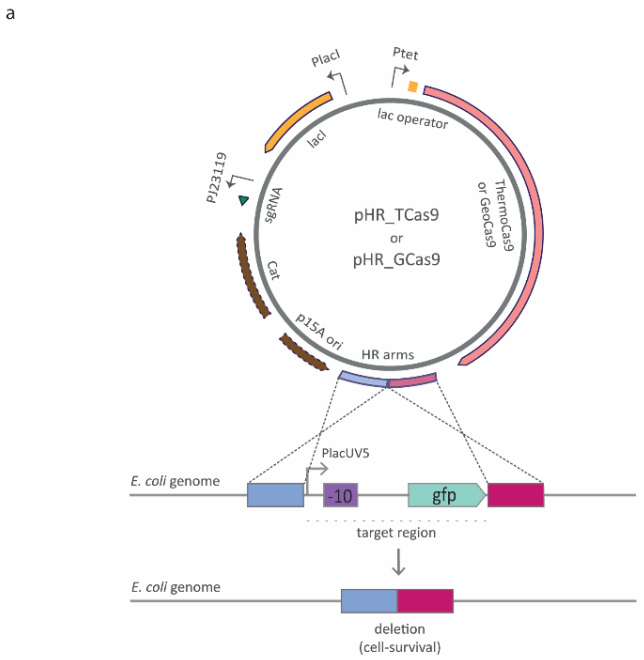


Supplementary Fig. 1 Schematic illustration of the genomic location of the targeted protospacers (gfp1/2/3). Protospacers gfp1 and gfp2 are located in the lacUV5 promoter of the *gfp* gene and are flanked by the 5'-N_4CACA-3' PAM, while protospacer gfp3 is found in the beginning of the *gfp* gene and is flanked by the 5'-N_4CCAA-3' PAM. Red dashed lines denote the position of each protospacer in the *E. coli* genome, while the purple square represents the -10 box of P_{lacUV5}.



Supplementary Fig. 2 Two-plasmid approach for AcrIIC1-mediated inhibition of ThermoCas9 and GeoCas9 *in vivo* cleavage activity. a Schematic illustration of the constructs transformed into the *E. coli_gfp* strain in killing-inhibition assays. **b, c** Transformation efficiency of *E. coli_gfp* cells that express AcrIIC1 for inhibition of either ThermoCas9- (**b**) or GeoCas9- (**c**) based cleavage activity. The symbols '+' and '-' represent induction (1000 µM IPTG for Cas9; 0.2% L-rhamnose for AcrIIC1) and absence of induction of protein expression, respectively. Bar graphs were created based on results from three independent biological replicates.

a



b

Replicate 1

ThermoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CCAA-3'	gfp3			20	20	100	100
	gfp4			20	20	100	100
5'-N4CGAA-3'	gfp5	1		12	13	92	92
	gfp6		9	13	22	100	59
5'-N4CAAA-3'	gfp7			8	8	100	100
	gfp8	13		3	16	19	19

Replicate 2

ThermoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CCAA-3'	gfp3			10	10	100	100
	gfp4			10	10	100	100
5'-N4CGAA-3'	gfp5			10	10	100	100
	gfp6			10	10	100	100
5'-N4CAAA-3'	gfp7	1		9	10	90	90
	gfp8	3	9	1	10	70	10

Replicate 3

ThermoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CCAA-3'	gfp3	1		9	10	90	90
	gfp4			10	10	100	100
5'-N4CGAA-3'	gfp5			10	10	100	100
	gfp6			10	10	100	100
5'-N4CAAA-3'	gfp7			10	10	100	100
	gfp8	2	6	8	16	88	50

c

Replicate 1

GeoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CGAA-3'	gfp5	1	2	17	20	95	85
	gfp6		9	11	20	100	55
5'-N4CAAA-3'	gfp7		3	15	18	100	83
	gfp8			20	20	100	100
5'-N4GAAA-3'	gfp9	12			12	0	0
	gfp10	18			18	0	0

Replicate 2

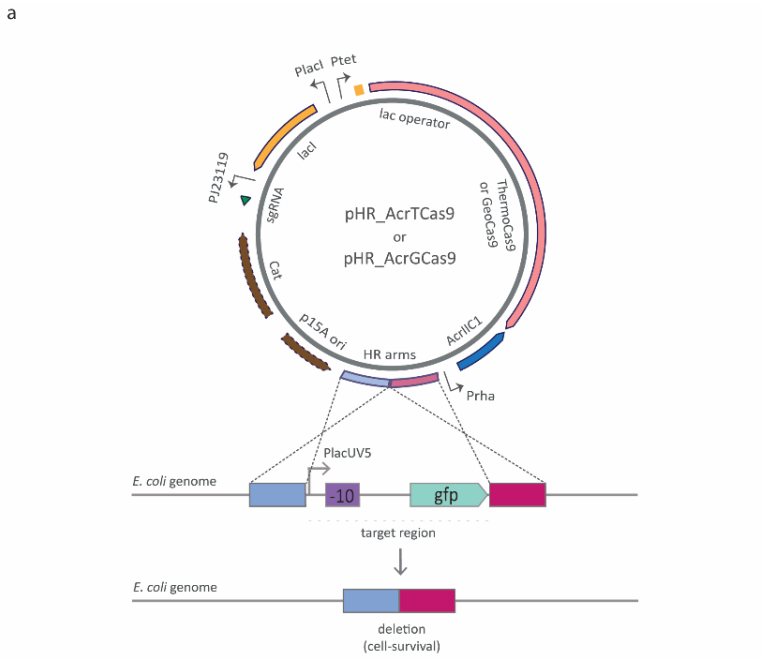
GeoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CGAA-3'	gfp5		6	4	10	100	40
	gfp6	2	3	5	10	80	50
5'-N4CAAA-3'	gfp7			10	10	100	100
	gfp8	1		9	10	90	90
5'-N4GAAA-3'	gfp9	5	5		10	50	0
	gfp10	6	3		9	33	0

Replicate 3

GeoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CGAA-3'	gfp5		9	1	10	100	10
	gfp6	1	2	7	10	90	70
5'-N4CAAA-3'	gfp7			10	10	100	100
	gfp8	1		9	10	90	90
5'-N4GAAA-3'	gfp9	6	4		10	40	0
	gfp10	5	5		10	50	0

Supplementary Fig. 3 ThermoCas9- and GeoCas9-based genome engineering in *E. coli*. a Schematic illustration of the pHRs_TCas9 and pHRs_GCas9 constructs for ThermoCas9- and GeoCas9-based genome editing in *E. coli_gfp* that carries the λ-red recombineering-expressing pKD46 plasmid. Homologous recombination arms (~600 bp each) of the target region were introduced to the previously constructed targeting plasmids to allow for λ-red-mediated deletion of the genomic *gfp* gene and along with the P_{lacUV5}, and subsequent counter-selection of the non-edited cells. Six spacers for each nuclease, designed to target protospacers within the sequence of the *gfp* gene and flanked by PAMs of variable preference, were inserted in the 5' end of the sgRNA-expressing module. **b**, **c** Overview of the number of screened colonies and their genotype for the ThermoCas9- (**b**) and GeoCas9- (**c**) mediated counter-selection approach. The genotype of the screened colonies is categorised into three groups regarding the deletion of the genomic *gfp* gene and its promoter; wild-type (WT), mixed, and clean knock-out (KO) mutant colonies. The results are based on three independent biological replicates, and the expression of Cas9 was always fully induced (1 mM IPTG).

a



b

Replicate 1

Acr + ThermoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CCAA-3'	gfp3	11	7	1	19	42	5
	gfp4	13	7		20	35	0
5'-N4CGAA-3'	gfp5	9	11		20	55	0
	gfp6	19			19	0	0
5'-N4CAAA-3'	gfp7	20			20	0	0
	gfp8	16		3	19	16	16

Replicate 2

Acr + ThermoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CCAA-3'	gfp3	6	1	2	9	33	22
	gfp4	6	4		10	40	0
5'-N4CGAA-3'	gfp5	4	6		10	60	0
	gfp6	10			10	0	0
5'-N4CAAA-3'	gfp7	10			10	0	0
	gfp8	9		1	10	10	10

Replicate 3

Acr + ThermoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CCAA-3'	gfp3	8		2	10	20	20
	gfp4	6	4		10	40	0
5'-N4CGAA-3'	gfp5	4	6		10	60	0
	gfp6	10			10	0	0
5'-N4CAAA-3'	gfp7	10			10	0	0
	gfp8	9		1	10	10	10

c

Replicate 1

Acr + GeoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CGAA-3'	gfp5	10	10		20	50	0
	gfp6	2	3	3	8	75	38
5'-N4CAAA-3'	gfp7	2	2	3	6	83	50
	gfp8	11	7		18	39	0
5'-N4GAAA-3'	gfp9	14	5		19	26	0
	gfp10	15	5		20	25	0

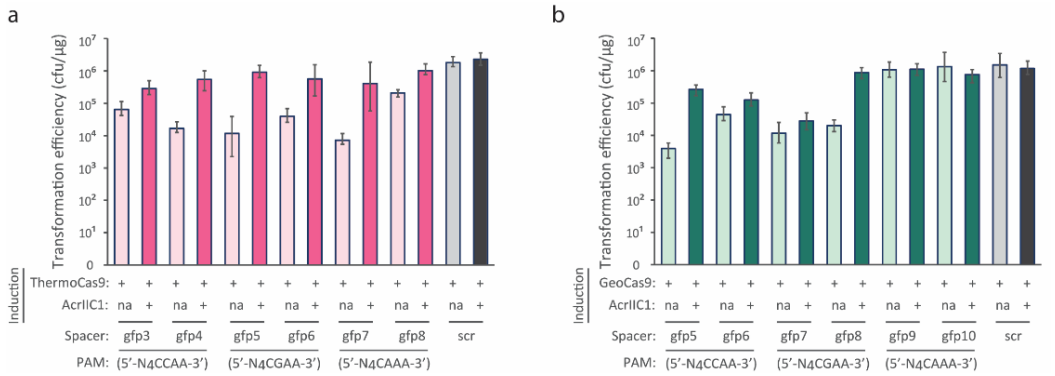
Replicate 2

Acr + GeoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CGAA-3'	gfp5	4	6		10	60	0
	gfp6	4	3	3	10	60	30
5'-N4CAAA-3'	gfp7	3	2	5	10	70	50
	gfp8	5	3		8	38	0
5'-N4GAAA-3'	gfp9	7	2		9	22	0
	gfp10	8	2		10	20	0

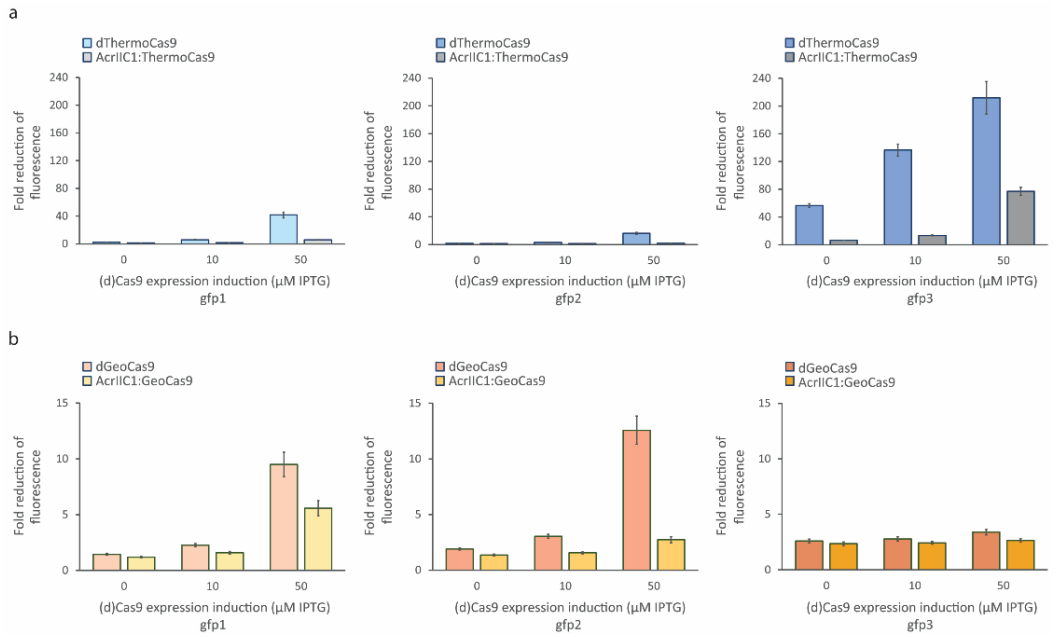
Replicate 3

Acr + GeoCas9							
PAM	Spacer	Genotype of screened colonies			Screened colonies	% edited colonies	% clean KO
		WT	Mixed	KO			
5'-N4CGAA-3'	gfp5	4	5		9	56	0
	gfp6	4	3	3	10	60	30
5'-N4CAAA-3'	gfp7	2	2	5	9	78	56
	gfp8	6	3		9	33	0
5'-N4GAAA-3'	gfp9	8	2		10	20	0
	gfp10	7	2		9	22	0

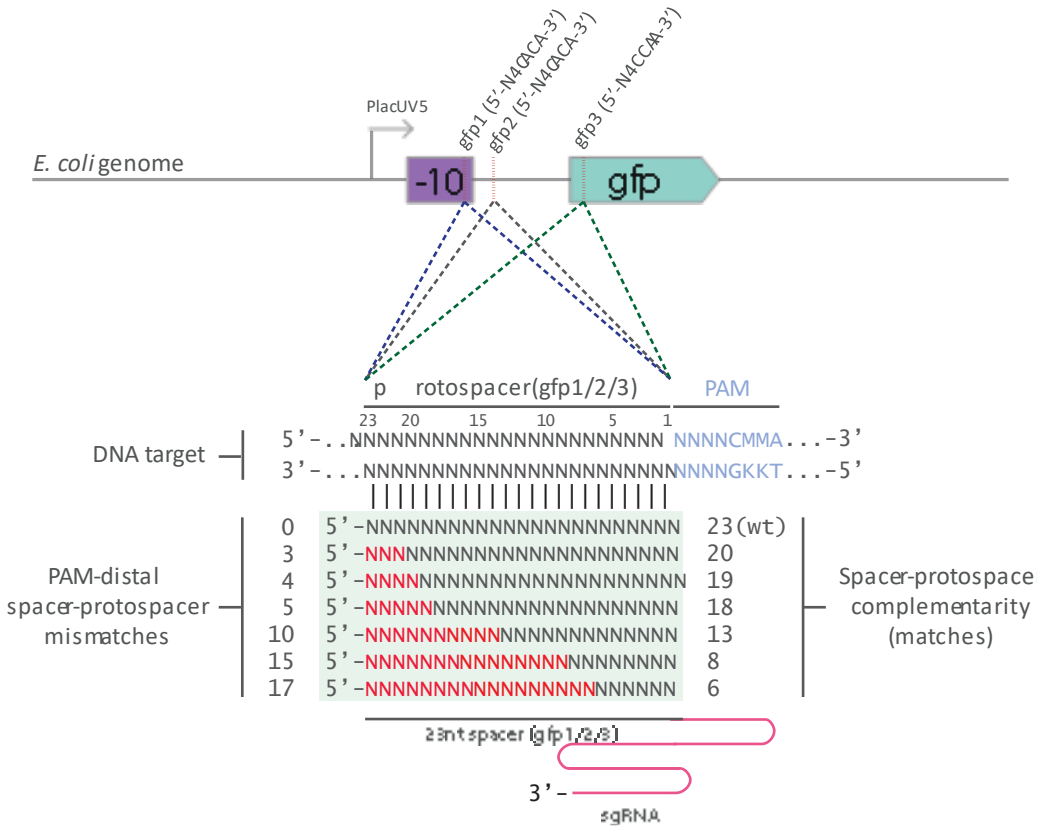
Supplementary Fig. 4 AcrIIC1-mediated inhibition of ThermoCas9- and GeoCas9-based genome engineering in *E. coli*. a Schematic illustration of the pHRs_AcrTCas9 and pHRs_AcrGCas9 constructs for AcrIIC1-mediated inhibition of ThermoCas9- and GeoCas9-based genome editing in *E. coli_gfp* that carries the λ-red recombineering-expressing pKD46 plasmid. The AcrIIC1 expressing module was inserted into the previously constructed editing plasmids to block Cas9-based counter-selection of non-edited cells. **b**, **c** Overview of the number of screened colonies and their genotype for the ThermoCas9- (**b**) and GeoCas9- (**c**) mediated counter-selection approach, in the presence of AcrIIC1. The genotype of the screened colonies is categorised into three groups regarding the deletion of the genomic *gfp* gene and its promoter; wild-type (WT), mixed, and clean knock-out (KO) mutant colonies. The expression of AcrIIC1 and Cas9 was fully induced. The results are based on three independent biological replicates.



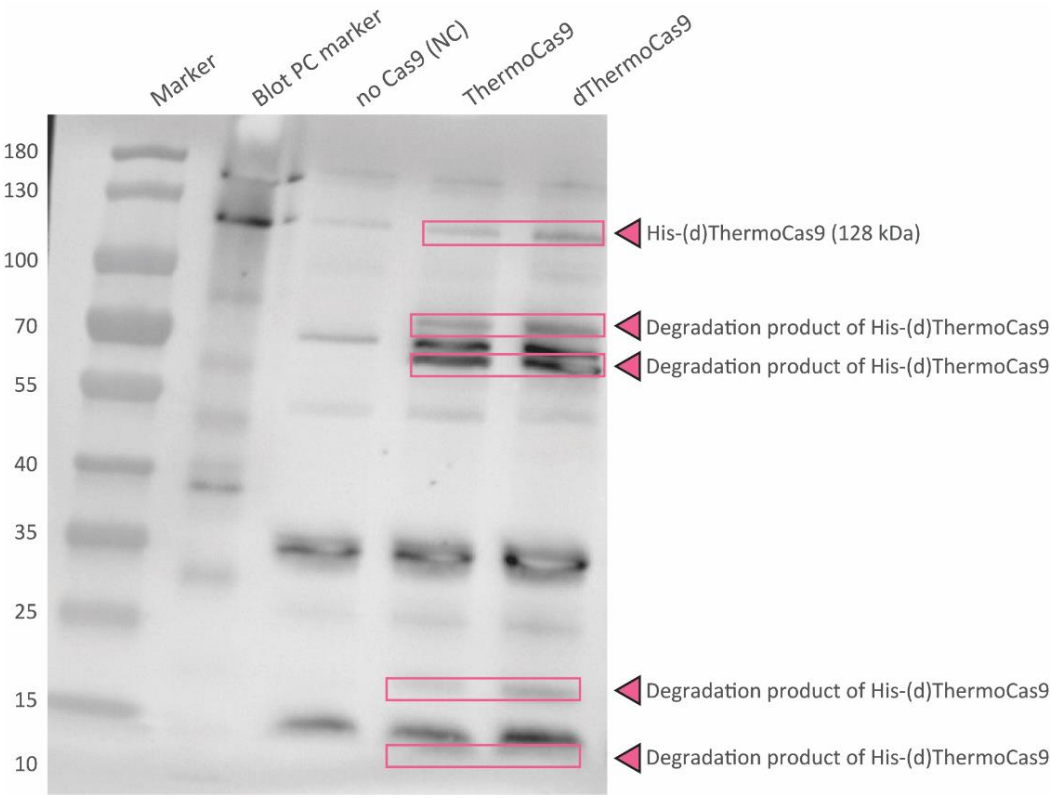
Supplementary Fig. 5 AcrIIC1-mediated inhibition of ThermoCas9- and GeoCas9-based cell killing during genome engineering in *E. coli*. Transformation efficiency of *E. coli_gfp* cells that express either ThermoCas9 (**a**) or GeoCas9 (**b**), guided by sgRNA with spacers gfp3-gfp8 and gfp5-gfp10, respectively. Light-coloured bars indicate full expression of Cas9 and absence of the *acriic1* gene ('not applicable' AcrIIC1 expression induction, 'na') in the genome editing plasmid, while dark-coloured bars indicate full expression of both Cas9 and AcrIIC1. Bar graphs were created based on results from three independent biological replicates.



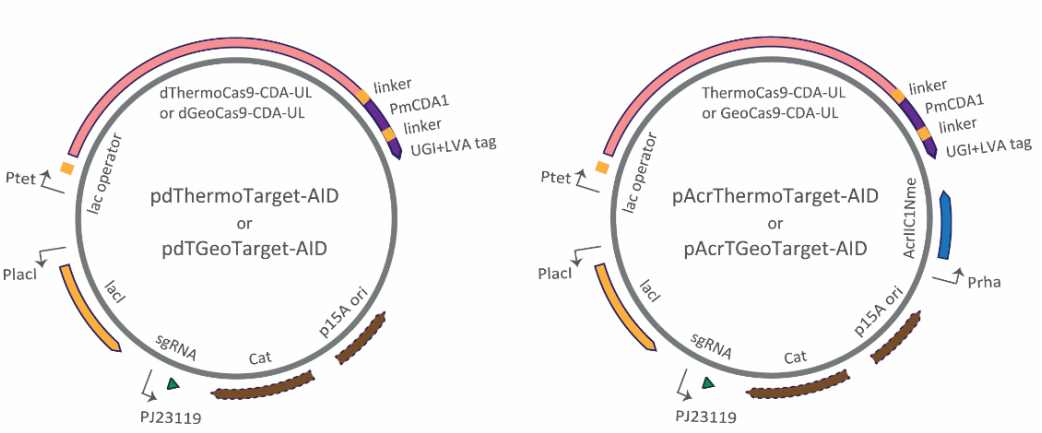
Supplementary Fig. 6 Comparison of dThermoCas9 and dGeoCas9 with their AcrIIC1:Cas9 complexes for silencing of the genome-integrated *gfp* gene in *E. coli*. Flow cytometry-based fluorescence loss assays show fold reduction of fluorescence of *E. coli_gfp* cells that express (**a**) dThermoCas9 (blue) or AcrIIC1:ThermoCas9 (grey); (**b**) dGeoCas9 (orange) or AcrIIC1:GeoCas9 (yellow), guided by sgRNA with spacers gfp1, gfp2, or gfp3. These spacers correspond to protospacers flanked by certain PAMs (5'-N4CACA-3' for gfp1 and gfp2; 5'-N4CCAA-3' for gfp3). The expression of the (d)Cas9 proteins was induced using variable IPTG concentrations (0, 10, 50 μM), while the expression of AcrIIC1 was always induced with 0.2% L-rhamnose in this two-plasmid approach. Bar graphs were created based on results from three independent biological replicates.



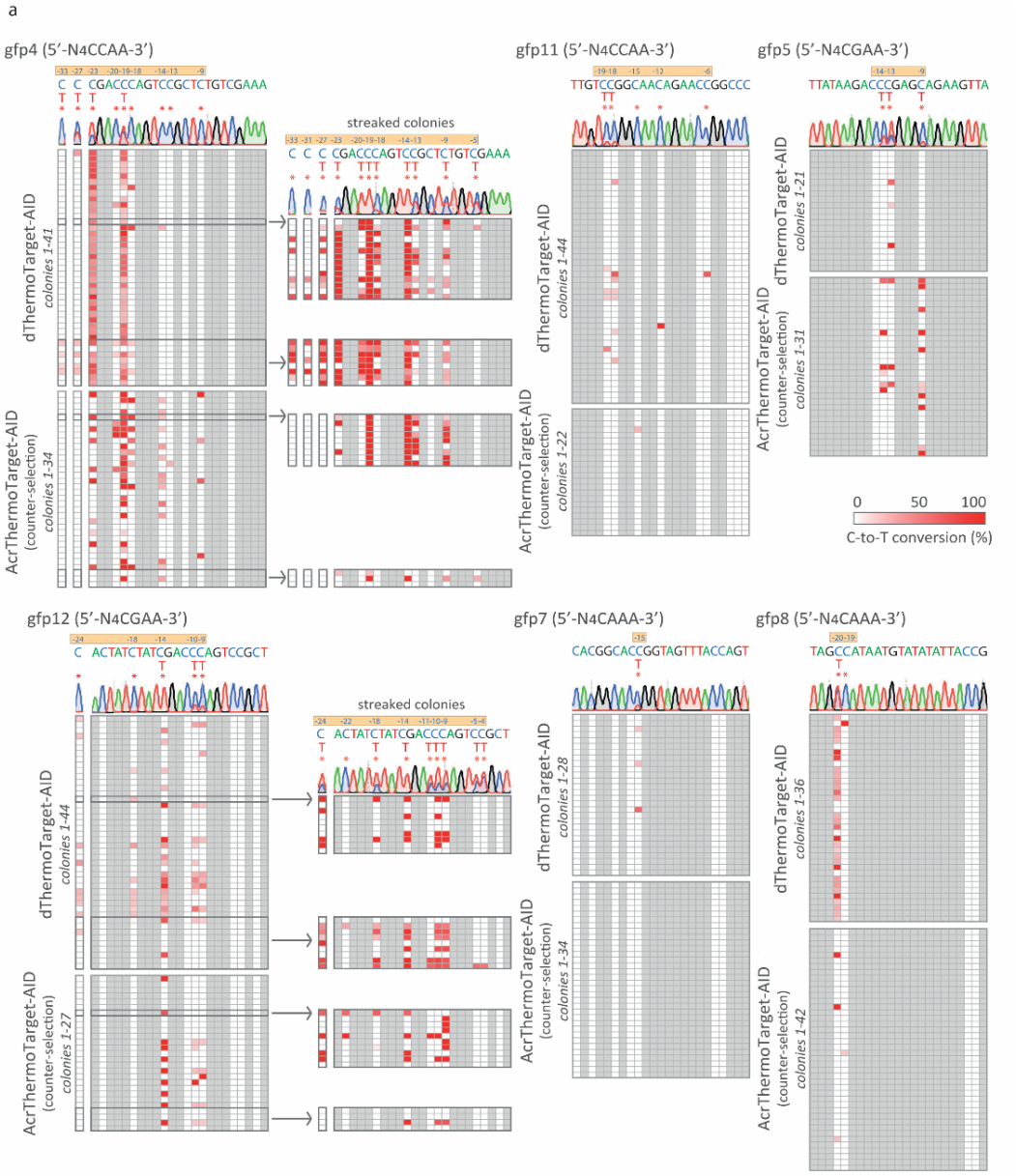
Supplementary Fig. 7 Spacer-protospacer mismatch library for examination of the targeting and binding specificity of ThermoCas9, GeoCas9, their deactivated variants, and their AcrIIC1:Cas9 complexes in *E. coli_gfp*. Multiple consecutive mismatches (red) were introduced in the 5'-end of the spacer (gfp1, gfp2, or gfp3), while the rest of the nucleotides of the spacer remained matched (black) to the corresponding protospacer. The PAM is shown in light blue, and the sgRNA-expressing module is depicted in pink.

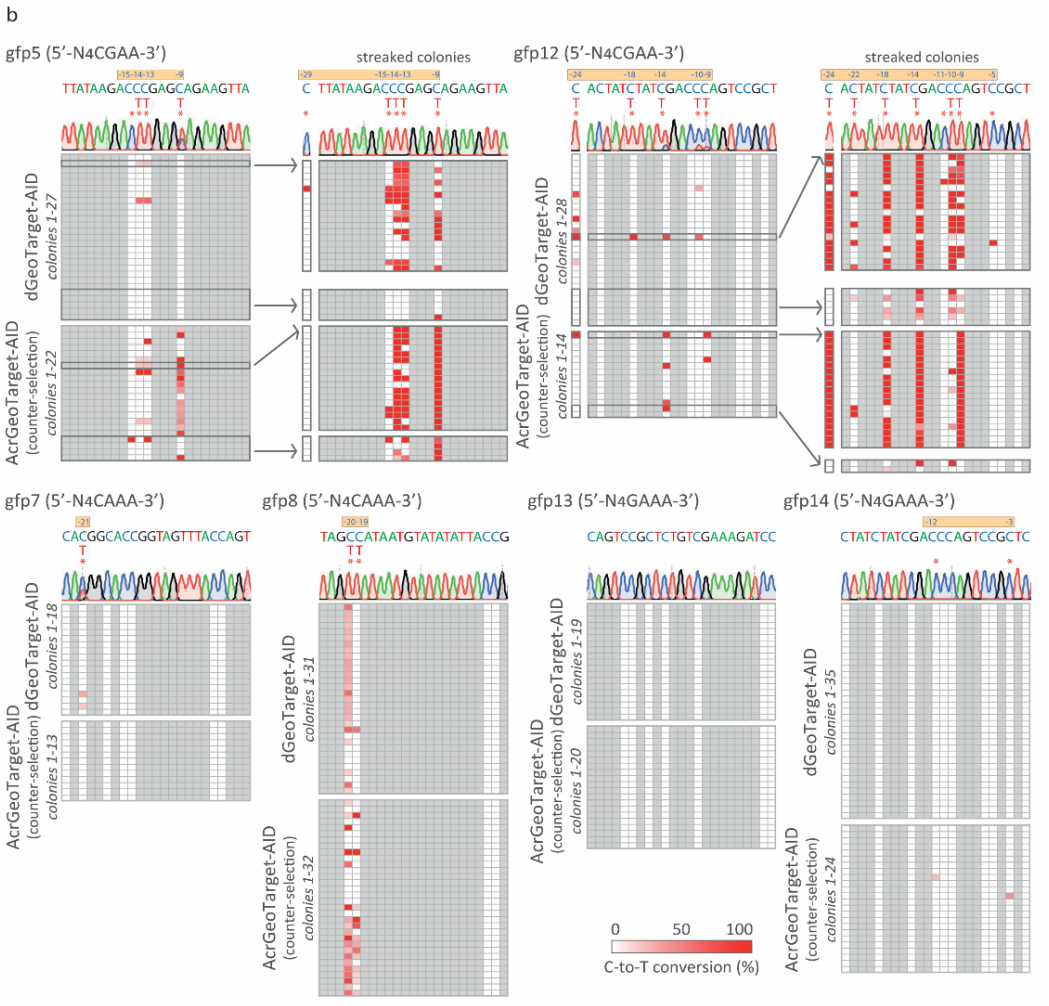


Supplementary Fig. 8 Comparison of His-ThermoCas9 and His-dThermoCas9 expression in *E. coli_gfp*. Western-blot assay shows the expression of His6-ThermoCas9, His6-dThermoCas9, and their degradation products (pink boxes). A band of approximately 128 kDa is observed not only in the presence (His-ThermoCas9, His-dThermoCas9) but also in the absence (no His-(d)ThermoCas9; NC) of His-(d)ThermoCas9. So, *E. coli_gfp* cells express a protein of the same size as the His-(d)ThermoCas9, not allowing us to draw any conclusions. However, His-(d)ThermoCas9 degradation products are present in the His-(d)ThermoCas9 samples, while they are absent in the no His-(d)ThermoCas9 sample. The intensity of these bands is comparable between the His-ThermoCas9 and the His-(d)ThermoCas9 samples.

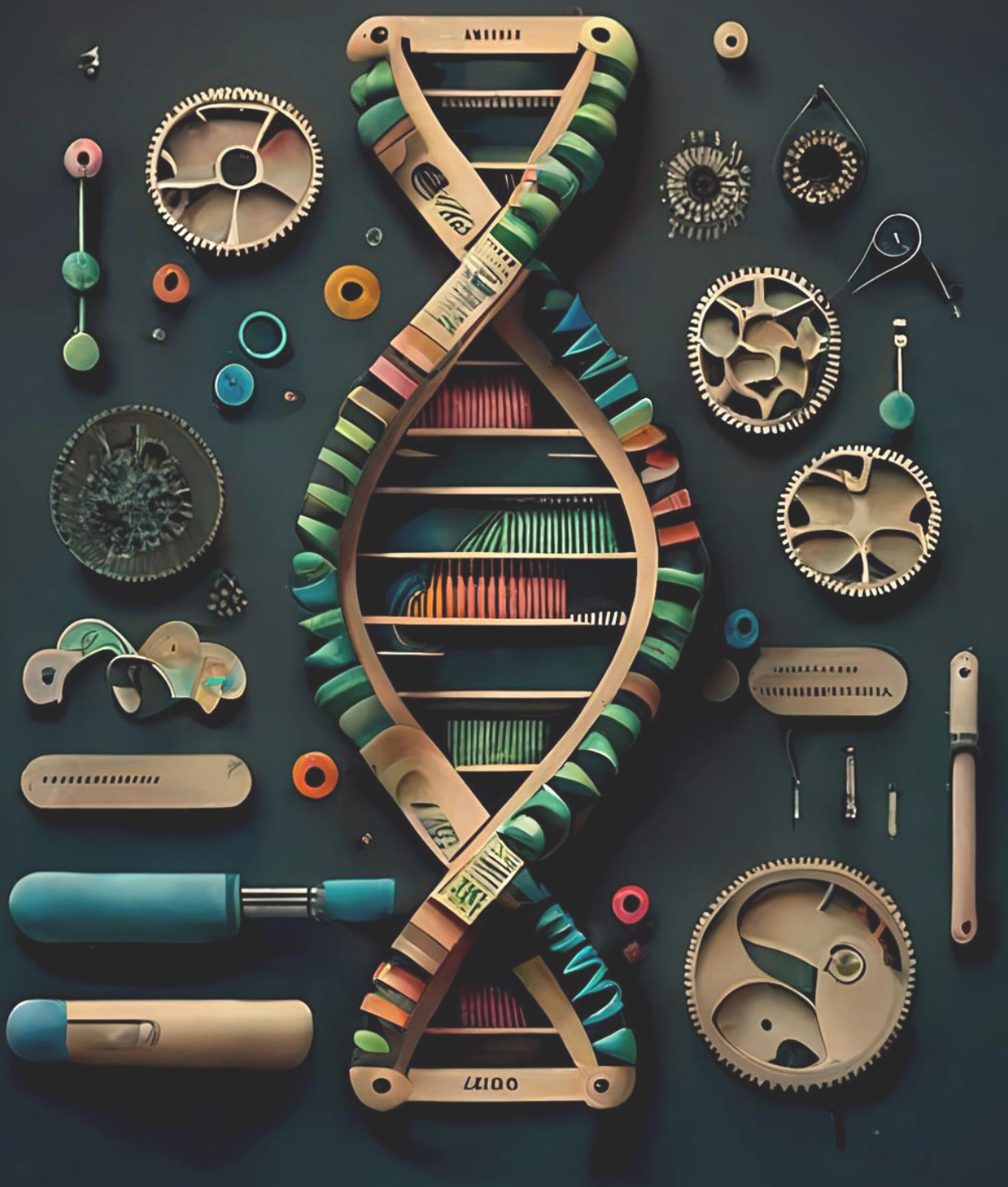


Supplementary Fig. 9 Base-editing in *E. coli* using dThermoCas9, dGeoCas9, and their AcrIIC1:Cas9 complexes. Schematic illustration of the dCas9 and AcrIIC1:Cas9 base-editing constructs. The *pmcda1* and *ugi* genes were fused to the 3' end of the *dcas9* or active *cas9* genes of the previously described pdTCas9, pdGCas9, pAcrTCas9, and pAcrGCas9 plasmids to enable base-editing. For each base-editor, six spacers targeting the *gfp* gene of the *E. coli_gfp* genome were applied.


a
gfp4 (5′-N4CCAA-3′)
gfp11 (5′-N4CCAA-3′)
gfp5 (5′-N4CGAA-3′)
gfp12 (5′-N4CGAA-3′)
gfp7 (5′-N4CAAA-3′)
gfp8 (5′-N4CAAA-3′)
streaked colonies
dThermoTarget-AID
colonies 1-41
AcrThermoTarget-AID
(counter-selection)
colonies 1-34
colonies 1-44
colonies 1-22
colonies 1-21
colonies 1-31
colonies 1-27
colonies 1-28
colonies 1-36
colonies 1-42
0
50
100
C-to-T conversion (%)



Supplementary Fig. 10 Characterisation of the dCas9 and AcrIIC1:Cas9 base-editors in *E. coli*. (a) ThermoTarget-AID, or **(b)** GeoTarget-AID editing heatmaps depicting the percentage of C•G to T•A conversion in every C-position within or immediately upstream of protospacers (x axis) located in the genomic *gfp* gene of *E. coli_gfp* single colonies transformed with the base-editing vectors (y axis). The percentages resulted from high-throughput Sanger sequencing analysis of several single colonies, employing a variation of the on-line tool 'EditR' and setting as threshold $p \leq 0.5$. White boxes represent no base-editing, light to darker pink boxes represent increasing base-editing efficiencies, and red boxes represent 100% base-editing efficiency. The red asterisks indicate edited Cs at a certain position in at least one of the screened colonies. The yellow boxes indicate the positions of edited Cs reported for the screened colonies. For each protospacer, the Sanger sequencing chromatogram of a random, screened single colony is presented as an example.



Chapter 6

Efficient genome and base editing in human cells using ThermoCas9

Despoina Trasanidou[1,2], Patrick Barendse[1,2,*,^], Evgenios Bouzetos[1,2,*], Laura de Haan[2], Hans Bouwmeester[2], Raymond H.J. Staals[1], Ioannis Mougiakos[1,**,#], & John van der Oost[1,**]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

[2]Laboratory of Toxicology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

*Equal contributions

**Corresponding authors

Accepted for publication (The CRISPR Journal)

Abstract

Most genetic engineering applications reported thus far rely on the type II-A CRISPR-Cas9 nuclease from *Streptococcus pyogenes* (SpyCas9), limiting the genome-targeting scope. Here, we demonstrate that a small, naturally accurate and thermostable type II-C Cas9 orthologue from *Geobacillus thermodenitrificans* (ThermoCas9) with alternative target site preference is active in human cells, and it can be used as an efficient genome editing tool, especially for gene disruption. In addition, we develop a ThermoCas9-mediated base-editor, called ThermoBE4, for programmable nicking and subsequent C-to-T conversions in human genomes. ThermoBE4 exhibits a three times larger window of activity compared to the corresponding SpyCas9 base-editor (BE4), which may be an advantage for gene mutagenesis applications. Hence, ThermoCas9 provides an alternative platform that expands the targeting scope of both genome- and base-editing in human cells.

Keywords

ThermoCas9, thermostable, genome editing, base editing, HEK293, human cells

Introduction

The CRISPR-Cas systems consist of clustered, regularly interspaced, short palindromic repeats (CRISPR) loci and CRISPR-associated (Cas) proteins. These adaptive immune systems protect their prokaryotic host against invading genetic elements[1-3]. Over the past decade, several CRISPR-Cas nucleases have been repurposed as genetic engineering tools for a broad range of organisms, from bacteria to humans[3-7]. The type II-A Cas9 from *Streptococcus pyogenes* (SpyCas9)[8] was the first Cas nuclease to be characterized in detail, and due to its consistently high genome-editing activity most widely used. Like other Cas9 variants, SpyCas9 requires two RNA molecules, a CRISPR RNA (crRNA) guide and a trans-activating CRISPR RNA (tracrRNA) anchor. For application purposes, a chimeric single guide RNA (sgRNA) molecule that carries an exchangeable sequence (called spacer) is used. The spacer is complementary to the desired DNA target site (called protospacer). Because target specificity relies on the spacer part of the sgRNA, it can easily be reprogrammed to introduce a double-stranded DNA break (DSDB) in any DNA sequence of interest. The only requirement is that the target sequence should be flanked downstream by a protospacer adjacent motif (called PAM), in case of SpyCas9 a 5'-NGG-3' (or less preferably 5'-NAG-3')[8-10].

In eukaryotes, the Cas9-mediated DSDBs are predominantly repaired by non-homologous end joining (NHEJ), which generates random insertions and/or deletions (indels) that often result in frameshifts and thus gene inactivation[8,11,12]. Alternatively, precise genetic modifications can be achieved by repairing the DSDB using homology-directed repair (HDR)[13]. However, HDR requires a repair template, is restricted to dividing cells (which is an issue for most therapeutically relevant cell types), and presents a strong DSDB-induced NHEJ background[14,15]. To introduce specific genomic alterations with minimum levels of undesired indels and without the need for DSDBs or donor DNA templates, the base-editing approach was recently developed. DNA base-editors are composed of a partially inactive Cas9 (nickase; nCas9) fused either to a cytidine deaminase (cytidine base-editors; CBEs), or to an evolved adenosine deaminase (adenosine base-editors; ABEs). The initial deamination of C-to-U (CBE) and of A-to-I (ABE) results in precise conversion of C•G to T•A and of A•T to G•C, respectively, through mismatch repair followed by DNA replication or repair. However, if base-excision takes place (e.g. by uracil DNA glycosylase) after the initial deamination, it may lead to reversion of deamination via base-excision repair, or to indels via AP site cleavage and mismatch repair[16-18]. One of the most efficient base-editors to date (known as BE4) is a nSpyCas9-CBE that additionally bears two copies of the uracil

DNA glycosylase inhibitor (*ugi*) gene in order to efficiently block excision of the uracil base, and thus increase the base-conversion rate[19]. Despite their high efficiencies, the majority of the current genome- and base-editing tools has been developed based on the SpyCas9 orthologue, providing a range of targetable sites limited to the 5'-NGG-3' PAM[8]. This is especially problematic in base-editing applications, where the PAM needs to be in a certain position adjacent to the targeted base[20].

Thousands of different Cas9 orthologues have been discovered in the past 10 years. Although only few have been characterized and applied for genome- or base-editing purposes to date[19,26-42], they present a tremendous untapped potential for alternative genetic engineering capabilities beyond those provided by SpyCas9[21-25]. First, these orthologues may provide distinct PAM specificities, expanding the targeting scope and the genome coverage. Second, most of them may show higher fidelity compared to SpyCas9, e.g. by recognizing longer PAM and protospacer sequences. Third, certain synthetic Cas9-Cas9 pairs have orthogonal guides, facilitating multiplexed applications[43-45]. Fourth, temporal, spatial, or conditional control of most of these variants can be achieved using natural anti-CRISPR inhibitors[46]. Finally, the type II-C Cas9 orthologues are typically smaller than SpyCas9, allowing for delivery through virus- or mRNA-based vectors, and some of them are resistant to harsh environmental conditions, such as higher stability in human plasma[42]. Hence, there is a plethora of Cas9 orthologues with potentially favorable properties and unexplored genetic engineering capabilities that could serve as alternatives for SpyCas9 guide and protein optimization procedures.

Here, we report the development of a compact, type II-C Cas9 orthologue from *Geobacillus thermodenitrificans* T12 (ThermoCas9) as a genetic engineering platform for human cells. ThermoCas9 is approximately 25% smaller (1,082 aa) than SpyCas9 (1,368 aa), and it recognizes a longer (23 nt) protospacer sequence as well as a distinct PAM (5'-N_4CRAA-3')[47]. We demonstrate that ThermoCas9 is able to induce NHEJ-based editing of human genomes at three different target genes (*dnmt1*, *emx1*, *vegfa*), with efficiencies similar to SpyCas9. In addition, we create a ThermoCas9-based cytidine base-editor (ThermoBE4) for human cells that exhibits a broader window of activity than the SpyCas9 base-editor (BE4). Overall, we propose ThermoCas9 as an alternative to SpyCas9 to expand the targeting scope and the repertoire of genome- and base-editing tools in eukaryotes.

Materials and Methods

Bacterial strains and growth conditions

E. coli DH5α was used in this study for cloning purposes. *E. coli* DH5α cells were cultured in Luria-Bertani broth (LB) or on LB agar plates, supplemented with ampicillin (100 μg/ml), at 37°C for 17 h. Liquid cultures were grown in a shaker incubator at 220 rpm.

HEK293T culturing and growth conditions

HEK293T (ATCC® CRL-3216™) cells were cultured at 37°C and 5% CO_2 in T75 bio-coated flasks containing 10 ml Dulbecco's Modified Eagle Medium (DMEM) supplemented with D-glucose (1 g/l), L-glutamine, HEPES (25 mM), pyruvate, 10% (v/v) fetal bovine serum (FBS), 2% penicillin (10,000 U/ml) and streptomycin (10,000 μg/ml). Cells were passaged at 70-90% confluency every 2-4 days, by washing with 5 ml Ca^{2+}/Mg^{2+}-free phosphate-buffered saline (PBS) and adding 1 ml 0.05% Trypsin-EDTA followed by incubation at 37°C for 5 min. The detached cells were suspended in DMEM growth medium for counting and further processing.

Construction of plasmids

Plasmids, primers or oligonucleotides, genes, and protein sequences used in this study are presented in Supplementary Tables 1, 3, 4, and 5, respectively. For the generation of genome editing plasmids (phTCas9_spT1-T9; phSCas9_spS1-S9), non-targeting plasmids (phTCas9_nt; phSCas9_nt) were first constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB). These plasmids contain a non-targeting spacer that can be excised and replaced using the type IIS restriction enzyme SapI (New England Biolabs). The fragments for assembling the plasmids were obtained through PCR with Q5® High-Fidelity 2X Master Mix (New England Biolabs). The PCR products were run on a 0.8% agarose gel and were subsequently purified using Zymogen gel DNA recovery kit (Zymo Research). The assembled plasmids were transformed to chemically competent *E. coli* DH5α cells[48], plated on LB agar containing ampicillin (100 μg/ml) and incubated overnight at 37°C. The next day, single colonies were inoculated in LB medium, grown overnight at 37°C (220 rpm) and the plasmids were isolated using the GeneJet plasmid Miniprep kit

(ThermoFisher Scientific) or the GenElute™ Endotoxin-free Plasmid Midiprep Kit (Sigma-Aldrich). All the constructs were verified using Sanger sequencing (Macrogen). A complementary pair of oligonucleotides was designed for each targeting spacer (spT1-T9; spS1-S9) to replace the non-targeting spacer. For annealing of the complementary oligos, 4 µl oligonucleotide pairs (100 µM each) were mixed in Milli-Q water to a final volume of 100 µl, heated at 95°C for 5 min, and slowly cooled to room temperature. The annealed oligonucleotides were ligated into the SapI-digested non-targeting plasmids following the T4 ligation protocol (NEB), and the assembled plasmids were isolated and verified using Sanger sequencing (Macrogen). For the generation of base-editing plasmids (phThermoBE4_spT1-T9; phBE4_spS1-S9), the human codon-optimized *thermocas9-linker-sv40nls* (Twist Bioscience) and *spycas9-linker-sv40nls*[19] genes in the genome editing plasmids were replaced by the human codon-optimized fusion genes *apobec1-linker-hnthermocas9(D8A)-linker-ugi-linker-ugi-linker-sv40nls*[19] (Twist Bioscience) and *apobec1-linker-hnspycas9(D10A)-linker-ugi-linker-ugi-linker-sv40nls*[19], respectively, using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB), and the assembled plasmids were isolated and confirmed using Sanger sequencing (Macrogen). Description of the assembled fragments used for the construction of all plasmids is detailed in Supplementary Table 1.

Genome- and base-editing

HEK293T cells were seeded on physically surface-treated 24-well plates (Greiner Bio-one) at a seeding density of 0.5 - 1.0 x 10^5 cells/well. After 24 h of incubation, 0.5-1.5 µg of genome- or base-editing/control plasmid were transfected into the HEK293T cells using Lipofectamine™ 3000 Transfection reagent (ThermoFisher Scientific), according to the user guide instructions. The transfected cells were cultured either 48 h for FACS or 72 h for screening of the population for genome- or base-editing events.

FACS and generation of single cell-derived knockout clones

After 48 h of incubation, the transfected cells were harvested (with Ca^{2+}/Mg^{2+}-free PBS, 0.05% Trypsin-EDTA, and transfection medium), centrifuged at 250 xg for 5 min, resuspended in 250 µl cell-sorting buffer (1% FBS in Ca/Mg2+-free PBS), and filtered through Bel-Art FLOWMI 40 micron cell strainers (Sigma). 96-well plates for cell-sorting were coated with FBS for 1-2 h. Fluorescent single cells were sorted into the coated 96-

wells plates containing 50% pre-conditioned culture medium using the Sony SH800 cell-sorter device (Sony) (488 nm laser, FITC detection channel for GFP fluorescence), centrifuged at 100 xg for 2 min, and cultured for approximately 3 weeks (37°C; 5% CO2) by adding fresh medium at least once per week to prevent evaporation. When confluency was reached, the clonally propagated cells were steadily passaged to T25 flasks, and screened for indels.

Screening for genome- and base-editing

HEK293T genomic DNA was isolated either from the population or from the clonal propagated cells, by removing the culture medium, washing with PBS (200 µl per sample), mixing with Lucigen Quick Extract™ DNA extraction solution (Lucigen) (100 µl per sample), transferring the cell lysate to PCR tubes and incubating the lysate in three consecutive steps (65°C for 15 min, 68°C for 15 min, and 98°C for 10 min). Genomic target regions (*dnmt1*, *emx1*, *vegfa*) were PCR amplified with Q5® High-Fidelity 2X Master Mix (New England Biolabs). The PCR products were verified on a 1% DNA agarose gel and they were subsequently purified with the DNA Clean and Concentrator kit (Zymo Research). To detect indel formation or cytidine deamination, the purified PCR products were subjected to Sanger sequencing (Macrogen). The sequencing results of the genome- and base-editing assays were analyzed using the Inference of CRISPR Edits tool (ICE, Synthego) and the free web tool EditR[49], respectively.

Results

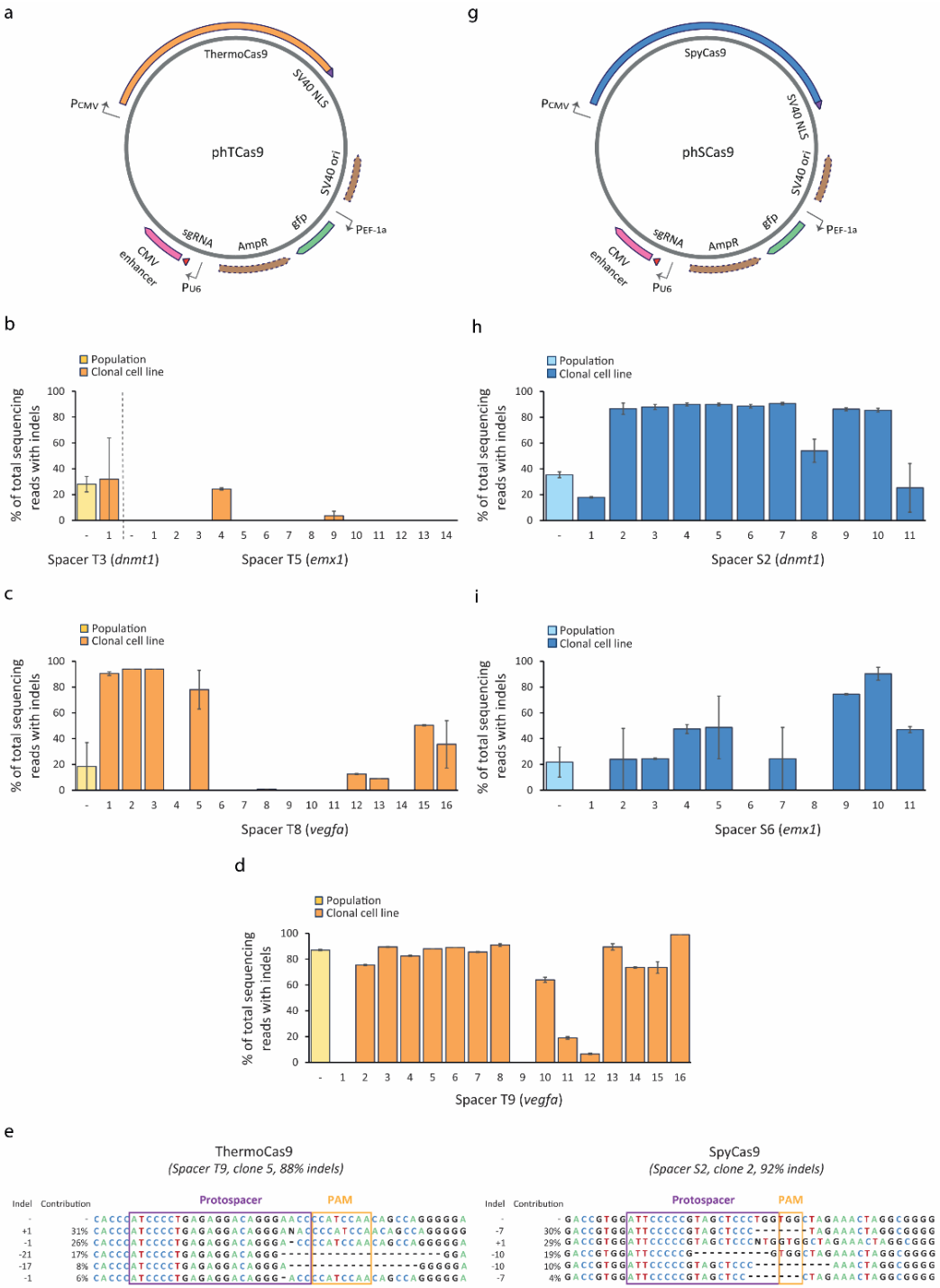
ThermoCas9 induces efficient editing of human genomes

We started by investigating the *in vivo* DNA cleavage activity of ThermoCas9 in human cells. As such, we constructed targeting ThermoCas9 plasmids (called phTCas9) encoding the human codon optimized *thermocas9-sv40nls* gene and its sgRNA module under the control of the constitutive Cytomegalovirus (P_{CMV}) and U6 RNA polymerase III (P_{U6}) promoters, respectively (Fig. 1a; Sup. Table 1). From the same plasmids, we co-expressed the *egfp* reporter gene under the constitutive elongation factor-1 alpha promoter ($P_{EF\text{-}1\alpha}$) to allow for sorting of the successfully transfected cells. We designed nine spacers that target protospacers in the chromosomal genes *dnmt1*, *emx1*, and *vegfa* (three spacers

per gene) (Sup. Table 2). All protospacers were flanked by a PAM that allows optimal cleavage activity of ThermoCas9 *in vitro* (5'-N_4CCAA-3')[47]. After transfection of the targeting plasmids in HEK293T cells (up to 70% efficiency), the transfected cells were sorted into single cells and populations of up to 1000 cells, via flow cytometry assisted cell sorting (FACS) based on EGFP fluorescence signal (Sup. Fig. 1, 2). The cells were cultured and screened for indels at the targeted sites through gene-specific PCR, and subsequent Sanger sequencing. Successfully transfected populations had average indel frequencies of 87% for the strongest spacer T9 as well as 28% and 18% for the medium strength spacers T3 and T8, respectively (Fig. 1b-d; Sup. Info). Moreover, indels were found in 25 out of 47 HEK293T clonal cell lines, presenting average indel frequencies of up to 99% for spacer T9, 32% for spacer T3, 94% for spacer T8, and 24% for spacer T5 (Fig. 1b-d; Sup. Info). The location of the indels suggests that ThermoCas9 created blunt-ended DSDB between the third and the fourth nucleotides immediately upstream of the PAM (Sup. Info), in agreement with our previous *in vitro* study[47]. Targeting of the human genes predominantly generated nucleotide deletions instead of insertions (Fig. 1e, f; Sup. Info). Overall, these observations suggest that ThermoCas9 is able to introduce DSDBs in human genomes at 37°C, and thus to induce indel formation to disrupt targeted genes for the generation of knock-out cell lines.

Comparison of ThermoCas9 and SpyCas9 *in vivo* cleavage activity in human cells

Next, we evaluated the *in vivo* DNA cleavage activity of SpyCas9 using the same setup for a direct comparison with ThermoCas9. For this purpose, we substituted the *thermocas9-sv40nls* gene and its sgRNA module in the phTCas9 plasmids with the human codon-optimized *spycas9-sv40nls* gene and its sgRNA, generating the phSCas9 plasmids (Fig. 1g; Sup. Table 1). Similar to the ThermoCas9 approach, we selected three protospacers, each flanked by an optimal PAM (5'-NGG-3')[8] in the chromosomal genes *dnmt1*, *emx1*, and *vegfa*, and individually incorporated the corresponding spacers into the phSCas9 plasmids (Sup. Table 2). Following the same procedure, in the successfully transfected populations we observed average indel frequencies of 35% for the stronger spacer S2, and 22% for the weaker spacer S6 (Fig. 1h, i; Sup. Info; Sup. Fig. 1, 2). In addition, SpyCas9-mediated cleavage of the human chromosomes led to the generation of indels in 19 out of 22 HEK293T clonal cell lines, presenting average indel frequencies of up to 90% for both


a
ThermoCas9
SV40 NLS
PCMV
phTCas9
SV40 ori
gfp
PEF-1a
sgRNA
AmpR
CMV enhancer
PU6
g
SpyCas9
SV40 NLS
PCMV
phSCas9
SV40 ori
gfp
PEF-1a
sgRNA
AmpR
CMV enhancer
PU6
b
Population
Clonal cell line
% of total sequencing reads with indels
Spacer T3 (dnmt1)
Spacer T5 (emx1)
h
Population
Clonal cell line
% of total sequencing reads with indels
Spacer S2 (dnmt1)
c
Population
Clonal cell line
% of total sequencing reads with indels
Spacer T8 (vegfa)
i
Population
Clonal cell line
% of total sequencing reads with indels
Spacer S6 (emx1)
d
Population
Clonal cell line
% of total sequencing reads with indels
Spacer T9 (vegfa)
e
ThermoCas9
(Spacer T9, clone 5, 88% indels)
Indel
Contribution
Protospacer
PAM
SpyCas9
(Spacer S2, clone 2, 92% indels)
Indel
Contribution
Protospacer
PAM

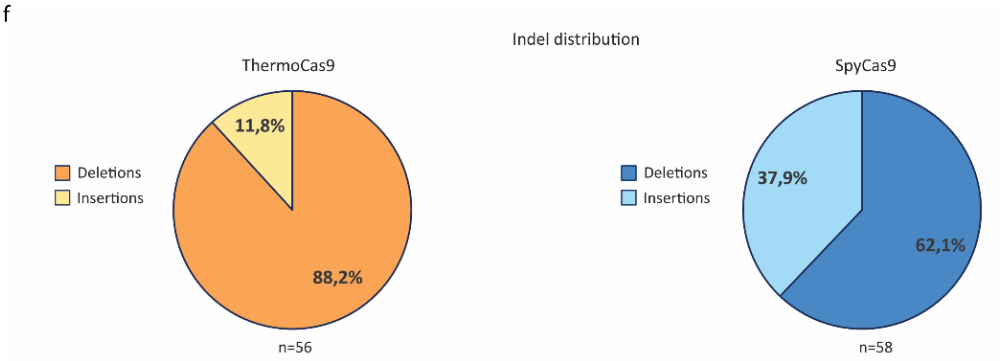


Fig. 1 Genome editing in human cells by ThermoCas9 or SpyCas9. **a.** Schematic illustration of the phTCas9 constructs for ThermoCas9-mediated cleavage of three genomic genes (*dnmt1*, *emx1*, and *vegfa*) in HEK293T cells. **b-d.** Indel frequencies in the targeted genomic genes of HEK293T cell populations and clonal cell lines, after successful transfection with the phTCas9 plasmids. The indel frequencies were obtained using the Inference of CRISPR Edits tool (ICE, Synthego) for multiple Sanger sequencing reactions of each target site, and error bars indicate the standard error of the mean. **e.** Types of insertions or deletions (referred to as 'indel') and their relative proportions (referred to as 'contribution') in the total number of cells screened via Sanger sequencing. Purple and orange boxes indicate the protospacer sequence and the PAM, respectively, while black, dotted lines represent the Cas9 cleavage site. **f.** Indel distribution in the total number of cells edited by ThermoCas9 (left) and SpyCas9 (right). All Sanger sequencing reactions (n=56 for ThermoCas9; n=58 for SpyCas9) were taken into consideration. **g.** Schematic illustration of the phSCas9 constructs for SpyCas9-mediated cleavage of three genomic genes (*dnmt1*, *emx1*, and *vegfa*) in HEK293T cells. **h, i.** Indel frequencies in the targeted genomic genes of HEK293T cell populations and clonal cell lines, after successful transfection with the phSCas9 plasmids. The indel frequencies were obtained using the ICE tool (Synthego) for multiple Sanger sequencing reactions of each target site, and error bars indicate the standard error of the mean.

spacers S2 and S6 (Fig. 1h, i; Sup. Info). Nucleotide deletions instead of insertions were mainly observed, as indicated for ThermoCas9 (Fig. 1e, f). Altogether, cleavage activities in HEK293T cells by ThermoCas9 are comparable to SpyCas9.

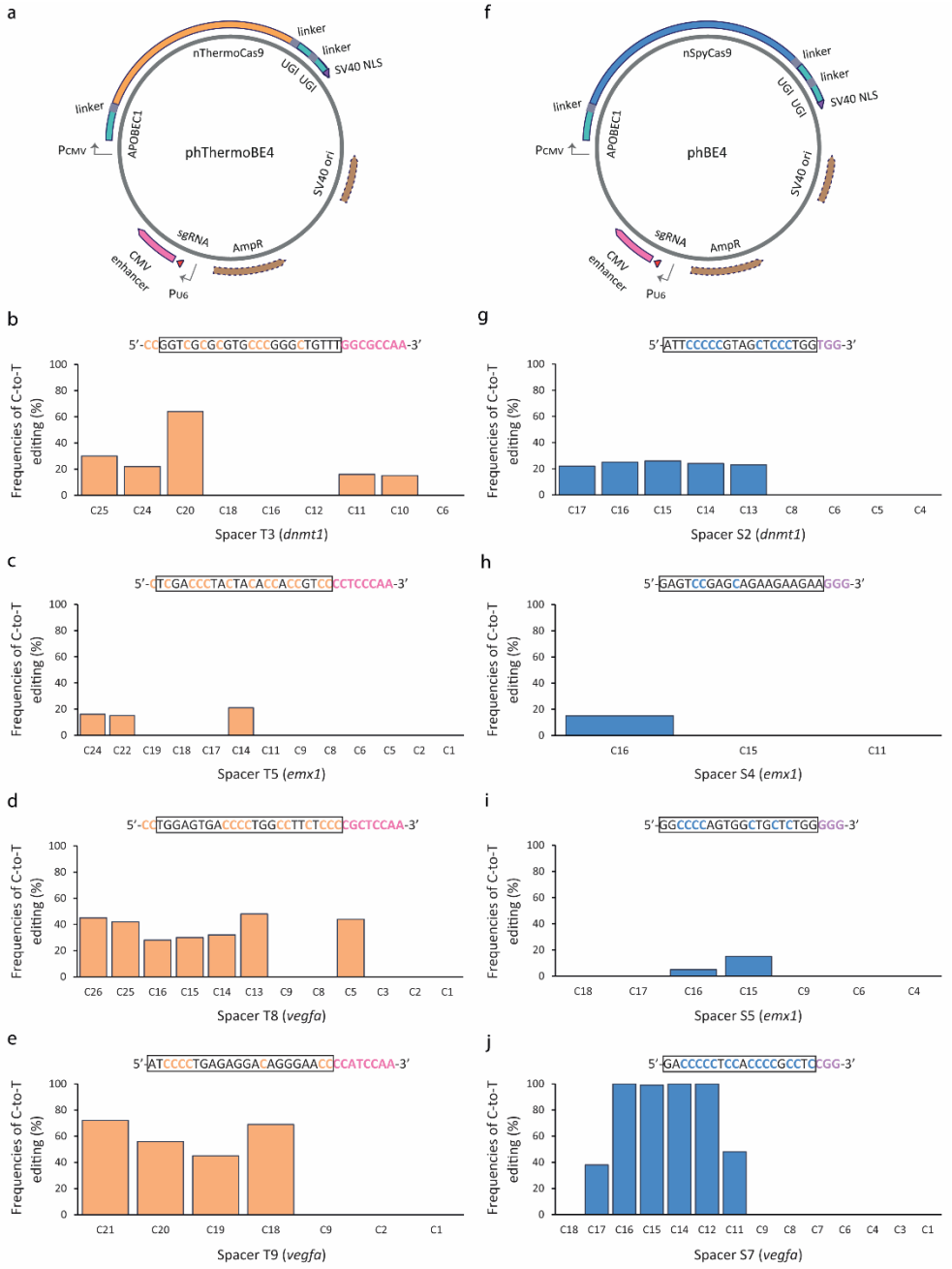
ThermoBE4 is an efficient base-editing tool for human cells

We then set out to develop a ThermoCas9-based base-editor for human cells. In the previously described phTCas9 plasmids (Fig. 1a; Sup. Table 1), we replaced the *thermocas9-sv40nls* gene with its D8A nickase (*nthermocas9*) gene. In addition, we fused the rat Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 1 (*rapobec1*)[50] gene, and two copies of the bacteriophage PBS2 uracil DNA glycosylase inhibitor (*ugi*)[51] gene

followed by the SV40 NLS signal to the N- and C-terminus of the *nthermocas9* gene, respectively. We transfected the constructed phThermoBE4 plasmids (Fig. 2a; Sup. Table 1) to HEK293T cells, and screened for cytidine deamination at the targeted sites, via gene-specific PCR followed by Sanger sequencing. We detected C•G to T•A conversions at four out of the nine protospacers we targeted with the ThermoBE4 system (Fig. 2b-e; Sup. Fig. 3). These protospacers (T3, T5, T8, and T9) are the same that were successfully used to induce indel formation (Fig. 1b-d). In all cases, the ThermoBE4 base-editor was able to edit Cs at multiple (up to seven) positions simultaneously, reaching editing efficiencies of up to 72% for Cs at the PAM-distal end of the protospacer (Fig. 2b-e; Sup. Fig. 3). Editing was observed not only within but also immediately upstream of the protospacer regions, presenting an editing window of up to 22 bp (from -5 to -26 positions relative to the PAM) (Fig. 2b-e; Sup. Fig. 4). Hence, our results demonstrate that the ThermoCas9 base-editor mediates efficient base-editing at multiple positions and in a relatively large activity window, providing an alternative approach for targeted gene mutagenesis in eukaryotes.

Comparison of ThermoBE4 and BE4 base-editing characteristics in human cells

Finally, we compared the ThermoCas9 base-editor with the previously reported SpyCas9 BE4 system[19]. For this reason, we substituted the *spycas9* gene from the phSCas9 plasmids (Fig. 1a; Sup. Table 1) with the synthetic *rapobec1-nspycas9-ugi-ugi-sv40nls* coding sequence, generating the phBE4 plasmids (Fig. 2f; Sup. Table 1). After transfection of the base-editing plasmids in HEK293T cells, our screening results showed that five out of nine protospacers were successfully base-edited at multiple (up to six) positions simultaneously (Fig. 2g-k; Sup. Fig. 3). Similar to ThermoBE4 and in agreement with a previous study[19], nSpyCas9-based BE4 preferentially edited C bases at the PAM-distal end of the protospacer (Fig. 2g-k). However, BE4 exhibited higher (absolute) editing efficiencies and narrower window of activity (7 bp; from -11 to -17 positions relative to the PAM) than ThermoBE4 (Fig. 2g-k; Sup. Fig. 3, 4). Remarkably, no editing was observed outside of the protospacer regions. On the whole, our results suggest that BE4 may be a better choice for site-specific mutagenesis, whereas ThermoBE4 could be considered for applications that require a wide activity window, such as gene mutagenesis.


a
linker
nThermoCas9
linker
UGI UGI
SV40 NLS
linker
APOBEC1
PCMV
phThermoBE4
SV40 ori
sgRNA
AmpR
CMV enhancer
PU6
f
linker
nSpyCas9
linker
UGI UGI
SV40 NLS
linker
APOBEC1
PCMV
phBE4
SV40 ori
sgRNA
AmpR
CMV enhancer
PU6
b
5′-CCGGTCGCGCGTGCCCGGGCTGTTTGGCGCCAA-3′
Frequencies of C-to-T editing (%)
Spacer T3 (dnmt1)
g
5′-ATTCCCCCGTAGCTCCCTGGTGG-3′
Frequencies of C-to-T editing (%)
Spacer S2 (dnmt1)
c
5′-CTCGACCCTACTACACCACCGTCCCCTCCCAA-3′
Frequencies of C-to-T editing (%)
Spacer T5 (emx1)
h
5′-GAGTCCGAGCAGAAGAAGAAGGG-3′
Frequencies of C-to-T editing (%)
Spacer S4 (emx1)
d
5′-CCTGGAGTGACCCCTGGCCTTCTCCCCGCTCCAA-3′
Frequencies of C-to-T editing (%)
Spacer T8 (vegfa)
i
5′-GGCCCCAGTGGCTGCTCTGGGGG-3′
Frequencies of C-to-T editing (%)
Spacer S5 (emx1)
e
5′-ATCCCCTGAGAGGACAGGGAACCCCATCCAA-3′
Frequencies of C-to-T editing (%)
Spacer T9 (vegfa)
j
5′-GACCCCCTCCACCCCGCCTCGGG-3′
Frequencies of C-to-T editing (%)
Spacer S7 (vegfa)

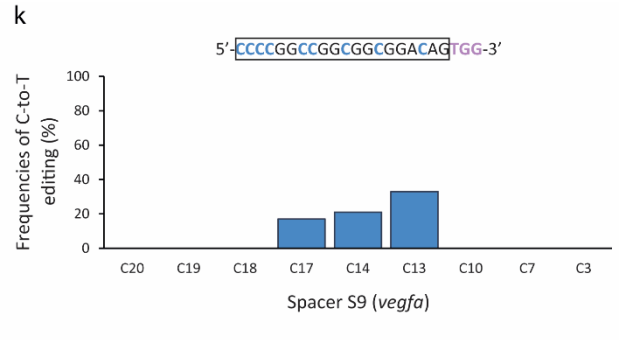


Fig. 2 Base-editing in human cells by ThermoBE4 or BE4. **a.** Schematic illustration of the phThermoBE4 constructs for nThermoCas9-mediated base-editing of three genomic genes (*dnmt1*, *emx1*, and *vegfa*) in HEK293T cells. **b-e.** C-to-T editing frequencies induced by the ThermoBE4 system on each cytosine at the targeted genomic genes of HEK293T cells, after transfection with the phThermoBE4 plasmids. The protospacer region is indicated with a grey box, while the target Cs and PAM are shown in yellow and pink, respectively. Editing frequencies were obtained using the EditR tool. **f.** Schematic illustration of the phBE4 constructs for nSpyCas9-mediated base-editing of three genomic genes (*dnmt1*, *emx1*, and *vegfa*) in HEK293T cells. **g-k.** C-to-T editing frequencies induced by the BE4 system on each cytosine at the targeted genomic genes of HEK293T cells, after transfection with the phBE4 plasmids. The protospacer region is indicated with a grey box, while the target Cs and PAM are shown in blue and purple, respectively. Editing frequencies were obtained using the EditR tool.

Discussion

In our previous studies, we demonstrated that a type II-C Cas9 endonuclease from *G. thermodenitrificans* T12 is active in a wide temperature range (20°C - 70°C) *in vitro*, and that it can be applied for genetic engineering purposes in both mesophilic and thermophilic bacteria[47]. ThermoCas9 can be regulated using an anti-CRISPR off-switch control (manuscript under revision), is considerably smaller than the type II-A SpyCas9, and it exhibits naturally higher fidelity (at 37°C) and stability[47], rendering it ideal for viral delivery and clinical applications. In addition, ThermoCas9 can reach different targets than SpyCas9 and other CRISPR nucleases, due to the recognition of alternative PAMs[47]. However, its activity has never been tested before in eukaryotes.

Here, we develop a ThermoCas9-based approach for NHEJ-mediated gene editing in HEK293T cells. Similar to SpyCas9, we achieve high indel frequencies both at population level and at clonal cell line level, despite the lower DNA cleavage rate of ThermoCas9 at 37°C[47]. We speculate that, like its close type II-C Cas9 homolog from *Geobacillus stearothermophilus* (GeoCas9), ThermoCas9 may have a longer half-life in the cells, thus

counterbalancing its lower cleavage activity at 37°C[42]. Cell populations generally presented a lower amount of detectable indels than clonally propagated cells, due to their higher heterogeneity and loss thereof during PCR and Sanger sequencing. Moreover, targeting efficiencies were (proto)spacer-specific, suggesting that epigenetic modifications, chromatin structures, DNA supercoiling, or guide/spacer secondary structures may affect the cleavage ability of ThermoCas9, as previously observed for SpyCas9 and other variants[52]. Indeed, negative supercoiling of the target DNA has been previously reported to improve both target accessibility to ThermoCas9[47], and R-loop formation by a type II-C Cas9 orthologue from *Acidothermus cellulolyticus* (AceCas9)[53]. Furthermore, in agreement to our previous *in vitro* findings and studies on other Cas9 variants[8,28,47,54,55], analysis of indels in HEK293T cells shows that the cleavage site of ThermoCas9 (and of SpyCas9) is between the third and the fourth nucleotide of the protospacer upstream of the PAM. The vast majority of generated indels were deletions, as also shown for SpyCas9 and the type II-A Cas9 nuclease from *Staphylococcus aureus* (SauCas9)[31]. To date, less than ten Cas9 orthologues (mainly of types II-A and II-C) have exhibited the ability to edit eukaryotic genomes[28,32-34,36-38,41,42]. Hence, ThermoCas9 may be a valuable addition to this Cas9 catalog.

To expand the targeting scope also for base-editing applications, we create an efficient ThermoCas9 base-editor (ThermoBE4) which facilitates C•G to T•A conversions in genomic loci with alternative PAMs. In accordance with the corresponding SauCas9 base-editor (SaBE4)[19], we enable editing in a larger window of activity compared to the SpyCas9 editor (BE4), with a preference for Cs located at the PAM-distal end of the protospacer. Intriguingly, ThermoBE4 provides a three-fold larger activity window that extends upstream of the protospacer region, and may facilitate gene mutagenesis applications in other eukaryotes, for example in plants. SauCas9 has a similar size as ThermoCas9, and they are both more compact than SpyCas9[31,47]. Therefore, we hypothesize that the smaller nuclease size leaves more nucleotide positions exposed for base-editing. However, structural analysis is required to fully understand the biochemical mechanism of ThermoBE4.

Conclusions

In summary, we demonstrate that ThermoCas9 can be used as an alternative platform for efficient genome- and base-editing of human genomes. The compact size, high stability

and accuracy, distinct PAM recognition, and anti-CRISPR off-switch control of ThermoCas9 render it a viable tool to complement the current CRISPR toolbox for a wide range of biotechnological and medical applications.

Acknowledgements

We would like to thank Simon Delphin Kool for his contribution to setting up the methods, Christian Südfeld for his technical assistance with the FACS as well as the Alexander S. Onassis Foundation, the Netherlands Organization for Scientific Research (NWO), and the European Research Council for the financial support.

Authorship contribution

Despoina Trasanidou: Conceptualization (supporting); Methodology (equal); Validation (lead); Formal analysis (equal); Investigation (equal); Data Curation (lead); Writing - Original Draft (lead); Visualization (lead); Funding acquisition (supporting). **Patrick Barendse:** Methodology (equal); Validation (supporting); Formal analysis (equal); Investigation (equal); Writing - Review & Editing (supporting). **Evgenios Bouzetos:** Methodology (equal); Validation (supporting); Formal analysis (equal); Investigation (equal). **Laura de Haan:** Writing - Review & Editing (supporting); Supervision (supporting). **Hans Bouwmeester:** Resources (supporting). **Raymond H.J. Staals:** Writing - Review & Editing (supporting); Supervision (supporting); Funding acquisition (supporting). **Ioannis Mougiakos:** Conceptualization (lead); Writing - Review & Editing (supporting); Supervision (lead); Project administration (lead); Funding acquisition (supporting). **John van der Oost:** Conceptualization (supporting); Resources (lead); Writing - Review & Editing (lead); Supervision (supporting); Project administration (supporting); Funding acquisition (lead).

Conflict of interest

J.v.d.O. is cofounder of NTrans Technologies, and member of the Scientific Advisory Board of NTrans Technologies and Hudson River Biotechnology. J.v.d.O. and R.H.J.S. are members of the Scientific Advisory Board of Scope Biosciences. I.M. is employed by the

commercial company SNIPR Biome (Copenhagen, Denmark). D.T., J.v.d.O., R.H.J.S. and I.M. are inventors on CRISPR-Cas related patents/patent applications.

Funding

D.T. is financially supported by the Alexander S. Onassis Foundation grant (F ZM 083-2/2018-2019). J.v.d.O. thanks the Dutch Research Council (NWO Spinoza grant SPI 93-537, and NWO Gravitation grant 024.003.019), and the European Research Council (ERC-AdG-834279) for financial support. R.H.J.S. is supported by a VIDI grant (VI.Vidi.203.074) from NWO.

References

1. Barrangou R, Fremaux C, Deveau H, et al.. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007;315:1709-1712. DOI: 10.1126/science.1138140.
2. Brouns SJ, Jore MM, Lundgren M, et al.. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008;321:960-964. DOI: 10.1126/science.1159689.
3. Mohanraju P, Saha C, van Baarlen P, et al.. Alternative functions of CRISPR-Cas systems in the evolutionary arms race. *Nature Reviews Microbiology* 2022;20:351-364. DOI: 10.1038/s41579-021-00663-z.
4. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;346:1258096. DOI: 10.1126/science.1258096.
5. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;157:1262-1278. DOI: 10.1016/j.cell.2014.05.010.
6. Wu WY, Lebbink JH, Kanaar R, et al.. Genome editing by natural and engineered CRISPR-associated nucleases. *Nature chemical biology* 2018;14:642-651. DOI: 10.1038/s41589-018-0080-x.
7. Nidhi S, Anand U, Oleksak P, et al.. Novel CRISPR-Cas systems: an updated review of the current achievements, applications, and future research perspectives. *International journal of molecular sciences* 2021;22:3327. DOI: 10.3390/ijms22073327.
8. Jinek M, Chylinski K, Fonfara I, et al.. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816-821. DOI: 10.1126/science.1225829. Epub 2012 Jun 28.
9. Deveau H, Barrangou R, Garneau JE, et al.. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of bacteriology* 2008;190:1390-1400. DOI: 10.1128/JB.01412-07.
10. Mojica FJ, Díez-Villaseñor C, García-Martínez J, et al.. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 2009;155:733-740. DOI: 10.1099/mic.0.023960-0.
11. Cong L, Ran FA, Cox D, Lin S, et al.. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819-823. DOI: 10.1126/science.1231143.
12. Mali P, Yang L, Esvelt KM, et al.. RNA-guided human genome engineering via Cas9. *Science* 2013;339:823-826. DOI: 10.1126/science.1232033.
13. Rouet P, Smih F, Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* 1994;91:6064-6068. DOI: 10.1073/pnas.91.13.6064.
14. Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell* 2012;47:497-510. DOI: 10.1016/j.molcel.2012.07.029.
15. Cox DBT, Platt RJ, Zhang, F. Therapeutic genome editing: prospects and challenges. *Nature medicine* 2015;21:121-131. DOI: 10.1038/nm.3793.

16. Komor AC, Kim YB, Packer MS, et al.. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016;533:420-424. DOI: 10.1038/nature17946.
17. Nishida K, Arazoe T, Yachie N, et al.. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 2016;353:aaf8729. DOI: 10.1126/science.aaf8729.
18. Gaudelli NM, Komor AC, Rees HA, et al.. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 2017;551:464-471. DOI: 10.1038/s41586-018-0070-x.
19. Komor AC, Zhao KT, Packer MS, et al.. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C: G-to-T: A base editors with higher efficiency and product purity. *Science advances* 2017;3:eaao4774. DOI: 10.1126/sciadv.aao4774.
20. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature reviews genetics* 2018;19:770-788. DOI: 10.1038/s41576-018-0059-1.
21. Makarova KS, Wolf YI, Iranzo J, et al.. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* 2020;18:67-83. DOI: 10.1038/s41579-019-0299-x.
22. Shmakov S, Smargon A, Scott D, et al.. Diversity and evolution of class 2 CRISPR-Cas systems. *Nature reviews microbiology* 2017;15:169-182. DOI: 10.1038/nrmicro.2016.184.
23. Chylinski K, Makarova KS, Charpentier E, et al.. Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research* 2014;42:6091-6105. DOI: 10.1093/nar/gku241.
24. Fonfara I, Le Rhun A, Chylinski K, et al.. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acids research* 2014;42:2577-2590. DOI: 10.1093/nar/gkt1074.
25. Gasiunas G, Young JK, Karvelis T, et al.. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nature communications* 2020;11:1-10. DOI: 10.1038/s41467-020-19344-1.
26. Kim YB, Komor AC, Levy JM, et al.. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature biotechnology* 2017;35:371-376. DOI: 10.1038/nbt.3803.
27. Liu Z, Chen S, Jia Y, et al.. Efficient and high-fidelity base editor with expanded PAM compatibility for cytidine dinucleotide. *Science China Life Sciences* 2021;64:1355-1367. DOI: 10.1007/s11427-020-1775-2.
28. Hou Z, Zhang Y, Propson NE, et al.. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences* 2013;110:15644-15649. DOI: 10.1073/pnas.1313587110.
29. Ibraheim R, Song CQ, Mir A, et al.. All-in-one adeno-associated virus delivery and genome editing by *Neisseria meningitidis* Cas9 *in vivo*. *Genome biology* 2018;19:1-11. DOI: 10.1186/s13059-018-1515-0.
30. Lee CM, Cradick TJ, Bao G. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Molecular Therapy* 2016;24:645-654. DOI: 10.1038/mt.2016.8.
31. Friedland AE, Baral R, Singhal P, et al.. Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome biology* 2015;16:1-10. DOI: 10.1186/s13059-015-0817-8.
32. Ran FACL, Cong L, Yan WX, et al.. *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 2015;520:186-191. DOI: 10.1038/nature14299.
33. Kim E, Koo T, Park SW, et al.. *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nature communications* 2017;8:1-12. DOI: 10.1038/ncomms14500.
34. Acharya S, Mishra A, Paul D, et al.. *Francisella novicida* Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proceedings of the National Academy of Sciences* 2019;116:20959-20968. DOI: 10.1073/pnas.1818461116.
35. Agudelo D, Carter S, Velimirovic M, et al.. Versatile and robust genome editing with *Streptococcus thermophilus* CRISPR1-Cas9. *Genome research* 2020;30:107-117. DOI: 10.1101/gr.255414.119.
36. Müller M, Lee CM, Gasiunas G, et al.. *Streptococcus thermophilus* CRISPR-Cas9 systems enable specific editing of the human genome. *Molecular Therapy* 2016;24:636-644. DOI: 10.1038/mt.2015.218.
37. Chatterjee P, Jakimo N, Jacobson JM. Minimal PAM specificity of a highly similar SpCas9 ortholog. *Science advances* 2018;4:eaau0766. DOI: 10.1126/sciadv.aau0766.
38. Hu Z, Wang S, Zhang C, et al.. A compact Cas9 ortholog from *Staphylococcus Auricularis* (SauriCas9) expands the DNA targeting scope. *PLoS biology* 2020;18:e3000686. DOI: 0.1371/journal.pbio.3000686.
39. Li X, Qian X, Wang B, et al.. Programmable base editing of mutated TERT promoter inhibits brain tumour growth. *Nature cell biology* 2020;22:282-288. DOI: 10.1038/s41556-020-0471-6.
40. Xu K, Ren C, Liu Z, et al.. Efficient genome engineering in eukaryotes using Cas9 from *Streptococcus thermophilus*. *Cellular and molecular life sciences* 2015;72:383-399. DOI: 10.1007/s00018-014-1679-z.

41. Fedorova I, Vasileva A, Selkova P, et al.. PpCas9 from *Pasteurella pneumotropica*-a compact Type II-C Cas9 ortholog active in human cells. *Nucleic acids research* 2020;48:12297-12309. DOI: 10.1093/nar/gkaa998.
42. Harrington LB, Paez-Espino D, Staahl BT, et al.. A thermostable Cas9 with increased lifetime in human plasma. *Nature communications* 2017;8:1-8. DOI: 10.1038/s41467-017-01408-4.
43. Bolukbasi MF, Liu P, Luk K, et al.. Orthogonal Cas9-Cas9 chimeras provide a versatile platform for genome editing. *Nature communications* 2018;9:1-12. DOI: 10.1038/s41467-018-07310-x.
44. Esvelt KM, Mali P, Braff JL, et al.. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods* 2013;10:1116-1121. DOI: 10.1038/nmeth.2681.
45. Takasugi PR, Wang S, Truong KT, et al.. Orthogonal CRISPR-Cas tools for genome editing, inhibition, and CRISPR recording in zebrafish embryos. *Genetics* 2022;220:iyab196. DOI: 10.1093/genetics/iyab196.
46. Trasanidou D, Gerós AS, Mohanraju P, et al.. Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS microbiology letters* 2019;366:fnz098. DOI: 10.1093/femsle/fnz098.
47. Mougiakos I, Mohanraju P, Bosma EF, et al.. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications* 2017;8:1-11. DOI: 10.1038/s41467-017-01591-4.
48. Green R, Rogers EJ. Transformation of chemically competent *E. coli*. *Methods Enzymol*ogy 2013;529:329-336. DOI: 10.1016/B978-0-12-418687-3.00028-8.
49. Kluesner MG, Nedveck DA, Lahr WS, et al.. EditR: a method to quantify base editing from Sanger sequencing. *The CRISPR journal* 2018;1:239-250. DOI: 10.1089/crispr.2018.0014.
50. Anant S, MacGinnitie AJ, Davidson NO. Apobec-1, the Catalytic Subunit of the Mammalian Apolipoprotein B mRNA Editing Enzyme, Is a Novel RNA-binding Protein. *Journal of Biological Chemistry* 1995;270:14762-14767. DOI: 10.1074/jbc.270.24.14762
51. Zhigang W, Smith DG, Mosbaugh DW. Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene* 1991;99:31-37. DOI: 10.1016/0378-1119(91)90030-f.
52. Chen F, Ding X, Feng Y, et al.. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nature communications* 2017;8:1-12. DOI: 10.1038/ncomms14958.
53. Tsui TKM, Hand TH, Duboy EC, et al.. The impact of DNA topology and guide length on target selection by a cytosine-specific Cas9. *ACS synthetic biology* 2017;6:1103-1113. DOI: 10.1021/acssynbio.7b00050.
54. Garneau JE, Dupuis MÈ, Villion M, et al.. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 2010;468:67-71. DOI: 10.1038/nature09523.
55. Gasiunas G, Barrangou R, Horvath P, et al.. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* 2012;109:E2579-E2586. DOI: 10.1073/pnas.1208507109.

Supplementary Table 1. Plasmids used in this study.

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pEGFP-C1	-	-	Positive control for HEK293T transfection and FACS.	Clontech
pHTTG	-	-	PCR template for phTCas9_SapI.	Lab stock
pHTSG	-	-	PCR template for phSCas9_SapI.	Lab stock
pSimpleII-U6-sgRNA-BsmBI-NLS-NmCas9-HA-NLS	-	-	PCR template for phTCas9_SapI and phSCas9_SapI.	Addgene #115694
pBE4	-	-	PCR template for phThermoBE4_T1.	Addgene #100802
pFYF1320 EGFP	-	-	PCR template for phThermoBE4_T1.	Addgene #47511
pdThermoTarget-AID_gfp4	-	-	PCR template for phThermoBE4_T1.	Lab stock
phTCas9_SapI	Gibson assembly	Fragment 1: *egfp* gene, SV40 PA terminator, backbone, PU6, SapI spacer, sgRNA, CMV enhancer, PCMV, and partly *hthermocas9-linker-sv40nls* gene from pHTTG amplified with primers BG18420 and BG19287.	ThermoCas9 and sgRNA containing non-targeting, excisable spacer with SapI restriction sites. Used for cloning of all the phTCas9 plasmids (through restriction enzyme digestion and ligation), and as negative control in genome editing assays.	This study
		Fragment 2: *hthermocas9-linker-sv40nls* gene, terminator, and backbone from pHTTG amplified with primers BG19289 and BG19291.		
		Fragment 3: SV40 ori from pHTTG amplified with primers BG19292 and BG19293.		
		Fragment 4: PEF-1α from pSimpleII-U6-sgRNA-BsmBI-NLS-NmCas9-HA-NLS amplified with primers BG19141 and BG19142.		
phTCas9_T1	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T1) targeting the *dnmt1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18732 and BG18741.		
phTCas9_T2	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T2) targeting the *dnmt1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18733 and BG18742.		
phTCas9_T3	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T3) targeting the *dnmt1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18734 and BG18743.		
phTCas9_T4	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T4) targeting the *emx1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18735 and BG18744.		
phTCas9_T5	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T5) targeting the *emx1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18736 and BG18745.		
phTCas9_T6	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T6) targeting the *emx1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18737 and BG18746.		
phTCas9_T7	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T7) targeting the *vegfa* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18738 and BG18747.		
phTCas9_T8	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T8) targeting the *vegfa* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18739 and BG18748.		
phTCas9_T9	T4 ligation	Backbone: SapI-digested phTCas9_nt.	ThermoCas9 and sgRNA (spacer T9) targeting the *vegfa* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18740 and BG18749.		
phSCas9_SapI	Gibson assembly	Fragment 1: *egfp* gene, SV40 PA terminator, backbone, PU6, SapI spacer, sgRNA, CMV enhancer, PCMV, and partly *hspycas9-linker-sv40nls* gene from pHTSG amplified with primers BG18420 and BG19288.	SpyCas9 and sgRNA containing non-targeting, excisable spacer with SapI restriction sites. Used for cloning of all the	This study

Supplementary Table 1. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 2: *hspycas9-linker-sv40nls* gene, terminator, and backbone from pHTSG amplified with primers BG19290 and BG19291.	phSCas9 plasmids (through restriction enzyme digestion and ligation), and as negative control in genome editing assays.	
		Fragment 3: SV40 ori from pHTSG amplified with primers BG19292 and BG19293.		
		Fragment 4: PEF-1α from pSimpleII-U6-sgRNA-BsmBI-NLS-NmCas9-HA-NLS amplified with primers BG19141 and BG19142.		
phSCas9_S1	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S1) targeting the *dnmt1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18750 and BG18759.		
phSCas9_S2	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S2) targeting the *dnmt1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18751 and BG18760.		
phSCas9_S3	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S3) targeting the *dnmt1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18752 and BG18761.		
phSCas9_S4	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S4) targeting the *emx1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18753 and BG18762.		
phSCas9_S5	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S5) targeting the *emx1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18754 and BG18763.		
phSCas9_S6	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S6) targeting the *emx1* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18755 and BG18764.		
phSCas9_S7	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S7) targeting the *vegfa* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18756 and BG18765.		
phSCas9_S8	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S8) targeting the *vegfa* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18757 and BG18766.		
phSCas9_S9	T4 ligation	Backbone: SapI-digested phSCas9_nt.	SpyCas9 and sgRNA (spacer S9) targeting the *vegfa* gene in the genome of HEK293T cells in genome editing assays.	This study
		Insert: Annealed oligonucleotides BG18758 and BG18767.		
phThermoBE4_T1	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls gene* from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T1) targeting the *dnmt1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T1 from pFYF1320 EGFP amplified with primers BG17307 and BG17310.		
		Fragment 4: Spacer T1, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17277 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T2	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T2) targeting the *dnmt1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T2 from pFYF1320 EGFP amplified with primers BG17307 and BG17312.		
		Fragment 4: Spacer T2, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17279 and BG17271.		

Supplementary Table 1. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T3	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T3) targeting the *dnmt1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T3 from pFYF1320 EGFP amplified with primers BG17307 and BG17315.		
		Fragment 4: Spacer T3, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17281 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T4	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T4) targeting the *emx1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T4 from pFYF1320 EGFP amplified with primers BG17307 and BG17316.		
		Fragment 4: Spacer T4, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17283 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T5	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T5) targeting the *emx1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T5 from pFYF1320 EGFP amplified with primers BG17307 and BG17319.		
		Fragment 4: Spacer T5, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17285 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T6	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T6) targeting the *emx1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T6 from pFYF1320 EGFP amplified with primers BG17307 and BG17321.		
		Fragment 4: Spacer T6, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17287 and BG17271.		

Supplementary Table 1. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T7	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T7) targeting the *vegfa* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T7 from pFYF1320 EGFP amplified with primers BG17307 and BG17322.		
		Fragment 4: Spacer T7, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17323 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T8	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T8) targeting the *vegfa* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T8 from pFYF1320 EGFP amplified with primers BG17307 and BG17327.		
		Fragment 4: Spacer T8, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17291 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phThermoBE4_T9	Gibson assembly	Fragment 1: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from hnThermoCas9 (D8A) G-block amplified with primers BG17303 and BG17304.	APOBEC1-linker-nThermoCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer T9) targeting the *vegfa* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17305 and BG17306.		
		Fragment 3: PU6, and spacer T9 from pFYF1320 EGFP amplified with primers BG17307 and BG17329.		
		Fragment 4: Spacer T9, and sgRNA from pdThermoTarget-AID_gfp4 with primers BG17293 and BG17271.		
		Fragment 5A: CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17308.		
		Fragment 5B: T-stretch, CMV enhancer, PCMV, and partly *apobec1-linker-hnthermocas9-linker-ugi-linker-ugi-linker-sv40nls* gene from Fragment 5A amplified with primers BG17274 and BG17309.		
phBE4_S1	Gibson assembly	Fragment 1: PU6 and spacer S1 from pFYF1320 EGFP amplified with primers BG17911 and BG17797.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S1) targeting the *dnmt1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S1, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17773 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		

Supplementary Table 1. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S2	Gibson assembly	Fragment 1: PU6 and spacer S2 from pFYF1320 EGFP amplified with primers BG17911 and BG17798.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S2) targeting the *dnmt1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S2, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17777 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S3	Gibson assembly	Fragment 1: PU6 and spacer S3 from pFYF1320 EGFP amplified with primers BG17911 and BG17799.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S3) targeting the *dnmt1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S3, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17779 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S4	Gibson assembly	Fragment 1: PU6 and spacer S4 from pFYF1320 EGFP amplified with primers BG17911 and BG17800.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S4) targeting the *emx1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S4, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17781 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S5	Gibson assembly	Fragment 1: PU6 and spacer S5 from pFYF1320 EGFP amplified with primers BG17911 and BG17801.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S5) targeting the *emx1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S5, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17783 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S6	Gibson assembly	Fragment 1: PU6 and spacer S6 from pFYF1320 EGFP amplified with primers BG17911 and BG17802.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S6) targeting the *emx1* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S6, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17785 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		

Supplementary Table 1. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
phBE4_S7	Gibson assembly	Fragment 1: PU6 and spacer S4 from pFYF1320 EGFP amplified with primers BG17911 and BG17803.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S7) targeting the *vegfa* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S4, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17787 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S8	Gibson assembly	Fragment 1: PU6 and spacer S5 from pFYF1320 EGFP amplified with primers BG17911 and BG17804.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S8) targeting the *vegfa* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S5, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17789 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		
phBE4_S9	Gibson assembly	Fragment 1: PU6 and spacer S6 from pFYF1320 EGFP amplified with primers BG17911 and BG17805.	APOBEC1-linker-nSpyCas9-linker-UGI-linker-UGI-linker-SV40NLS and sgRNA (spacer S9) targeting the *vegfa* gene in the genome of HEK293T cells in base-editing assays.	This study
		Fragment 2: Spacer S6, sgRNA, and partly CMV enhancer from pFYF1320 EGFP amplified with primers BG17791 and BG17912.		
		Fragment 3: Partly CMV enhancer, PCMV, and partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene from pBE4 amplified with primers BG17272 and BG17793.		
		Fragment 4: Partly *apobec1-linker-hnspycas9-linker-ugi-linker-ugi-linker-sv40nls* gene and backbone from pBE4 amplified with primers BG17794 and BG17795.		
		Fragment 5: Backbone from pBE4 amplified with primers BG17796 and BG17913.		

Supplementary Table 2. Spacers used in this study.

Spacer ID	Target strand	Targeting location in the genome (5' -> 3')	3' PAM (5' -> 3')	Sequence of spacer (23 nt) (5' -> 3')	Protein	Application
SapI	X	X	X	A**GAAGAGC**GCGCCAT**GCTCTTC**G	hThermoCas9, hSCas9	-
T1	-	*dnmt1* gene (14645 - 14622)	NNNNCCAA	CCCGGTTGTAAGCATGAGCACCG	hThermoCas9	Genome & base editing assays
T2	+	*dnmt1* gene (35086 - 35109)	NNNNCCAA	TGGCTCGCGCCAAAACAGTCATG	hThermoCas9	Genome & base editing assays
T3	+	*dnmt1* gene (480 - 503)	NNNNCCAA	GGTCGCGCGTGCCCGGGCTGTTT	hThermoCas9	Genome & base editing assays
T4	+	*emx1* gene (22941 - 22964)	NNNNCCAA	ACCGGTGGCGCATTGCCACGAAG	hThermoCas9	Genome & base editing assays
T5	+	*emx1* gene (14661 - 14684)	NNNNCCAA	TCGACCCTACTACACCACCGTCC	hThermoCas9	Genome & base editing assays
T6	+	*emx1* gene (20785 - 20808)	NNNNCCAA	CCTCACTCCATCCATCGAGGCCG	hThermoCas9	Genome & base editing assays
T7	+	*vegfa* gene (-911 - -888)	NNNNCCAA	CCCGTTCTCAGCTCCACAAACTT	hThermoCas9	Genome & base editing assays
T8	-	*vegfa* gene (-427 - -450)	NNNNCCAA	GGAGTGACCCCTGGCCTTCTCCC	hThermoCas9	Genome & base editing assays
T9	-	*vegfa* gene (4232 -4209)	NNNNCCAA	ATCCCCTGAGAGGACAGGGAACC	hThermoCas9	Genome & base editing assays
S1	+	*dnmt1* gene (581 - 601)	NGG	CCCCAAATACAGCAAGCTTT	hSpyCas9	Genome & base editing assays
S2	+	*dnmt1* gene (523 - 543)	NGG	ATTCCCCCGTAGCTCCCTGG	hSpyCas9	Genome & base editing assays
S3	+	*dnmt1* gene (35230 - 35250)	NGG	CATTCAGTGCGGGCAGTACC	hSpyCas9	Genome & base editing assays
S4	+	*emx1* gene (22904 - 22924)	NGG	GAGTCCGAGCAGAAGAAGAA	hSpyCas9	Genome & base editing assays
S5	+	*emx1* gene (22773 - 22793)	NGG	GGCCCCAGTGGCTGCTCTGG	hSpyCas9	Genome & base editing assays
S6	+	*emx1* gene (22827 - 22847)	NGG	CCCTGGCCCAGGTGAAGGTG	hSpyCas9	Genome & base editing assays
S7	-	*vegfa* gene (633 - 613)	NGG	GACCCCCTCCACCCCGCCTC	hSpyCas9	Genome & base editing assays
S8	+	*vegfa* gene (-744 - 724)	NGG	TGGCCCCTGTGCCCAGCCCT	hSpyCas9	Genome & base editing assays
S9	+	*vegfa* gene (548 - 568)	NGG	CCCCGGCCGGCGGCGGACAG	hSpyCas9	Genome & base editing assays

***Red color**: restriction sites of SapI to create sticky ends (during restriction digestion) and insert a spacer targeting a gene of interest.*

Supplementary Table 3. Primers/Oligonucleotides used in this study.

	Primer/ Oligo ID	Sequence (5' -> 3')	Function
Genome editing assays	BG18420	ATGGTGAGCAAGGGCGAG	Construction of phTCas9_SapI.
	BG19287	GATGTATTCATGCTCGAACGCCTC	
	BG19289	GGCGTTCGAGCATGAATACATC	
	BG19291	ATGACCAAAATCCCTTAACGTGAGTTTT	
	BG19292	aaaactcacgttaagggattttggtcatCTAACTCCGCCCATCCC	
	BG19293	GGATATACAAGCTCCCGGG	
	BG19141	aaagctcccgggagcttgtatatccGCTGAGTAGTGCGCGAGC	
	BG19142	acagctcctcgcccttgctcaccatGGTGGCGGTG	
	BG18732	ACCCCCGGTTGTAAGCATGAGCACCG	Construction of phTCas9_T1.
	BG18741	GACCGGTGCTCATGCTTACAACCGGG	
	BG18733	ACCTGGCTCGCGCCAAAACAGTCATG	Construction of phTCas9_T2.
	BG18742	GACCATGACTGTTTTGGCGCGAGCCA	
	BG18734	ACCGGTCGCGCGTGCCCGGGCTGTTT	Construction of phTCas9_T3.
	BG18743	GACAAACAGCCCGGGCACGCGCGACC	
	BG18735	ACCACCGGTGGCGCATTGCCACGAAG	Construction of phTCas9_T4.
	BG18744	GACCTTCGTGGCAATGCGCCACCGGT	
	BG18736	ACCTCGACCCTACTACACCACCGTCC	Construction of phTCas9_T5.
	BG18745	GACGGACGGTGGTGTAGTAGGGTCGA	
	BG18737	ACCCCTCACTCCATCCATCGAGGCCG	Construction of phTCas9_T6.
	BG18746	GACCGGCCTCGATGGATGGAGTGAGG	
	BG18738	ACCCCCGTTCTCAGCTCCACAAACTT	Construction of phTCas9_T7.
	BG18747	GACAAGTTTGTGGAGCTGAGAACGGG	
	BG18739	ACCGGAGTGACCCCTGGCCTTCTCCC	Construction of phTCas9_T8.
	BG18748	GACGGGAGAAGGCCAGGGGTCACTCC	
	BG18740	ACCATCCCCTGAGAGGACAGGGAACC	Construction of phTCas9_T9.
	BG18749	GACGGTTCCCTGTCCTCTCAGGGGAT	
	BG18420	ATGGTGAGCAAGGGCGAG	Construction of phSCas9_SapI.
	BG19288	CGTCCCATCCATCTTCTCTAATATGGG	
	BG19290	CCATATTAGAGAAGATGGATGGGACG	
	BG19291	ATGACCAAAATCCCTTAACGTGAGTTTT	
	BG19292	aaaactcacgttaagggattttggtcatCTAACTCCGCCCATCCC	
	BG19293	GGATATACAAGCTCCCGGG	
	BG19141	aaagctcccgggagcttgtatatccGCTGAGTAGTGCGCGAGC	
	BG19142	acagctcctcgcccttgctcaccatGGTGGCGGTG	
	BG18750	ACCCCCCAAATACAGCAAGCTTT	Construction of phSCas9_S1.
	BG18759	AACAAAGCTTGCTGTATTTGGGG	
	BG18751	ACCATTCCCCCGTAGCTCCCTGG	Construction of phSCas9_S2.
	BG18760	AACCCAGGGAGCTACGGGGGAAT	
	BG18752	ACCCATTCAGTGCGGGCAGTACC	Construction of phSCas9_S3.

Supplementary Table 3. Primers/Oligonucleotides used in this study. *(continuation)*

	BG18761	AACGGTACTGCCCGCACTGAATG	
	BG18753	ACCGAGTCCGAGCAGAAGAAGAA	Construction of phSCas9_S4.
	BG18762	AACTTCTTCTTCTGCTCGGACTC	
	BG18754	ACCGGCCCCAGTGGCTGCTCTGG	Construction of phSCas9_S5.
	BG18763	AACCCAGAGCAGCCACTGGGGCC	
	BG18755	ACCCCCTGGCCCAGGTGAAGGTG	Construction of phSCas9_S6.
	BG18764	AACCACCTTCACCTGGGCCAGGG	
	BG18756	ACCGACCCCCTCCACCCCGCCTC	Construction of phSCas9_S7.
	BG18765	AACGAGGCGGGGTGGAGGGGGTC	
	BG18757	ACCTGGCCCCTGTGCCCAGCCCT	Construction of phSCas9_S8.
	BG18766	AACAGGGCTGGGCACAGGGGCCA	
	BG18758	ACCCCCCGGCCGGCGGCGGACAG	Construction of phSCas9_S9.
	BG18767	AACCTGTCCGCCGCCGGCCGGGG	
Base-editing assays	BG17303	ttctggtggttcttctggtggttctAAATACAAGATTGGCCTCGCC	Construction of phThermoBE4_T1/T2/T3/T4/T5/T6/T7/T8/T9.
	BG17304	ctccagaaccaccagagtcacccccccAGGGGCCGAATTGTCTCG	
	BG17305	gggggtgactctggtg	
	BG17306	atatctggcccgtacatcg	
	BG17307	gcttcgcgatgtacgggccagatatGAGGGCCTATTTCCCATGATTC	
	BG17310	cggtgctcatgcttacaaccgggGGTGTTTCGTCCTTTCCAC	
	BG17277	cccggttgtaagcatgagcaccgGTCATAGTTCCCCTGAGATTATCG	
	BG17271	taatcaatgtcaacgcgtaaaaaaaCGCCTAAGAGTGGGGAATG	
	BG17272	tacgcgttgacattgattattgac	
	BG17308	caccagaactttcgggtg	
	BG17274	ttttttttACGCGTTGACATTGATTATTGAC	
	BG17309	agaaccaccagaagaacCACCAGAACTTTCGGGGTG	
	BG17312	catgactgttttggcgcgagccaGGTGTTTCGTCCTTTCCAC	
	BG17279	tggctcgcgccaaaacagtcatgGTCATAGTTCCCCTGAGATTATCG	
	BG17315	aaacagcccgggcacgcgcgaccGGTGTTTCGTCCTTTCCAC	
	BG17281	ggtcgcgcgtgcccgggctgtttGTCATAGTTCCCCTGAGATTATCG	
	BG17316	cttcgtggcaatgcgccaccggtGGTGTTTCGTCCTTTCCAC	
	BG17283	accggtggcgcattgccacgaagGTCATAGTTCCCCTGAGATTATCG	
	BG17319	ggacggtggtgtagtagggtcgaGGTGTTTCGTCCTTTCCAC	
	BG17285	tcgaccctactacaccaccgtccGTCATAGTTCCCCTGAGATTATCG	
	BG17321	cggcctcgatggatggagtgaggGGTGTTTCGTCCTTTCCAC	
	BG17287	cctcactccatccatcgaggccgGTCATAGTTCCCCTGAGATTATCG	
	BG17322	aagtttgtggagctgagaacgggGGTGTTTCGTCCTTTCCAC	
	BG17323	cccgttctcagctccacaaacttGTCATAGTTCCCCTGAGATTATCG	
	BG17327	gggagaaggccaggggtcactccGGTGTTTCGTCCTTTCCAC	
	BG17291	ggagtgacccctggccttctcccGTCATAGTTCCCCTGAGATTATCG	
	BG17329	ggttccctgtcctctcaggggatGGTGTTTCGTCCTTTCCAC	
	BG17293	atcccctgagaggacagggaaccGTCATAGTTCCCCTGAGATTATCG	
	BG17911	gcttcgcgatgtacgggccagatatgaggggcctatttcccatgattc	Construction of phBE4_S1/S2/S3/S4/S5/S6/S7/S8/S9.
	BG17797	aaagcttgctgtatttggggGGTGTTTCGTCCTTTCCACAAG	

Supplementary Table 3. Primers/Oligonucleotides used in this study. *(continuation)*

	BG17773	ccccaaatacagcaagctttGTTTTAGAGCTAGAAATAGCAAG	
	BG17912	tagtcaataatcaatgtcaacgcgtAAAAAAAGCACCGACTC	
	BG17272	TACGCGTTGACATTGATTATTGAC	
	BG17793	TTCGTGAGTTCATTGTACAC	
	BG17794	CACAGTGTACAATGAACTCACG	
	BG17795	CAACTAGAAGGCACAGTCG	
	BG17796	CTCGACTGTGCCTTCTAGTTG	
	BG17913	ATATCTGGCCCGTACATCG	
	BG17798	ccagggagctacgggggaatGGTGTTTCGTCCTTTCCACAAG	
	BG17777	attcccccgtagctccctggGTTTTAGAGCTAGAAATAGCAAG	
	BG17799	ggtactgcccgcactgaatgGGTGTTTCGTCCTTTCCACAAG	
	BG17779	cattcagtgcgggcagtaccGTTTTAGAGCTAGAAATAGCAAG	
	BG17800	ttcttcttctgctcggactcGGTGTTTCGTCCTTTCCACAAG	
	BG17781	gagtccgagcagaagaagaaGTTTTAGAGCTAGAAATAGCAAG	
	BG17801	ccagagcagccactggggccGGTGTTTCGTCCTTTCCACAAG	
	BG17783	ggccccagtggctgctctggGTTTTAGAGCTAGAAATAGCAAG	
	BG17802	caccttcacctgggccagggGGTGTTTCGTCCTTTCCACAAG	
	BG17785	ccctggcccaggtgaagggtgGTTTTAGAGCTAGAAATAGCAAG	
	BG17803	gaggcggggtggaggggggtcGGTGTTTCGTCCTTTCCACAAG	
	BG17787	gacccccctccaccccgccctcGTTTTAGAGCTAGAAATAGCAAG	
	BG17804	agggctgggcacaggggccaGGTGTTTCGTCCTTTCCACAAG	
	BG17789	tggcccctgtgcccagccctGTTTTAGAGCTAGAAATAGCAAG	
	BG17805	ctgtccgccgccggccggggGGTGTTTCGTCCTTTCCACAAG	
	BG17791	ccccggccggcggcggacagGTTTTAGAGCTAGAAATAGCAAG	
Sequencing	BG18480	TCTGGTGGTTCTCCCAAGAAG	Sequencing of phTCas9_SapI, and phTCas9_T1/T2/T3/T4/T5/T6/T7/T8/T9.
	BG18482	GTATTGGGCGATCTTCCGCT	
	BG18596	GCAAACAAACCACCGCTG	
	BG19312	GGGGAGAACCGTATATAAGTGC	
	BG19313	GGCACCAGTTGCGTGAG	
	BG18597	GTGACCACCCTGACCTAC	
	BG18598	AATGAATGCAATTGTTGTTGTTAACTTG	
	BG18599	ATTCTCTTACTGTCATGCCATCC	
	BG18600	TTAGGCGTTTTGCGCTGC	
	BG18601	ATTTACGGTAAACTGCCCACTTG	
	BG18602	CTTTTCGAGAAGAAACATGAGATTGATG	
	BG18603	CAACCCTCAAAGAGAACGAAAAG	
	BG18604	GAATGGTAAGTGCGCATACTC	
	BG18605	AATAGATGAGAGATCTGGCAAGATTC	
	BG18480	TCTGGTGGTTCTCCCAAGAAG	Sequencing of phSCas9_SapI, and phSCas9_S1/S2/S3/S4/S5/S6/S7/S8/S9.
	BG18482	GTATTGGGCGATCTTCCGCT	
	BG18596	GCAAACAAACCACCGCTG	
	BG19312	GGGGAGAACCGTATATAAGTGC	
	BG19313	GGCACCAGTTGCGTGAG	

Supplementary Table 3. Primers/Oligonucleotides used in this study. *(continuation)*

	BG18597	GTGACCACCCTGACCTAC	
	BG18598	AATGAATGCAATTGTTGTTGTTAACTTG	
	BG18599	ATTCTCTTACTGTCATGCCATCC	
	BG18600	TTAGGCGTTTTGCGCTGC	
	BG18601	ATTTACGGTAAACTGCCCACTTG	
	BG18606	AAGAAATTTTTAGCAATGAGATGGCC	
	BG18607	GCCGTTATCCGCTTCAATGATC	
	BG18608	TACGAGTATTTCACAGTGTACAATGAAC	
	BG18609	CAAGGGGACTCATTGCACG	
	BG18610	TTATTAAACGTCAGCTCGTGGAAAC	
	BG18611	CCTACAGTTGCCTATTCTGTCCTAG	
	BG17091	CTACAACAAGGCAAGGCTTG	Sequencing of phThermoBE4_T1/T2/T3/T4/T5/T6/T7/T8/T9.
	BG17092	TAACAACTCCGCCCCATTG	
	BG17093	CATCACTGAATTCCTGTCAAGG	
	BG17330	AAATACAAGATTGGCCTCGCC	
	BG17295	GAGCGCAGGTTGATTTATAAACAG	
	BG17296	AGACAGCAATCAGACAGCTG	
	BG17297	CCCTAAAGAATCCATTAAGGCAC	
	BG17331	CACGCGAGAAGACCATTAAAAC	
	BG17305	GGGGGTGACTCTGGTG	
	BG17107	CGCAACGCAATTAATGTGAG	
	BG17332	ATGGTGATGGTGATGATGACC	
	BG17298	CAAGTACGTCCACCTGATAC	
	BG17299	TACCGCCTAAGAGTCTCCTG	
	BG17300	GAGAGCGAGAATAGGGGATG	
	BG17301	TTGTGGTATGCACCAAGCTC	
	BG17302	GCCACACATCAATCTCATGTTTC	
	BG17114	GCTAGAACCACCAGAAGAAC	
	BG17115	GCTCTCTCGGATCGAAGAATAC	
	BG17271	TAATCAATGTCAACGCGTAAAAAAACGCCTAAGAGTGGGGAATG	
	BG17091	CTACAACAAGGCAAGGCTTG	Sequencing of phBE4_S1/S2/S3/S4/S5/S6/S7/S8/S9.
	BG17092	TAACAACTCCGCCCCATTG	
	BG17093	CATCACTGAATTCCTGTCAAGG	
	BG17094	GGTATACACGTCGCAAGAAC	
	BG17095	TTAGCGATGCAATCCTCCTATC	
	BG17096	ACTGAGGGCATGCGTAAAC	
	BG17097	GTAGTGGATGAGCTAGTTAAGGTC	
	BG17098	ATTGGTGTCGGACTTCAGAAAG	
	BG17099	CCGAAAACGGATGTTGGC	
	BG17100	GGATAGATTTGTCACAGCTTGG	
	BG17995	CAACTAGAAGGCACAGTCG	
	BG17108	ATAACCAGTTGCTTACCGGTC	
	BG17109	TCCTTGTAACCTTTCGCCTC	

Supplementary Table 3. Primers/Oligonucleotides used in this study. *(continuation)*

	BG17110	TCGTCGTATTTCGTATTCATTCG	
	BG17111	AGTCATCATGGATCAGCTGC	
	BG17112	TTGTCATCCATGCGAACC	
	BG17113	TAGCCGTCGGGATTTAGAG	
	BG17114	GCTAGAACCACCAGAAGAAC	
	BG17115	GCTCTCTCGGATCGAAGAATAC	
Screening of genomic sites	BG19826	TGAATCTCTTGCACGAATTTCTGC	Population PCR and Sanger sequencing of the genomic target site T1 to screen for genome or base-editing after transfection with phTCas9_T1 or phThermoBE4_T1, respectively.
	BG19827	CCGCCTGTAATACTAGCTACTTGG	
	BG19296	CAGTAATTGCCAGAGCAACCTTAG	Population PCR and Sanger sequencing of the genomic target site T2 to screen for genome or base-editing after transfection with phTCas9_T2 or phThermoBE4_T2, respectively.
	BG19297	CCAGAGTTCAAGAAATTCTCCTGC	
	BG19294	GCTGAAGCCTCCGAGATG	Population PCR and Sanger sequencing of the genomic target site T3 to screen for genome or base-editing after transfection with phTCas9_T3 or phThermoBE4_T3, respectively.
	BG19295	GCTCCCACTGATCCTTGTG	
	BG19298	CTTTGCTGGGGCTAGAGG	Population PCR and Sanger sequencing of the genomic target site T4 to screen for genome or base-editing after transfection with phTCas9_T4 or phThermoBE4_T4, respectively.
	BG19299	TGAATTACCCTTGACCCCCTC	
	BG19306	GAAGGAATAAACTCCAACGGCG	Population PCR and Sanger sequencing of the genomic target site T5 to screen for genome or base-editing after transfection with phTCas9_T5 or phThermoBE4_T5, respectively.
	BG19307	CAGGCAGCTAGGTTCAGTTTG	
	BG19308	GGAAGGTACAGAAAAGGGCC	Population PCR and Sanger sequencing of the genomic target site T6 to screen for genome or base-editing after transfection with phTCas9_T6 or phThermoBE4_T6, respectively.
	BG19309	CAGGGCACCTCTTGGATTC	
	BG19770	CTACAGACGTTCCTTAGTGCTGG	Population PCR and Sanger sequencing of the genomic target sites T7 and T8 to screen for genome or base-editing after transfection with phTCas9_T7 and phTCas9_T8 (for genome editing) or phThermoBE4_T7 and phThermoBE4_T8 (for base-editing), respectively.
	BG19771	GCAATGAAGGGGAAGCTCG	
	BG19310	GGAAGTGTGGGGAAGGTACAG	Population PCR and Sanger sequencing of the genomic target site T9 to screen for genome or base-editing after transfection with phTCas9_T9 or phThermoBE4_T9, respectively.
	BG19311	GAAATGGGAGGATCCAAACCAC	
	BG19294	GCTGAAGCCTCCGAGATG	Population PCR and Sanger sequencing of the genomic target sites S1 and S2 to screen for genome or base-editing after transfection with phSCas9_S1 and phSCas9_S2 (for genome editing) or phBE4_S1 and phBE4_S2 (for base-editing), respectively.
	BG19295	GCTCCCACTGATCCTTGTG	
	BG19296	CAGTAATTGCCAGAGCAACCTTAG	Population PCR and Sanger sequencing of the genomic target site S3 to screen for genome or base-editing after transfection with phSCas9_S3 or phBE4_S3, respectively.
	BG19297	CCAGAGTTCAAGAAATTCTCCTGC	
	BG19298	CTTTGCTGGGGCTAGAGG	Population PCR and Sanger sequencing of the genomic target site S4, S5 and S6 to screen for genome or base-editing after transfection with phSCas9_S4, phSCas9_S5, and phSCas9_S6 (for genome editing) or phBE4_S4, phBE4_S5, and phBE4_S6 (for base-editing), respectively.
	BG19299	TGAATTACCCTTGACCCCCTC	
	BG19849	TTTGGAAACCAGCAGAAAGAGG	Population PCR and Sanger sequencing of the genomic target site S7, and S9 to screen for genome or base-editing after transfection with phSCas9_S7, and phSCas9_S9 (for genome editing) or phBE4_S7, and phBE4_S9 (for base-editing), respectively.
	BG19850	CTTACCTTGGCATGGTGGAG	
	BG19770	CTACAGACGTTCCTTAGTGCTGG	Population PCR and Sanger sequencing of the genomic target site S8 to screen for genome or base-editing after transfection with phSCas9_S8 or phBE4_S8, respectively.
	BG19771	GCAATGAAGGGGAAGCTCG	

***Lowercase**: overhang sequence for Gibson assembly*
***Uppercase**: sequence annealing to the PCR template DNA*
***Red color**: spacer-protospacer mismatch (nucleotide substitution in the spacer sequence)*

Supplementary Table 4. DNA sequences of cas9 (fusion) genes used in this study.

Gene/ Fusion gene	Sequence (5' -> 3')
hthermocas9-linker-sv40nls	atgaaatacaagattggccctcgacattggcataacctctataggctgggccgtgatcaaccttgatatacccccgaatagaggaccttggcgttaggatcttcgatcgcgcagagaatcccaagacaggcgaaagcctcgccctgcctaggcgtttggcacgttcagcaaggagacgct tgaggcgcaggaagcacaggcttgaacgaatccgaaggctctttgtgcgagaggggatcttgacaaaagaggaacttaataaacttttcgagaagaaacatgagattgatgtgtggcaactgagagtcgaggccctcgaccgcaagctcaacaatgacgagctggcaagaattttg ttgcacttggcaaagaggcgagggttccggtcaaatcgaaaatcagaacgaactaataaagagaattcaaccatgttgaagcacatagaggagaatcagagcatattgtctagctatcgtacagtggctgagatggtcgttaaagaccccaagttcagcttgcataaaaggaacaa ggaagacaactatacgaataccgtggctcgggatgacctcgagcgcgagattaagctcatctttgctaagcaaagggagtacggcaatattgtatgtactgaggcgttcgagcatgaatacatcagtatctgggccagccagcggccctttcgcctcaaaagacgacattgaaaagaa ggtgggcttttgcaccttcgaacccaaggagaagcgtgctccgaaggccacctatacttttcaatctttttacagtgtgggagcacatcaataagctcaggctggttagccccggcgggatacgagctctcacagacgacgagcgcaggttgatttataaacaggccttccacaagaac aagattacttttcacgacgtgagaaccccttctcaatctgcccgatgatacgcgcttcaagggcctcttgtacgatcggaatacaaccctcaaagagaacgaaaaggtcagatttctggagcttggtgcatacccacaagatccgcaaggcaattgattctgtatacgggaagggtgctgc taagagcttcagacccatcgacttcgacacccttcggatatgcccttactatgttcaaggatgatacggatatccgatcttatctgaggaatgagtatgagcagaacgggaagcgcatggagaacctggccgacaaggtttacgacgaggagctcatagaggagctgcttaatctgagc ttcagcaaattcgggcacttgtctctgaaggccttgcggaatattctcccttacatggagcagggtgaggtgtatagcacggcatgcgagcgggctgggtacaccttcactggccccaagaagaagcaaaagacagtgtcgcttcccaatatcccaccaatagcaaacccgtggtga tgagagctttgacccaagctcggaaggtcgtgaacgcaataatttaagaaatatgggagccccgtgtcaatacacattgagctcgcgcgtgagctcagtcagtcattcgacgagagacggaagatgcaaaaggagcaagaggggaatcggaagaagaatgagacagcaatcagac agctggtggagtacggactcaccctgaaccctaccggcctggatatcgtcaagtttaagctgtggtccgagcagaatggtaagtgcgcatactctctgcagcccatagagattgaaagacttctggagcctggatacactgaggtggatcacgtcatcccctattctcgctctcttgatg actcatacacacaaacaaggtgctggtactcactaaagaaaatagggagaagggtaataggacgcctgcagagtacctgggggttgggaagcgagcggtggcagcaatttgaaaccttcgtgctcaccaacaaacaattcagcaagaagaaaagggacaggctgctgcgactgcact atgacgagaatgaggagaacgaattcaagaacagaaacctgaacgacacacgctacatatctcggtttctggccaatttcatccgtgagcacctgaagtttgcagattccgacgataagcagaaggtctatacagtgaatggaaggatcactgctcacctgaggtctcggtggaact tcaataagaatcgtgaggagagtaacctgcaccacgctgtagacgccgctatagttgcatgtactactccctctgacattgctcgtgtgacggcgttttaccagagaagagagcagaacaaggagctttcaaagaaaaccgaccctcaattcccccaaccctggccacatttcgccga cgagcttcaagccaggctgagcaagaacctaaagaatccattaaggcactgaacctcggcaactacgacaatgaaaagcttgagtcactgcaacctgtcttcgtgagcagaatgcccaaaagatcaattaccggtgctgcccaccaggagactcttaggcggtacattggaatag atgaagatctggcaagattcaaactgttgttaagaagaagcttagtgaaattcagctcgacaagaccggacactttcccatgtatggaaaggagtctgaccctcgcacttacgaggcaatacggcagcgcctgctggagcacaataacgatcctaagaaagcattccaggaacccc tctacaagcccaagaagaatggggagcttgggccgattattagaactattaagataattgacaccaccaaccaggtgatcccactgaatgacgggaagaccgtagcatataattctaatattgttaggttgatgtgttcgaaaaggacggtaagtactactgcgttcccatatacact attgacatgatgaagggcattctgcccaataaagcaatagaacctaataaagccctatagcgaatggaaagagatgactgaagattacaccttcggttagcctgtatcccaacgacctgattaggatagagttcccacgcgagaagaccattaaaacagccgttggagaggagatt aagatcaaagacttgtttgcatactaccagacgatagatagcagcaacggcggcttgagccttgtgtcccacgacaataatttctccctgagaagtattggcagccgcactctgaagcgctttgaaaagtatcaggtggacgtacttggaaatatttataaggttcggggagagaaaa ggctgggcgtcgcttccagttcacactccaaggcaggcgagacaattcggccccttgtctggtggttctcccaagaagaagaggaaagtctaa
hspycas9-linker-sv40nls	atggataaaaagtattctattggtttagacatcggcactaattccgttggatgggctgtcataaccgatgaatacaaagtaccttcaaagaaatttaaggtgttgggggaacacagaccgtcattcgattaaaaagaatcttatcggtgccctcctattcgatagtggcgaaacggcaga ggcgactcgcctgaaacgaaccgctcggagaaggtatacacgtcgcaagaaccgaatatgttacttacaagaaatttttagcaatgagatggccaaagttgacgattctttcctttcaccgtttggaagaagtccttcctttgtcgaagaggacaagaaacatgaacggcaccccatctttg gaaacatagtagatgaggtggcatatcatgaaaagtacccaacgatttatcacctcagaaaaaagctagttgactcaactgataaagcggacctgaggttaatctacttggctcttgcccatatgataaagttccgtgggcactttctcattgagggtgatctaaatccggacaactcg gatgtcgacaaactgttcatccagttagtacaaacctataatcagttgtttgaagagaaccctataaatgcaagtggcgtggatgcgaaggctattcttagcgcccgcctctctaaatcccgacggctagaaaacctgatcgcacaattacccggagagaagaaaaatgggttgttcg gtaaccttatagcgctctcactaggcctgacaccaaatttaagtcgaacttcgacttagctgaagatgccaaattgcagcttagtaaggacacgtacgatgacgatctcgacaatctactggcacaaattggcgatcagtatgcggacttatttttggctgccaaaaaccttagcgatg caatcctcctatctgacatactgagagttaatactgagattaccaaggcgccgttatccgcttcaatgatcaaaaggtacgatgaacatcaccaagacttgacacttctcaaggccctagtccgtcagcaactgcctgagaaatataaggaaatattctttgatcagtcgaaaaacgggg tacgcaggttatattgacggcggagcgagtcaagaggaattctacaagtttatcaaacccatattagagaagatggatgggacggaagagttgcttgtaaaactcaatcgcgaagatctactgcgaaagcagcggacttttcgcaacggtagcattccacatcaaatccacttaggc gaattgcatgctatacttagaaggcaggaggatttttatccgttcctcaaagacaatcgtgaaaagattgagaaaatcctaacctttcgcataccttactatgtgggaccccgggcccgagggaactctcggttcgcatggatgacaagaaagtccgaagaaacgattactccatggaa ttttgaggaagttgtcgataaaggtgcgtcagctcaatcgttcatcgagaggatgaccaacttttgacaagaatttaccgaacgaaaaagtattgcctaagcacagtttacttttacgagtatttcacagtgtacaatgaactcacgaaagttaagtatgtcactgagggcatgcgtaaac ccgcctttctaagcggagaacagaagaaagcaatagtagatctgttattcaagaccaaccgcaaagtgacagttaagcaattgaaagaggactactttaagaaaattgaatgcttcgattctgtcgagatctccggggtagaagatcgatttaatgcgtcacttggtacgtatcatga cctcctaaagataattaaagataaggacttcctggataacgaagagaatgaagatatcttagaagatatagtgttgactcttaccctctttgaagatcgggaaatgattgaggaaagactaaaaacatacgctcacctgttcgacgataaggttatgaaacagttaaagaggcgtcgc tatacgggctgggacgattgtcgcggaaacttatcaacgggataagagacaagcaaagtggtaaaactattctcgattttctaaagagcgacggcttcgccaatagggaactttatgcagctgatccatgatgactctttaaccttcaaagaggatatacaaaggcacagtttccg gacaagggactcattgcacgaacatattgcgaatcttgctggttcgccagccatcaaaaagggcatactccagacagtcaaagtagtggatgagctagttaaggtcatgggacgtcacaaacccggaaaacattgtaatcgagatggcacgcgaaaatcaaacgactcagaaggg gcaaaaaaacagtcgagagcggatgaagagaatagaagaggggtattaaagaactgggcagccgagatcttaaaggagcatcctgtggaaaatacccaattgcagaacgagaaactttacctctattacctacaaaatggaagggacatgtatgttgatcaggaactggacataaa ccgtttatctgattacgacgtcgatcacattgtaccccaatcctttttgaaggacgattcaatcgacaataaagtcttacacgctcggataagaaccgagggaaaagtgacaatgttccaagcgaggaagtcgtaaagaaaatgaagaactattggcggcagctcctaaatgcgaa actgataacgcaaagaaagttcgataacttaactaaagctgagaggggtggcttgtctgaacttgacaaggccggatttattaaacgtcagctcgtggaaacccgccaaatcacaaagcatgttgcacagatactagattcccgaatgaatacgaaatacgacgagaacgataagc tgattcgggaagtcaaagtaatcacttttaaagtcaaaattggtgtcggacttcagaaaggatttttcaattctataaagttagggagataaataactaccaccatgcgcacgacgcttatcttaatgccgtcgtagggaccgcactcattaagaaatacccgaagctagaaagtgagttt gtgtatggtgattacaaagtttatgacgtccgtaagatgatcgcgaaaagcgaacaggagatagggcaaggctacagccaaatacttctttttattctaacatatgaatttctttaagacggaaatcactctggcaaacggagagatacgcaaacgaccttttaattgaaaccaatgggg agacaggtgaaatcgtatgggataagggccgggacttcgcgacggtgagaaaagttttgtccatgccccaagtcaacatagtaaagaaaactgaggtgcagaccggaggggttttcaaaggaatcgattcttccaaaaaggaatagtgataagctcatcgctcgtaaaaaggactgg gacccgaaaaagtacggtggcttcgatagccctacagttgcctattctgtcctagtagtggcaaaagttgagaagggaaaatccaagaaactgaagtcagtcaaagaattattggggataacgattatggagcgctcgtctttttgaaaagaaccccatcgacttccttgaggcgaaag gttacaaggaagtaaaaaaggatctcataattaaactaccaaagtatagtctgtttgagttagaaaatggccgaaaacggatgttggctagcgccggagagcttcaaaaggggaacgaactcgcactaccgtctaaatacgtgaatttcctgtatttagcgtcccattacgagaagtt gaaaggttcacctgaagataacgaacagaagcaacttttgttgagcagcacaaacattatctcgacgaaatcatagaggcaaatttcggaattcagtaagagagtcatcctagctgatgccaatctggacaaagtattaagcgcatacaacaagcacagggataaacccatacgtg

Supplementary Table 4. DNA sequences of cas9 (fusion) genes used in this study. *(continuation)*

	agcaggcggaaaatattatccatttgtttactcttaccaacctcggcgctccagccgcattcaagtattttgacacaacgatagatcgcaaacgatacacttctaccaaggaggtgctagacgcgacactgattcaccaatccatcacgggattatatgaaactcggatagatttgtcac agcttgggggtgactctggtggttctcccaagaagaagaggaaagtctaa
apobec1-linker-hnthermocas9(D8A)-linker-ugi-linker-ugi-linker-sv40nls	atgagctcagagactggcccagtggctgtggaccccacattgagacggcggatcgagccccatgagtttgaggtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggccggcactccatttggcgacatacatcacagaacactaaca agcacgtcgaagtcaacttcatcgagaagttcacgacagaaagatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggagcccatgcggcgaatgtagtagggccatcactgaattcctgtcaaggtatccccacgtcactctgtttatttacatcgcaaggctgt accaccacgctgaccccgcaatcgacaaggcctgcgggatttgatctcttcaggtgtgactatccaaattatgactgagcaggagtcaggatactgctggaggaaacttttgtgaattatagcccgagtaatgaagcccactggcctaggtatccccatctgtgggtacgactgtacgttc ttgaactgtactgcatcatactgggcctgcctccttgtctcaacatttctgagaaggaagcagccacagctgacattctttaccatcgctcttcagtcttgtcattaccagcgactgcccccacacattctctgggccaccgggttgaaatctggtggttcttctggtggttctagcggcagcg agactcccgggacctcagagtccgccacacccgaaagttctggtggttcttctggtggttctaaatacaagattggcctgCcattggcataacctctatagggctgggccgtgatcaaccttgatatacccccgaatagaggacctggcgttaggatcttcgatcgcgcagagaatccc aagacaggcgaaagcctcgccctgcctaggcgtttggcacgttcagcaaggagacgcttgaggcgcaggaagcacaggcttgaacgaatccgaaggctctttgtgcgagaggggatcttgacaaaagaggaacttaataaacttttcgagaagaaacatgagattgatgtgtggca actgagagtcgaggccctcgaccgcaagctcaacaatgacgagctggcaagaattttgttgcacttggcaaagaggcgagggttccggtcaaatcgaaaatcagaacgaactaataaagagaattcaaccatgttgaagcacatagaggagaatcagagcatattgtctagctatc gtacagtggctgagatggtcgttaaagaccccaagttcagcttgcataaaaggaacaaggaagacaactatacgaatacgtggctcgggatgacctcgagcgcgagattaagctcatctttgctaagcaaaggggagtacggcaatattgtatgtactgaggcgttcgagcatgaat acatcagtatctgggccagccagcggcccttcgccctcaaaagacgacattgaaaagaaggtgggcttttgcaccttcgaaccccaaggagaagcgtgctccgaaggccacctatacttttcaatcttttacagtgtgggagcacatcaataagctcaggctggttagcccggcgggat acgagctctcacagacgacgagcgcaggttgatttatataaacaggccttccacaagaacaagattacttttcacgacgtgagaacccttctcaatctgcccgatgatacgcgcttcaagggcctcttgtacgatcggaatacaacctcaaagagaacgaaaaggtcagatttctggag cttggtgcataccacaagatccgcaaggcaattgattctgtatacgggaaggtgctgctaagagcttcagaccccatcgacttcgacaccttcggatatgcccttactatgttcaaggatgatacggatatccgatcttatctgaggaatgagtatgagcagaacgggaagcgcatgga gaacctggccgacaaggtttacgacgaggagctcatagaggagctgcttaatctgagcttcagcaaattcgggcacttgtctctgaaggccttgcggaatattctcccttacatggagcagggtgaggtgatagcacggcatgcgagcgggctgggtacacctcactggccccaag aagaagcaaaagacagtgctgcttcccaatatcccaccaatagcaaacccgtggtgatgagagcttgacccaagctcggaaggtcgtgaacgcaataatgaagaaatatgggagccccgtgtcaatacacattgagctcgcgcgtgagctcagtcagtcattcgacgagagacg gaagatgcaaaaggagcaagaggggaatcggaagaagaatgagacagcaatcagacagctggtggagtacggactcaccctgaaccctaccggcctggatatcgtcaagtttaagctgtggtccgagcagaatggtaagtgcgcatactctctgcagcccatagagattgaaaga cttctggagcctggatacactgaggtggatcacgtcatcccctattctcgctctcttgatgactcatacacaaacaaggtgctggtactcactaaagaaaatagggagaagggtaataggacgcctgcagagtacctggggttgggaagcgagcggtggcagcaatttgaaaccttcg tgctcaccaacaaacaattcagcaagaagaaaagggacaggctgctgcgactgcactatgacgagaatgaggagaacgaattcaagaacagaaacctgaacgacacacgctacatatctcggtttctggccaatttcatccgtgagcacctgaagtttgcagattccgacgataag cagaaggtctatacagtgaatggaaggatcactgctcacctgaggtctcggtggaacttcaataagaatcgtgaggagagtaacctgcaccacgctgtagacgccgctatagttgcatgtactactcccctctgacattgctcgtgtgacggcgtttttaccagagaagaagagcagaaca aggagctttcaaagaaaaccgcaccctcaattcccccaaccctggccacatttccgccgacgagcttcaagccaggctgagcaagaaccctaaagaatccattaaggcactgaacctcggcaactacgacaatgaaaagcttgagtcactgcaacctgtcttcgtgagcagaatgccca aaagatcaattaccggtgctgcccaccaggagactcttaggcggtacattggaatagatgagagatctggcaagattcaaactgttgttaagaagaagcttagtgaaattcagctcgacaagaccggacactttccatgtatggaaaggagtctgaccctcgcacttacgaggcaat acggcagcgcctgctggagcacaataacgatcctaagaaagcattccaggaacccctctacaagccaaagaagaatggggagcttgggccgattattagaactattaagatcaattgaacacccaccaacaggtgatccccactgaatgacggggaagaccgtagcatataatttctaata ttgttagggttgatgtgttcgaaaaggacggtaagtactactgcgttcccatatacactattgacatgatgaagggcattctgcccaataaagcaatagaacctaataagcctatagcgaatggaaaagagatgactgaagattacaccttttcggtttagcctgtatcccaacgacctg attaggatagagttcccacgcgagaagaccattaaaacagccgttggagaggagattaagatcaaagacttgtttgcatactaccagacgatagatagcagcaacggcggcttgagcctttgtgtcccacgacaataatttctccctgagaagtattggcagccgcactctgaagcgct ttgaaaagtatcaggtggacgtacttggaaatatttataaggttcggggagagaaaagggtgggcgtcgcttccagttcacactccaaggcaggcgagacaattcggccccctgggggtgactctggtggttctggaggatctggtggttctactaatctgtcagatattattgaaaag gagaccggtaagcaactggttatccaggaatccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccctgaatacaagcctgggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttctcccaagaagaagaggaaa gtctaa
apobec1-linker-hnspycas9(D10A)-linker-ugi-linker-ugi-linker-sv40nls	atgagctcagagactggcccagtggctgtggaccccacattgagacggcggatcgagccccatgagtttgaggtattcttcgatccgagagagctccgcaaggagacctgcctgctttacgaaattaattgggggggccggcactccatttggcgacatacatcacagaacactaaca agcacgtcgaagtcaacttcatcgagaagttcacgacagaaagatatttctgtccgaacacaaggtgcagcattacctggtttctcagctggagcccatgcggcgaatgtagtagggccatcactgaattcctgtcaaggtatccccacgtcactctgtttatttacatcgcaaggctgt accaccacgctgaccccgcaatcgacaaggcctgcgggatttgatctcttcaggtgtgactatccaaattatgactgagcaggagtcaggatactgctggaggaaacttttgtgaattatagcccgagtaatgaagcccactggcctaggtatccccatctgtgggtacgactgtacgttc ttgaactgtactgcatcatactgggcctgcctccttgtctcaacattctgagaaggaagcagccacagctgacattctttaccatcgctcttcagtcttgtcattaccagcgactgcccccacacattctctgggccaccgggttgaaatctggtggttcttctggtggttctagcggcagcg agactcccgggacctcagagtccgccacacccgaaagttctggtggttcttctggtggttctgataaaagattatctattggtttagCcatcggcactaattccgttggatgggctgtcataaccgatgaatacaaagtaccttcaaagaaatttaaggtgttggggaacacagaccgtc attcgattaaaaagaatcttatcggtgccctcctattcgatagtggcgaaacggcagaggcgactcgcctgaaacgaaccgctcggagaaggtatacacgtcgcaagaaccgaatatgttacttacaagaaatttttagcaatgagatggccaaagttgacgattctttcttccgct ttggaagagtccttccttgtcgaagaggacaagaaacatgaacggcacccatctttggaaacatagtagatgaggtggcatatcatgaaaagtacccaacgatttatcacctcagaaaaagctagttgactcaactgataaagcggacctgaggttaatctacttggctcttgccc atatgataaagttccgtgggcactttctcattgagggtgatctaaatccggacaactcggatgtcgacaaactgttcatccagttagtacaaacctataatcagttgtttgaagagaaccctataaatgcaagtggcgtggatgcgaaggctattcttagcgcccgcctctctaaatcccg acggctagaaaacctgatcgcacaattacccggagagaagaaaaatgggttgttcggtaacctttatagcgctctcactaggcctgacaccaaattttaagtcgaacttcgacttagctgaagatgccaaattgcagcttagtaaggacacgtacgatgacgatctcgacaatctactg gcacaaattggagatcagtatgcggacttattttggctgccaaaacctgagcgatcaatcctcctatctgacatactgagagttaatactgagattaccaaggcgccgttatccgcttcaatgatcaaaaggtacgatgaacatcaccaagacttgacacttctcaaggccctagtc cgtcagcaactgcctgagaaatataaggaaatattcttttgatcagtcgaaaaacgggtacgcaggttatattgacggcggagcgagtcaagaggaattctacaagtttatcaaacccatattagagaagatggatgggacggaagagttgcttgtaaaactcaatcgcgaagatcta ctgcgaaagcagcggacttttcgacaacggtagcattccacatcaaatccacttaggcgaattgcatgctatacttagaaggcaggaggatttttatccgttcctcaaagacaatcgtgaaaagattgagaaaatcctaaccttttcgcatacctactatgtgggaccccgccggcgagg gaactctcggttcgcatggatgacaagaaagtccgaagaaacgattactccatggaatttttgaggaagttgtcgataaaggtgcgtcagctcaatcgttcatcgagaggatgaccaacttttgacaagaatttaccgaacgaaaaagtattgccttagcagtatttcacttatcgagtatt tcacagtgtacaatgaactcacgaaagttaagtatgtcactgagggcatgcgtaaaccccgccttttctaagcggagaacagaagaaagcaatagtagatctgttattcaagaccaaccgcaaagtgacagttaagcaattgaaagaggactactttaagaaaattgaatgcttcgatt ctgtcgagatctccgggtagaagatcgatttaatgcgtcacttggtacgtatcatgacctcctaaagataatttaaagataaggacttcctggataacgaagagaatgaagatatcttagaagatatagtgttgactcttaccctcttttgaagatcgggaaatgattgaggaaagacta aaaacatacgctcacctgttcgacgataaggttatgaaacagttaaagaggcgtcgctatacgggctggggacgattgtcgcggaaacttatcaacgggataagagacaagcaaagtggtaaaactattctcgattttctaaagagcgacggcttcgccaatagggaactttatgcag

Supplementary Table 4. DNA sequences of cas9 (fusion) genes used in this study. *(continuation)*

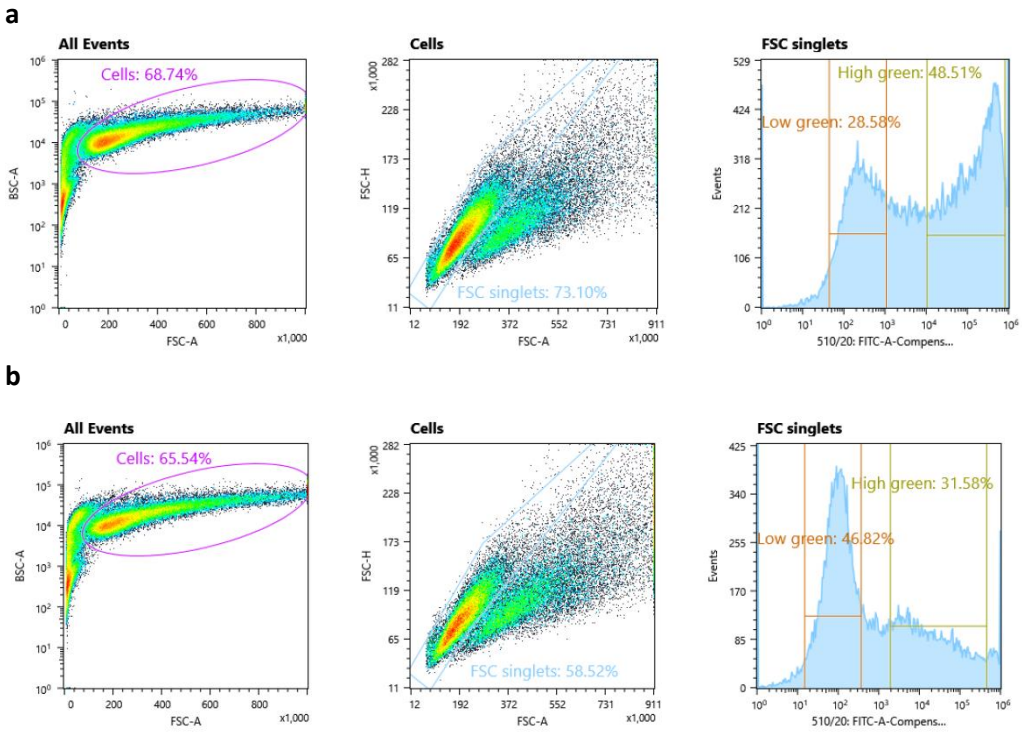
	ctgatccatgatgactctttaaccttcaaagaggatatacaaaaggcacaggtttccggacaaggggactcattgcacgaacatattgcgaatcttgctggttcgccagccatcaaaaagggcatactccagacagtcaaagtagtggatgagctagttaaggtcatgggacgtcaca aaccggaaaacattgtaatcgagatggcacgcgaaaatcaaacgactcagaaggggcaaaaaaacagtcgagagcggatgaagagaatagaagagggtattaaagaactgggcagccagatcttaaaggagcatcctgtggaaaatacccaattgcagaacgagaaacttta cctctattacctacaaaatggaaggggacatgtatgttgatcaggaactggacataaaccgtttatctgattacgacgtcgatcacattgtaccccaatccttttttgaaggacgattcaatcgacaataaagtgcttacacgctcggataagaaccgagggaaaagtgacaatgttccaa gcgaggaagtcgtaaagaaaatgaagaactattggcggcagctcctaaatgcgaaactgataacgcaaagaaagttcgataacttaactaaagctgagaggggtggcttgtctgaaccttgacaaggccggatttattaaacgtcagctcgtggaaacccgccaaatcacaaagcat gttgcacagatactagattcccgaatgaatacgaaatacgacgagaacgataagctgattcgggaagtcaaagtaatcactttaaagtcaaaattggtgtcggacttcagaaaggattttcaattctataaagttagggagataaataactaccaccatgcgcacgacgcttatctta atgccgtcgtagggaccgcactcattaagaaatacccgaagctagaaagtgagtttgtgtatggtgattacaaagtttatgacgtccgtaagatgatcgcgaaaagcgaacaggagataggcaaggctacagccaaatacttctttttattctaacattatgaatttctttaagacggaa atcactctggcaaacggagagatacgcaaacgacctttaattgaaaccaatgggggagacaggtgaaatcgtatgggataagggccgggacttcgcgacggtgagaaaagttttgtccatgccccaagtcaacatagtaaagaaaactgaggtgcagaccggagggttttcaaagg aatcgattcttccaaaaaggaatagtgataagctcatcgctcgtaaaaaggactgggacccgaaaaagtacggtggcttcgatagccctacagttgcctattctgtcctagtagtggcaaaagttgagaagggaaaatccaagaaactgaagtcagtcaaagaattattggggataa cgattatggagcgctcgtctttttgaaaagaaccccatcgacttccttgaggcgaaaggttacaaggaagtaaaaaaggatctcataattaaactaccaaagtatagtctgtttgagttagaaaatggccgaaaacggatgttggctagcgccggagagcttcaaaaggggaacgaac tcgcactaccgtctaaatacgtgaatttcctgtatttagcgtcccattacgagaagttgaaaggttcacctgaagataacgaacagaagcaacttttttgttgagcagcacaaacattatctcgacgaaatcatagagcaaatttcggaattcagtaagagagtcatcctagctgatgcca atctggacaaagtattaagcgcatacaacaagcacagggataaacccatacgtgagcaggcggaaaatattatccatttgtttactcttaccaacctcggcgctccagccgcattcaagtattttgacacaacgatagatcgcaaacgatacacttctaccaaggaggtgctagacgc gacactgattcaccaatccatcacgggattatatgaaactcggatagatttgtcacagcttggggtgactctggtggttctggaggatctggtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggtttatccaggaatccatcctcatgctcccagagga ggtggaagaagtcattgggaacaagccggaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacgagaatgtcatgcttctgactagcgacgcccctgaatacaagcctttggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtg gttctggaggatctggtggttctactaatctgtcagatattattgaaaaggagaccggtaagcaactggtttatccaggaatccatcctcatgctcccagaggaggtggaagaagtcattgggaacaagccggaaagcgatatactcgtgcacaccgcctacgacgagagcaccgacg agaatgtcatgcttctgactagcgacgcccctgaatacaagccttgggctctggtcatacaggatagcaacggtgagaacaagattaagatgctctctggtggttctcccaagaagaaggaggaaagtctaa

Pink, uppercase and underlined: *nucleotide substitution*

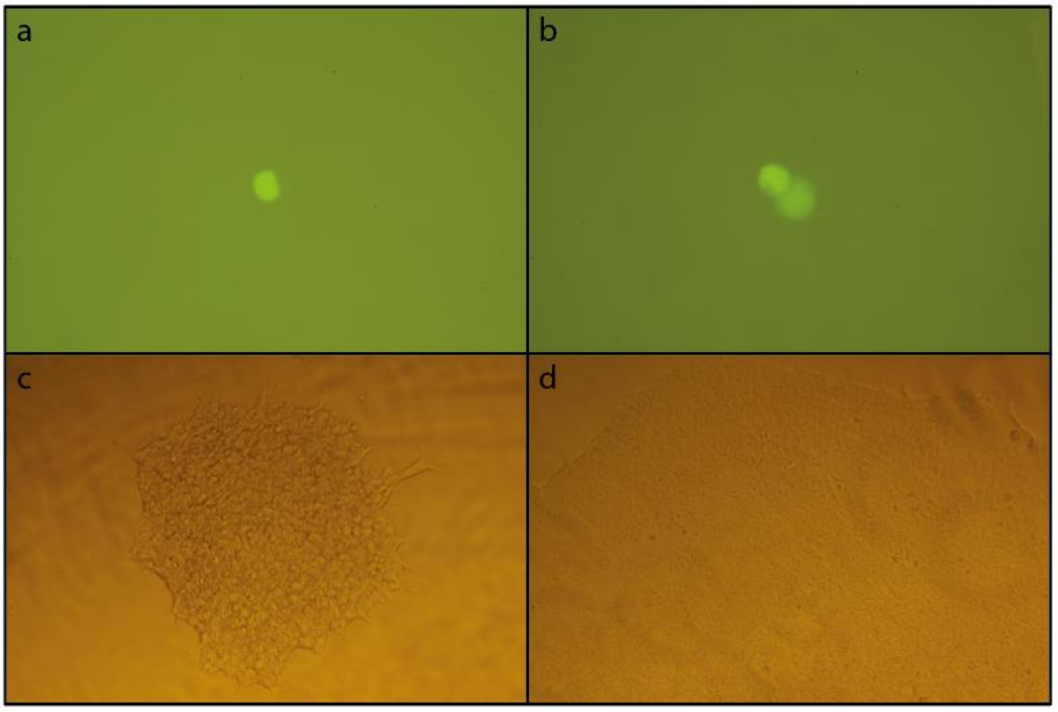
Supplementary Table 5. Amino acid sequences of Cas9 (fusion) proteins used in this study.

Protein/ Fusion protein	Sequence (5' -> 3')
hThermoCas9-linker-SV40NLS	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSIL SSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQ KTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKG NRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDP QFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTN QVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGE KRVGVASSSHSKAGETIRPLSGGSPKKKRKV*
hSpyCas9-linker-SV40NLS	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRP PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLG GDSGGSPKKKRKV*
APOBEC1-linker-hnThermoCas9(D8A)-linker-UGI-linker-UGI-linker-SV40NLS	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYC WRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSKYKIGLAIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGES LALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIF AKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYG KGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELAR ELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDEN EENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPV FVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILP NKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPLGGDSGGSGGSGGSTNLSDIIEKETGKK QLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVI QDSNGENKIKMLSGGSPKKKRKV*
APOBEC1-linker-hnSpyCas9(D10A)-linker-UGI-linker-UGI-linker-SV40NLS	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYC WRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI KKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRR YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGN KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKK RKV*

Pink and underlined: *amino acid substitution*



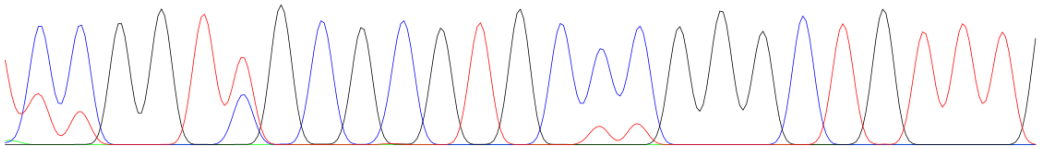
Sup. Fig. 1 Flow cytometry assisted cell-sorting (FACS) representative gating. HEK293T cells transfected with the GFP-expressing constructs (**a**) phTCas9_spT3 and (**b**) phSCas9_spS6. The gate 'high green' indicates EGFP fluorescent cells, whereas 'low green' indicates non-transfected cells causing background fluorescent signal.



Sup. Fig. 2 Representative examples of HEK293T cell expansion after FACS. (**a**) Fluorescence microscopy image of a single cell 2 days after FACS. (**b**) Fluorescence microscopy image of a dividing single cell 5 days after FACS. (**c, d**) Microscopy image of clonally propagated cells 13 days after FACS.

a

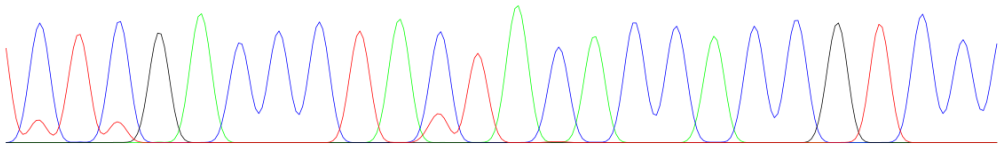
ThermoBE4: Spacer T3 *(Protospacer: 5'-GGTCGCGCGTGCCCGGGCTGTTT-3')*



	C	C	G	G	T	C	G	C	G	C	G	T	G	C	C	C	G	G	G	C	T	G	T	T	T
T	30	22	0	0	100	64	0	0	0	1	0	100	0	0	16	15	0	0	0	0	100	0	100	100	100
G	0	0	100	100	0	0	100	0	100	0	100	0	100	0	0	0	100	100	100	0	0	100	0	0	0
C	70	78	0	0	0	36	0	100	0	99	0	0	0	100	84	85	0	0	0	100	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

b

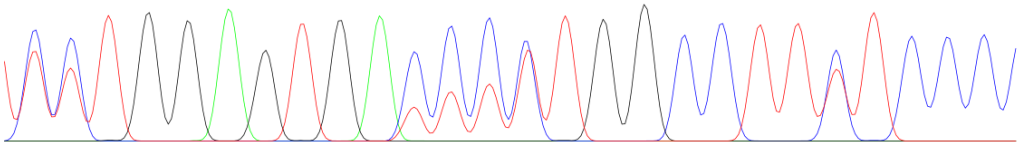
ThermoBE4: Spacer T5 *(Protospacer: 5'-TCGACCCTACTACACCACCGTCC-3')*



	C	T	C	G	A	C	C	C	T	A	C	T	A	C	A	C	C	A	C	C	G	T	C	C
T	16	99	15	0	0	0	0	0	100	0	21	100	0	1	0	0	0	0	0	0	0	100	0	0
G	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
C	84	1	85	0	0	100	100	100	0	0	79	0	0	99	0	100	100	0	100	100	0	0	100	100
A	0	0	0	0	100	0	0	0	0	100	0	0	100	0	100	0	0	100	0	0	0	0	0	0

c

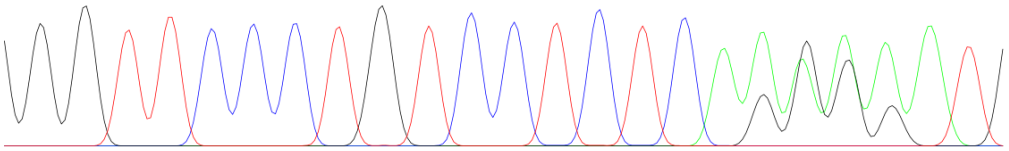
ThermoBE4: Spacer T8 *(Protospacer: 5'-GGAGTGACCCCTGGCCTTCTCCC-3')*



	C	C	T	G	G	A	G	T	G	A	C	C	C	C	T	G	G	C	C	T	T	C	T	C	C	C
T	45	42	99	0	0	0	0	99	0	0	28	30	32	48	99	0	0	0	0	100	100	44	100	0	0	0
G	0	0	1	100	100	0	100	1	100	0	0	0	0	0	1	100	100	0	0	0	0	0	0	0	0	0
C	55	58	0	0	0	0	0	0	0	0	72	70	68	52	0	0	0	100	100	0	0	56	0	100	100	100
A	0	0	0	0	0	100	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

d

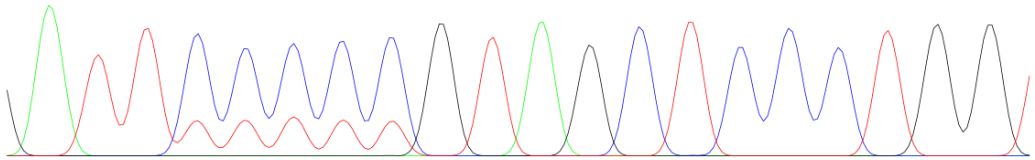
ThermoBE4: Spacer T9 *(Protospacer: 5'-ATCCCCTGAGAGGACAGGGAACC-3'; reverse reaction)*



	G	G	T	T	C	C	C	T	G	T	C	C	T	C	T	C	A	G	G	G	G	A	T
T	0	0	100	100	0	0	0	100	0	100	0	0	100	0	100	0	0	0	0	0	0	0	100
G	100	100	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	31	55	44	28	0	0
C	0	0	0	0	100	100	100	0	0	0	100	100	0	100	0	100	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	69	45	56	72	100	0

e

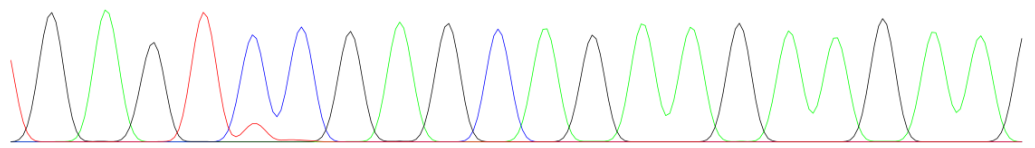
Spacer S2 *(Protospacer: 5'-ATTCCCCCGTAGCTCCCTGG-3')*



	A	T	T	C	C	C	C	C	G	T	A	G	C	T	C	C	C	T	G	G
T	0	100	100	22	25	26	24	23	0	100	0	0	0	100	0	0	0	100	0	0
G	0	0	0	0	0	0	0	0	100	0	0	100	0	0	0	0	0	0	100	100
C	0	0	0	78	75	74	76	77	0	0	0	0	100	0	100	100	100	0	0	0
A	100	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0

f

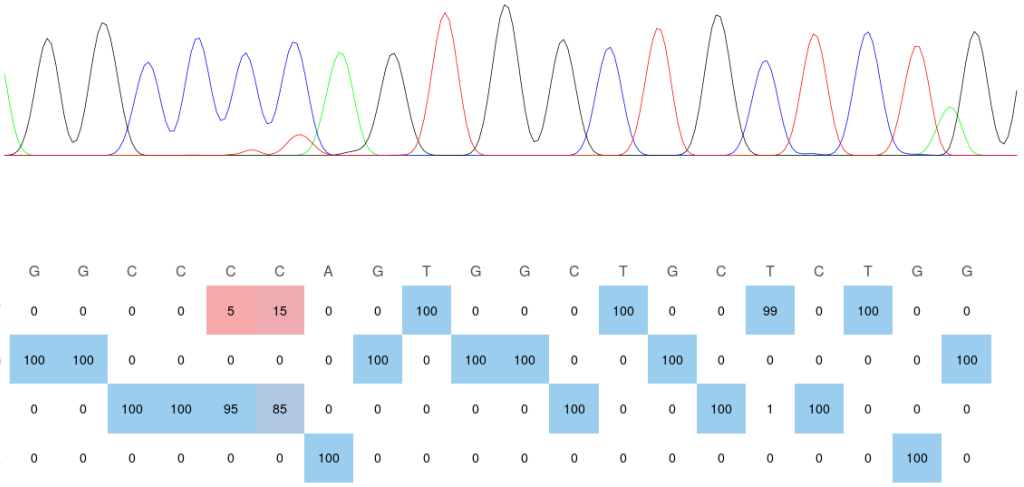
Spacer S4 *(Protospacer: 5'-GAGTCCGAGCAGAAGAAGAA-3')*



	G	A	G	T	C	C	G	A	G	C	A	G	A	A	G	A	A	G	A	A
T	0	0	0	100	15	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	100	0	100	0	0	0	100	1	100	0	0	100	0	0	100	0	0	100	0	0
C	0	0	0	0	85	98	0	0	0	100	0	0	0	0	0	0	0	0	0	0
A	0	100	0	0	0	0	0	99	0	0	100	0	100	100	0	100	100	0	100	100

g

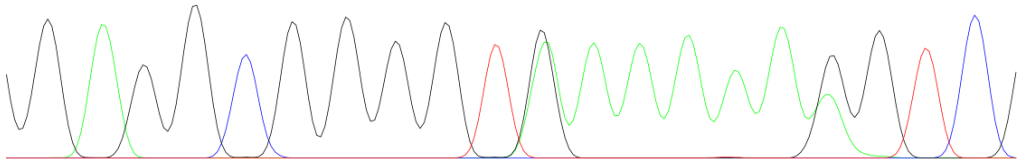
Spacer S5 *(Protospacer: 5'-GGCCCCAGTGGCTGCTCTGG-3')*



	G	G	C	C	C	C	A	G	T	G	G	C	T	G	C	T	C	T	G	G
T	0	0	0	0	5	15	0	0	100	0	0	0	100	0	0	99	0	100	0	0
G	100	100	0	0	0	0	0	100	0	100	100	0	0	100	0	0	0	0	0	100
C	0	0	100	100	95	85	0	0	0	0	0	100	0	0	100	1	100	0	0	0
A	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100	0

h

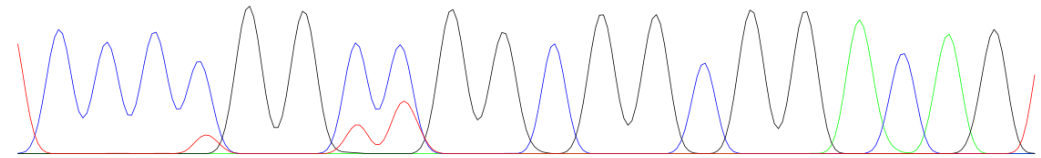
Spacer S7 *(Protospacer: 5'-GACCCCCTCCACCCCGCCTC-3'; reverse reaction)*



	G	A	G	G	C	G	G	G	G	T	G	G	A	G	G	G	G	G	T	C
T	0	0	0	0	0	0	0	0	0	99	0	0	0	0	1	0	0	0	100	0
G	100	0	100	100	1	100	100	100	100	1	52	0	0	0	0	0	62	100	0	0
C	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100
A	0	100	0	0	0	0	0	0	0	0	48	100	100	100	99	100	38	0	0	0

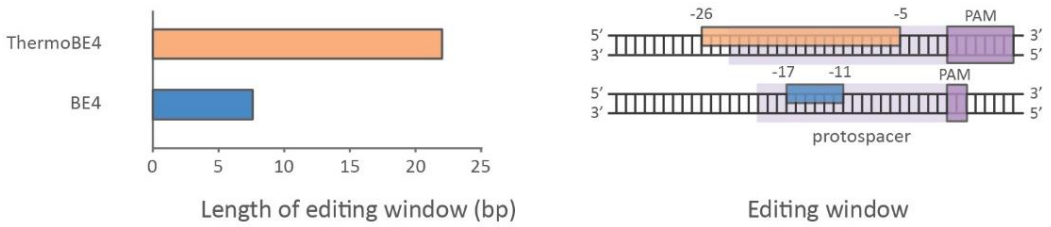
i

Spacer S9 *(Protospacer: 5'-CCCCGGCCGGCGGCGGACAG-3')*



	C	C	C	C	G	G	C	C	G	G	C	G	G	C	G	G	A	C	A	G
T	0	0	0	17	0	0	21	33	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	100	100	0	0	100	100	0	100	100	0	100	100	0	0	0	100
C	100	100	100	83	0	0	79	67	0	0	100	0	0	100	0	0	0	100	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	100	0

Sup. Fig. 3 Percentages of nucleotide conversions in the targeted protospacers, induced by (**a-d**) ThermoBE4 and (**e-i**) BE4. The percentages resulted from Sanger sequencing analysis of the tested populations, employing the on-line tool "EditR".



Sup. Fig. 4 Comparison of the length (left) and spectrum (right) of the editing window for ThermoBE4 and BE4. The dark purple boxes indicate the PAM region, while the light purple areas depict the protospacer region. Orange and blue boxes represent the overall activity window of the base-editors.

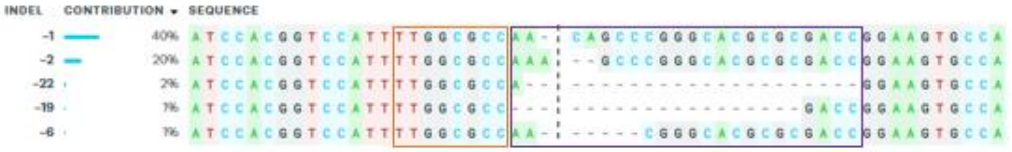
ThermoCas9-mediated genome editing

Spacer T3 (***Protospacer:*** *5′-GGTCGCGCGTGCCCGGGCTGTTT-3′*; ***PAM:*** *5′-N_4CCAA-3′*)

Clone 1

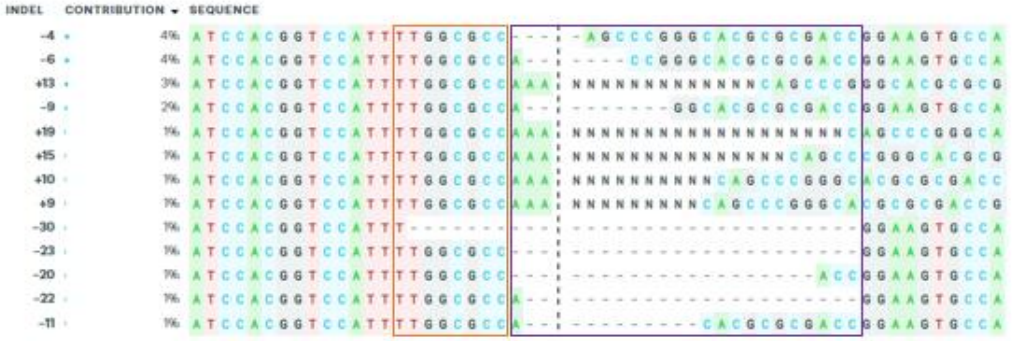
Reaction 1 (reverse): 0% indels

Reaction 2 (reverse): 64% indels

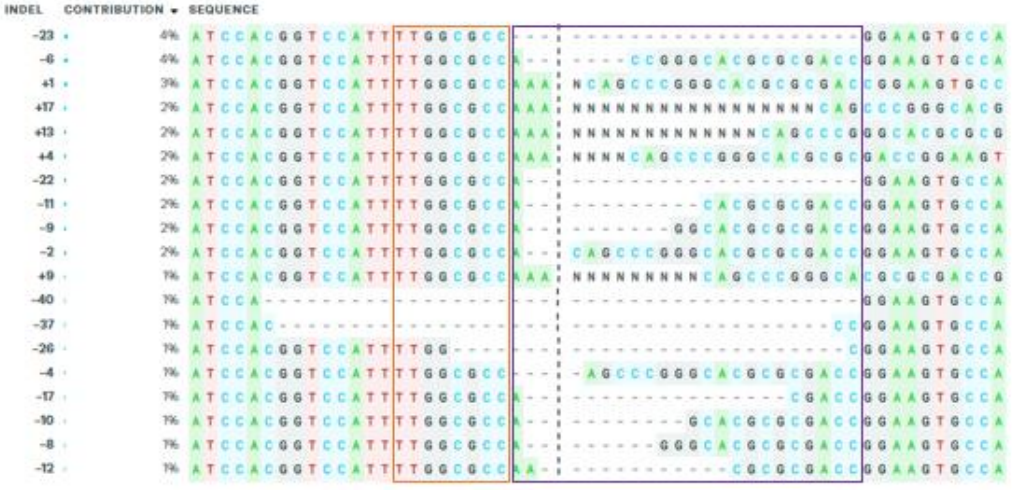


Population

Reaction 1 (reverse): 22% indels



Reaction 2 (reverse): 34% indels



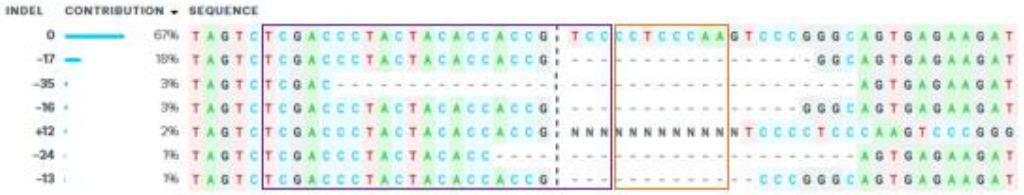
Spacer T5 (***Protospacer:*** *5'-TCGACCCTACTACACCACCGTCC-3'*; ***PAM:*** *5'-N_4CCAA-3'*)

Clone 1-3, 5-8, 10-14

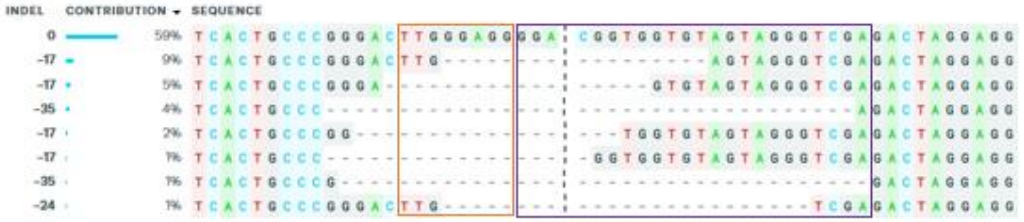
Reactions 1 (forward), 2 (reverse) & 3 (reverse): 0% indels

Clone 4

Reaction 1 (forward): 26% indels



Reaction 2 (reverse): 23% indels



Reaction 3 (reverse): 24% indels

Clone 9

Reaction 1 (reverse): 0% indels

Reaction 2 (reverse): 7% indels

Population

Reactions 1 (forward), 2 (reverse), 3 (forward) & 4 (reverse): 0% indels

Spacer T8 (***Protospacer:*** *5′-GGAGTGACCCCTGGCCTTCTCCC-3′*; ***PAM:*** *5′-N₄CCAA-3′*)

Clone 1

Reaction 1 (forward): 89% indels

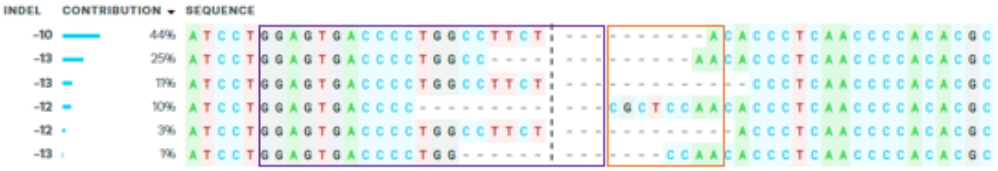
Reaction 1 (forward): 92% indels

Clone 2

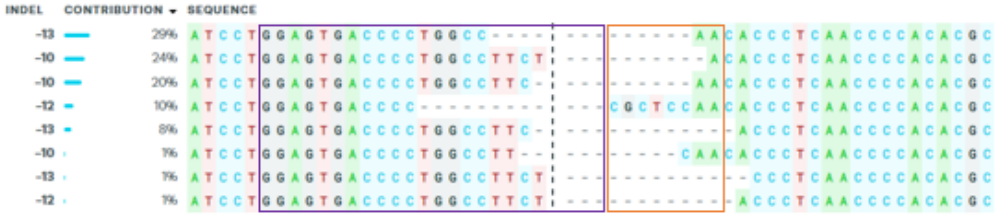
Reaction 1 & 2 (forward): 94% indels

Clone 3

Reaction 1 (forward): 94% indels



Reaction 2 (forward): 94% indels



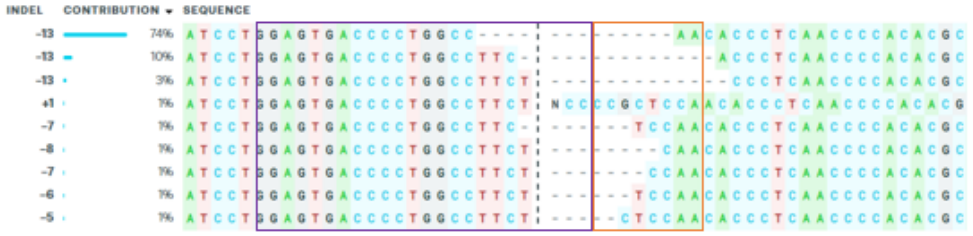
Clone 4, 6, 7, 9-11, 14

Reactions 1 (forward) & 2 (forward): 0% indels

Clone 5

Reaction 1 (forward): 63% indels

Reaction 2 (forward): 93% indels



Clone 8

Reaction 1 & 2 (forward): 1% indels

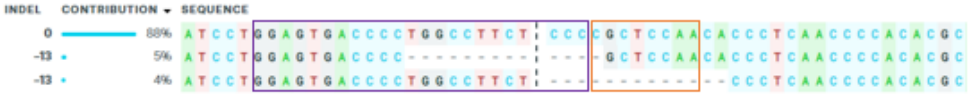
Clone 12

Reaction 1 (forward): 13% indels

Reaction 2 (forward): 12% indels

Clone 13

Reaction 1 & 2 (forward): 9% indels



Clone 15

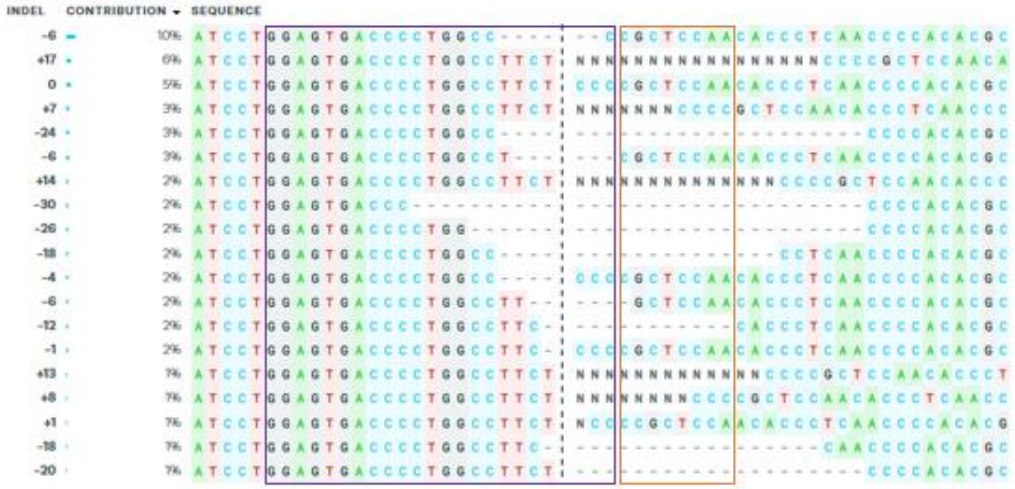
Reaction 1 (forward): 51% indels

Reaction 2 (forward): 50% indels

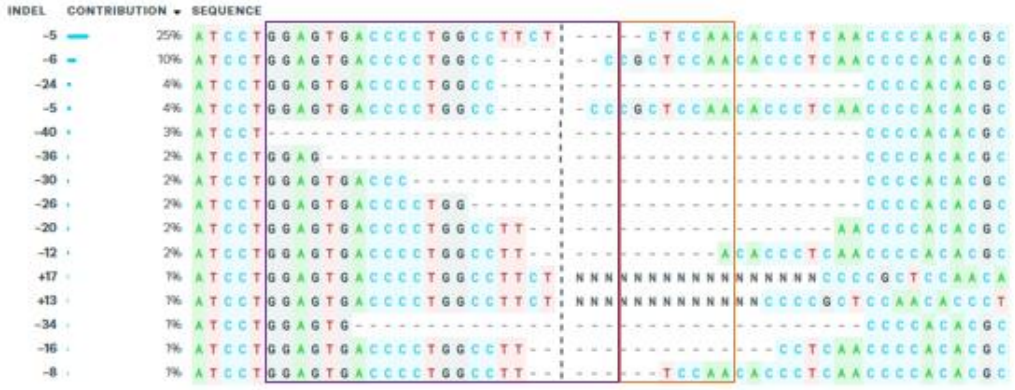
Clone 16

Reaction 1 (forward): 0% indels

Reaction 2 (forward): 46% indels



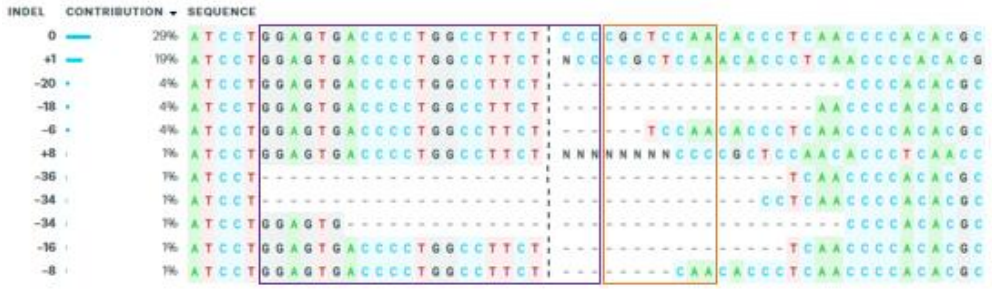
Reaction 3 (forward): 61% indels



Population

Reaction 1 (forward): 0% indels

Reaction 2 (forward): 37% indels



Spacer T9 ***(Protospacer:** 5'-ATCCCCTGAGAGGACAGGGAACC-3'; **PAM:** 5'-N$_4$CCAA-3')*

Clone 1 & 9

Reactions 1 (forward) & 2 (forward): 0% indels

Clone 2

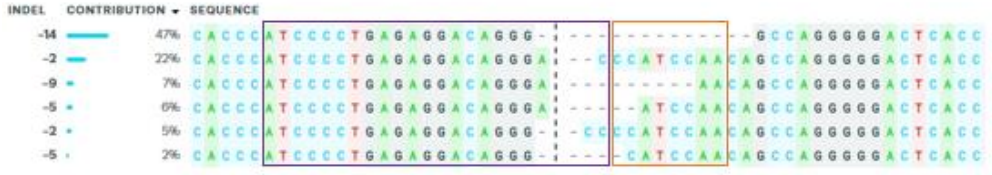
Reaction 1 (forward): 76% indels

Reaction 2 (forward): 75% indels

Clone 3

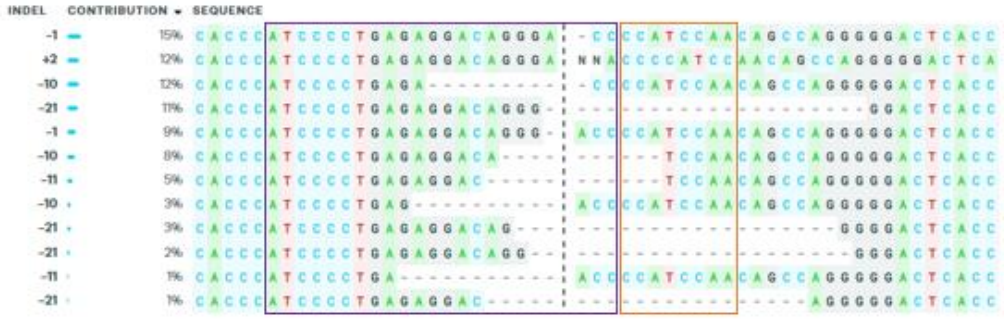
Reaction 1 (forward): 90% indels

Reaction 2 (forward): 89% indels



Clone 4

Reaction 1 (forward): 82% indels



Reaction 2 (forward): 83% indels

Clone 5

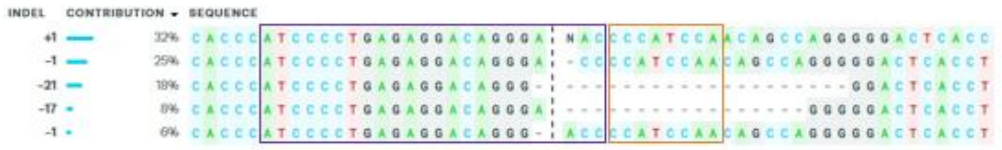
Reaction 1 (forward): 88% indels

Reaction 2 (forward): 88% indels

Clone 6

Reaction 1 (forward): 89% indels

Reaction 2 (forward): 89% indels



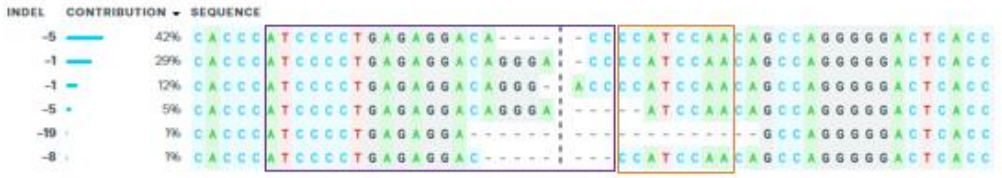
Clone 7

Reaction 1 (forward): 86% indels

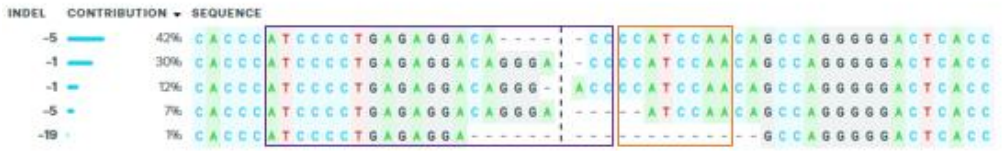
Reaction 2 (forward): 85% indels

Clone 8

Reaction 1 (forward): 90% indels

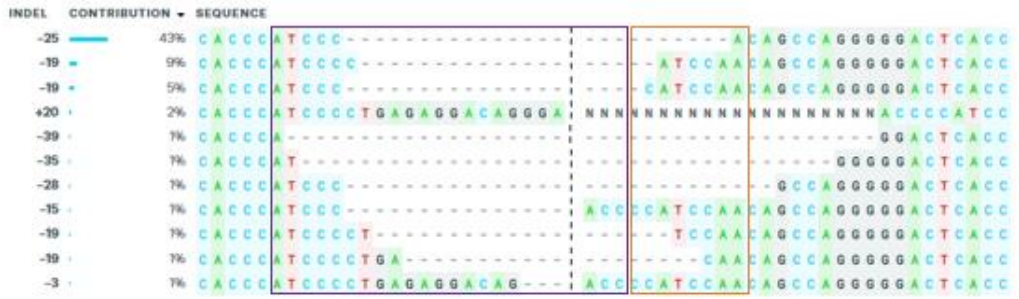


Reaction 2 (forward): 92% indels



Clone 10

Reaction 1 (forward): 66% indels

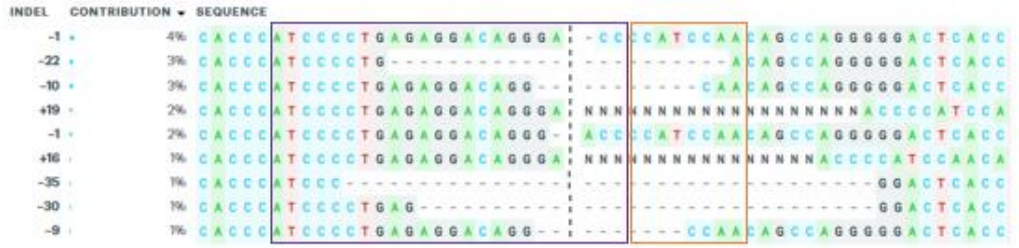


Reaction 2 (forward): 62% indels

Clone 11

Reaction 1 (forward): 20% indels

Reaction 2 (forward): 18% indels



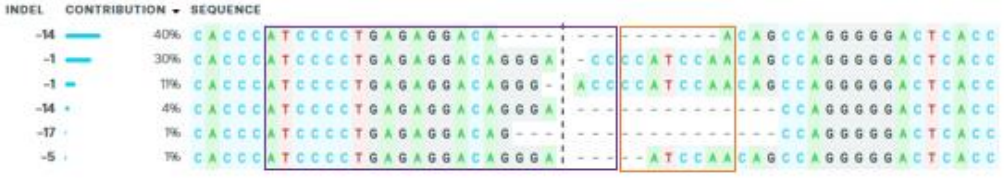
Clone 12

Reaction 1 (forward): 6% indels

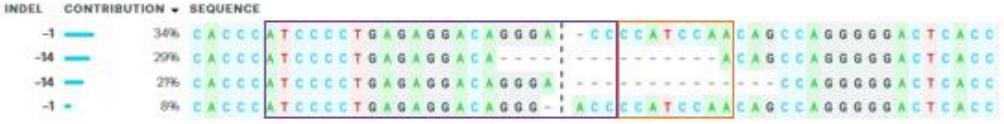
Reaction 2 (forward): 7% indels

Clone 13

Reaction 1 (forward): 87% indels

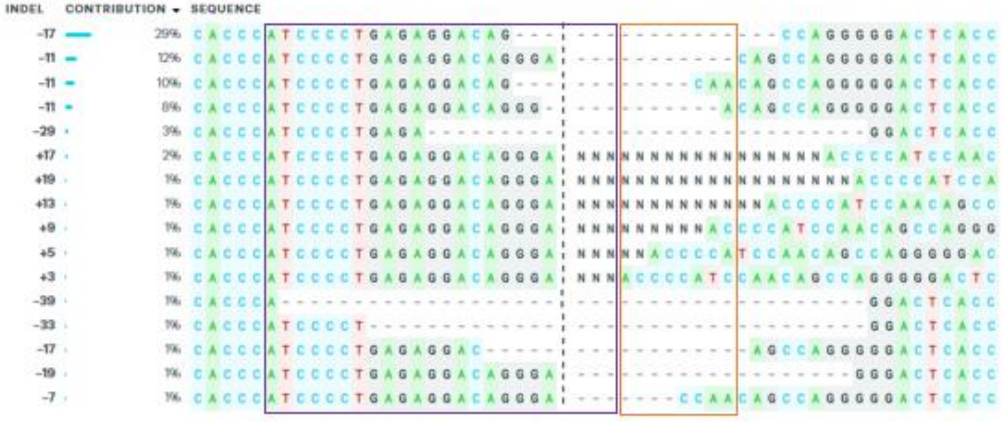


Reaction 2 (forward): 92% indels

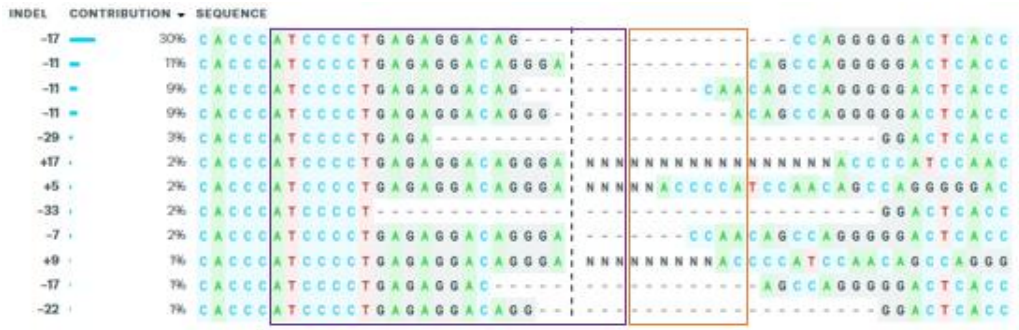


Clone 14

Reaction 1 (forward): 74% indels

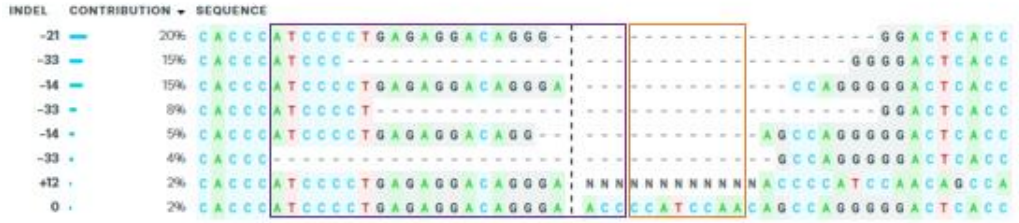


Reaction 2 (forward): 73% indels

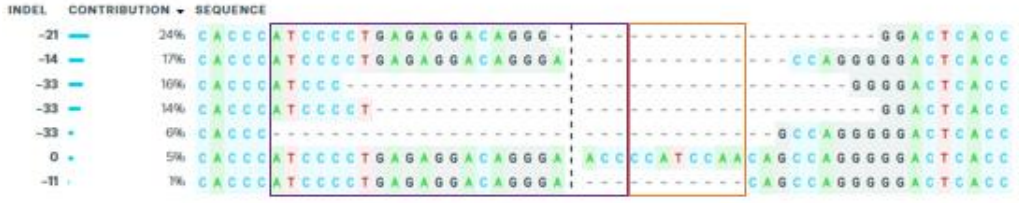


Clone 15

Reaction 1 (forward): 69% indels



Reaction 2 (forward): 78% indels



Clone 16

Reaction 1 (forward): 99% indels

Reaction 2 (forward): 99% indels

Population

Reaction 1 (forward): 86% indels

Reaction 2 (forward): 87% indels

Reaction 3 (forward): 88% indels

SpyCas9-mediated genome editing

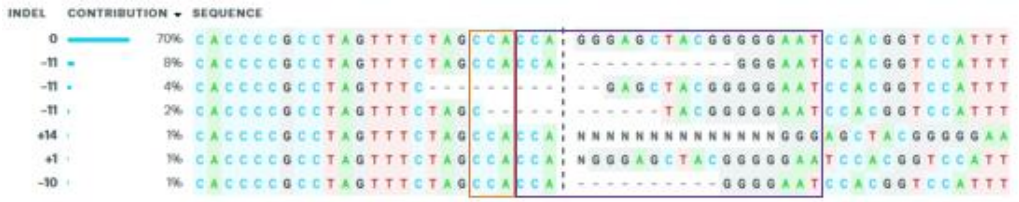
<u>Spacer S2</u> (***Protospacer:*** *5'-ATTCCCCCGTAGCTCCCTGG-3'*; ***PAM:*** *5'-NGG-3'*)

Clone 1

Reaction 1 (forward): 19% indels

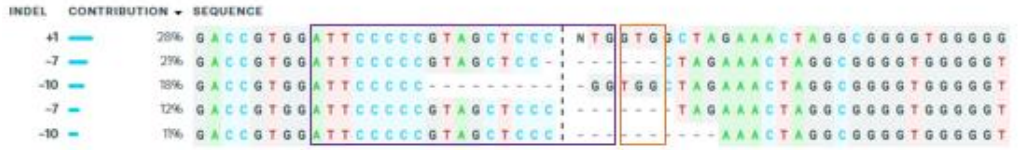
Reaction 2 (forward): 18% indels

Reaction 3 (reverse): 17% indels

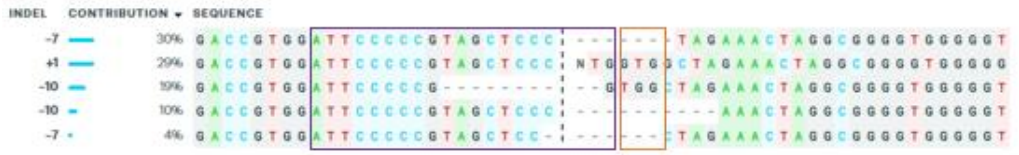


Clone 2

Reaction 1 (forward): 90% indels



Reaction 2 (forward): 92% indels

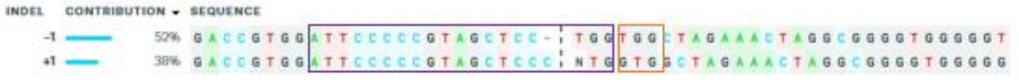


Reaction 3 (reverse): 78% indels

Clone 3

Reaction 1 (forward): 90% indels

Reaction 2 (forward): 90% indels



Reaction 3 (reverse): 84% indels

Clone 4

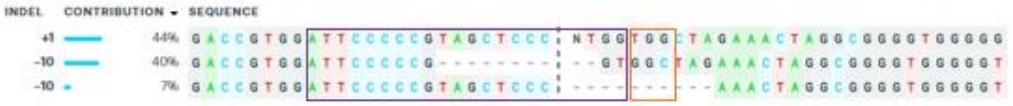
Reaction 1 (forward): 90% indels

Reaction 2 (forward): 92% indels

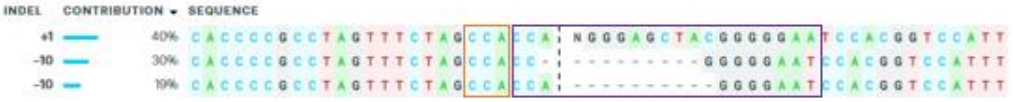
Reaction 3 (reverse): 88% indels

Clone 5

Reaction 1 (forward): 91% indels

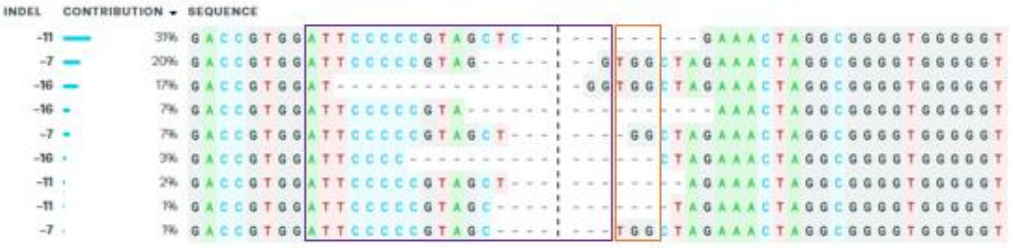


Reaction 2 (reverse): 89% indels

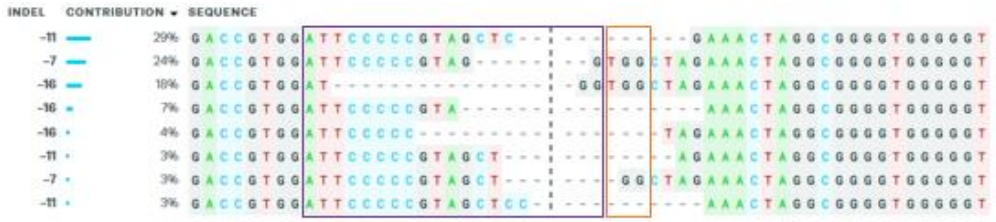


Clone 6

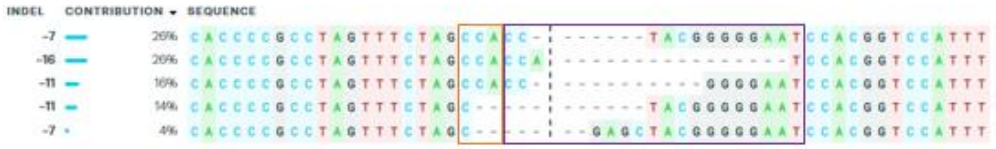
Reaction 1 (forward): 89% indels



Reaction 2 (forward): 91% indels



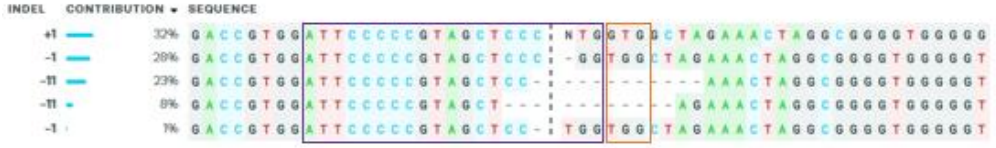
Reaction 3 (reverse): 86% indels



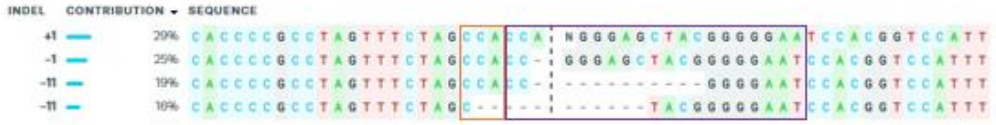
Clone 7

Reaction 1 (forward): 91% indels

Reaction 2 (forward): 92% indels



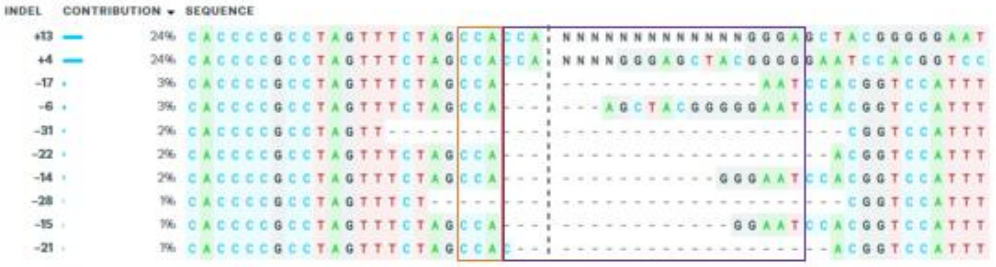
Reaction 3 (reverse): 89% indels



Clone 8

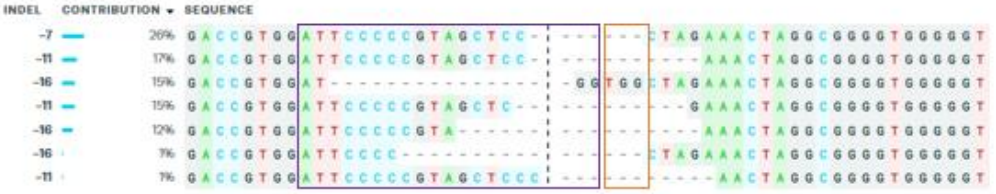
Reaction 1 (forward): 45% indels

Reaction 2 (reverse): 63% indels



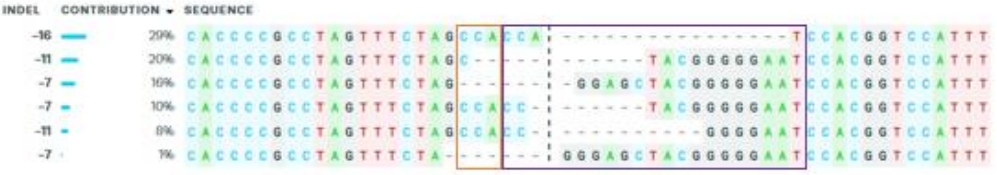
Clone 9

Reaction 1 (forward): 87% indels



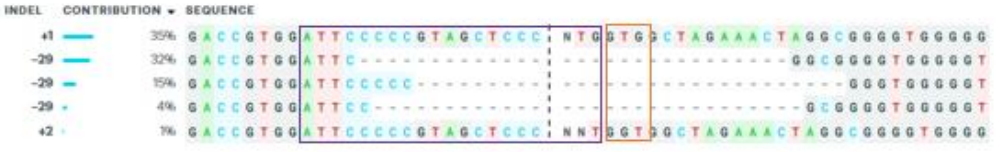
Reaction 2 (forward): 88% indels

Reaction 3 (reverse): 84% indels

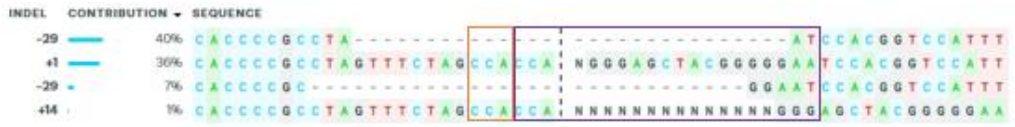


Clone 10

Reaction 1 (forward): 87% indels

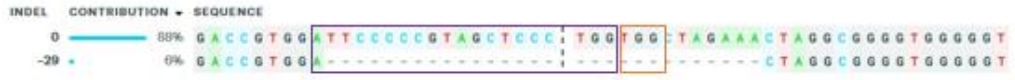


Reaction 2 (reverse): 84% indels



Clone 11

Reaction 1 (forward): 6% indels



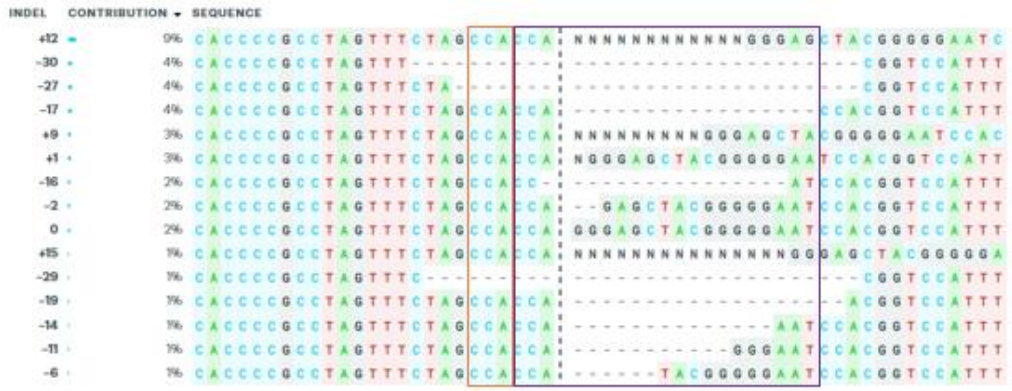
Reaction 2 (forward): 7% indels

Reaction 3 (reverse): 63% indels

Population

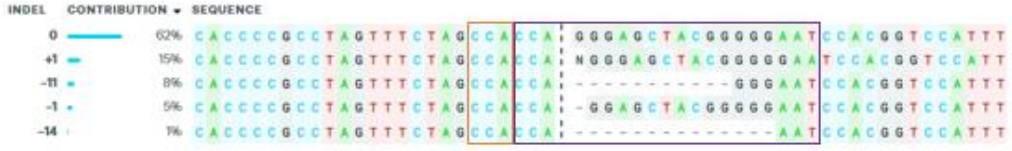
Reaction 1 (forward): 42% indels

Reaction 2 (reverse): 37% indel

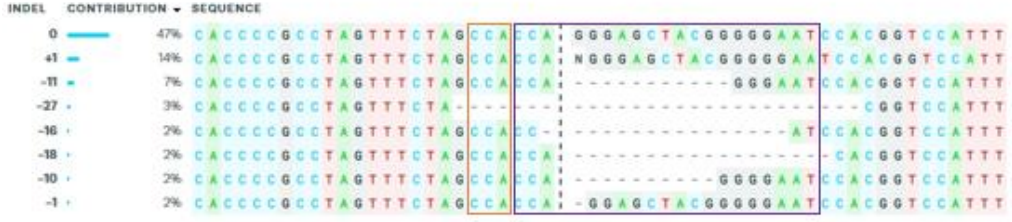


Reaction 3 (reverse): 37% indels

Reaction 4 (reverse): 29% indels



Reaction 5 (reverse): 32% indels



Spacer S6 (***Protospacer:*** *5′-CCCTGGCCCAGGTGAAGGTG-3′*; ***PAM:*** *5′-NGG-3′*)

Clone 1, 6 & 8

Reactions 1 (forward), 2 (reverse) & 3 (reverse): 0% indels

Clone 2

Reactions 1 (forward) & 2 (reverse): 0% indels

Reaction 3 (reverse): 72% indels

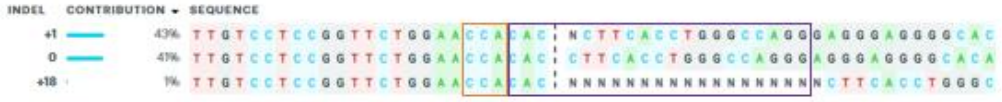
Clone 3

Reaction 1 (reverse): 25% indels

Reaction 2 (reverse): 24% indels

Clone 4

Reaction 1 (reverse): 44% indels

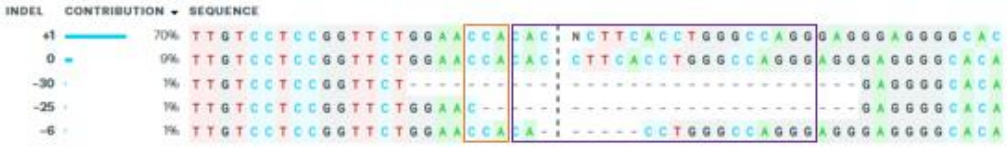


Reaction 2 (reverse): 51% indels

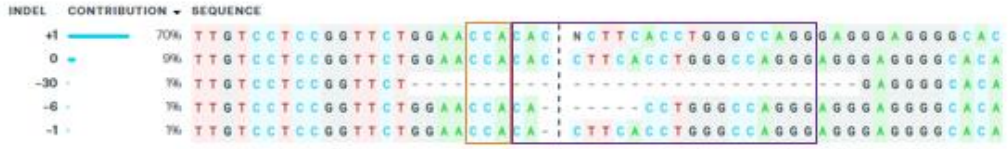
Clone 5

Reaction 1 (forward): 0% indels

Reaction 2 (reverse): 73% indels



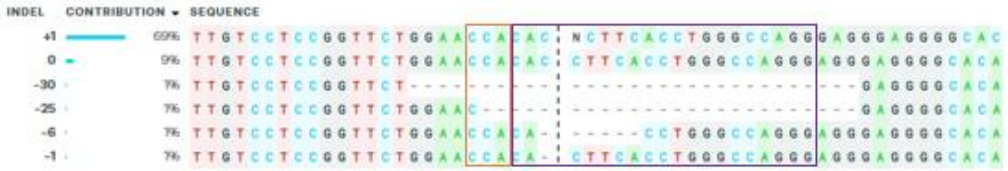
Reaction 2 (reverse): 73% indels



Clone 7

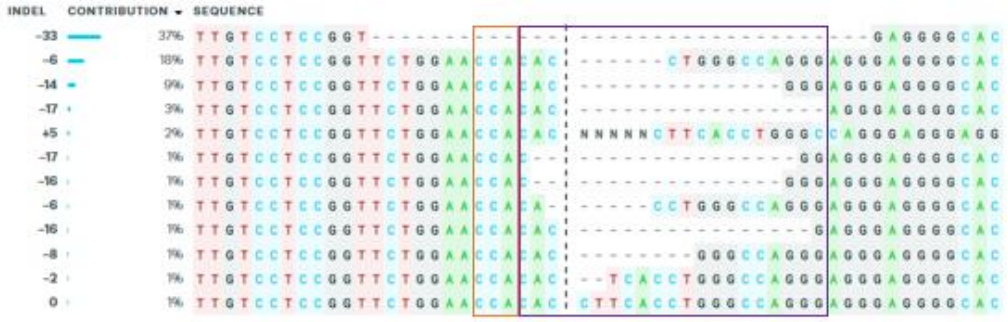
Reactions 1 (forward) & 2 (reverse): 0% indels

Reaction 3 (reverse): 73% indels

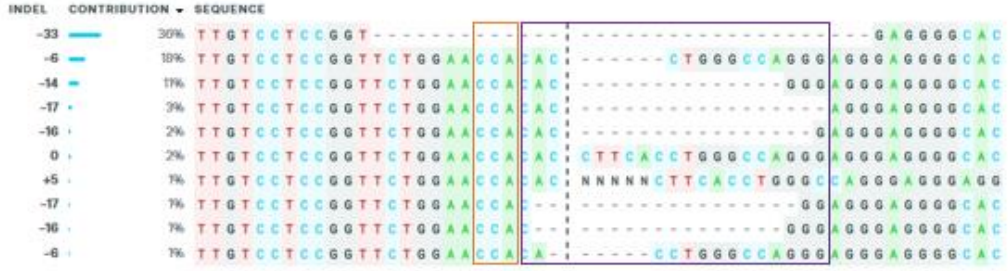


Clone 9

Reaction 1 (reverse): 75% indels



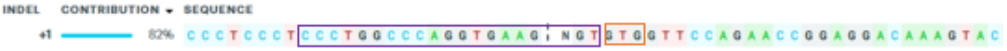
Reaction 2 (reverse): 74% indels



Clone 10

Reaction 1 (forward): 81% indels

Reaction 2 (forward): 82% indels



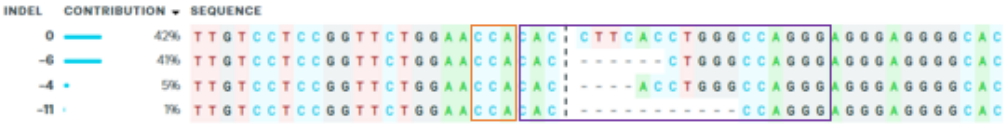
Reaction 3 (reverse): 99% indels

Reaction 4 (reverse): 99% indels

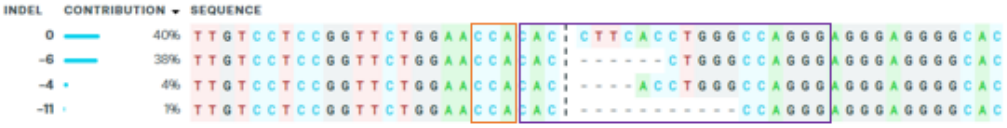
Clone 11

Reaction 1 (forward): 51% indels

Reaction 2 (reverse): 47% indels



Reaction 3 (reverse): 43% indels



Population

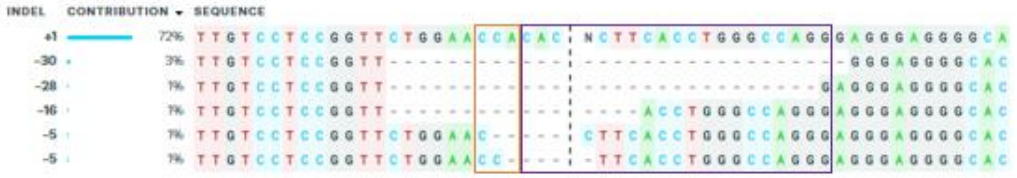
Reaction 1 (forward): 17% indels

Reaction 2 (forward): 17% indels

Reaction 3 (reverse): 5% indels

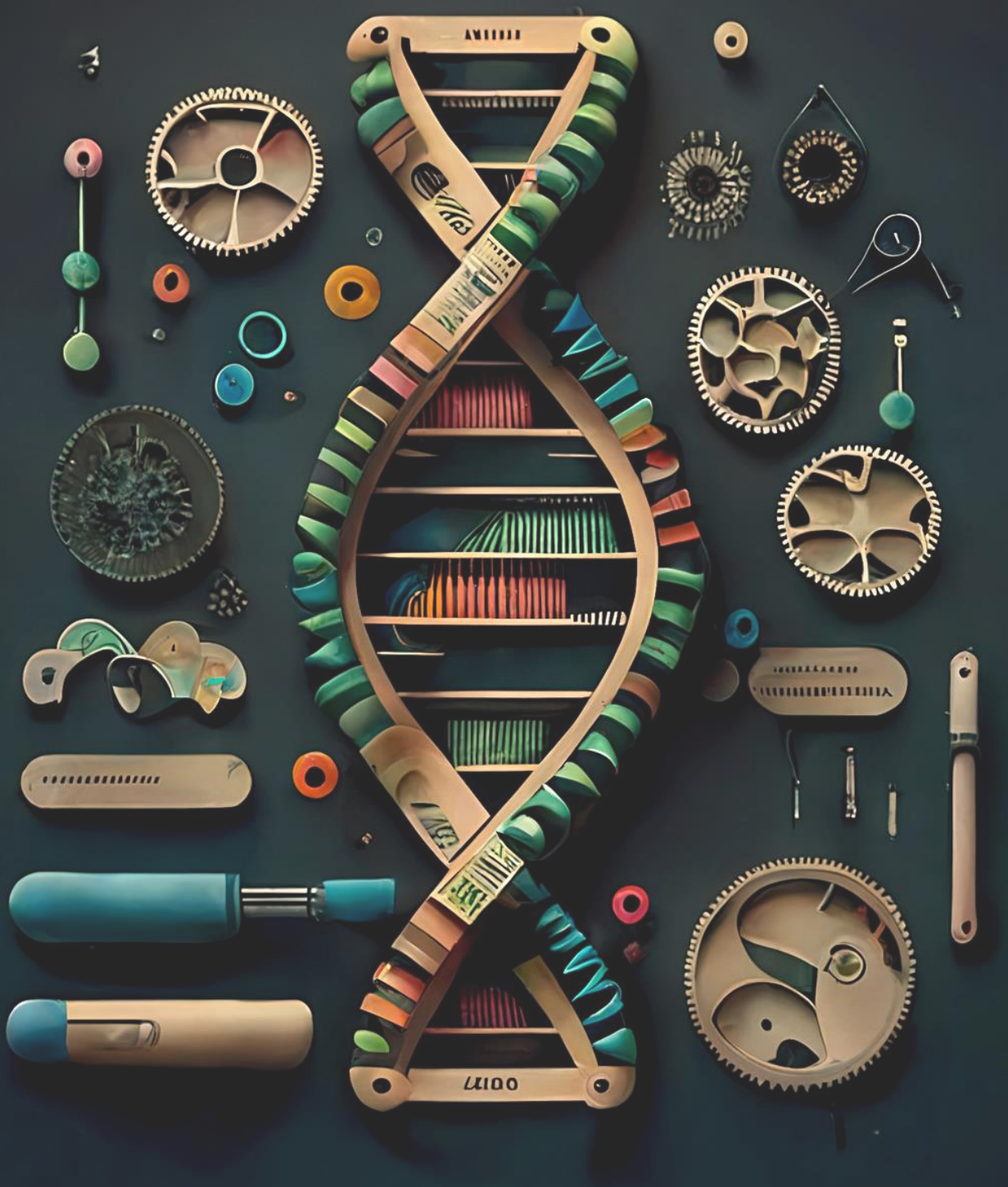
Reaction 4 (reverse): 5% indels

Reaction 5 (reverse): 79% indels



Reaction 6 (reverse): 8% indels

Sup. Info Analysis of indel frequencies using Sanger sequencing and Inference of CRISPR Edits (ICE) tool.



Chapter 7

Rational and random engineering of ThermoCas9 for optimal activity at moderate temperatures

Evgenios Bouzetos*, Despoina Trasanidou*, Youri Kobus, Ricardo Villegas Warren, Raymond H. J. Staals, & John van der Oost**

Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

*Equal contributions

**Corresponding author

Abstract

Several type II-C Cas9 nucleases from thermophilic bacteria have recently been characterized. Despite their relatively high stability and fidelity at moderate temperatures, the relatively low catalytic activity of these Cas9 variants limits their use for genome editing in mesophilic hosts and/or cells at temperatures of 37°C or below. Here, we describe some first steps towards optimizing the cleavage activity of the thermostable type II-C Cas9 orthologue from *Geobacillus thermodenitrificans* (ThermoCas9). The goal is to obtain rationally and/or randomly generated ThermoCas9 variants with enhanced activity at lower temperatures. For *in vivo* selection a previously described toxin-targeting system in *E. coli* is used between 20-37°C. Although we show that ThermoCas9 is active along the entire temperature range in case of sgRNAs with longer guides and the most preferred PAM, we also observed that no ThermoCas9 nuclease activity is detected under sub-optimal conditions, i.e. when using truncated guides and non-perfect PAMs. Hence, the latter conditions substantially reduce the background level of survival by the wild type ThermoCas9, hence allowing for selection of surviving *E. coli* clones only when they express optimized ThermoCas9 variants. We used *in vitro* random mutagenesis (error-prone PCR) to generate a ThermoCas9 library that enables *in vivo* selection of catalytically enhanced ThermoCas9 variants by using truncated sgRNAs to target one or more protospacers with suboptimal PAM; this library will be screened in the near future. In addition, we will use a rational engineering approach, for which we performed *in silico* structural analyses of ThermoCas9 to identify residues that may play a role in stabilizing the protein, and as such key targets for substituting in an attempt to generate structural flexibility; the latter mutations potentially contribute to lowering the temperature optimum of the enzyme. Our initial findings provide the basis for the random and/or rational generation of ThermoCas9 variants with optimal activity at low temperatures, and thus with potential in biotechnological and medical applications.

Keywords

ThermoCas9, laboratory evolution, ccdb toxin, rational engineering

Introduction

Initially discovered as a crucial element of the prokaryotic clustered, regularly interspaced, short palindromic repeat (CRISPR) immune system, the CRISPR-associated Cas9 nuclease from *Streptococcus pyogenes* (SpyCas9) has been the subject of numerous biochemical and biophysical studies that revealed its mechanism of action and enabled its exploitation in a range of genetic engineering applications[1-9]. During the last decade, several protein engineering efforts have been made to optimize the specificity[10-18], to alter the target site recognition[15,18-22], to reduce the size[3,13,23], or to re-establish the function of SpyCas9[24-31]. These studies are based on random mutagenesis ('random engineering'), structure-based directed mutagenesis ('rational engineering'), or combination thereof. As an alternative approach, novel naturally existing Cas9 orthologues have been identified and characterized[33-42]. The majority of them belong to the CRISPR-Cas type II-C clade, exhibiting smaller size, higher precision, distinct target site recognition, enhanced stability at high temperature/salinity, but reduced cleavage activities compared to the type II-A SpyCas9 variant[43-45]. Given that studies on improving the catalytic efficiency of Cas9 proteins are rare[46], a combination of rational and random engineering would be the most suitable approach to enhance the cleavage activities of type II-C Cas9 orthologues, in an attempt to optimize their genetic engineering potential.

Cas9 nuclease activity is guided by a CRISPR RNA (crRNA) and a transactivating crRNA (tracrRNA), which are often fused to a single guide RNA (sgRNA) molecule[1,47-49]. The sgRNA contains a short (typically 20–24 nucleotides) sequence at the 5' end (spacer) that is adjusted to base pair with the target DNA sequence (protospacer) of interest[1]. Target site recognition also requires a short, Cas9-specific protospacer adjacent motif (PAM, typically 3-7 nt) immediately downstream of the protospacer[1,50,51]. Assembly of the Cas9:sgRNA complex triggers structural rearrangement of the Cas9 REC lobe from an inactive to a DNA recognition–competent conformation[52,53]. Cas9 scans the DNA via three-dimensional (3D) diffusion, locates a PAM through its PAM-interacting (PI) domain, and unwinds the PAM-adjacent nucleation site ('seed') at the target strand (TS) DNA, allowing for sgRNA strand invasion and DNA:sgRNA heteroduplex formation at the seed region of the protospacer[54,55]. In the initial stage of R-loop formation, the guide-TS heteroduplex is being expanded from the seed, and the (partly) displaced non-target strand (NTS) DNA threads into the NUC lobe which is in a disordered, cleavage-incompetent state[56]. Progressive extension of base-pairing towards the 5' end of the guide results in a major reorientation of the HNH domain towards the TS, which triggers conformational

activation of the RuvC domain towards the NTS, through extreme folding-unfolding rearrangement of the hinge loop that connects the two catalytic domains[54-58]. This arginine-rich α helical linker additionally stabilizes the R-loop, hampering strand rehybridization[56]. Mutations in the linker have been connected to altered catalytic efficiency, suggesting it as a potential target for generation of catalytically enhanced Cas9 variants[59].

Here, we describe initial steps towards optimizing the catalytic efficiency of the thermostable type II-C Cas9 orthologue from *Geobacillus thermodenitrificans* T12 (ThermoCas9) at mesophilic temperatures[41]. We employ a directed evolution approach that couples the ThermoCas9 nuclease activity with bacterial cell survival under ccdB toxin-expression conditions. We show that ThermoCas9 is active at temperatures below 37°C but mainly when combined with a full-length spacer and/or an optimal PAM. We generate a random mutagenesis library of ThermoCas9 for *in vivo* selection of catalytically enhanced mutants, and we analyze its complexity. In addition, to allow for a complementary approach, we perform an *in silico* analysis to predict the residues responsible for thermostability and low cleavage efficiency, and we propose targets for engineering ThermoCas9 variants with increased nuclease activity at low temperatures.

Results

In vivo cleavage activity of ThermoCas9 at mesophilic temperatures

The goal of this study is to optimize the DNA cleavage activity of ThermoCas9 at temperatures below 37°C. To analyze the activity of the wild type ThermoCas9 enzyme under these conditions in *Escherichia coli*, we constructed a ThermoCas9-expression plasmid (pCas9) and a guide RNA-expression plasmid (pGuide_T.23) encoding the codon-harmonised *thermocas9* gene and its sgRNA module, respectively, both under the control of the anhydrotetracycline (aTC)-inducible promoter L-tetO1 ($P_{L\text{-}tetO1}$) (Fig. 1a; Supplementary Table 2). This promoter allows for regulation of expression by up to 5,000-fold, enabling adjustment of the selection force[60]. In a third plasmid (pTarget_CCAA), we designed a protospacer matching the guide RNA (spacer T.23) as well as a downstream PAM (5'-N_4CCAA-3') that enables optimal cleavage activity of ThermoCas9 in *E. coli*[41] (Fig. 1a; Supplementary Tables 2, 3; see also **Chapter 5**). In the pTarget plasmid, we also cloned

the *ccdB* toxin gene under the control of the arabinose-inducible promoter (P_{araBAD}). In case of arabinose-induced expression of the toxin gene, the result will be cell death, unless the ThermoCas9 activity is sufficient to cleave the toxin-producing pTarget plasmid (Fig. 1a, b; Supplementary Table 2). We chose P_{araBAD} because of its rapid kinetics[61], high induction ratio[62], tight regulation and glucose-mediated repression[63]. Moreover, we expressed the mutant *lacY(a177c)* transporter gene from the IPTG-inducible *lac* promoter (P_{lac}) to ensure robust arabinose import, *ccdB* expression, and killing of cells in which the ThermcoCas9 activity is too low, thus minimising background survival of toxin-containing cells (false positives) (Fig. 1a, b). As control, a sgRNA bearing a non-targeting (NT) spacer was applied (Supplementary Tables 2, 3).

First, we evaluated the signal-to-background ratio of the toxicity-based bacterial survival assay by transforming the pTarget_CCAA construct in the *E. coli* BW25141 strain in the presence of arabinose and IPTG ('toxin-induced, selective condition') or in the presence of glucose ('non-selective condition') (Supplementary Fig. 1a; Supplementary Table 1). In parallel, we examined the effect of different temperatures (37°C, 30°C, 25°C, 20°C) in the toxic activity of CcdB. We found that the survival rate (number of colonies on the selective condition/ number of colonies on the non-selective condition) was only 0-10%, verifying the low background survival in our positive selection system (Supplementary Fig. 1a, b). Moreover, we demonstrated that the temperature has no effect on the CcdB toxicity (Supplementary Fig. 1b). We then assessed the nuclease activity of ThermoCas9 at 30°C and 20°C by co-transforming the pGuide_T.23 and pCas9 constructs in the *E. coli* BW25141:pTarget_CCAA strain (Supplementary Table 1). In agreement with our previous *in vitro* study[41], we observed that ThermoCas9 was able to efficiently cleave plasmid DNA even at 20°C, exhibiting survival rates of 85-87% (Fig. 1b, c; Supplementary Fig. 1c). Overall, we conclude that, under these conditions, ThermoCas9 is active *in vivo* at mesophilic temperatures.

Next, we examined the *in vivo* cleavage activity of ThermoCas9 at mesophilic temperatures under sub-optimal conditions. One way to do this is by varying the size of the spacer part of the guide. ThermoCas9 has been previously reported to exhibit optimal cleavage activity in the presence of 23 bp spacer-protospacer complementarity, while no activity is observed when the complementarity drops to 17 bp or lower[41](see also **Chapter 5**). We verified these findings in our bacterial selection system at 30°C and 20°C, showing that pairing of ThermoCas9 with a truncated 17 nt spacer results in complete cell death, while pairing with longer spacers (18 nt, 19 nt, 20 nt) leads to a survival rate of up to 10% (Fig. 1c; Supplementary Fig. 1d; Supplementary Table 2, 3). Although this selection trait

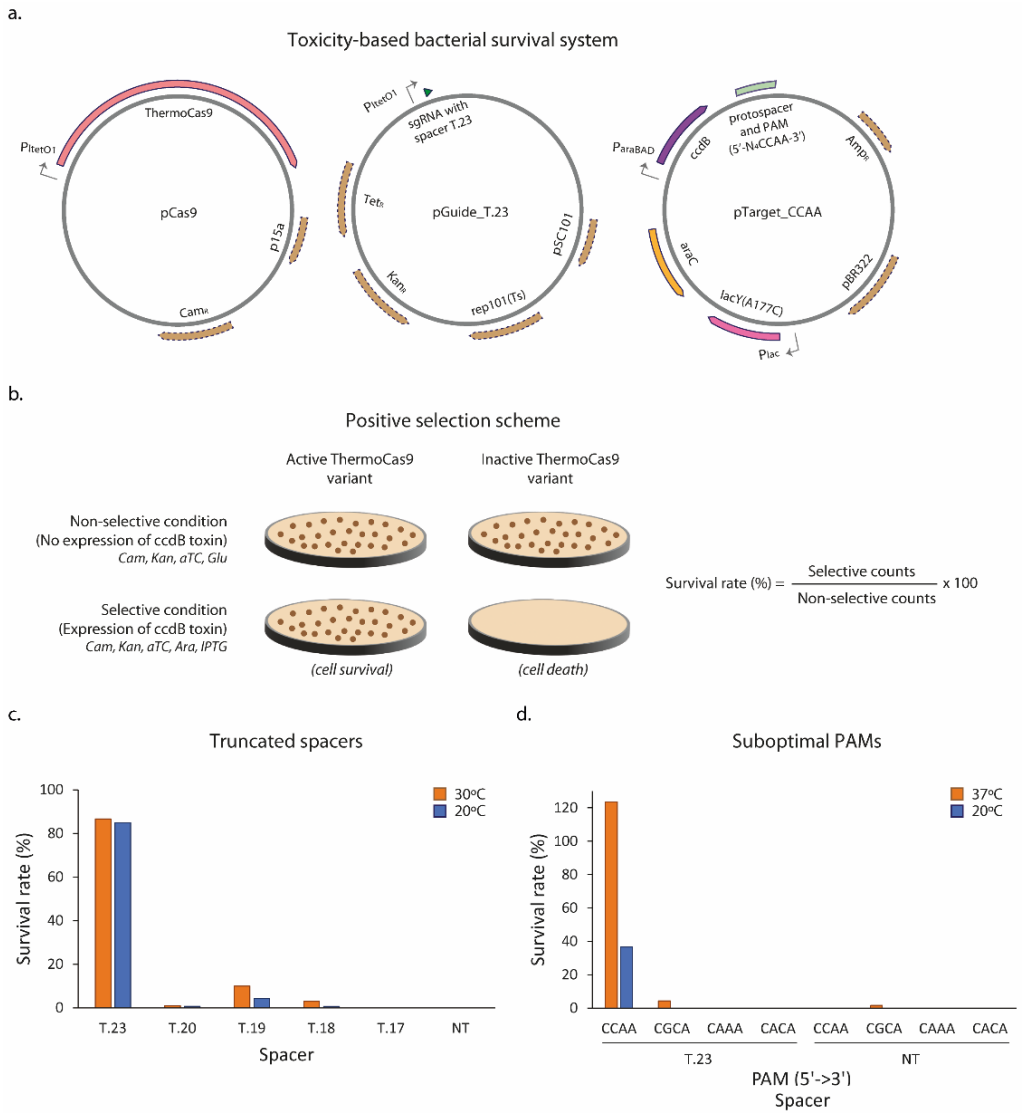


Fig. 1 *In vivo* cleavage activity of ThermoCas9 at mesophilic temperatures. **a.** Schematic illustration of the pCas9, pGuide_T.23, and pTarget_CCAA constructs used in the toxicity-based bacterial survival assays. **b.** Positive selection scheme of the survival assays. When the expression of the ccdB toxin is induced, the active ThermoCas9 variants cleave the ccdB-expressing plasmid allowing for cell survival while cell death is observed in the case of the inactive ThermoCas9 variants. When the expression of the ccdB toxin is not induced, all cells survive as there is no selection pressure from the toxin. The survival rate is calculated by dividing the selective to the non-selective counts. **c.** Survival rate (%) of *E. coli* BW25141:pTarget_CCAA cells co-transformed with pCas9 and pGuide containing a full-length (T.23) or truncated (T.20, T.19, T.18, T.17) targeting spacer. The spot assays were performed at 30°C or 20°C. **d.** Survival rate (%) of *E. coli* BW25141 that carries a target plasmid with optimal (5'-N_4CCAA-3') or suboptimal (5'-N_4CGCA-3', 5'-N_4CAAA-3', 5'-N_4CACA-3') PAM, and is transformed with pCRISPR_T.23 containing a full-length targeting plasmid. The spot assays were performed at 37°C or 20°C.

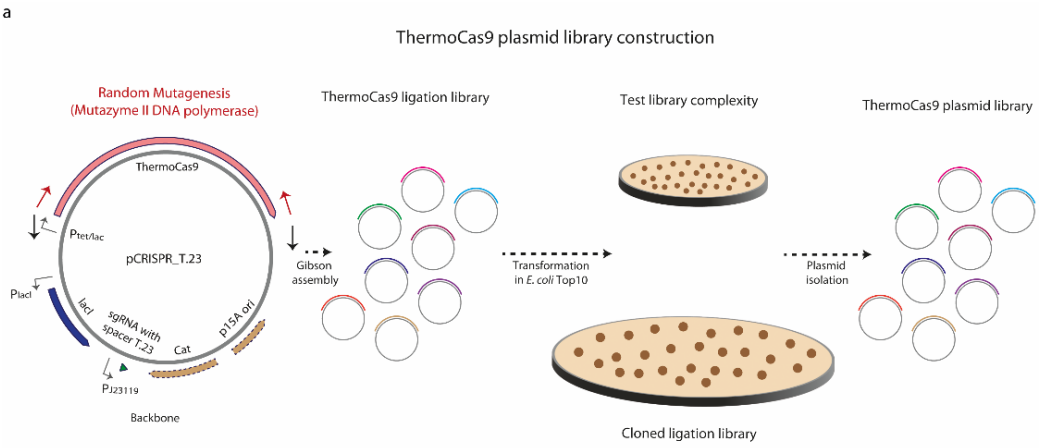
could be used to generate catalytically-enhanced variants of ThermoCas9, no connection between the temperature and the nuclease activity was observed. A second way to generate sub-optimal conditions is based on a previously reported PAM screen. ThermoCas9 exhibits optimal cleavage activity *in vitro* in the presence of a 5'-N_4CCAA-3' PAM, and decreased nuclease efficiencies in the presence of less preferred PAMs (5'-N_4CAAA-3', 5'-N_4CACA-3', 5'-N_4CGCA-3')[41]. Drop of the temperature severely limits the functionality of ThermoCas9 with these sub-optimal PAMs[41], hence providing a means to select for variants with adjusted PAM specificity and/or enhanced activity. An analogous approach was recently used successfully to obtain Cas12a variants with enhanced activity[64].

Here, we generated a dual-plasmid approach (Supplementary Fig. 2) to simplify our *E. coli* selection system, and we verified the connection between ThermoCas9 activity at low temperatures and PAM preference. We observed that ThermoCas9 retains its activity at both 37°C and 20°C in the presence of the optimal 5'-N_4CCAA-3' PAM, though losing more than half of its cleavage efficiency at 20°C (Fig. 1d; Supplementary Fig. 1e). However, in the presence of sub-optimal PAMs, the 5'-N_4CGCA-3' PAM allowed for low ThermoCas9-mediated cleavage but only at 37°C (Fig. 1d; Supplementary Fig. 1e; Supplementary Table 2). Hence, based on this analysis, it is concluded that either the optimal 5'-N_4CCAA-3' PAM or the suboptimal 5'-N_4CGCA-3' PAM may be exploited to create catalytically enhanced variants of ThermoCas9 at 20°C.

In vivo random engineering of ThermoCas9

To randomly optimize the cleavage activity of ThermoCas9 at mesophilic temperatures, we generated a library of the entire *thermocas9* gene with a relatively low mutation rate, and we cloned it in the pCRISPR_T.23 backbone containing the sgRNA with spacer T.23 (Fig. 2a). After transformation of the library in the One Shot™ TOP10 Electrocomp™ *E. coli* strain, we screened several single colonies through PCR and Sanger sequencing to get an impression of the composition of the library. We observed that 20 out of the 27 screened colonies contained the randomly mutagenized pCRISPR_T.23 construct (Fig. 2b). The mutation frequency of the ligation library was 4.8 mutations/gene, with up to 5 different mutations being reported in individual clones (Fig. 2b), a frequency that is commonly used in directed evolution studies[65-68]. Although non-synonymous amino acid substitutions were observed for 8 out of 20 screened colonies, the majority of colonies (11 out of 20)

a



b

Colony No.	Amino acid	Domain
1	A249V	REC
	A278V	REC
	H765	RuvCIII
	I826F	PI
	D1017E	PI
2	I903 (Indel)	PI
3	A278T	REC
	V738I	RuvCIII
	K749	RuvCIII
	A921V	PI
4	Y855 (Indel)	PI
5	Q290E	REC
	T323I	REC
	R427H	REC
	G452R	REC
	R571W	HNH
6	L80*	REC
7	R36I	RuvCI
	P466S	RuvCII
8	FAILED ASSEMBLY	
9	A222L	REC
	P467R	RuvCII
	K659N	HNH
	P807S	RuvCIII
	P861 (Indel)	PI

Colony No.	Amino acid	Domain
10	H173R	REC
	L462M	RuvCII
	K682G	RuvCIII
	I699N	RuvCIII
11	Y217*	REC
12	Q389 (Indel)	REC
13	Q1014H	PI
	Y1048*	PI
14	L46	RuvCI
	A161T	REC
15	N39 (Indel)	RuvCI
16	W16*	RuvCI
17	S506	RuvCII
	R606C	HNH
18	T1041 (Indel)	PI
19	E388L	REC
	R644*	RuvCIII
20	I194	REC
	R269C	REC
	L545F	HNH
	G1061W	PI

c

Fig. 2 *In vivo* random mutagenesis library of ThermoCas9. **a.** Schematic illustration of the construction of the ThermoCas9 plasmid library. **b.** Analysis of the ThermoCas9 library, revealing the amino acid modifications detected from screening of 27 single colonies of the test library plate. **c.** Characterization of the amino acid modifications found in the ThermoCas9 library.

presented deactivation of the *thermocas9* gene, due to the generation of insertions/deletions (indels) or stop codons (Fig. 2b, c). This suggests a toxic effect that may be caused by overexpression of ThermoCas9:sgRNA, and possibly by off-targeting of the chromosomal DNA. Hence, future libraries should be stored under non-inducing conditions to prevent expression of the genes encoding the ThermoCas9 enzyme and its guide. In addition, we will consider using a shorter (17 nt) spacer with a matching protospacer flanked by the suboptimal 5'-N_4CGCA-3' PAM (Supplementary Fig. 2b) in the

toxin-encoding target plasmid of *E. coli* BW25141 in order to enable positive selection of ThermoCas9 variants with improved catalytic activities at low temperatures.

Rational engineering of ThermoCas9

Alternatively, we aim to reveal the structural characteristics of ThermoCas9 that are responsible for its thermostable nature. These insights would allow for introducing directed amino acid substitutions that may enhance its nuclease activity at mesophilic temperatures. ThermoCas9 is a 1,082 amino acid type II-C Cas9 orthologue active at a wide temperature range (from 20°C to 70°C *in vitro*), with optimal activity at 50-60°C[41] (Fig. 3a). Thermostable proteins typically share some protein sequence and structural features that differentiate them from their mesophilic counterparts. In general, there seem to be a wide range of possibilities that are used in different combinations by different proteins. Specifically, it has been observed that thermophilic proteins exhibit higher numbers of smaller amino acids and lower numbers of larger amino acids[69-73]. Moreover, thermostable proteins on average possess elevated numbers (or even networks) of salt-bridges (between positively and negatively charged amino acids), hydrophobic interactions, hydrogen bond network, and disulfide bonds. In addition, shorter loops and lower flexibility of α-helical segments may collectively provide a higher degree of rigidity[74-81]. Hence, identification and replacement of rigid, stabilizing sites would weaken the strong intramolecular interactions, decreasing thermostability[82] and probably shifting the optimal activity down to mesophilic temperatures.

We first used the Phyre2 software[83] to perform a structural analysis for the detection of the closest mesophilic structural homolog of ThermoCas9 (Supplementary Fig. 3). Our findings indicated that the closest structural homolog of ThermoCas9 is the Cas9 protein from the mesophilic bacterium *Neisseria meningitidis* (Nme1Cas9; 41% identity) (Supplementary Fig. 3). ThermoCas9 and Nme1Cas9 both belong to the same CRISPR-Cas subtype II-C, and have an identical length[32] (Fig. 3a). An AlphaFold2[84] prediction of the ThermoCas9 3D structure also revealed a high similarity to the experimentally determined structure of Nme1Cas9[85] (PDB code: 6JDV, Fig. 3b), supporting the Phyre2 findings. According to Chimera[86], ThermoCas9 has more hydrogen bonds compared to Nme1Cas9, which is also reflected by its higher number of residues with polar uncharged side chain calculated by EMBOSS Pepstats[87] (Fig. 3c; Supplementary Fig. 4).

Next, we identified the rigid sites of ThermoCas9 using PredyFlexy[88] and PredictProtein[89], and we examined the predicted flexibility of the corresponding Nme1Cas9 residues (Fig. 3d; Supplementary Fig. 5, 6). We found eight rigid site residues of ThermoCas9 (I33, S55, L121, I497, L599, Y616, C940, L1025) with the highest difference in predicted flexibility compared to the corresponding residues of Nme1Cas9 (Fig. 3d; Supplementary Fig. 5, 6). PredictProtein analysis also showed that seven of these sites were deeply buried in the protein, underlying their high rigidity (Fig. 3d; Supplementary Fig. 7). Substitution of these amino acids may destabilise the structure of ThermoCas9, and as such increase conformational flexibility (and enhance catalytic activity) at mesophilic temperatures[75,82,90-92]. In addition, Hingeprot[93] analysis identified hinge residues of ThermoCas9 (R50, A56, A446, K455, M806) that probably participate in large conformational changes of the protein's nuclease domains essential for the catalytic activity (Fig. 3e; Supplementary Fig. 8). Hence, we propose that replacement of R50, K455 and M806 by alanines may facilitate hinge motions and increase the overall structural flexibility. Furthermore, it was previously reported that arginine and glutamic acid residues in thermophilic proteins are substituted by residues that provide less stability and fewer electrostatic interactions (lysine and aspartic acid) in their mesophilic counterparts[94,95]. This strategy could be followed to decrease the thermostability of ThermoCas9, by replacing selected arginine and some glutamate residues that may form salt bridges, with lysine and aspartic acid residues, respectively (Fig. 3d). Due to the high biochemical similarities of these amino acids, no major disruptions in the overall structure and function are expected. Altogether, we conclude that relaxing of rigid sites and hinge residues may increase the structural flexibility of ThermoCas9 and thus its enzymatic activity at mesophilic temperatures.

Except for residues responsible for the thermostability of ThermoCas9, we also tried to identify amino acids that may influence its catalytic efficiency. The thermostable AceCas9, a type II-C Cas9 orthologue from *Acidothermus cellulolyticus* (optimum growth temperature 55°C) has recently been engineered to acquire higher catalytic activity at 37°C *in vivo*[46] (Supplementary Fig. 9). The randomly obtained AceCas9(G685A) and AceCas9(V709A) mutants exhibited faster HNH conformational sampling, and thus quicker RuvC-mediated cleavage, and more efficient introduction of DSDBs: up to 4-fold increase of *in vitro* cleavage efficiency compared to the wild-type nuclease[46]. The G685A mutated residue belongs to an allosteric hinge helix of the loop that connects the HNH domain with the RuvC-III domain, while the V709A mutated residue belongs to a β-sheet structure of the RuvC-III domain that interacts with the hinge helix (PDB code: 6WBR)[46,96].

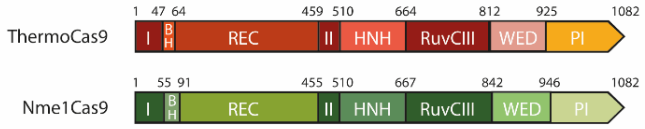
ClustalO[87] alignment of ThermoCas9 with AceCas9 revealed the corresponding residues in the protein sequence of ThermoCas9 (Supplementary Fig. 10). Specifically, R667 resembles the G685A residue (both present in a hinge helix), while the T694 resembles the V709A residue (both present on a β-sheet), according to Chimera (Fig. 3f; Supplementary Fig. 10). Given that engineering of loops and of their interacting structural elements may alter protein stability[97-99], we additionally propose the generation of the ThermoCas9(R667A) and ThermoCas9(T694A) mutants with theoretically enhanced catalytic activity at mesophilic temperatures.

Discussion

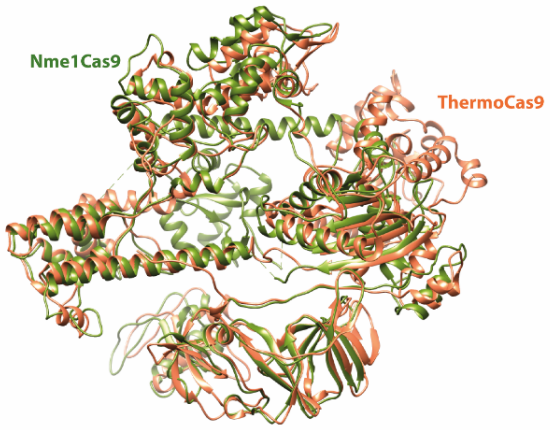
In previous studies, we and others showed that the type II-C Cas9 orthologue from the thermophilic bacterium *G. thermodenitrificans* T12 is active in a wide temperature range (20°C - 70°C) *in vitro*, and that it can be exploited as genetic engineering tool for not only thermophilic but also mesophilic bacteria and even human cells[41,100,101](see also **Chapters 5** and **6**). ThermoCas9 is an attractive CRISPR nuclease because of its small size, high fidelity, variable PAM preferences, stability, and off-switch control by anti-CRISPR proteins (Acrs)[41](see also **Chapter 5**). However, because of its thermophilic origin, it exhibits lower *in vitro* dsDNA cleavage activity at mesophilic temperatures (20°C - 40°C) compared to the commonly used type II-A SpyCas9 orthologue[41,44]. The functionality of ThermoCas9 has never been examined before *in vivo* at temperatures lower than 37°C. Hence, study and optimization of its nuclease activity at mesophilic temperatures may not only enhance its applicability in prokaryotes and human cells but also expand its potential in other sectors, such as plant biotechnology.

Here, we employed a bacterial survival assay that couples nuclease activity with cell survival under ccdB toxin-expression conditions[102-108] in order to test the activity of ThermoCas9 at moderate temperatures and to select the most potent nuclease variants. In agreement with our previous *in vitro* study[41], ThermoCas9 was found to be active at 20°C - 37°C under optimal conditions (full length targeting spacer T.23 and 5'-N_4CCAA-3' PAM), while its activity was completely lost under sub-optimal conditions (using the 17 nt truncated spacer T.17, or less preferred PAMs). We took advantage of this selection trait to create sub-optimal conditions where the ThermoCas9 variants with the highest cleavage efficiency would thrive (compared to wild-type or inactive variants) through cell

a



b



c

Residues	ThermoCas9	Nme1Cas9	Difference
positively charged	220	219	1
R	93	92	1
H	24	26	-2
K	103	101	2
negatively charged	149	150	-1
D	58	58	0
E	91	92	-1
polar uncharged	219	184	35
S	61	46	15
T	60	48	12
N	64	48	16
Q	34	42	-8
hydrophobic	394	410	-16
A	57	82	-25
I	73	51	22
L	99	109	-10
M	13	18	-5
F	46	48	-2
W	9	11	-2
Y	44	30	14
V	53	61	-8
special cases	100	119	-19
C	6	7	-1
G	50	68	-18
P	44	44	0
HYDROGEN BONDS	1,107	769	-767,893

d

ThermoCas9			Nme1Cas9
Rigid site residue	Solvent Accessibillity	Domain	Flexible residue
I19	Buried	RuvCI	
I33	Buried	RuvCI	E
A47	Exposed	RuvCI	
P49	Exposed	BH	
R50	Exposed	BH	
L52	Exposed	BH	
R54	Exposed	BH	
S55	Exposed	BH	G
A56	Buried	BH	
R57	Exposed	BH	
R58	Exposed	BH	
R59	Exposed	BH	
E68	Exposed	REC	
I70	Buried	REC	
R101	Buried	REC	
*A115	Buried	REC	
*R116	Buried	REC	
*I117	Buried	REC	
*L118	Buried	REC	
*L119	Buried	REC	
*H120	Buried	REC	
L121	Buried	REC	K
K195	Exposed	REC	
I220	Buried	REC	
G238	Buried	REC	
F256	Buried	REC	
R332	Exposed	REC	
F333	Buried	REC	
E335	Exposed	REC	
L370	Buried	REC	

ThermoCas9			Nme1Cas9
Rigid site residue	Solvent Accessibillity	Domain	Flexible residue
I407	Buried	REC	
E408	Exposed	REC	
E409	Exposed	REC	
A442	Buried	REC	
L461	Exposed	RuvCII	
L462	Exposed	RuvCII	
*I497	Buried	RuvCII	S
Q516	Exposed	HNH	
I530	Exposed	HNH	
R531	Exposed	HNH	
Q532	Exposed	HNH	
L533	Buried	HNH	
I547	Buried	HNH	
V548	Exposed	HNH	
*F550	Buried	HNH	
P585	Buried	HNH	
*L599	Buried	HNH	S
*V600	Buried	HNH	
Y616	Buried	HNH	N
T629	Exposed	HNH	
F630	Buried	HNH	
V631	Buried	HNH	
L648	Exposed	HNH	
I669	Buried	RuvCIII	
*S670	Buried	RuvCIII	
F672	Buried	RuvCIII	
L673	Buried	RuvCIII	
T694	Buried	RuvCIII	
*V722	Buried	RuvCIII	
*V727	Buried	RuvCIII	

ThermoCas9			Nme1Cas9
Rigid site residue	Solvent Accessibillity	Domain	Flexible residue
V738	Buried	RuvCIII	
T739	Buried	RuvCIII	
A740	Exposed	RuvCIII	
K783	Exposed	RuvCIII	
N786	Exposed	RuvCIII	
V801	Buried	RuvCIII	
L820	Buried	PI	
G825	Exposed	PI	
V836	Buried	PI	
V837	Exposed	PI	
M854	Exposed	PI	
A866	Buried	PI	
I867	Buried	PI	
Q869	Exposed	PI	
I899	Buried	PI	
R900	Exposed	PI	
K903	Exposed	PI	
V927	Buried	PI	
*V929	Buried	PI	
D930	Buried	PI	
*C940	Buried	PI	G
*V941	Buried	PI	
*R987	Exposed	PI	
F1010	Exposed	PI	
S1024	Exposed	PI	
L1025	Buried	PI	R
G1066	Exposed	PI	
A1068	Exposed	PI	

e

Hinge site residue	Solvent Accessibillity	Domain
R50	Exposed	BH
A56	Exposed	BH
A446	Buried	REC
K455	Exposed	REC
M806	Buried	RuvCIII

f

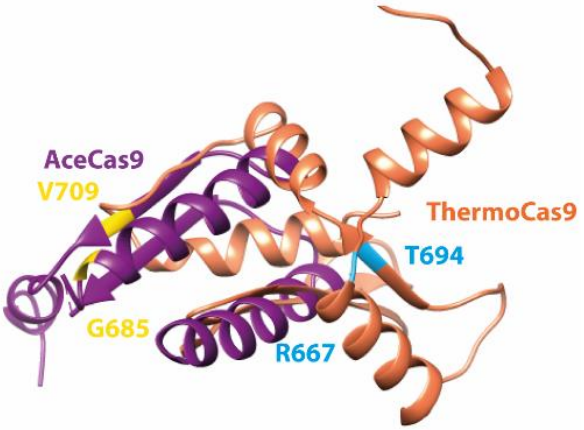


Fig. 3 Rational engineering of ThermoCas9. **a** Comparison of ThermoCas9 and Nme1Cas9 domain architectures. **b** Comparison of ThermoCas9 and Nme1Cas9 3D structures using Chimera. **c** Comparison of ThermoCas9 and Nme1Cas9 amino acid composition using EMBOSS Pepstats. **d** Rigid sites of ThermoCas9 and their replacement with corresponding flexible sites of Nme1Cas9, based on PredyFlexy analysis. Blue colour indicates the arginine and glutamic acid residues, while yellow colour represents polar uncharged residues. The asterisk (*) indicates rigid sites verified by PredictProtein analysis. **e** Hinge site residues of ThermoCas9 provided by Hingeprot study. **f** Comparison of ThermoCas9 and AceCas9 hinge loop residues that affect catalytic activity. Blue and yellow colours represent amino acid residues of ThermoCas9 and AceCas9, respectively.

survival, as previously described for the type II-C AceCas9 orthologue[46]. In the here described initial phase of this project, we generated a low mutation frequency Thermocas9-expressing ligation library via *in vivo* random mutagenesis, which presented mostly non-synonymous amino acid substitutions. Nevertheless, the majority of screened colonies showed deactivation of the *thermocas9* gene, suggesting toxicity of ThermoCas9 under the used conditions, and when combined with spacer T.23. Possible explanations for selecting inactivation of the gene include burden of overproduction and/or cleavage of chromosomal off-targets. Therefore, we suggest to prevent induction of Thermocas9 expression and to apply the above-mentioned spacer and PAM suboptimal conditions.

As a complementary approach, we performed an *in silico* analysis of structural features of ThermoCas9 that may be responsible for its thermostability and its decreased activity at moderate temperatures. We predicted *in silico* that the thermostability of ThermoCas9 originates from its high number of polar charged or uncharged amino acids and of hydrogen bonds, as often observed for thermophilic proteins[74-81]. We detected the rigid

and hinge sites of ThermoCas9 and we proposed specific amino acid substitutions that would increase its structural flexibility and thus its catalytic activity at mesophilic temperatures[75,82,100-102]. Finally, we found two amino acid residues that may play a crucial role in the allosteric activation of the enzymatic activity. The first residue is part of the hinge loop between the HNH and the RuvCIII domains and it interacts with the second residue that is part of the RuvCIII domain. We suggested specific substitutions of these amino acids with others that destabilize the structure, and increase the nuclease activity, as previously reported for AceCas9[46]. We also believe that *in vivo* random mutagenesis of only the HNH-RuvCIII loop region, instead of the entire *thermocas9* gene, may reveal novel and probably more efficient ThermoCas9 variants, underlying the power of this combined random-rational engineering approach.

Conclusion

In summary, we demonstrate that ThermoCas9 is active *in vivo* at temperatures below 37°C but only under optimal conditions. We designed a combined random-rational approach to optimize its nuclease activity, using both *in vivo* random mutagenesis and structure-based engineering, and we envision its implementation for the detection of catalytically enhanced ThermoCas9 variants.

Materials and Methods

Bacterial strains and growth conditions

Bacterial strains used in this study are listed in Supplementary Table 1. All *E. coli* strains were cultured in Luria-Bertani broth (LB) or on LB agar plates, supplemented when necessary with chloramphenicol (15 μg/ml), kanamycin (25 μg/ml), and/or ampicillin (75 μg/ml). In the two-plasmid approach, L-arabinose (1.5 mg/mL), IPTG (1 mM), and/or aTC (100 ng/ml) were additionally used to induce the expression of ccdB, lacY(A177C), and ThermoCas9:sgRNA, respectively. Glucose (0.1% w/v) was applied to inhibit the expression of ccdB. In the two-plasmid approach, IPTG (1 mM) was used to induce the expression of both lacY(A177C) and ThermoCas9, while L-arabinose (1.5 mg/mL) or glucose (0.1% w/v) were applied to induce or inhibit the expression of ccdB. The *E. coli* DH5α strain was grown overnight at 37°C (or 30°C in the presence of pGuide plasmids due

to thermosensitive origin of replication), the *E. coli* BW25141 strain was cultured at 30°C, 25°C, or 20°C for 24 h, 48 h, or 72 h, respectively, and the One Shot™ TOP10 Electrocomp™ *E. coli* strain was cultured overnight at 37°C. Liquid cultures were grown in a shaker incubator at 220 rpm.

Construction of plasmids

Plasmids, primers/oligonucleotides/G-blocks, genes, and protein sequences used in this study are presented in Supplementary Tables 2, 4, 5, and 6, respectively. The codon-harmonised *thermocas9* gene was designed and generated by Twist. The pCas9, the pGuide_BsaI, and the pTarget_CAAA/CACA/CGCA were constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB). The fragments for assembling the plasmids were obtained through PCR with Q5® High-Fidelity 2X Master Mix (NEB). The PCR products were run on a 0.8% agarose gel and were subsequently purified using Zymogen gel DNA recovery kit (Zymo Research). The assembled plasmids were transformed to chemically competent *E. coli* DH5α cells and plated on LB agar containing chloramphenicol (15 µg/ml), kanamycin (25 µg/ml), or ampicillin (75 µg/ml) and incubated overnight at 37°C (or 30°C in the case of pGuide_BsaI). The next day, single colonies were inoculated in LB medium, grown overnight at 37°C or 30°C (220 rpm) and the plasmids were isolated using the GeneJet plasmid Miniprep kit (ThermoFisher Scientific). All the constructs were verified using Sanger sequencing (Macrogen). The pGuide_T.23/T.20/T.19/T.18/T.17/NT, and pTarget_CCAA were constructed using T4 ligation (NEB). The inserts were created by mixing 4 µl oligonucleotide pairs (100 µM each) in Milli-Q water to a final volume of 100 µl, heated at 95°C for 5 min, and slowly cooled to room temperature. The inserts were ligated into the BsaI-digested pGuide backbone (in the case of pGuide constructs) or the XbaI/SphI-double digested p11-LacY-wtx1 backbone (in the case of pTarget_CCAA). Description of the assembled fragments used for the construction of each plasmid is detailed in Supplementary Table 2.

Survival assays

To assess the signal-to-background ratio of the assay, electrocompetent *E. coli* BW25141 cells were transformed with 100 ng of the ccdB-expressing plasmid (pTarget_CCAA), and cultured at different mesophilic temperatures (30°C for 24 h, 25°C for 48 h, 20°C for 72 h)

on LB agar supplemented with ampicillin (75 µg/ml) in the presence of arabinose (1.5 mg/mL) and IPTG (1 mM) (selective condition) or in the presence of glucose (0.1% w/v; non-selective condition) (Supplementary Fig. 1a). After plating 100 µl of undiluted biological triplicates (from 500 µl recovery), colony forming units (CFUs) were counted to calculate the survival rate (number of colonies on the selective plate/ number of colonies on the non-selective plate). To examine the cleavage activity of ThermoCas9, electrocompetent *E. coli* BW25141:pTarget_CCAA cells were co-transformed with 100 ng of targeting (pGuide_T.23/T.20/T.19/T.18/T.17) or non-targeting (pGuide_NT) sgRNA-expressing plasmid, and 100 ng of ThermoCas9-expressing plasmid (pCas9), and cultured at 30°C (24 h) or 20°C (72 h) on LB agar supplemented with kanamycin (25 µg/ml), and chloramphenicol (15 µg/ml) in the presence of arabinose (1.5 mg/mL), IPTG (1 mM), and aTC (100 ng/ml) (selective condition), or in the presence of glucose (0.1% w/v) and aTC (100 ng/ml) (non-selective condition) (Fig. 1b). Serial dilutions (0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}) were spotted (2 µl) on each plate, and CFUs were counted to calculate the survival rate. To test the cleavage activity of ThermoCas9 with ccdB-expressing plasmid containing suboptimal PAM (pTarget_CAAA, pTarget_CACA, pTarget_CGCA) or optimal PAM as control (pTarget_CCAA), electrocompetent *E. coli* BW25141:pTarget_CCAA/CAAA/CACA/CGCA cells were transformed with 100 ng of ThermoCas9:sgRNA_T.23/NT-expressing plasmid (pCRISPR-T.23/NT), and cultured 37°C (24 h) or 20°C (72 h) on LB agar supplemented with chloramphenicol (15 µg/ml) in the presence of arabinose (1.5 mg/mL) and IPTG (1 mM) (selective condition), or in the presence of glucose (0.1% w/v) and IPTG (1 mM) (non-selective condition) (Supplementary Fig. 2b). The spot assay was then performed as described above.

ThermoCas9 library construction

Error-prone PCR was performed on the entire *thermocas9* gene of the pCRISPR_T.23 construct using GeneMorph II Random Mutagenesis Kit (Agilent Technologies) under low error rate (0–5 mutations per kb) conditions with primers designed for Gibson Assembly (Supplementary Table 2, 4). The backbone of pCRISPR_T.23 was amplified using Q5® High-Fidelity 2X Master Mix (NEB) and digested with the restriction enzymes BamHI and XbaI. The PCR product of the randomly mutagenized library and the digested backbone of pCRISPR_T.23 were run on a 0.8% agarose gel and were purified using Zymogen gel DNA recovery kit (Zymo Research). The ThermoCas9 ligation library (100 ng) was transformed to One Shot™ TOP10 Electrocomp™ *E. coli* cells, which were incubated in 250 µl SOC for 1

hour at 37°C and plated on LB agar containing chloramphenicol (15 µg/ml) at a 245 mm^2 plate (Corning; library plate) and a 88 mm^2 plate (Fischer Scientific; library complexity test plate). After overnight incubation at 37°C, *thermocas9*-specific colony PCR (Q5® High-Fidelity 2X Master Mix, NEB) was performed for 27 single colonies to test the library complexity. PCR products were run on a 0.8% agarose gel, purified using Zymogen gel DNA recovery kit (Zymo Research), and subjected to Sanger sequencing (Macrogen). Finally, the library composition and complexity was analysed based on the sequencing results.

In silico analysis

The primary structures of ThermoCas9 and Nme1Cas9 were analyzed using EMBOSS Pepstats[87], and the ridig sites were identified using PredyFlexy[88] and PredictProtein[89]. Hingeprot[93] and Phyre2[83] analyses were performed to detect the hinge amino acid residues and the closest mesophilic structural homolog of ThermoCas9. AlphaFold[84] was applied for prediction of the 3D structure of ThermoCas9, while the crystal structures of Nme1Cas9 and AceCas9 were extracted from the PDB database (6JDV and 6WBR, respectively). Chimera 1.15[86] was used to visualize all 3D structures and to identify amino acid interactions within the protein structures. Finally, ClustalO[87] was applied for alignment of the ThermoCas9 and AceCas9 primary structures.

References

1. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096), 816-821.
2. Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., ... & Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935-949.
3. Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, *513*(7519), 569-573.
4. Kamiya, Y., & Asanuma, H. (2015). Pre-organized Guide RNA in the Cas9 Complex Is Ready for the Selection of Target Double-Stranded DNA. *ChemBioChem*, *16*(16), 2273-2275.
5. Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096.
6. Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262-1278.
7. Wu, W. Y., Lebbink, J. H., Kanaar, R., Geijsen, N., & Van Der Oost, J. (2018). Genome editing by natural and engineered CRISPR-associated nucleases. *Nature chemical biology*, *14*(7), 642-651.
8. Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., ... & Tripathi, V. (2021). Novel CRISPR–Cas systems: an updated review of the current achievements, applications, and future research perspectives. International journal of molecular sciences, 22(7), 3327.
9. Mohanraju, P., Saha, C., van Baarlen, P., Louwen, R., Staals, R. H., & van der Oost, J. (2022). Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nature Reviews Microbiology*, *20*(6), 351-364.

10. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490-495.
11. Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84-88.
12. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., ... & Cereseto, A. (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nature biotechnology*, *36*(3), 265-271.
13. Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., ... & Doudna, J. A. (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*, *550*(7676), 407-410.
14. Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., ... & Kim, J. S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications*, *9*(1), 1-10.
15. Tsai, S. Q., Nguyen, N., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets. *Nature*, *529*(7587), 490-495.
16. Choi, G. C., Zhou, P., Yuen, C. T., Chan, B. K., Xu, F., Bao, S., ... & Wong, A. S. (2019). Combinatorial mutagenesis en masse optimizes the genome editing activities of SpCas9. *Nature Methods*, *16*(8), 722-730.
17. Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., ... & Behlke, M. A. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nature medicine*, *24*(8), 1216-1224.
18. Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., ... & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, *556*(7699), 57-63.
19. Hirano, S., Nishimasu, H., Ishitani, R., & Nureki, O. (2016). Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. *Molecular cell*, *61*(6), 886-894.
20. Hu, X., Meng, X., Liu, Q., Li, J., & Wang, K. (2018). Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice. *Plant biotechnology journal*, *16*(1), 292-297.
21. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., ... & Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, *361*(6408), 1259-1262.
22. Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, *368*(6488), 290-296.
23. Sung, K., Park, J., Kim, Y., Lee, N. K., & Kim, S. K. (2018). Target specificity of Cas9 nuclease via DNA rearrangement regulated by the REC2 domain. *Journal of the American Chemical Society*, *140*(25), 7778-7781.
24. Bester, A. C., Lee, J. D., Chavez, A., Lee, Y. R., Nachmani, D., Vora, S., ... & Pandolfi, P. P. (2018). An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance. *Cell*, *173*(3), 649-664.
25. Wang, X., Ma, S., Liu, Y., Lu, W., Sun, L., Zhao, P., & Xia, Q. (2019). Transcriptional repression of endogenous genes in BmE cells using CRISPRi system. *Insect biochemistry and molecular biology*, *111*, 103172.
26. Stepper, P., Kungulovski, G., Jurkowska, R. Z., Chandra, T., Krueger, F., Reinhardt, R., ... & Jurkowski, T. P. (2017). Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. *Nucleic acids research*, *45*(4), 1703-1713.
27. Xiao, B., Yin, S., Hu, Y., Sun, M., Wei, J., Huang, Z., ... & Jiang, L. (2019). Epigenetic editing by CRISPR/dCas9 in Plasmodium falciparum. *Proceedings of the National Academy of Sciences*, *116*(1), 255-260.
28. Dreissig, S., Schiml, S., Schindele, P., Weiss, O., Rutten, T., Schubert, V., ... & Houben, A. (2017). Live-cell CRISPR imaging in plants reveals dynamic telomere movements. *The Plant Journal*, *91*(4), 565-573.
29. Ma, H., Tu, L. C., Naseri, A., Chung, Y. C., Grunwald, D., Zhang, S., & Pederson, T. (2018). CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. *Nature methods*, *15*(11), 928-931.
30. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424.
31. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464-471.
32. Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., & Thomson, J. A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences, 110(39), 15644-15649.
33. Friedland, A. E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., ... & Bumcrot, D. (2015). Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome biology, 16(1), 1-10.
34. Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., ... & Kim, J. S. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nature communications, 8(1), 1-12.

35. Acharya, S., Mishra, A., Paul, D., Ansari, A. H., Azhar, M., Kumar, M., ... & Chakraborty, D. (2019). Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. Proceedings of the National Academy of Sciences, 116(42), 20959-20968.
36. Müller, M., Lee, C. M., Gasiunas, G., Davis, T. H., Cradick, T. J., Siksnys, V., ... & Mussolino, C. (2016). Streptococcus thermophilus CRISPR-Cas9 systems enable specific editing of the human genome. Molecular Therapy, 24(3), 636-644.
37. Chatterjee, P., Jakimo, N., & Jacobson, J. M. (2018). Minimal PAM specificity of a highly similar SpCas9 ortholog. Science advances, 4(10), eaau0766.
38. Hu, Z., Wang, S., Zhang, C., Gao, N., Li, M., Wang, D., ... & Wang, Y. (2020). A compact Cas9 ortholog from Staphylococcus Auricularis (SauriCas9) expands the DNA targeting scope. PLoS biology, 18(3), e3000686.
39. Fedorova, I., Vasileva, A., Selkova, P., Abramova, M., Arseniev, A., Pobegalov, G., ... & Severinov, K. (2020). PpCas9 from Pasteurella pneumotropica-a compact Type II-C Cas9 ortholog active in human cells. Nucleic acids research, 48(21), 12297-12309.
40. Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C., & Doudna, J. A. (2017). A thermostable Cas9 with increased lifetime in human plasma. Nature communications, 8(1), 1-8.
41. Mougiakos, I., Mohanraju, P., Bosma, E. F., Vrouwe, V., Finger Bou, M., Naduthodi, M. I., ... & Van Der Oost, J. (2017). Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications*, *8*(1), 1-11.
42. Tsui, T. K. M., Hand, T. H., Duboy, E. C., & Li, H. (2017). The impact of DNA topology and guide length on target selection by a cytosine-specific Cas9. *ACS synthetic biology*, *6*(6), 1103-1113.
43. Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, *42*(10), 6091-6105.
44. Mir, A., Edraki, A., Lee, J., & Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS chemical biology*, *13*(2), 357-365.
45. Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K., & Doudna, J. A. (2015). Single-stranded DNA cleavage by divergent CRISPR-Cas9 enzymes. *Molecular cell*, *60*(3), 398-407.
46. Hand, T. H., Roth, M. O., Smith, C. L., Shiel, E., Klein, K. N., Gilbert, D. M., & Li, H. (2021). Catalytically Enhanced Cas9 through Directed Protein Evolution. *The CRISPR Journal*, *4*(2), 223-232.
47. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., ... & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602-607.
48. Chylinski, K., Le Rhun, A., & Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology*, *10*(5), 726-737.
49. Charpentier, E., Richter, H., van der Oost, J., & White, M. F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS microbiology reviews*, *39*(3), 428-441.
50. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733-740.
51. Shah, S. A., Erdmann, S., Mojica, F. J., & Garrett, R. A. (2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA biology*, *10*(5), 891-899.
52. Jiang, F., Zhou, K., Ma, L., Gressel, S., & Doudna, J. A. (2015). A Cas9–guide RNA complex preorganized for target DNA recognition. *Science*, *348*(6242), 1477-1481.
53. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., ... & Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, *343*(6176), 1247997.
54. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62-67.
55. Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., ... & Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences*, *111*(27), 9798-9803.
56. Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., ... & Doudna, J. A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, *351*(6275), 867-871.
57. Josephs, E. A., Kocak, D. D., Fitzgibbon, C. J., McMenemy, J., Gersbach, C. A., & Marszalek, P. E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic acids research*, *43*(18), 8924-8941.
58. Cencic, R., Miura, H., Malina, A., Robert, F., Ethier, S., Schmeing, T. M., ... & Pelletier, J. (2014). Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PloS one*, *9*(10), e109213.
59. Zeng, Y., Cui, Y., Zhang, Y., Zhang, Y., Liang, M., Chen, H., ... & Lou, J. (2018). The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic acids research*, *46*(1), 350-361.

60. Lutz, R., & Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research*, *25*(6), 1203-1210.
61. Johnson, C. M., & Schleif, R. F. (1995). In vivo induction kinetics of the arabinose promoters in Escherichia coli. *Journal of bacteriology*, *177*(12), 3438-3442.
62. Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. O. N. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology*, *177*(14), 4121-4130.
63. Haldimann, A., Daniels, L. L., & Wanner, B. L. (1998). Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. *Journal of bacteriology*, *180*(5), 1277-1286.
64. Zhang, L., Zuris, J. A., Viswanathan, R., Edelstein, J. N., Turk, R., Thommandru, B., ... & Vakulskas, C. A. (2021). AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. *Nature communications*, *12*(1), 3908.
65. Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A., & Pedersen, A. H. (1999). Directed evolution of a fungal peroxidase. *Nature biotechnology*, *17*(4), 379-384.
66. Wan, L., Twitchett, M. B., & Eltis, L. D. Mauk. AG, and Smith, M. 1998. *vitro evolution of horse heart myoglobin to increase peroxidase activity. Proc. Natl. Acad. Sci. USA*, *95*, 12825-12831.
67. You, L., & Arnold, F. H. (1996). Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Engineering, Design and Selection*, *9*(1), 77-83.
68. Shafikhani, S., Siegel, R. A., Ferrari, E., & Schellenberger, V. (1997). Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization. *Biotechniques*, *23*(2), 304-310.
69. Meruelo, A. D., Han, S. K., Kim, S., & Bowie, J. U. (2012). Structural differences between thermophilic and mesophilic membrane proteins. *Protein Science*, *21*(11), 1746-1753.
70. Schneider, D., Liu, Y., Gerstein, M., & Engelman, D. M. (2002). Thermostability of membrane protein helix–helix interaction elucidated by statistical analysis. *FEBS letters*, *532*(1-2), 231-236.
71. Szilágyi, A., & Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, *8*(5), 493-504.
72. Spassov, V. Z., Karshikoff, A. D., & Ladenstein, R. (1995). The optimization of protein-solvent interactions: Thermostability and the role of hydrophobic and electrostatic interactions. *Protein science*, *4*(8), 1516-1527.
73. Vogt, G., & Argos, P. (1997). Protein thermal stability: hydrogen bonds or internal packing?. *Folding and Design*, *2*, S40-S46.
74. Vieille, C., & Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and molecular biology reviews*, *65*(1), 1-43.
75. Mamonova, T. B., Glyakina, A. V., Galzitskaya, O. V., & Kurnikova, M. G. (2013). Stability and rigidity/flexibility—Two sides of the same coin?. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1834*(5), 854-866.
76. Reetz, M. T., Carballeira, J. D., & Vogel, A. (2006). Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie*, *118*(46), 7909-7915.
77. Kumar, S., Tsai, C. J., & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein engineering*, *13*(3), 179-191.
78. Russell, R. J., & Taylor, G. L. (1995). Engineering thermostability: lessons from thermophilic proteins. *Current Opinion in Biotechnology*, *6*(4), 370-374.
79. Querol, E., Perez-Pons, J. A., & Mozo-Villarias, A. (1996). Analysis of protein conformational characteristics related to thermostability. *Protein Engineering, Design and Selection*, *9*(3), 265-271.
80. Jaenicke, R., & Böhm, G. (1998). The stability of proteins in extreme environments. *Current opinion in structural biology*, *8*(6), 738-748.
81. Ladenstein, R., & Antranikian, G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Biotechnology of Extremophiles*, 37-85.
82. Yu, H., & Huang, H. (2014). Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnology advances*, *32*(2), 308-315.
83. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), 845-858.
84. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589.
85. Sun, W., Yang, J., Cheng, Z., Amrani, N., Liu, C., Wang, K., ... & Wang, Y. (2019). Structures of Neisseria meningitidis Cas9 complexes in catalytically poised and anti-CRISPR-inhibited states. *Molecular cell*, *76*(6), 938-952.
86. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., & Ferrin, T. E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC bioinformatics*, *7*, 1-10.

87. Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., ... & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic acids research*, *50*(W1), W276-W279.
88. de Brevern, A. G., Bornot, A., Craveur, P., Etchebest, C., & Gelly, J. C. (2012). PredyFlexy: flexibility and local structure prediction from sequence. *Nucleic acids research*, *40*(W1), W317-W322.
89. Bernhofer, M., Dallago, C., Karl, T., Satagopam, V., Heinzinger, M., Littmann, M., ... & Rost, B. (2021). PredictProtein-predicting protein structure and function for 29 years. *Nucleic acids research*, *49*(W1), W535-W540.
90. Karshikoff, A., Nilsson, L., & Ladenstein, R. (2015). Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *The FEBS journal*, *282*(20), 3899-3917.
91. Paul, M., Hazra, M., Barman, A., & Hazra, S. (2014). Comparative molecular dynamics simulation studies for determining factors contributing to the thermostability of chemotaxis protein "CheY". *Journal of Biomolecular Structure and Dynamics*, *32*(6), 928-949.
92. Yu, H., Yan, Y., Zhang, C., & Dalby, P. A. (2017). Two strategies to engineer flexible loops for improved enzyme thermostability. *Scientific reports*, *7*(1), 1-15.
93. Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R., & Haliloglu, T. (2008). HingeProt: automated prediction of hinges in protein structures. *Proteins: Structure, Function, and Bioinformatics*, *70*(4), 1219-1227.
94. Zhou, X. X., Wang, Y. B., Pan, Y. J., & Li, W. F. (2008). Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino acids*, *34*(1), 25-33.
95. Lee, D. Y., Kim, K. A., Yu, Y. G., & Kim, K. S. (2004). Substitution of aspartic acid with glutamic acid increases the unfolding transition temperature of a protein. *Biochemical and biophysical research communications*, *320*(3), 900-906.
96. Das, A., Hand, T. H., Smith, C. L., Wickline, E., Zawrotny, M., & Li, H. (2020). The molecular basis for recognition of 5′-NNNCC-3′ PAM and its methylation state by Acidothermus cellulolyticus Cas9. *Nature communications*, *11*(1), 1-11.
97. Damnjanović, J., Nakano, H., & Iwasaki, Y. (2014). Deletion of a dynamic surface loop improves stability and changes kinetic behavior of phosphatidylinositol-synthesizing Streptomyces phospholipase D. *Biotechnology and bioengineering*, *111*(4), 674-682.
98. Yedavalli, P., & Madhusudhana Rao, N. (2013). Engineering the loops in a lipase for stability in DMSO. *Protein Engineering, Design & Selection*, *26*(4), 317-324.
99. Yu, H., Zhao, Y., Guo, C., Gan, Y., & Huang, H. (2015). The role of proline substitutions within flexible regions on thermostability of luciferase. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1854*(1), 65-72.
100. Ganguly, J., Martin-Pascual, M., & van Kranenburg, R. (2020). CRISPR interference (CRISPRi) as transcriptional repression tool for Hungateiclostridium thermocellum DSM 1313. *Microbial biotechnology*, *13*(2), 339-349.
101. Gallo, G., Mougiakos, I., Bianco, M., Carbonaro, M., Carpentieri, A., Illiano, A., ... & Fiorentino, G. (2021). A hyperthermoactive-Cas9 editing tool reveals the role of a unique arsenite methyltransferase in the arsenic resistance system of Thermus thermophilus HB27. *Mbio*, *12*(6), e02813-21.
102. Chen, Z., & Zhao, H. (2005). A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic acids research*, *33*(18), e154-e154.
103. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., & Joung, J. K. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature biotechnology*, *33*(12), 1293-1298.
104. Hand, T. H., Das, A., Roth, M. O., Smith, C. L., Jean-Baptiste, U. L., & Li, H. (2018). Phosphate lock residues of Acidothermus cellulolyticus Cas9 are critical to its substrate specificity. *ACS synthetic biology*, *7*(12), 2908-2917.
105. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., ... & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481-485.
106. Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., ... & Kim, J. S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications*, *9*(1), 1-10.
107. Lee, J., Jung, M. H., Jeong, E., & Lee, J. K. (2019). Using Sniper-Cas9 to minimize off-target effects of CRISPR-Cas9 without the loss of on-target activity via directed evolution. *JoVE (Journal of Visualized Experiments)*, (144), e59202.
108. Hand, T. H., Das, A., & Li, H. (2019). Directed evolution studies of a thermophilic Type II-C Cas9. In *Methods in enzymology* (Vol. 616, pp. 265-288). Academic Press.

Supplementary tables & figures

Supplementary Table 1. Strains used in this study.

Strain	Description	Plasmid	Antibiotic resistance	Reference
E. coli DH5α	-	-	-	Lab stock
E. coli BW25141	-	-	-	46
E. coli BW25141: pTarget_CCAA	*E. coli* BW25141 carrying a plasmid that contains the ccdb gene and a ThermoCas9 protospacer flanked by the 5'-N4CCAA-3' PAM.	pTarget_CCAA	Ampicillin	This study
E. coli BW25141: pTarget_CAAA	*E. coli* BW25141 carrying a plasmid that contains the ccdb gene and a ThermoCas9 protospacer flanked by the 5'-N4CAAA-3' PAM.	pTarget_CAAA	Ampicillin	This study
E. coli BW25141: pTarget_CACA	*E. coli* BW25141 carrying a plasmid that contains the ccdb gene and a ThermoCas9 protospacer flanked by the 5'-N4CACA-3' PAM.	pTarget_CACA	Ampicillin	This study
E. coli BW25141: pTarget_CGCA	*E. coli* BW25141 carrying a plasmid that contains the ccdb gene and a ThermoCas9 protospacer flanked by the 5'-N4CGCA-3' PAM.	pTarget_CGCA	Ampicillin	This study
One Shot™ TOP10 Electrocomp™ *E. coli*	Electrocompetent cells that provide transformation efficiency of 1 x 10^{10} cfu/ µg supercoiled DNA.	-	-	ThermoFischer Scientific (C404050)

Supplementary Table 2. Plasmids used in this study.

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pSGuide	-	-	PCR template for pGuide_BsaI.	Lab stock
p11-LacY-wtx1	-	-	PCR template for pTarget_CCAA.	Addgene (Plasmid #69056)
pACYC184	-	-	PCR template for pCas9.	Lab stock
pCRISPR_T.23	-	-	ThermoCas9- and targeting (23 nt) sgRNA- expressing plasmid used in survival assays with pTarget_CCAA, as a dual-plasmid approach.	64
pCRISPR_NT	-	-	ThermoCas9- and non-targeting sgRNA- expressing plasmid used in survival assays with pTarget_CCAA, as a dual-plasmid approach.	64
pCas9	Gibson assembly	Fragment 1: Annealed oligonucleotides BG26061 and BG26062.	ThermoCas9-expressing plasmid used in survival assays.	This study
		Fragment 2: Backbone from pACYC184 amplified with primers BG26063 and BG26064.		
		Fragment 3: Backbone, and PletO1 from G-block A (BG26065) amplified with primers BG26066 and BG26067.		
		Fragment 4: Partly thermocas9 gene from G-block B (BG26070) amplified with primers BG26072 and BG26073.		
		Fragment 5: Partly thermocas9 gene from G-block C (BG26071) amplified with primers BG26074 and BG26075.		
pGuide_BsaI	Gibson assembly	Fragment 1: sgRNA of ThermoCas9 from pTCas9 amplified with primers BG26041 and BG26042.	sgRNA containing non-targeting, excisable spacer with BsaI restriction sites. Used for cloning of all pGuide plasmids (through restriction enzyme digestion and ligation).	This study
		Fragment 2: Partly PletO1, BsaI spacer, sgRNA of ThermoCas9, and partly terminator from fragment 1 amplified with primers BG26409 and BG26410.		
		Fragment 3: Terminator, and backbone from pSGuide amplified with primers BG26367 and BG26018.		
		Fragment 4: Backbone, PletO1, and BsaI spacer from pSGuide amplified with primers BG26022 and BG26046.		
pGuide_T.23	T4 ligation	Backbone: BsaI-digested pGuide.	sgRNA (23 nt spacer) targeting the pTarget_CCAA, pTarget_CAAA, or pTarget_CGCA in survival assays.	This study
		Insert: Annealed oligonucleotides BG26033 and BG26034.		
pGuide_T.20	T4 ligation	Backbone: BsaI-digested pGuide.	sgRNA (20 nt spacer) targeting the pTarget_CCAA, pTarget_CAAA, or pTarget_CGCA in survival assays.	This study
		Insert: Annealed oligonucleotides BG27390 and BG27391.		
pGuide_T.19	T4 ligation	Backbone: BsaI-digested pGuide.	sgRNA (19 nt spacer) targeting the pTarget_CCAA, pTarget_CAAA, or pTarget_CGCA in survival assays.	This study
		Insert: Annealed oligonucleotides BG27392 and BG27393.		
pGuide_T.18	T4 ligation	Backbone: BsaI-digested pGuide.	sgRNA (18 nt spacer) targeting the pTarget_CCAA, pTarget_CAAA, or pTarget_CGCA in survival assays.	This study
		Insert: Annealed oligonucleotides BG27394 and BG27395.		
pGuide_T.17	T4 ligation	Backbone: BsaI-digested pGuide.	sgRNA (17 nt spacer) targeting the pTarget_CCAA, pTarget_CAAA, or pTarget_CGCA in survival assays.	This study
		Insert: Annealed oligonucleotides BG27396 and BG27397.		
pGuide_NT	T4 ligation	Backbone: BsaI-digested pGuide.	sgRNA (non-targeting spacer) used as negative control in survival assays.	This study
		Insert: Annealed oligonucleotides BG26470 and BG26471.		
pTarget_CCAA	T4 ligation	Backbone: XbaI- and SphI-digested p11-LacY-wtx1.	CcdB-expressing plasmid that contains a 23 nt protospacer flanked by the 5'-N4CCAA-3' PAM. PCR template for pTarget-CAAA, pTarget-CACA, and pTarget-CGCA.	This study
		Insert: Annealed oligonucleotides BG26324 and BG26325.		

Supplementary Table 2. Plasmids used in this study. *(continuation)*

Plasmid ID	Cloning strategy	Description of fragments	Function	Reference
pTarget_CAAA	Gibson assembly	Fragment 1: CAAA PAM, terminator, and backbone from pTarget_CCAA amplified with primers BG27746 and BG27747.	CcdB-expressing plasmid that contains a 23 nt protospacer flanked by the 5'-N4CAAA-3' PAM.	This study
		Fragment 2: Backbone, lacy(a177c) gene, *arac* gene, Para, ccdb gene, protospacer, and CAAA PAM from pTarget_CCAA amplified with primers BG27748 and BG27749.		
pTarget_CACA	Gibson assembly	Fragment 1: CACA PAM, terminator, and backbone from pTarget_CCAA amplified with primers BG27750 and BG27747.	CcdB-expressing plasmid that contains a 23 nt protospacer flanked by the 5'-N4CACA-3' PAM.	This study
		Fragment 2: Backbone, lacy(a177c) gene, *arac* gene, Para, ccdb gene, protospacer, and CAAA PAM from pTarget_CCAA amplified with primers BG27748 and BG27751.		
pTarget_CGCA	Gibson assembly	Fragment 1: CGCA PAM, terminator, and backbone from pTarget_CCAA amplified with primers BG27784 and BG27747.	CcdB-expressing plasmid that contains a 23 nt protospacer flanked by the 5'-N4CGCA-3' PAM.	This study
		Fragment 2: Backbone, lacy(a177c) gene, *arac* gene, Para, ccdb gene, protospacer, and CAAA PAM from pTarget_CCAA amplified with primers BG27748 and BG27792.		
ThermoCas9 plasmid library	Gibson assembly	Fragment 1: Backbone from pCas9 amplified with primers BG28246 and BG28247.	Plasmid library of randomly mutagenized thermocas9 gene.	This study
		Fragment 2: Thermocas9 gene from pCas9 amplified with primers BG28248 and BG28249.		

Supplementary Table 3. Spacers used in this study.

Spacer ID	Target strand	3' PAM (5' -> 3')	Sequence of spacer (23 nt) (5’ -> 3’)	Application
spBsaI	X	X	TGAGACCGGATCCGGTCTCC	Survival assay
spT.23	+	N4CCAA, N4CAAA, N4CACA, N4CGCA	AGAATTTATGCCCATTCACATCC	Survival assay
spT.20	+	N4CCAA	ATTTATGCCCATTCACATCC	Survival assay
spT.19	+	N4CCAA	TTTATGCCCATTCACATCC	Survival assay
spT.18	+	N4CCAA	TTATGCCCATTCACATCC	Survival assay
spT.17	+	N4CCAA	TATGCCCATTCACATCC	Survival assay
spNT	X	X	CTAGATCCGCAGTAACCCCATGG	Survival assay

Supplementary Table 4. Primers/Oligonucleotides/G-blocks used in this study.

	Primer/ Oligo ID	Sequence (5' -> 3')	Function
Survival assays	BG26061	taactcgagcatgcatctagagggccctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgacttgctagcatcacctgtaagtcggacgaattc	Construction of pCas9.
	BG26062	gaattcgtccgacttacaggtgatgctagcaagtcgaggctgatcagcgagctctagcatttaggtgacactatagaatagggccctctagatgcatgctcgagtta	
	BG26063	catcacctgtaagtcggacgaattcTTAATAAGATGATCTTCTTGAGATCGTTTTGG	
	BG26064	TTCGAATTTCTGCCATTCATCCG	
	BG26065 (G-block A)	GACACGAAGCATGAGTACTGTTCAGCCAGCAAGACAGCGATCCCGGGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTCTCTGGCT AACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTTCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCACA TCAGCAGGACGCACTGACCAGGGAGACCCAAGCTTGCCACCATGGTGTACCCCTACGACGTGCCCGACTACGCCGAATTGCCTCCAAAAAA GAAGAGAAAGGTAGGGATCCGAACCTGCATGACAGGAACTAGGCG	
	BG26066	agcggatgaatggcagaaattcgaaTTCAGCCAGCAAGACAGC	
	BG26067	GGTTCGGATCCCTACCTTTC	
	BG26070 (G-block B)	GACACGAAGCATGAGTACTGATGAAGTATAAAATCGGGCTGGATATCGGCATTACCTCAATCGGGTGGGCTGTGATTAATCTGGACATTCCCCGCATCGA AGATTTAGGGGTGCGCATTTTTGACCGAGCGGAAAACCCGAAAACTGGTGAGTCTCTTGCTCTGCCACGGCGCTTAGCCCGCTCCGCCCGGCGGCGGTTA CGTCGGCGCAAACATCGGTTAGAGCGCATTCGCCGCTTATTCGTGCGCGAAGGTATTTTAACCAAGGAAGAGTTAAACAAGTTATTTGAAAAAAAGCACG AAATCGACGTGTGGCAACTGCGGGTCGAAGCTTTAGATCGGAAACTTAATAACGATGAATTAGCCCGCATCCTGCTGCATTTAGCTAAACGTCGGGGTTT TCGATCCAACCGCAAGTCAGAGCGCACTAACAAAGAAAACTCAACCATGTTAAAACATATTGAAGAAAACCAGTCCATTCTGTCTTCATACCGGACCGTCG CTGAAATGGTCGTGAAGGATCCGAAATTTTCCTTACACAAGCGGAATAAAGAGGATAATTACACTAACACAGTCGCCCGCGACGATCTGGAACGTGAAAT CAAATTAATTTTCGCCAAACAACGCGAATATGGTAACATCGTCTGCACTGAAGCTTTTGAACACGAGTATATTTCCATTTGGGCTAGCCAGCGCCCCTTTGC TTCAAAGGATGATATCGAGAAAAAAGTGGGGTTCTGTACCTTTGAGCCCAAAGAAAAACGCGCGCCAAAAGCTACTTACACTTTCCAATCCTTCACTGTGT GGGAACATATTAACAAACTGCGGCTGGTGTCCCCGGGTGGCATCCGTGCTCTTACTGATGATGAACGGCGGCTGATATACAAGCAGGCTTTTCATAAAAA TAAAATCACTTTCCATGATGTCCGGACTCTGCTGAACCTGCCCGACGACACTCGGTTTAAAGGGCTGTTATATGACCGGAACACTACCTTAAAGGAAAATG AGAAAGTCCGCTTCCTGGAATTAGGCGCCTATCATAAAATACGTAAAGCGATCGACAGCGTGTATGGCAAAGGTGCTGCTAAATCTTTTCGGCCGATTGA TTTT GATACTTTTGGCTACGCTTTAACCATGTTTAAAGACGACACTGACATTCGCTCATACCTGCGGAACGAATACGAACAGAATGGTAAACGGATGGAA AATCTTGCGGATAAAGTGTATGATGAAGAACTGATTGAAGAACTGTTAAACTTAAGCTTTTCAAAGTTTGGGCATCTTTCCCTGAAAGCGCTGCGCAACA TCCTGCCATATATGGAACAGGGCGAAGTGTACTCTACTGCTTGTGAACGGGCTGGTTATACTTTTACTGGTCCAAAGAAAAAACAAAAAACCGTACTGTT ACCGAACATTCCGCCGATCGCCAATCCGGTGGTGATGCGCGCTTTAACTCAAGCTCGCAAAGTGGTGAATGCCATTATCAAAAAGTACGGCTCTCCGGTG TCCATCCATATCGAATTAGCCCGTGAACTTTCTCAGTCCTTTGATGCGCATGACAGGAACTAGGCG	
	BG26071 (G-block C)	GACACGAAGCATGAGTACTGCGTGAACTTTCTCAGTCCTTTGATGAACGGCGGAAAATGCAAAAAGAACAAGAAGGTAACCGGAAGAAAAACGAAACA GCCATTCGCCAGCTGGTCGAATATGGTTTAACCTTAAATCCAACAGGTCTGGACATTGTGAAATTCAAACTTTGGAGCGAACAGAACGGTAAATGTGCCT ATTCTTTACAGCCGATCGAAATCGAGCGTCTGTTAGAACCAGGCTATACTGAAGTGGACCATGTGATTCCATACAGCCGGAGCCTGGACGATAGCTATAC TAATAAAGTCCTGGTGCTGACTAAGGAGAACCGGGAAAAAGGTAACCGCACTCCAGCTGAATATTTAGGTTTAGGCTCTGAACGGTGGCAGCAATTCGA GACCTTTGTGCTGACTAATAAGCAATTTAGCAAAAAGAAGCGTGATCGGTTACTGCGTCTGCATTACGATGAAAACGAAGAAAATGAGTTTAAAAATCG GAATCTTAATGATACTCGGTATATCTCTCGCTTCCTGGCTAACTTTATTCGCGAACATTTAAAATTCGCCGACAGCGATGACAAACAGAAAGTATACACCG TGAACGGCCGGATTACTGCCCATTTACGCAGCCGGTGGAATTTTAACAAAAACCGTGAAGAAAGCAATCTGCATCATGCCGTGGATGCTGCCATCGTGG CCTGCACTACCCCGAGCGATATCGCCCGGGTGACTGCCTTCTATCAGCGTCGCGAACAGAACAAAGAATTATCCAAAAAGACCGATCCGCAATTTCCGC AACCCTGGCCGCACTTTGCTGATGAATTACAAGCGCGGTTATCTAAAAATCCAAAGGAGTCAATAAAAGCTTTAAATCTGGGTAATTATGATAACGAGA AATTAGAAAGCCTGCAACCGGTCTTTGTGTCCCGGATGCCGAAGCGTAGCATAACTGGTGCGGCTCATCAGGAAACTCTGCGTCGGTATATCGGCATCG ACGAACGTAGCGGTAAAATACAAACCGTGGTGAAAAAGAAACTTTCCGAGATCCAGTTAGATAAAACTGGGCATTTCCCAATGTACGGTAAAGAAAGCG ATCCAAGAACTTATGAAGCCATTCGCCAGCGTCTGCTGGAACATAACAATGACCCAAAAAAGGCGTTTCAGGAGCCCTTATATAAACCGAAGAAGAACGG TGAACTTGGGCCCATCATCCGGACTATCAAAATCATCGATACCACTAATCAGGTCATTCCGTTAAACGATGGCAAAACTGTGGCCTACAACAGCAACATCG TGCGTGTGGACGTGTTTGAGAAAGATGGCAAATATTATTGTGTGCCCATCTATACTATAGATATGATGAAAGGTATCCTGCCAAACAAGGCGATCGAGCC GAACAAACCGTACTCAGAGTGGAAGGAAATGACCGAGGACTATACTTTCCGGTTCTCACTTTACCCAAATGATCTGATCCGGATCGAATTTCCCCGGGAAA	

Supplementary Table 4. Primers/Oligonucleotides/G-blocks used in this study. *(continuation)*

		AAACTATAAAGACAGCTGTGGGTGAAGAAATCAAAATTAAGGATTTATTCGCCTATTATCAGACTATCGACTCCTCCAATGGTGGTTTATCACTGGTCAG CCATGATAACAACTTTAGCTTACGCAGCATCGGGTCTCGAACTTTAAAACGGTTCGAGAAATACCAGGTAGATGTGCTTGGCAACATCTACAAAGTGCG AGGTGAAAAGCGAGTCGGTGTGGCGTCTTCATCACATAGCAAAGCCGGTGAAACAATCCGGCCGTTATAACGCATGACAGGAACTAGGCG	
	BG26072	gaagagaaaggtagggatccgaaccATGAAGTATAAAATCGGGCTGGATATC	
	BG26073	CATCAAAGGACTGAGAAAGTTCACG	
	BG26074	CGTGAACTTTCTCAGTCCTTTGATG	
	BG26075	gccctctagatgcatgctcgagttaTTATAACGGCCGGATTGTTTCAC	
	BG26018	CGGCTGAAAGCGCTATTTC	Construction of pGuide_BsaI/T.23/T.20/T.19/T.18/T.17/NT.
	BG26022	ctggaagaaatagcgctttcagccgTCTGACGCTCAGTGGAAC	
	BG26033	gcacAGAATTTATGCCCATTCACATCC	
	BG26034	tgacGGATGTGAATGGGCATAAATTCT	
	BG26041	agcactgagaccggatccggtctccGTCATAGTTCCCCTGAGATTATCG	
	BG26042	actgagcctttcgttttataaaaaaCGCCTAAGAGTGGGGAATG	
	BG26046	GGAGACCGGATCCGG	
	BG26367	CGAAAGACTGGGCCTTTCG	
	BG26409	AGCACTGAGACCGGATC	
	BG26410	ataaaacgaaaggcccagtctttcgACTGAGCCTTTCGTTTTATAAAAAACG	
	BG26470	gcacCTAGATCCGCAGTAACCCCATGG	
	BG26471	tgacCCATGGGGTTACTGCGGATCTAG	
	BG27390	gcacATTTATGCCCATTCACATCC	
	BG27391	tgacGGATGTGAATGGGCATAAAT	
	BG27392	gcacTTTATGCCCATTCACATCC	
	BG27393	tgacGGATGTGAATGGGCATAAA	
	BG27394	gcacTTATGCCCATTCACATCC	
	BG27395	tgacGGATGTGAATGGGCATAA	
	BG27396	gcacTATGCCCATTCACATCC	
	BG27397	tgacGGATGTGAATGGGCATA	
	BG26324	ctagaGAGTCCGAGCAGAAAGAGAAGGGAGAATTTATGCCCATTCACATCCCCATCCAAGTCTTGGTTTTCGCTCCGAAGGTAAAAGAAAGTCGCGC AACAAAACCCCGCTACTCGGCGATTgcatg	Construction of pTarget_CCAA/CAAA/CACA/CGCA.
	BG26325	cAATCGCCGAGTAGCGGGGTTTTGTTGCGCGACTTTCTTTTACCTTCGGAGCGAAAACCAAGACTTGGATGGGGATGTGAATGGGCATAAATTCTCC CTTCTCTTTCTGCTCGGACTCt	
	BG27746	ACATCCCCATCAAAGTCTTGG	
	BG27747	AGTGAATCCGTAATCATGGTCATAG	
	BG27748	TATGACCATGATTACGGATTCACTG	
	BG27749	CAAGACTTTGATGGGGATGTG	
	BG27750	ACATCCCCATCACAGTCTTGG	
	BG27751	CAAGACTGTGATGGGGATGTG	
	BG27784	TCCCCATCGCAGTCTTG	
	BG27792	CAAGACTGCGATGGGGATG	
Library	BG28246	CTTCATAAGCAGGCCATTTTGTCTG	Construction of ThermoCas9 plasmid library.
	BG28247	TTAAATTGCTAACGCAGTCAGGCAC	
	BG28248	GTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACGGAATTGTTATCCG	
	BG28249	CAGACAAAATGGCCTGCTTATGAAGCGGGCCATTTTTGTTTAATCC	
	BG26250	TAACTCGAGCATGCATCTAGAG	Sequencing of pCas9.

Supplementary Table 4. Primers/Oligonucleotides/G-blocks used in this study. *(continuation)*

Sequencing	BG26251	ATAGTCCTGTCGGGTTTCG	
	BG26252	GGATATACCACCGTTGATATATCCC	
	BG26253	TGTCGGCAGAATGCTTAATG	
	BG26261	CGCAAACATCGGTTAGAGC	
	BG26262	TCCAATCCTTCACTGTGTGG	
	BG26263	CCGTACTGTTACCGAACATTCC	
	BG26264	CTTAATGATACTCGGTATATCTCTCGC	
	BG26265	GCGATCCAAGAACTTATGAAGC	
	BG26047	CGACCTCATTAAGCAGCTCTAATG	Sequencing of pGuide_BsaI/T.23/T.20/T.19/T.18/T.17/NT.
	BG26048	GCTCACTCAAAGGCGGTAATAC	
	BG26049	CAGAGGGTCTAGCAGAATTTACAAG	
	BG26050	TCTATGAGGATTTATGAGTGGTTATTAAAAGAAC	
	BG26051	TAGAACAACTGTTCACCGTTACATATC	
	BG26052	GTAATGAAGGAGAAAACTCACCG	
	BG26053	ACATTATCGCGAGCCCATTTATAC	
	BG26278	TGTTATTGGCGTTTCCTCC	Sequencing of pTarget_CCAA/CAAA/CACA/CGCA.
	BG26279	AGCTTAAGACGCTAATCCCTAAC	
	BG26280	AGAAAAGTCCACATTGATTATTTGC	
	BG26281	TTGAGGCCAACGGTTATCTC	
	BG26282	TTGGTAACGAATCAGACAATTGAC	
	BG26283	GCTTAAGGCTAAATGCCGAATG	
Screening	BG27865	GAGAACCCACTGCTTACTG	Colony PCR for screening of ThermoCas9 plasmid library complexity.
	BG27866	CAGAGCAAGAGATTACGCG	
	BG26261	CGCAAACATCGGTTAGAGC	Sequencing of ThermoCas9 plasmid library.
	BG26262	TCCAATCCTTCACTGTGTGG	
	BG26263	CCGTACTGTTACCGAACATTCC	
	BG26264	CTTAATGATACTCGGTATATCTCTCGC	
	BG26265	GCGATCCAAGAACTTATGAAGC	
	BG27865	GAGAACCCACTGCTTACTG	

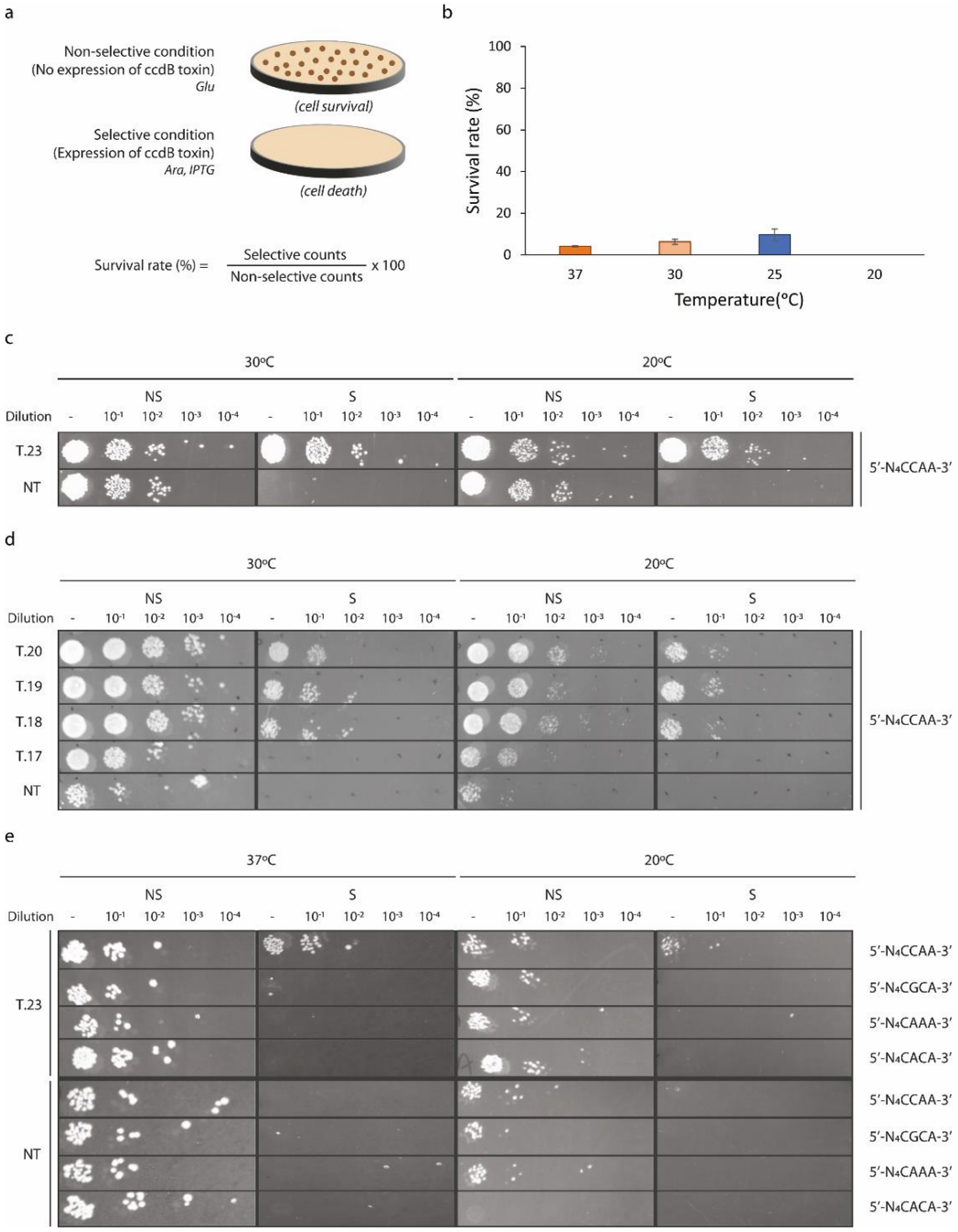
***Lowercase**: overhang sequence for Gibson assembly*
***Uppercase**: sequence annealing to the PCR template DNA*
***Yellow colour**: random sequence for G-block synthesis*
***Red color**: BsaI restriction site*
***Blue color**: XbaI restriction site*
***Purple color**: SphI restriction site*

Supplementary Table 5. DNA sequences of genes used in this study.

Gene/ Fusion gene	Sequence (5' -> 3')
thermocas9 (codon harmonized) gene	atgaagtataaaatcgggctggatatcggcattacctcaatcgggtgggctgtgattaatctggacattccccgcatcgaagatttaggggtgcgcatttttgaccgagcggaaaacccgaaaactggtgagtctcttgctctgccacggcgcttagcccgctccgcccggcggcggttacgtcggcgcaaacatcggttagagcgcattcgccgcttattcgtgcgcgaaggtattttaaccaaggaagagttaaacaagttatttgaaaaaaagcacgaaatcgacgtgtggcaactgcgggtcgaagctttagatcggaaacttaataacgatgaattagcccgcatcctgctgcatttagctaaacgtcgggggttttcgatccaaccgcaagtcagagcgcactaacaaagaaaactcaaccatgttaaaacatattgaagaaaaccagtccattctgtcttcatacccgaccgtcgctgaaatggtcgtgaaggatccgaaatttttccttacacaagcggaataaagaggataattacactaacacagtcgcccgcgacgatctggaacgtgaaatcaaattaatttttcgccaaacaacgcgaatatggtaacatcgtctgcactgaagcttttgaacacgagtatatttccatttgggctagccagcgccccttgcttcaaaggatgatatcgagaaaaaagtggggttctgtacctttgagcccaaagaaaaacgcgcgccaaaagctacttacactttccaatccttcactgtgtgggaacatattaacaaactgcggctggtgtccccgggtggcatccgtgctcttactgatgatgaacggcggctgatatacaagcaggcttttcataaaaatgaaaatcactttccatgatgtccggactctgctgaacctgcccgacgacactcggtttaaaggctgttatatgaccggaacactaccttaaaggaaaatgagaaagtccgcttcctggaattaggcgcctatcataaaatacgtaaagcgatcgacagcgtgtatggcaaaggtgctgctaaatcttttcggccgattgattttgatacttttggctacgctttaaccatgtttaaagacgacactgacattcgctcatacctgcggaacgaatacgaacagaatggtaaacggatggaaaatcttgcggataaagtgtatgatgaagaactgattgaagaactgttaaacttaagcttttcaaagttttggcatctttccctgaaagcgctgcgcaacatcctgccatatatggaacagggcgaagtgtactctactgcttgtgaacgggctggttatacttttactggtccaaagaaaaaacaaaaaaccgtactgttaccgaacattccgccgatcgccaatccggtggtgatgcgcgctttaactcaagctcgcaaagtggtgaatgccattatcaaaaagtacggctctccggtgtccatccatatcgaattagcccgtgaacttttctcagtcctttgatgaacggcggaaaatgcaaaaagaacaagaaggtaaccggaagaaaaacgaaacagccattcgccagctggtcgaatatggttttaaccttaaatccaacaggtctggacattgtgaaattcaaacttttggagcgaacagaacggtaaatgtgcctattctttacagccgatcgaaatcgagcgtctgttagaaccaggctatactgaagtggaccatgtgattccatacagccggagcctggacgatagctatactaataaagtcctggtgctgactaaggagaaccgggaaaaaggtaaccgcactccagctgaatatttaggtttaggctctgaacggtggcagcaattcgacacgtttgtgctgactaataagcaatttagcaaaaagaagcgtgatcggttactgcgtctgcattacgatgaaaacgaagaaaatgagtttaaaaatcggaatcttaatgatactcggtatatctctcgcttcctggctaacttattcgcgaacatttaaaattcgccgacagcgatgacaaacagaaagtatacaccgtgaacggccggattactgcccatttacgcagccggtggaatttttaacaaaaaccgtgaagaaagcaatctgcatcatgccgtggatgctgccatcgtggcctgcactaccccgagcgatatcgcccgggtgactgccttctatcagcgtcgcgaacagaacaaagaattatccaaaaagaccgatccgcaatttccgcaaccctggccgcacctttgctgatgaattacaagcgcggttatctaaaaatccaaaggagtcaataaaagctttaaatctgggtaatttatgataacgagaaattagaaagcctgcaaccggtctttgtgtcccggatgccgaagcgtagcataactggtgcggctcatcaggaaactctgcgtcggtatatcggcatcgacgaacgtagcggtaaaatacaaaccgtggtgaaaaagaaactttccgagatccagttagataaaactgggcatttcccaatgtacggtaaagaaagcgatccaagaacttatgaagccattcgccagcgtctgctggaacataacaatgacccaaaaaaggcgtttcaggagccctttatataaaccgaagaagaacggtgaacttgggcccatcatccggactatcaaaatcatcgatacccactaatcaggtcattccgttaaacgatggcaaaactgtggcctacaacagcaacatcgtgcgtgtggacgtgtttgagaaagatggcaaatattattgtgtgcccatctatactatagatatgatgaaaggtatcctgccaaacaaggcgatcgagccgaacaaaccgtactcagagtggaaggaaatgaccgaggactatactttccggttctcactttacccaaatgatctgatccggatcgaatttccccgggaaaaaactataaagacagctgtgggtgaagaaatcaaaattaaggatttattcgcctattatcagactatcgactcctccaatggtggtttatcactggtcagccatgataacaactttagcttacgcagcatcggctcccgaactttaaaacggttcgagaaataccaggtagatgtgcttggcaacatctacaaagtgcgaggtgaaaagcgagtcggtgtggcgtcttcatcacatagcaaagccggtgaaacaatccggccgttataa
ccdb gene	atgcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtggatgtacagagtgatattattgacacgcccgggcgacggatggtgatcccccctggccagtgcacgtctgctgtcagataaagtctcccgtgaactttacccggtggtgcatatcggggatgaaagcttggcgcatgatgaccaccgatatggccagtgtgccggtctccgttatcgggggaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttttggggaatataa
lacy(a177c) gene	atgtactatttgaaaaacacaaacttttggatgttcggtttattcttttttcttttacttttttatcatgggagcctacttcccgttttcccgatttggctacatgatatcaacctatatcagcaaaagtgatacgggtattatttttgccgctatttctctgttctcgctattattccaaccgctgtttggtctgctttctgacaaactcggtctacgcaaatacctgctgtggattattaccggcatgttagtgatgtttgcgccgttctttatttttatcttcgggccactgctgcagtacaacatttagtaggatcgattgttggtggtatttatctaggctttagttttaacgccggtgcgccagcagtagaggcatttattgagaaagtcagccggcgcagtaatttcgaatttggtcgcgcgcggatgtttggcagtgttggctgggcgctggttgcctcgattgtcgggatcatgttcaccattaataatcagtttgttttctggctgggctctggcagttgtctcatcctcgccgttttactctttttcgccaaaacggacgcgccctcgagtgccacggttgccaatgcggtaggtgccaaccattcggcatttagccttaagctggcactggaactgttcagacagccaaaactgtggttttgtcactgtatgttattggcgtttcctccacctacgatgttttgaccaacagtttgctaatttctttacttcgttctttgctaccggtgaacagggtacccgcgtatttggctacgtaacgacaatgggcgaattacttaacgcctcgattatgttctttgcgccactgatcattaatcgcatcggtgggaagaatgccctgctgctggctggcactattatgtctgtacgtattatttggctcatcgttcgccacctcagcgctggaagtggttattctgaaaacgctgcatatgtttgaagtaccgttcctgctggtgggctcctttaaatatattactagtcagtttgaagtgcgtttttcagcgacgatttatctggtcagtttcagcttctttaagcaactggcgatgatttttatgtctgtactggcgggcaatatgtatgaaagcataggtttccaaggcgcttatctggtgctgggtctggtggcgctgggcttcaccttaatttccgtgttcacgcttagcggcccgggcccgctttccctgctgcgtcgtcaggtgaatgaagtcgcttaa

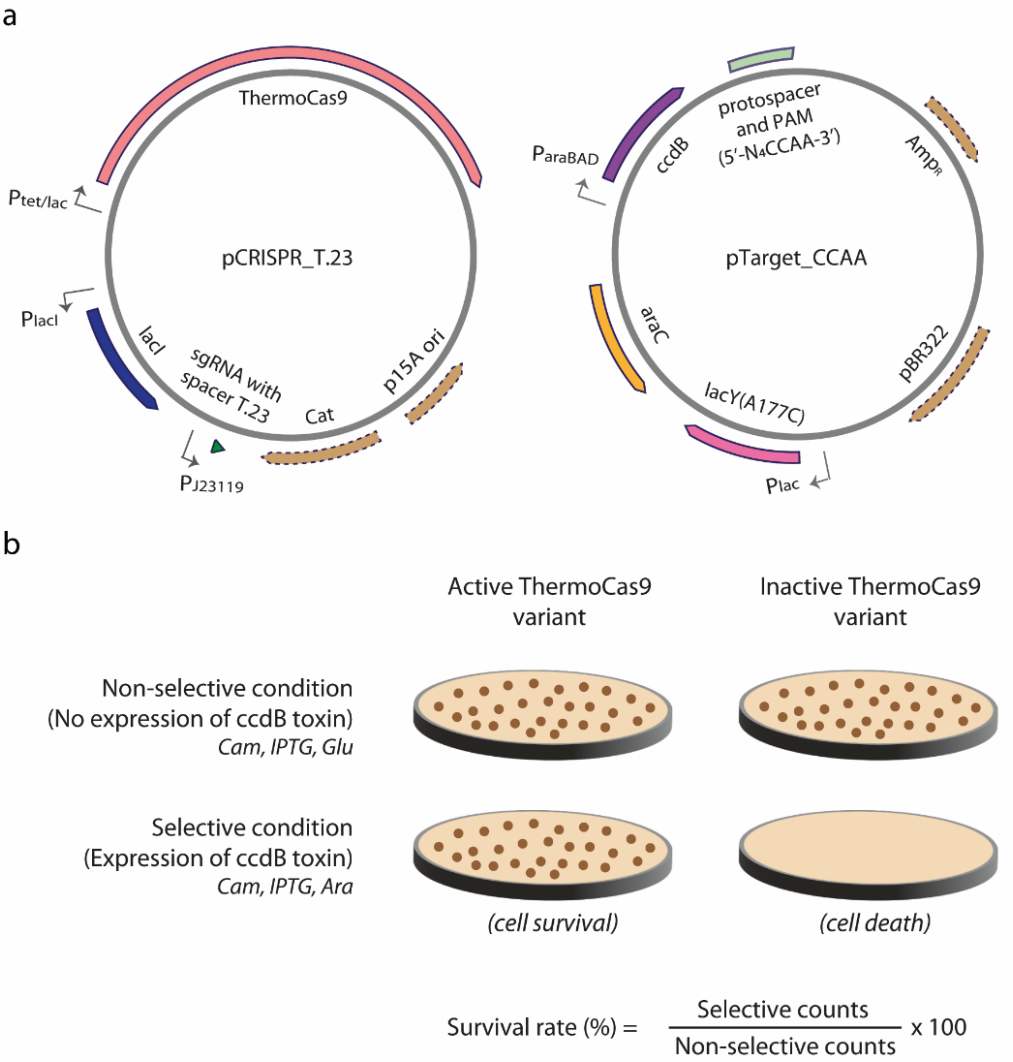
Supplementary Table 6. Amino acid sequences of proteins used in this study.

Protein/ Fusion protein	Sequence (5' -> 3')
ThermoCas9 protein	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEENQSIL SSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLL NLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQ KTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKG NRTPAEYLGLGSERWQQFDTFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDP QFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTN QVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGE KRVGVASSSHSKAGETIRPL*
CcdB protein	MQFKVYTYKRESRYRLFVDVQSDIIDTPGRRMVIPLASARLLSDKVSRELYPVVHIGDESWRMMTTDMASVPVSVIGEEVADLSHRENDIKNAINLMFWGI*
LacY(A177C) protein	MYYLKNTNFWMFGLFFFYFFIMGAYFPFFPIWLHDINHISKSDTGIIFAAISLFSLLFQPLFGLLSDKLGLRKYLLWIITGMLVMFAPFFIFIFGPLLQYNILVGSIVGGIYLGFSFNAGAPAVEAFIEKVSRRSNFEFGRARMFGSVGWAL VASIVGIMFTINNQFVFWLGSGSCLILAVLLFFAKTDAPSSATVANAVGANHSAFSLKLALELFRQPKLWFLSLYVIGVSSTYDVFDQQFANFFTSFFATGEQGTRVFGYVTTMGELLNASIMFFAPLIINRIGGKNALLLAGTIMSVRIIG SSFATSALEVVILKTLHMFEVPFLLVGSFKYITSQFEVRFSATIYLVSFSFFKQLAMIFMSVLAGNMYESIGFQGAYLVLGLVALGFTLISVFTLSGPGPLSLLRRQVNEVA*

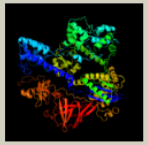
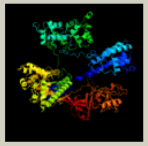
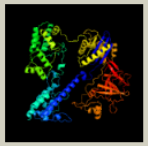
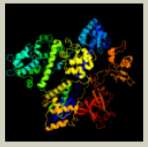
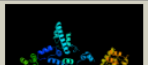


Supplementary Fig. 1 Temperature effect on the activity of ccdB and ThermoCas9. **a.** Positive selection scheme of the toxicity-based bacterial survival assay in the absence of ThermoCas9:sgRNA. When the expression of the ccdB toxin is induced cell death is observed, while when the expression of the ccdB toxin is not induced the cells survive. The survival rate is calculated by dividing the selective to the non-selective

counts. **b.** Survival rate (%) of *E. coli* BW25141 cells transformed with pTarget_CCAA at different mesophilic temperatures (37°C, 30°C, 25°C, 20°C), defining the signal-to-background ratio. **c.** Spot assays depicting the survival rate of *E. coli* BW25141:pTarget_CCAA cells co-transformed with pCas9 and pGuide containing a full-length (T.23) targeting spacer or a non-targeting (NT) spacer. Assays were performed at 30°C or 20°C. **d.** Spot assays depicting the survival rate of *E. coli* BW25141:pTarget_CCAA cells co-transformed with pCas9 and pGuide containing a full-length (T.23) or truncated (T.20, T.19, T.18, T.17) targeting spacer. Assays were performed at 30°C or 20°C. **e.** Spot assays depicting the survival rate of of *E. coli* BW25141 that carries a target plasmid with optimal (5'-N_4CCAA-3') or suboptimal (5'-N_4CGCA-3', 5'-N_4CAAA-3', 5'-N_4CACA-3') PAM, and is transformed with pCRISPR_T.23 containing a full-length targeting plasmid. Assays were performed at 37°C or 20°C.



Supplementary Fig. 2 Toxicity-based bacterial survival assay using a two-plasmid approach. **a.** Schematic illustration of the Pcrispr_T.23, and pTarget_CCAA constructs used in the survival assays. **b.** Positive selection scheme of the survival assays. When the expression of the ccdB toxin is induced, the active ThermoCas9 variants cleave the ccdB-expressing plasmid allowing for cell survival while cell death is observed in the case of the inactive ThermoCas9 variants. When the expression of the ccdB toxin is not induced, all cells survive. The survival rate is calculated by dividing the selective to the non-selective counts.

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c6jdvA_	Alignment		100.0	41	**PDB header:**hydrolase/rna/dna **Chain:** A: **PDB Molecule:**crispr-associated endonuclease cas9; **PDBTitle:** crystal structure of nme1cas9 in complex with sgrna and target dna2 (atatgatt pam) in catalytic state **PDB Entry:** PDBe RCSB PDBj Run Investigator
2	c5axwA_	Alignment		100.0	29	**PDB header:**hydrolase/rna/dna **Chain:** A: **PDB Molecule:**crispr-associated endonuclease cas9; **PDBTitle:** crystal structure of staphylococcus aureus cas9 in complex with sgrna2 and target dna (ttgggt pam) **PDB Entry:** PDBe RCSB PDBj Run Investigator
3	c5x2hA_	Alignment		100.0	38	**PDB header:**hydrolase/rna/dna **Chain:** A: **PDB Molecule:**crispr-associated endonuclease cas9; **PDBTitle:** crystal structure of campylobacter jejuni cas9 in complex with sgrna2 and target dna (agaaaca pam) **PDB Entry:** PDBe RCSB PDBj Run Investigator
4	c6m0wA_	Alignment		100.0	27	**PDB header:**hydrolase/dna/rna **Chain:** A: **PDB Molecule:**crispr-associated endonuclease cas9 1; **PDBTitle:** crystal structure of streptococcus thermophilus cas9 in complex with2 the agaa pam **PDB Entry:** PDBe RCSB PDBj Run Investigator
5	c6rjgC_	Alignment		100.0	27	**PDB header:**hydrolase **Chain:** C: **PDB Molecule:**cas 9; **PDBTitle:** cryo-em structure of st1cas9-sgrna-acriia6-tdna59-ntpam complex. **PDB Entry:** PDBe RCSB PDBj Run Investigator
						**PDB header:**hydrolase **Chain:** B: **PDB Molecule:**crispr-associated endonuclease cas9/csn1; **PDBTitle:** crystal structure of mn-bound s.pyogenes cas9

Supplementary Fig. 3 Phyre2 structural analysis of ThermoCas9. The closest structural homologue is the Nme1Cas9 protein (41% structural identity).

a

PEPSTATS of EMBOSS_001 from 1 to 1082

Molecular weight = 126343.37 Residues = 1082
Average Residue Weight = 116.768 Charge = 59.0
Isoelectric Point = 10.1109
A280 Molar Extinction Coefficients = 115060 (reduced) 115435 (cystine bridges)
A280 Extinction Coefficients 1mg/ml = 0.911 (reduced) 0.914 (cystine bridges)
Improbability of expression in inclusion bodies = 0.733

Residue	Number	Mole%	DayhoffStat
A = Ala	57	5.268	0.613
B = Asx	0	0.000	0.000
C = Cys	6	0.555	0.191
D = Asp	58	5.360	0.975
E = Glu	91	8.410	1.402
F = Phe	46	4.251	1.181
G = Gly	50	4.621	0.550
H = His	24	2.218	1.109
I = Ile	73	6.747	1.499
J = ---	0	0.000	0.000
K = Lys	103	9.519	1.442
L = Leu	99	9.150	1.236
M = Met	13	1.201	0.707
N = Asn	64	5.915	1.376
O = ---	0	0.000	0.000
P = Pro	44	4.067	0.782
Q = Gln	34	3.142	0.806
R = Arg	93	8.595	1.754
S = Ser	61	5.638	0.805
T = Thr	60	5.545	0.909
U = ---	0	0.000	0.000
V = Val	53	4.898	0.742
W = Trp	9	0.832	0.640
X = Xaa	0	0.000	0.000
Y = Tyr	44	4.067	1.196
Z = Glx	0	0.000	0.000

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	234	21.627
Small	(A+B+C+D+G+N+P+S+T+V)	453	41.867
Aliphatic	(A+I+L+V)	282	26.063
Aromatic	(F+H+W+Y)	123	11.368
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	494	45.656
Polar	(D+E+H+K+N+Q+R+S+T+Z)	588	54.344
Charged	(B+D+E+H+K+R+Z)	369	34.104
Basic	(H+K+R)	220	20.333
Acidic	(B+D+E+Z)	149	13.771

b

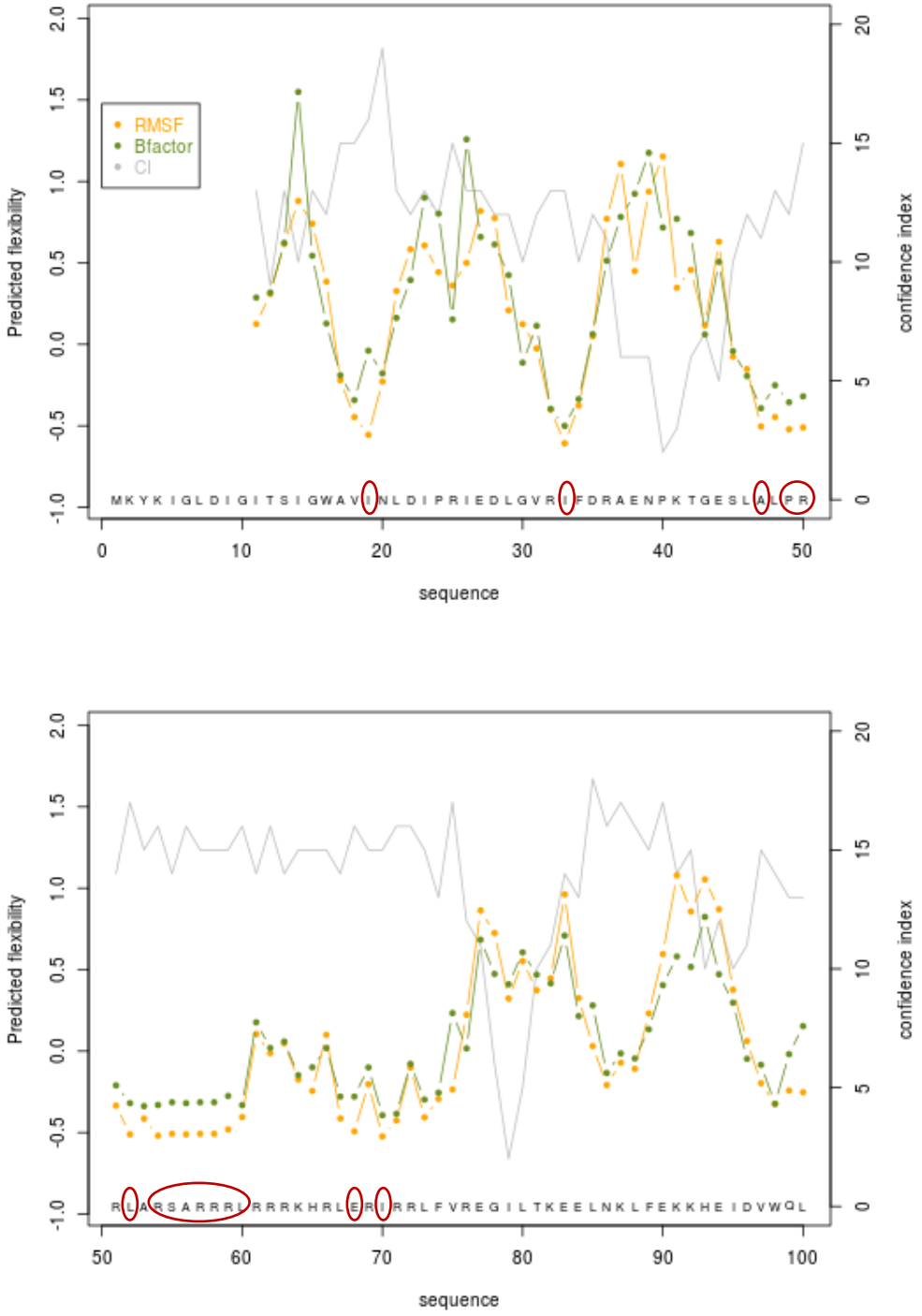
PEPSTATS of EMBOSS_001 from 1 to 1082

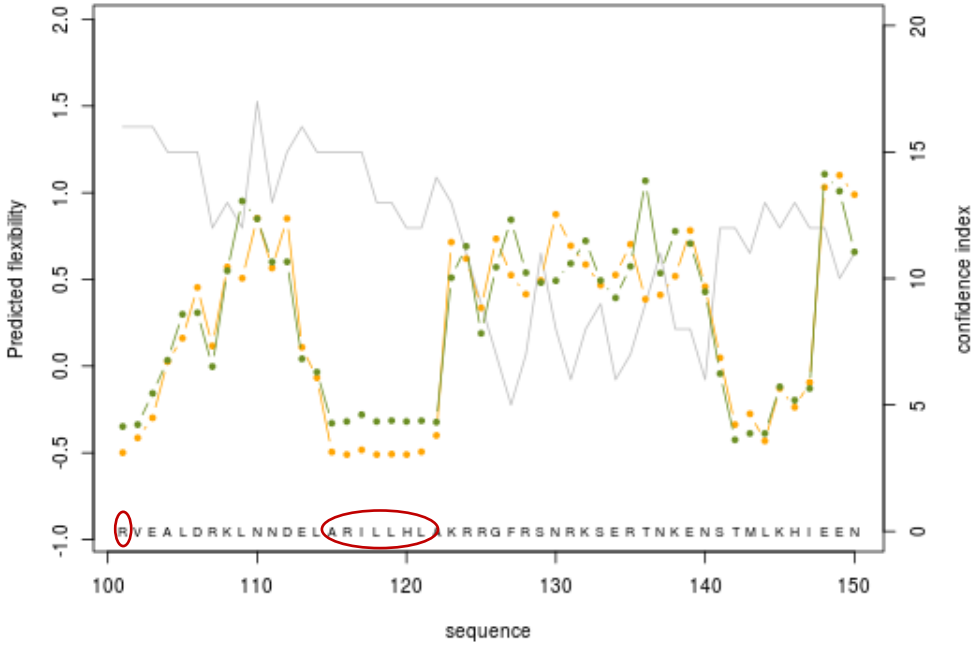
Molecular weight = 124394.66 Residues = 1082
Average Residue Weight = 114.967 Charge = 56.0
Isoelectric Point = 10.1577
A280 Molar Extinction Coefficients = 105200 (reduced) 105575 (cystine bridges)
A280 Extinction Coefficients 1mg/ml = 0.846 (reduced) 0.849 (cystine bridges)
Improbability of expression in inclusion bodies = 0.718

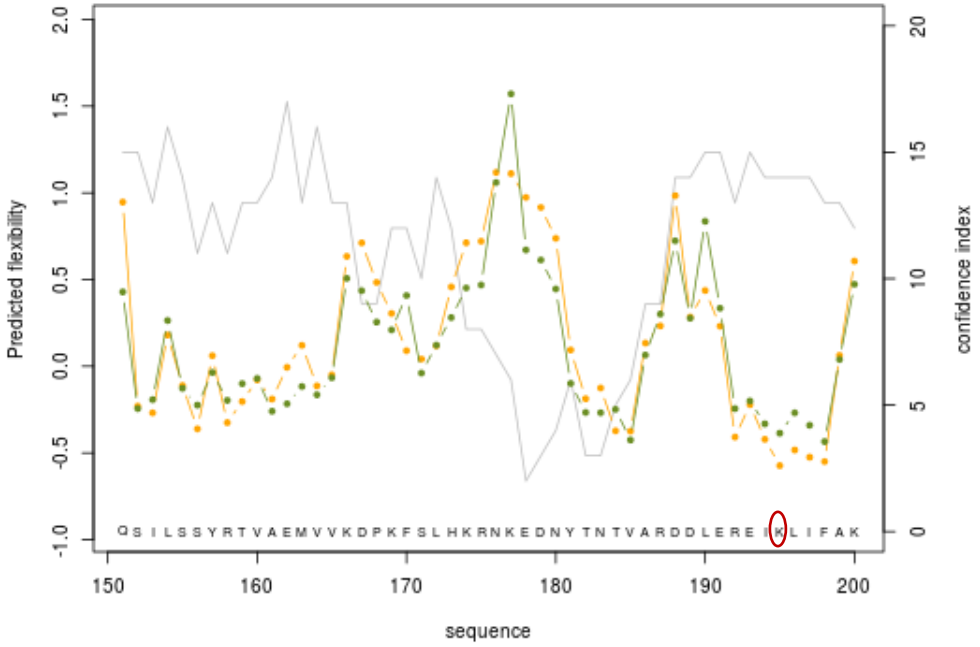
Residue	Number	Mole%	DayhoffStat
A = Ala	82	7.579	0.881
B = Asx	0	0.000	0.000
C = Cys	7	0.647	0.223
D = Asp	58	5.360	0.975
E = Glu	92	8.503	1.417
F = Phe	48	4.436	1.232
G = Gly	68	6.285	0.748
H = His	26	2.403	1.201
I = Ile	51	4.713	1.047
J = ---	0	0.000	0.000
K = Lys	101	9.335	1.414
L = Leu	109	10.074	1.361
M = Met	18	1.664	0.979
N = Asn	48	4.436	1.032
O = ---	0	0.000	0.000
P = Pro	44	4.067	0.782
Q = Gln	42	3.882	0.995
R = Arg	92	8.503	1.735
S = Ser	46	4.251	0.607
T = Thr	48	4.436	0.727
U = ---	0	0.000	0.000
V = Val	61	5.638	0.854
W = Trp	11	1.017	0.782
X = Xaa	0	0.000	0.000
Y = Tyr	30	2.773	0.815
Z = Glx	0	0.000	0.000

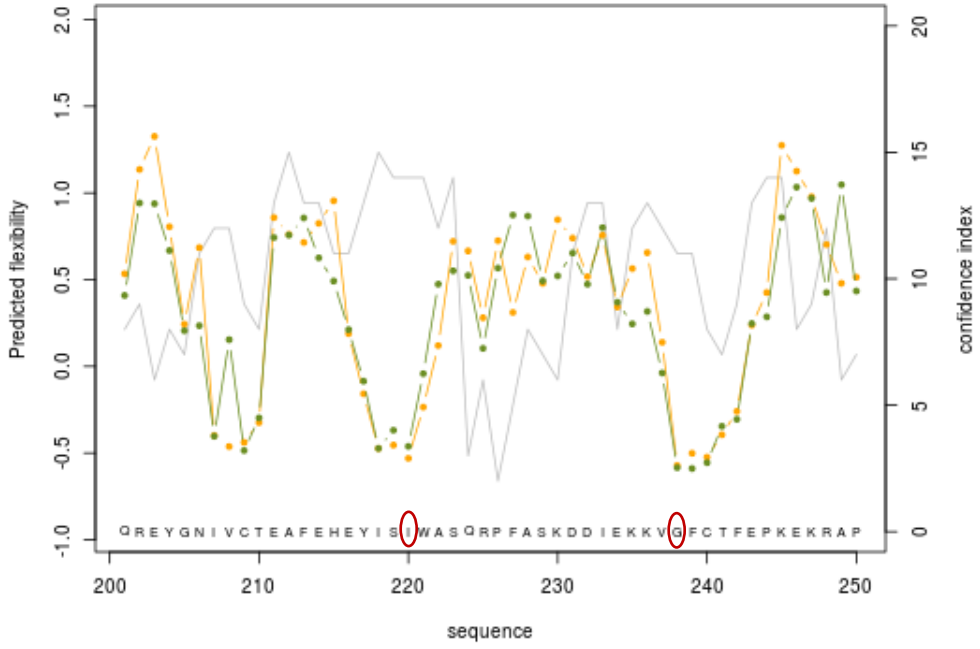
Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	251	23.198
Small	(A+B+C+D+G+N+P+S+T+V)	462	42.699
Aliphatic	(A+I+L+V)	303	28.004
Aromatic	(F+H+W+Y)	115	10.628
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	529	48.891
Polar	(D+E+H+K+N+Q+R+S+T+Z)	553	51.109
Charged	(B+D+E+H+K+R+Z)	369	34.104
Basic	(H+K+R)	219	20.240
Acidic	(B+D+E+Z)	150	13.863

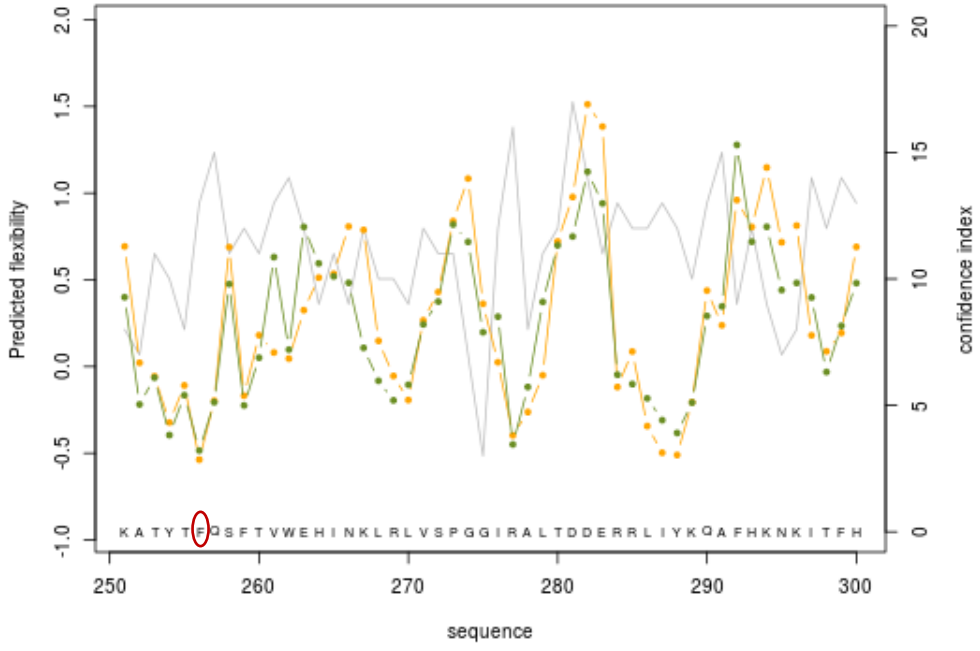
Supplementary Fig. 4 EMBOSS PEPSTATS analysis of the amino acid composition of ThermoCas9 (**a**) and Nme1Cas9 (**b**).

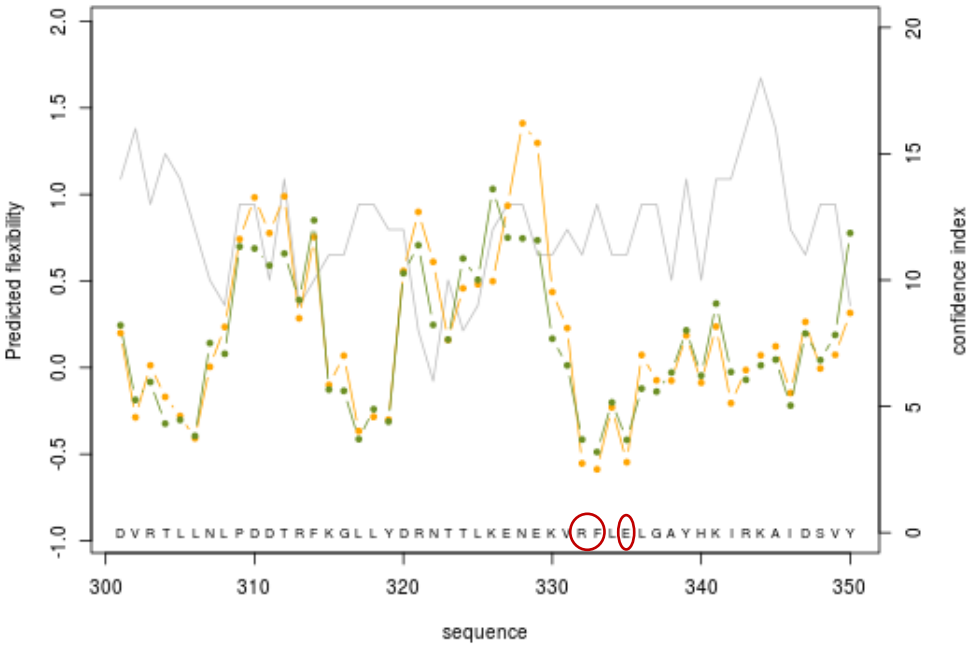

RMSF
Bfactor
CI
Predicted flexibility
confidence index
sequence
MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPR
RLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQL

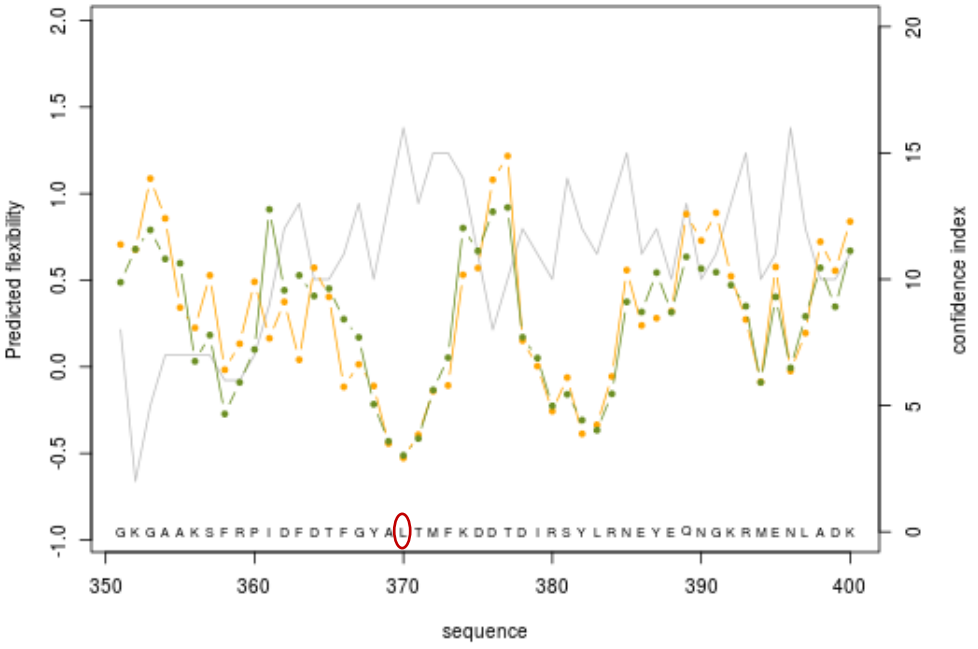

Predicted flexibility
confidence index
sequence
R V E A L D R K L N N D E L A R I L L H L A K R R G F R S N R K S E R T N K E N S T M L K H I E E N
100
110
120
130
140
150

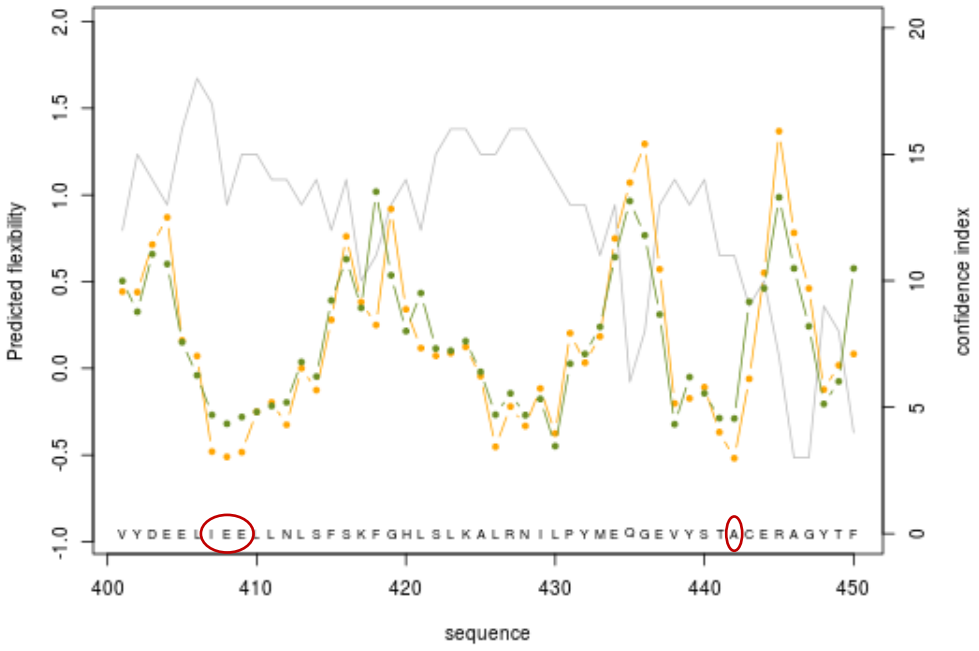

Predicted flexibility
confidence index
sequence
Q S I L S S Y R T V A E M V V K D P K F S L H K R N K E D N Y T N T V A R D D L E R E I K L I F A K
150
160
170
180
190
200

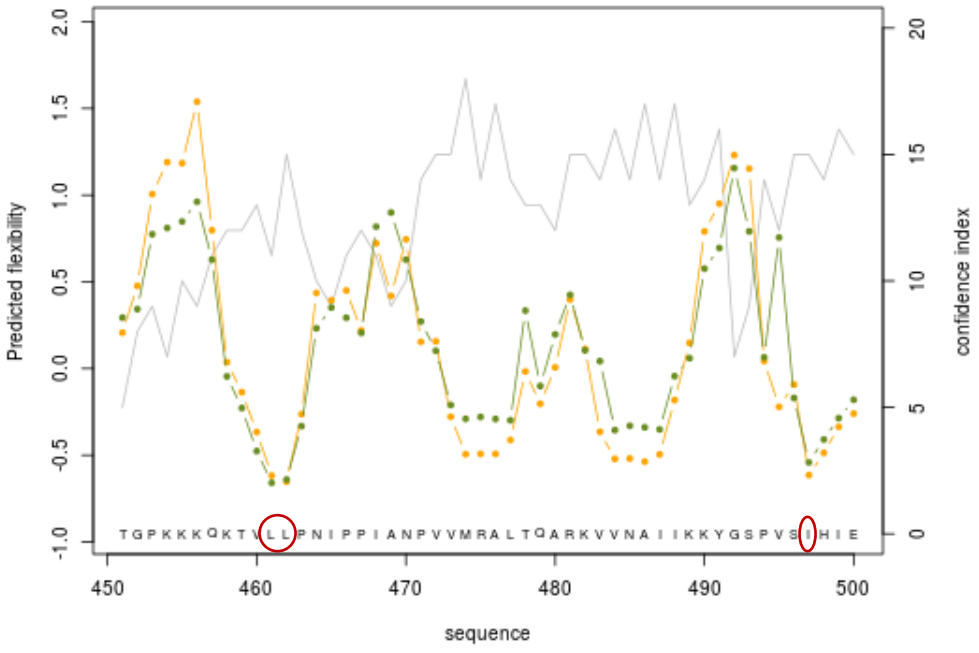

Predicted flexibility
confidence index
sequence
QREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAP
200
210
220
230
240
250

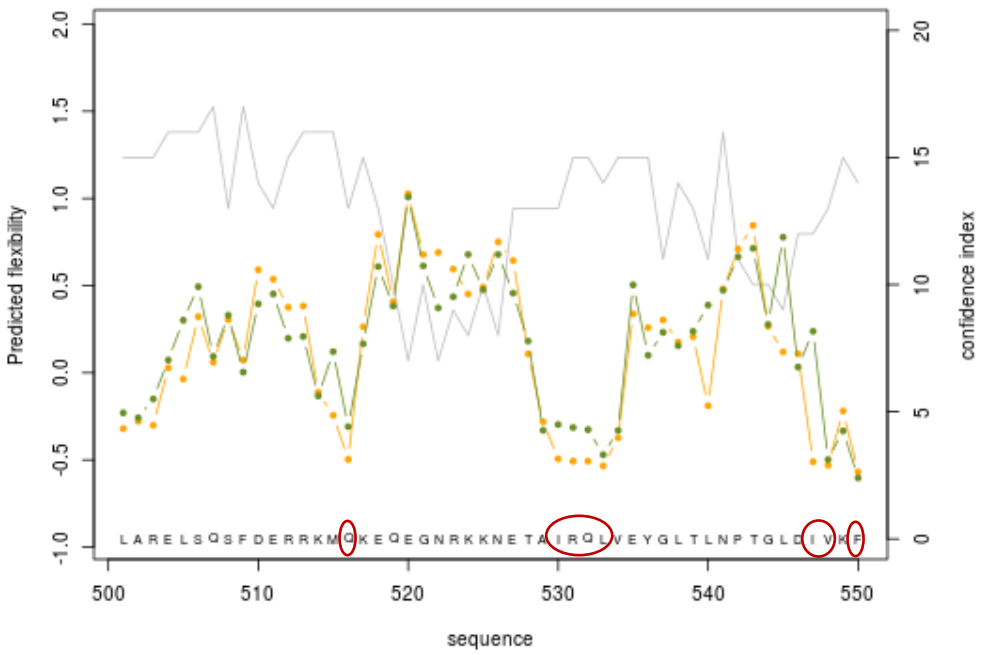

Predicted flexibility
confidence index
sequence
KATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFH
250
260
270
280
290
300

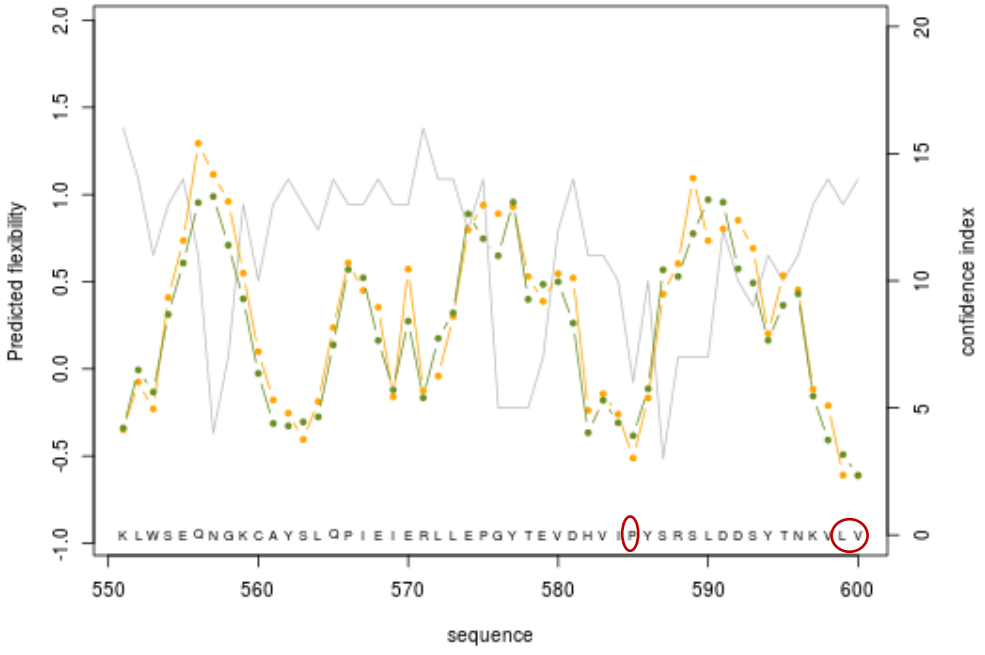

Predicted flexibility
confidence index
sequence
D V R T L L N L P D D T R F K G L L Y D R N T T L K E N E K V R F L E L G A Y H K I R K A I D S V Y

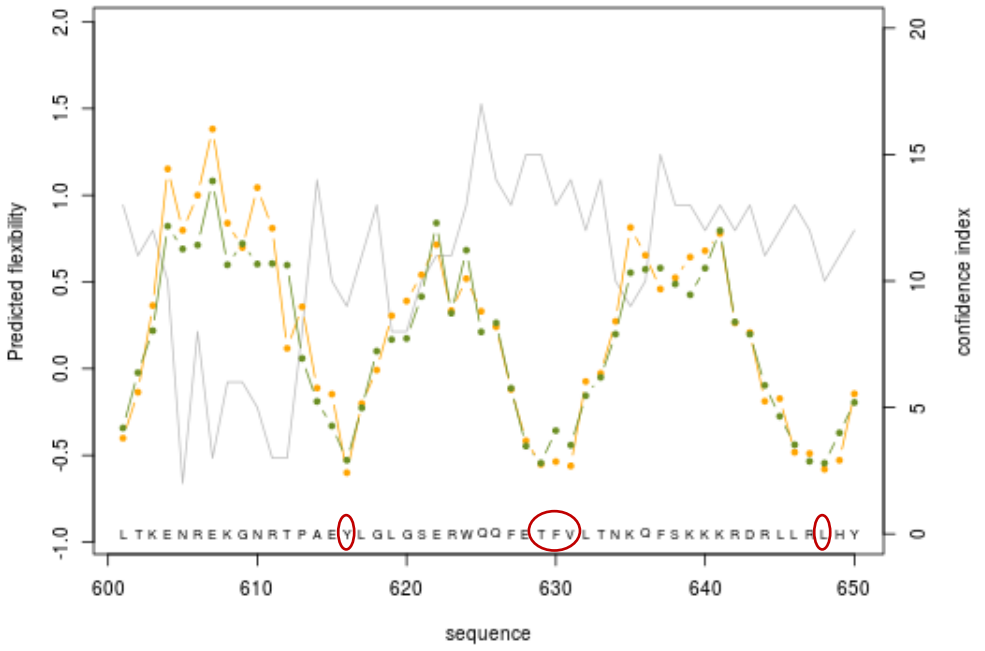

Predicted flexibility
confidence index
sequence
G K G A A K S F R P I D F D T F G Y A L T M F K D D T D I R S Y L R N E Y E Q N G K R M E N L A D K

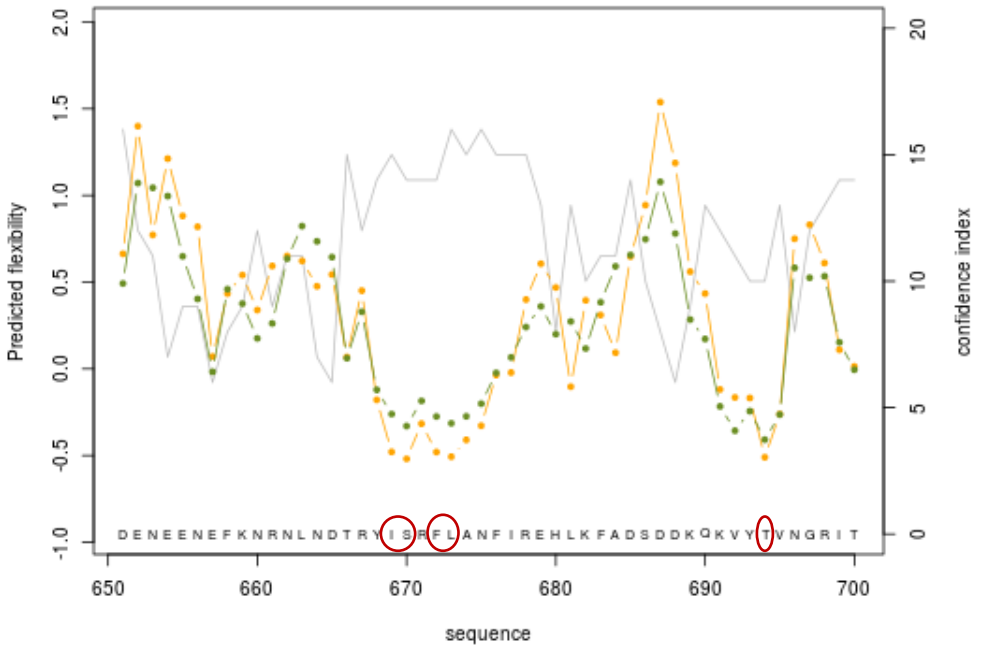

Predicted flexibility
confidence index
sequence
400
410
420
430
440
450

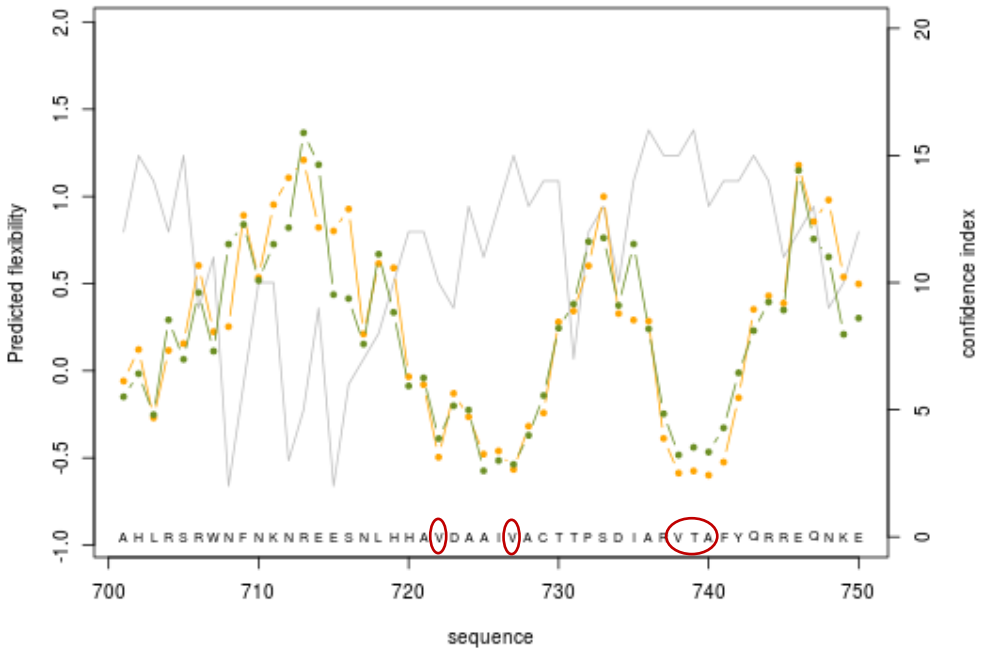

Predicted flexibility
confidence index
sequence
450
460
470
480
490
500

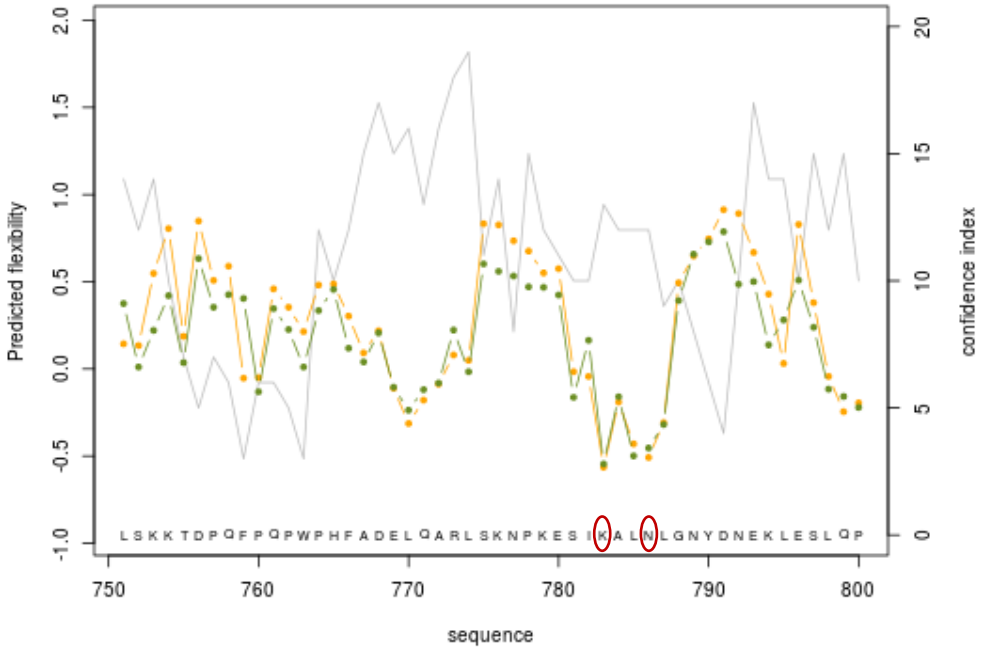

Predicted flexibility
confidence index
sequence
500
510
520
530
540
550

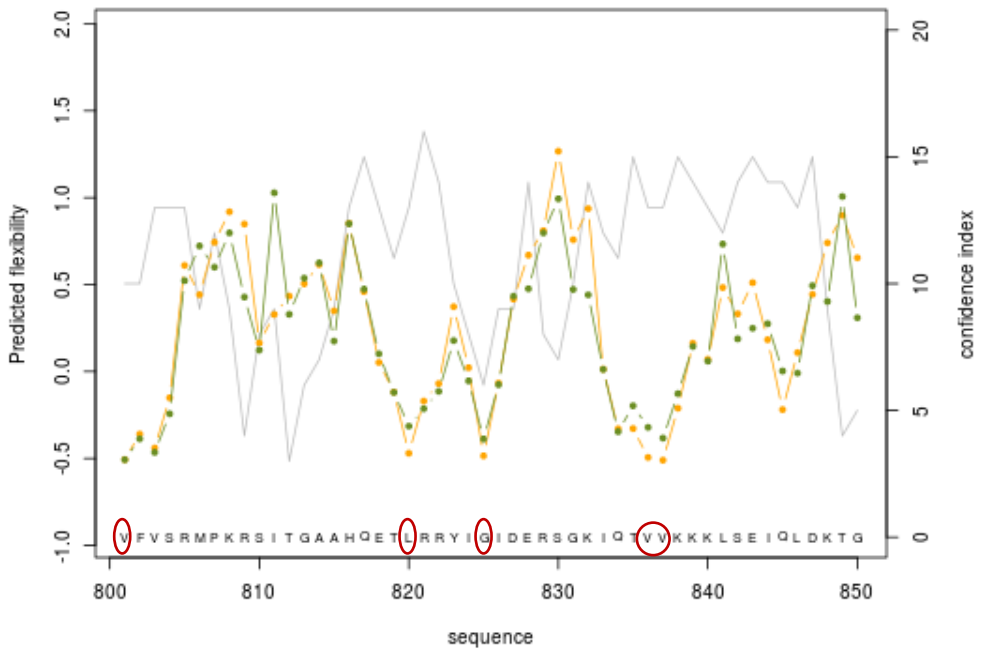

Predicted flexibility
confidence index
sequence
550
560
570
580
590
600

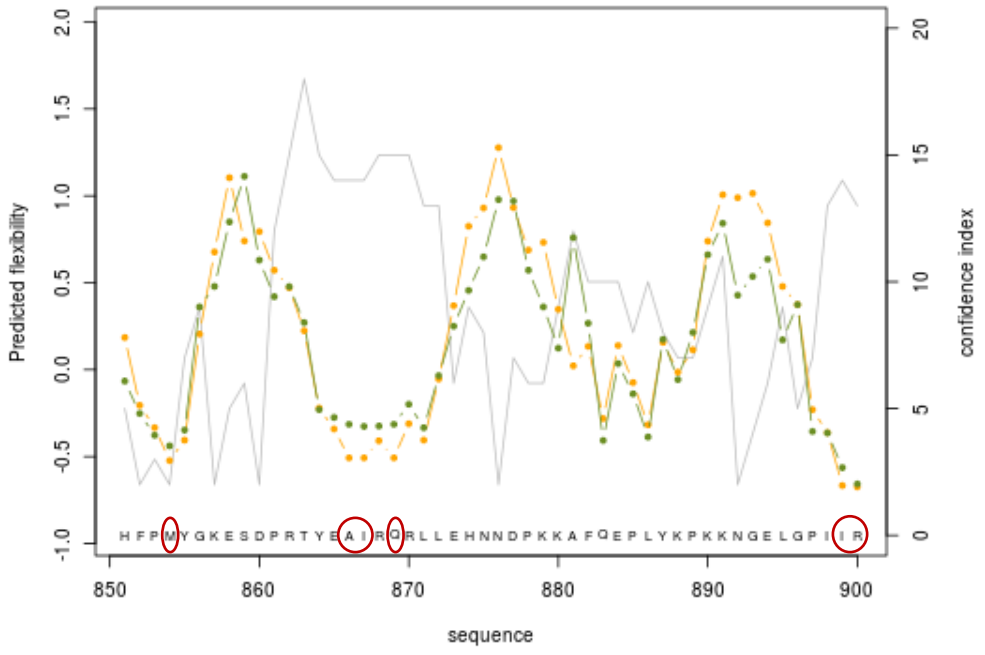

Predicted flexibility
confidence index
sequence
600
610
620
630
640
650
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20

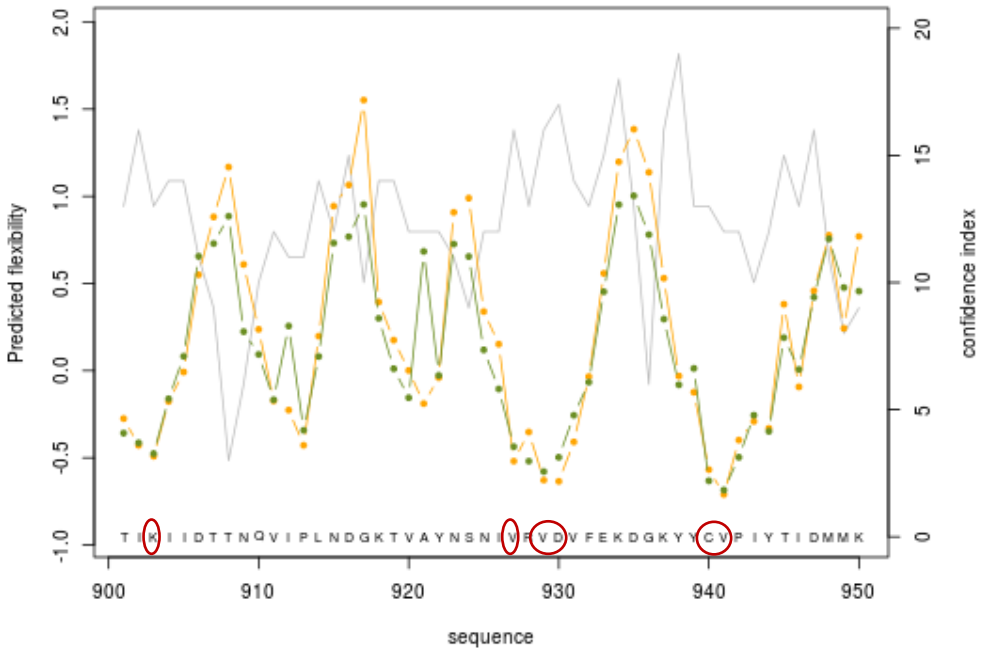

Predicted flexibility
confidence index
sequence
650
660
670
680
690
700
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20

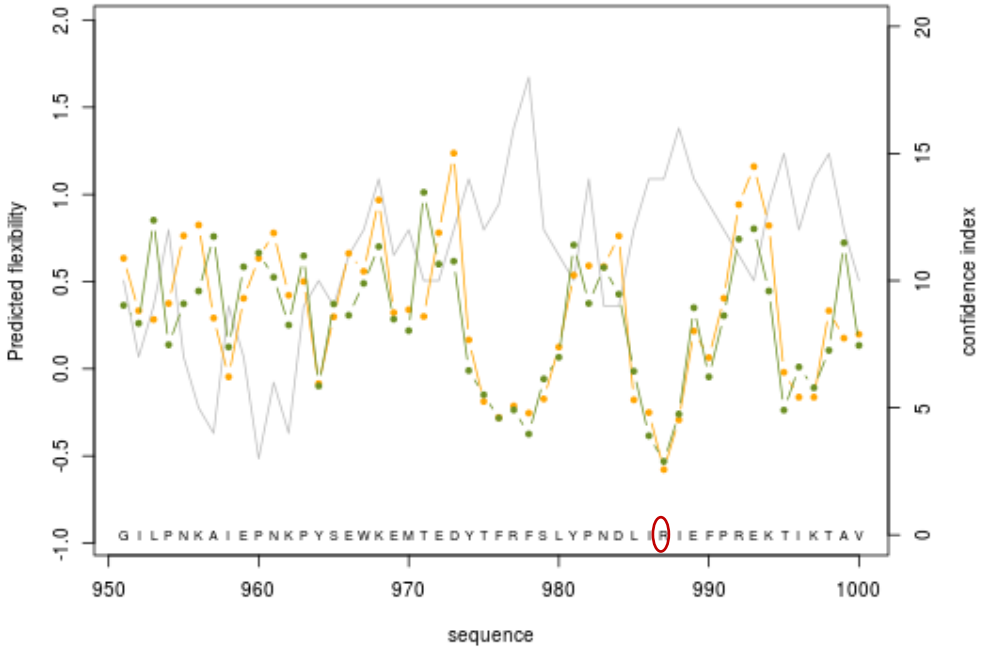

Predicted flexibility
confidence index
sequence
A H L R S R W N F N K N R E E S N L H H A V D A A I V A C T T P S D I A R V T A F Y Q R R E Q N K E
700
710
720
730
740
750

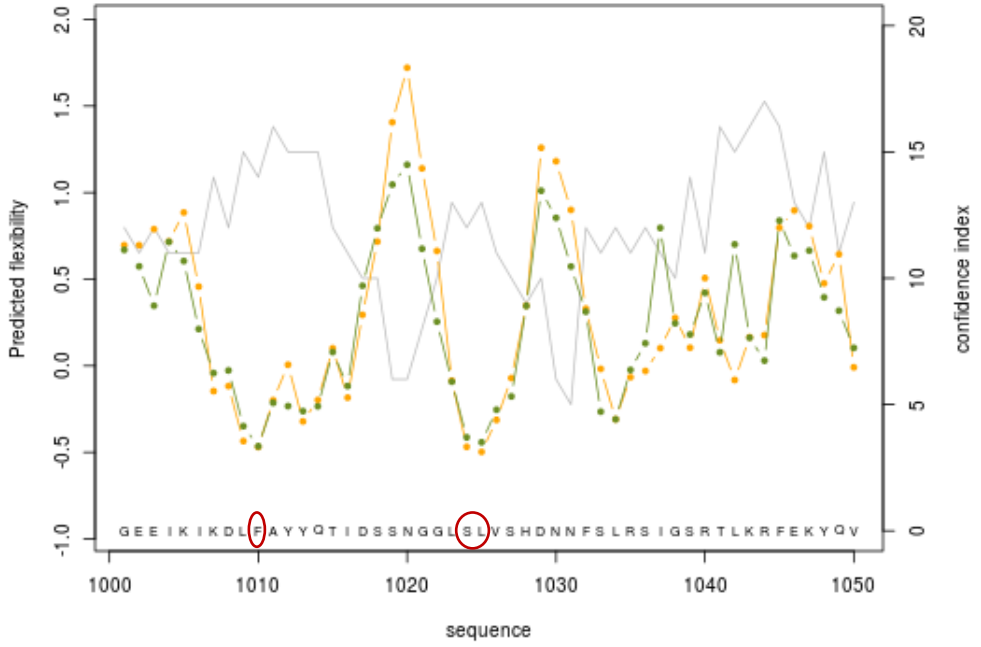

Predicted flexibility
confidence index
sequence
L S K K T D P Q F P Q P W P H F A D E L Q A R L S K N P K E S I K A L N L G N Y D N E K L E S L Q P
750
760
770
780
790
800

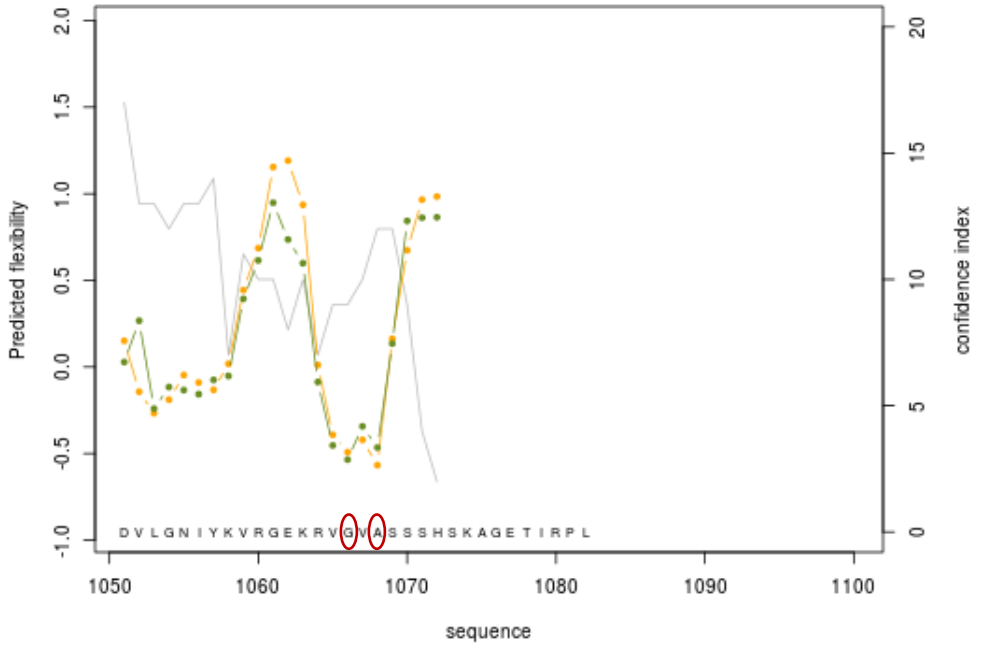

Predicted flexibility
confidence index
sequence
800
810
820
830
840
850


Predicted flexibility
confidence index
sequence
850
860
870
880
890
900

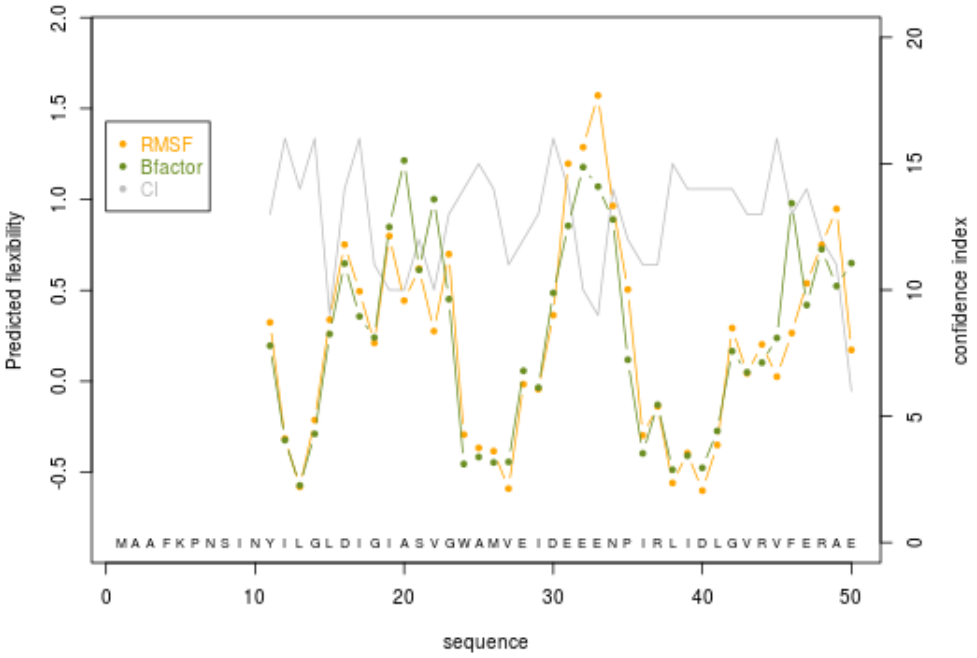

Predicted flexibility
confidence index
sequence
900
910
920
930
940
950

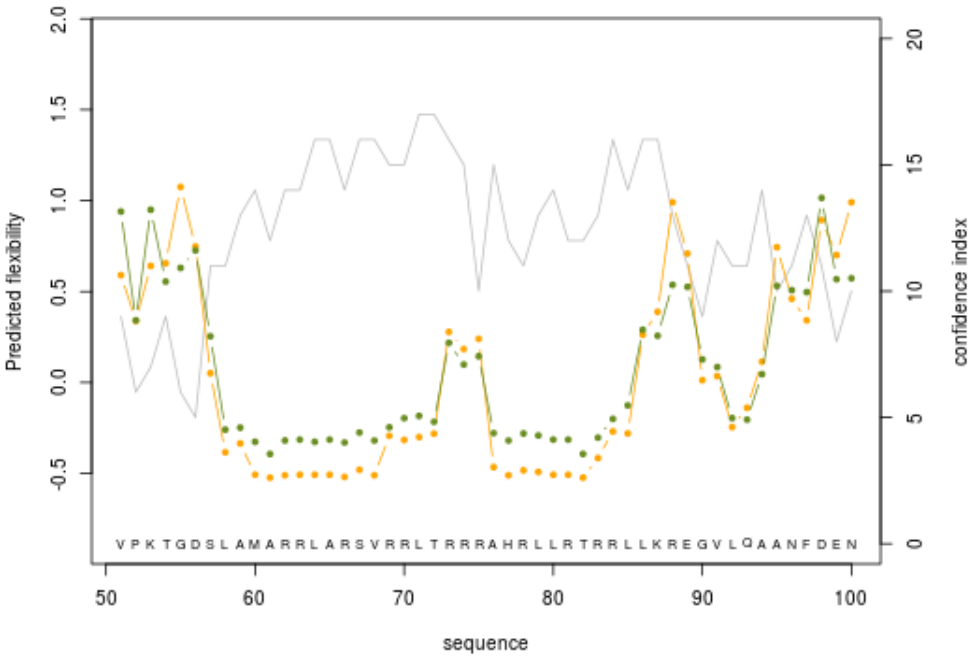

Predicted flexibility
confidence index
sequence
950
960
970
980
990
1000

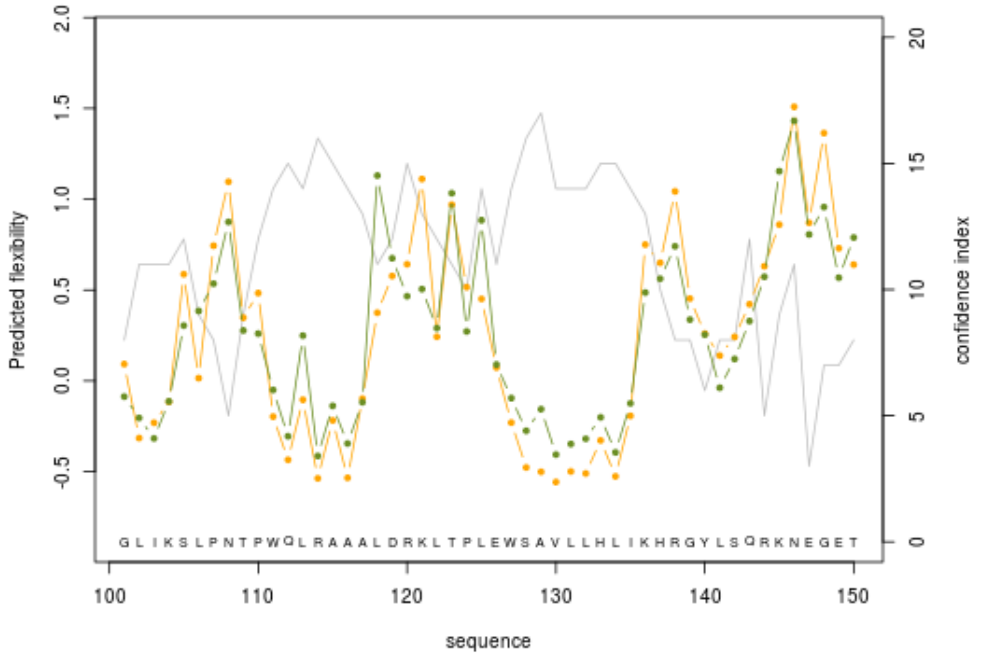

Predicted flexibility
confidence index
sequence
GEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQV
1000
1010
1020
1030
1040
1050

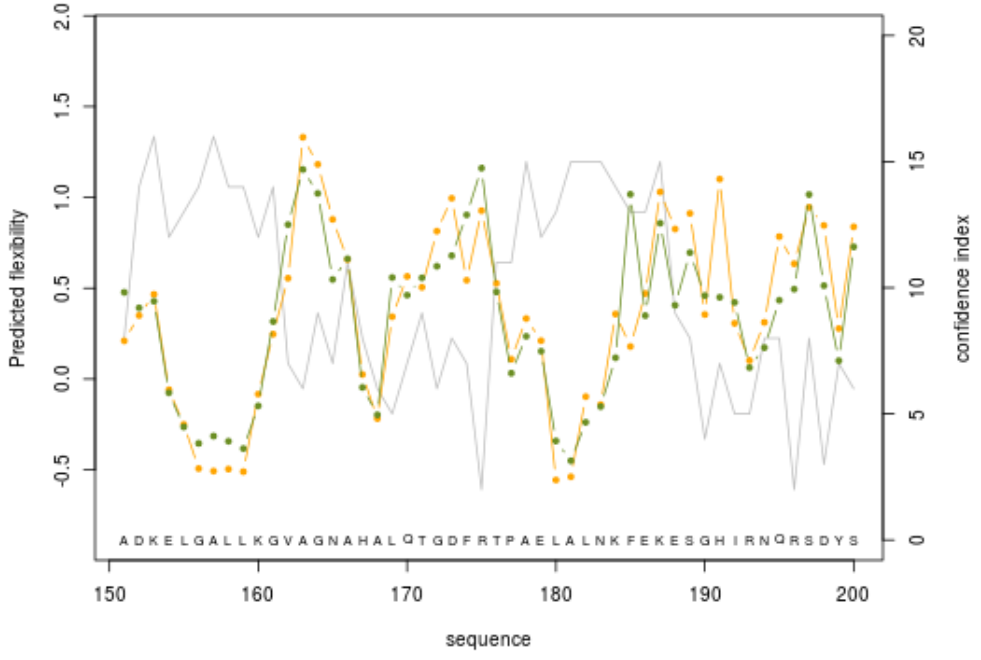

Predicted flexibility
confidence index
sequence
DVLGNIYKVRGEKRVGVASSSHSKAGETIRPL
1050
1060
1070
1080
1090
1100

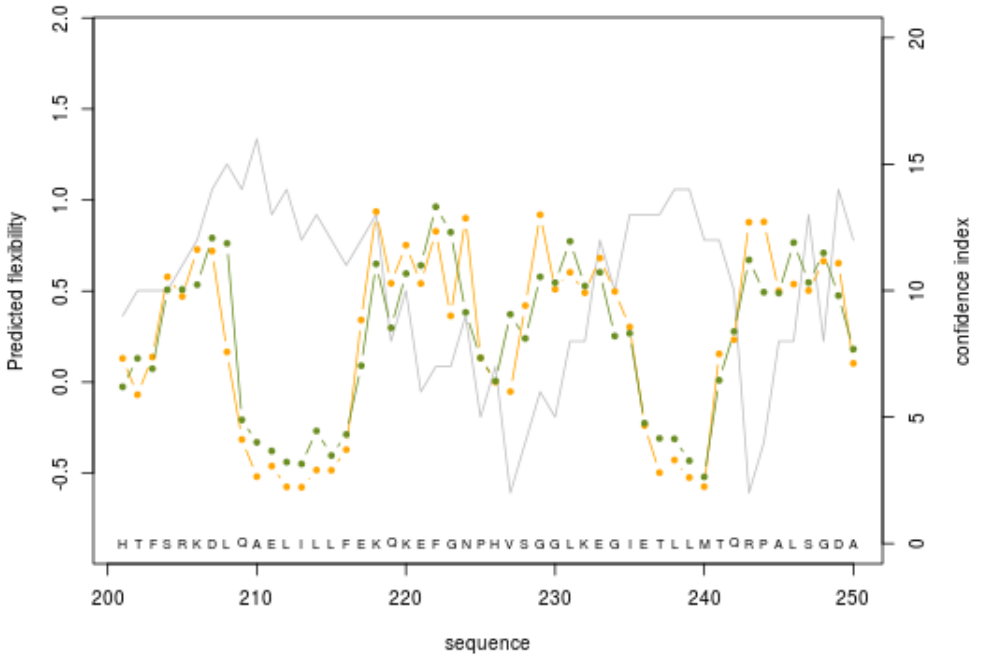
Supplementary Fig. 5 PredyFlexy analysis of the rigid site residues present in the ThermoCas9 protein. Graphs exhibit the predicted flexibility of each amino acid along the protein sequence. Dots depict individual measure points. Green and orange lines represent B-factor and RMSF prediction respectively deduded from flexibility class prediction. Gray line represents the confidence index. Accurate predictions show a high confidence index. Residues with predicted flexibility equal to or lower than 0.5 were considered rigid sites (red circles).

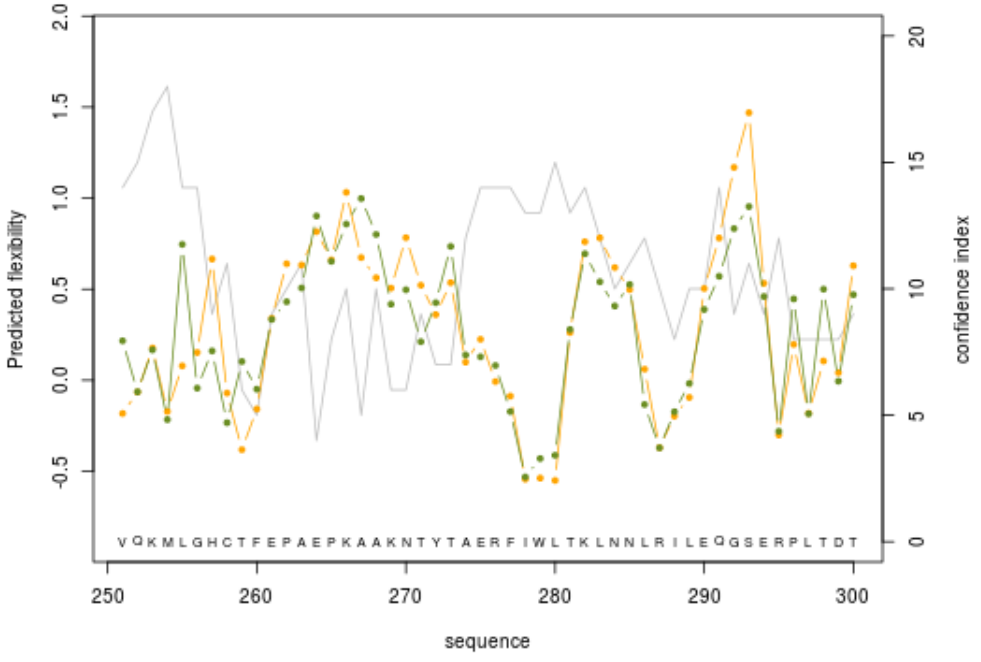

RMSF
Bfactor
CI
Predicted flexibility
confidence index
sequence
MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAE

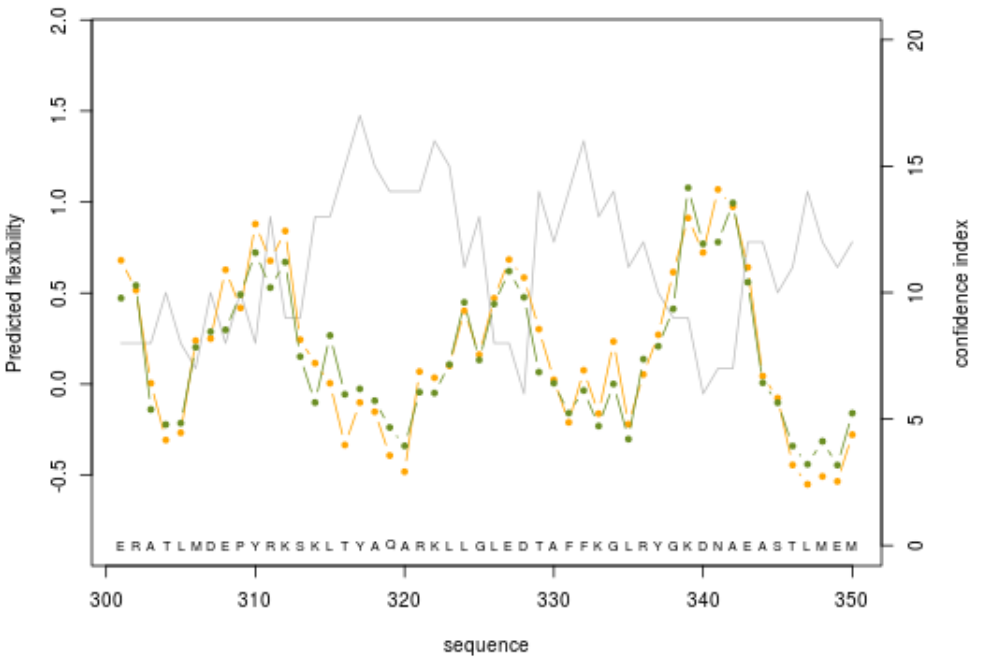

Predicted flexibility
confidence index
sequence
VPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDEN

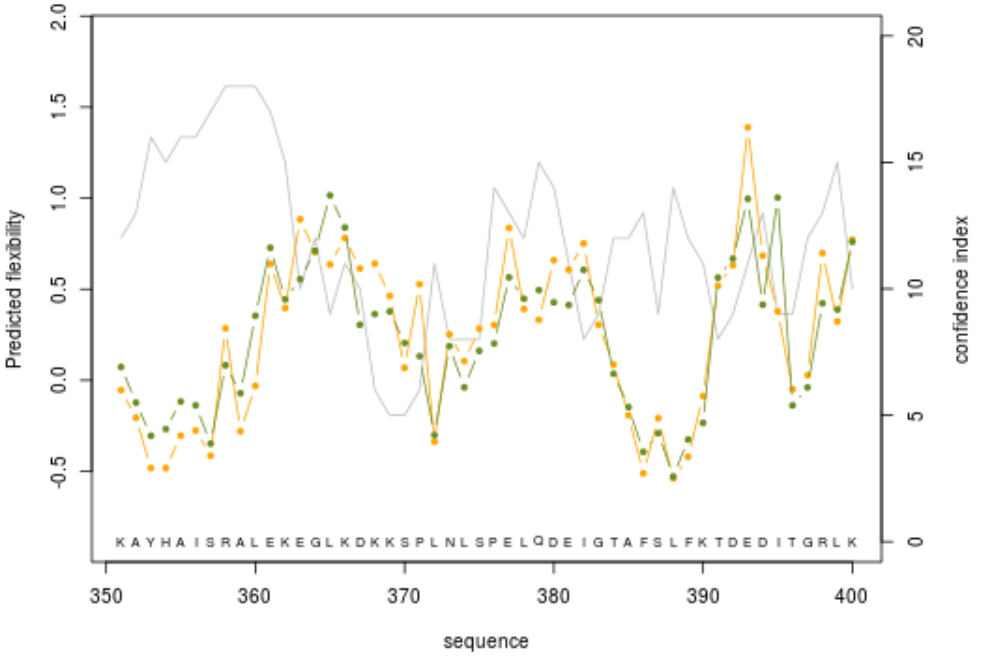

Predicted flexibility
confidence index
sequence
G L I K S L P N T P W Q L R A A A L D R K L T P L E W S A V L L H L I K H R G Y L S Q R K N E G E T
100
110
120
130
140
150

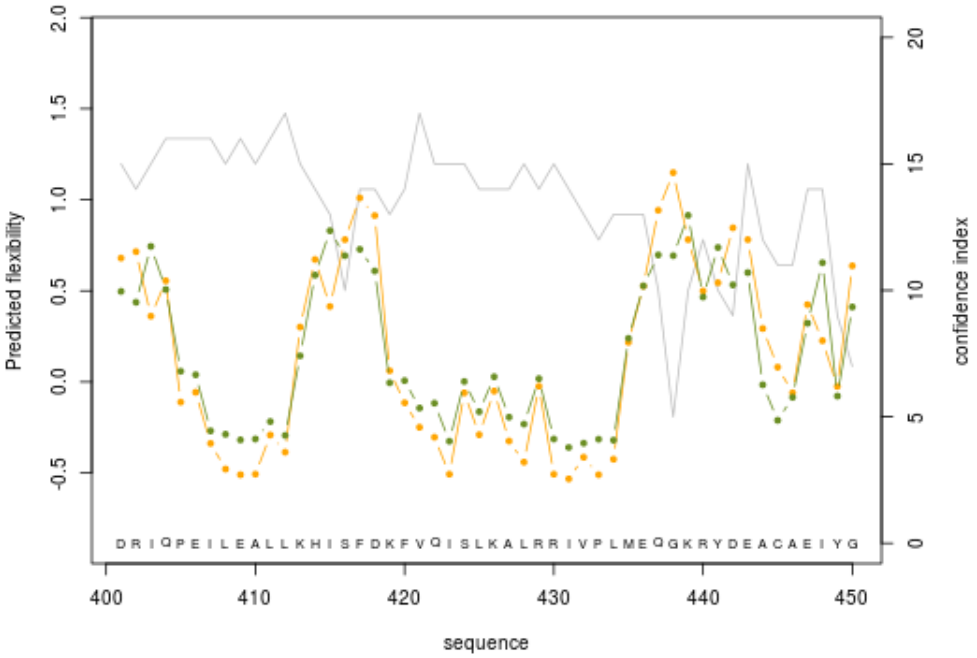

Predicted flexibility
confidence index
sequence
A D K E L G A L L K G V A G N A H A L Q T G D F R T P A E L A L N K F E K E S G H I R N Q R S D Y S
150
160
170
180
190
200

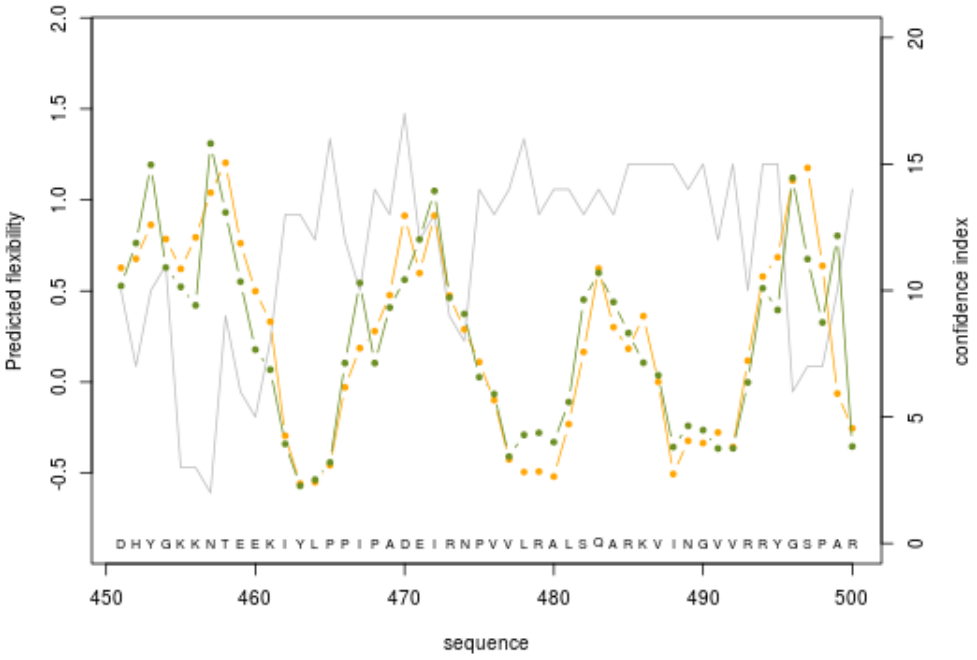

Predicted flexibility
confidence index
sequence
HTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA
200
210
220
230
240
250

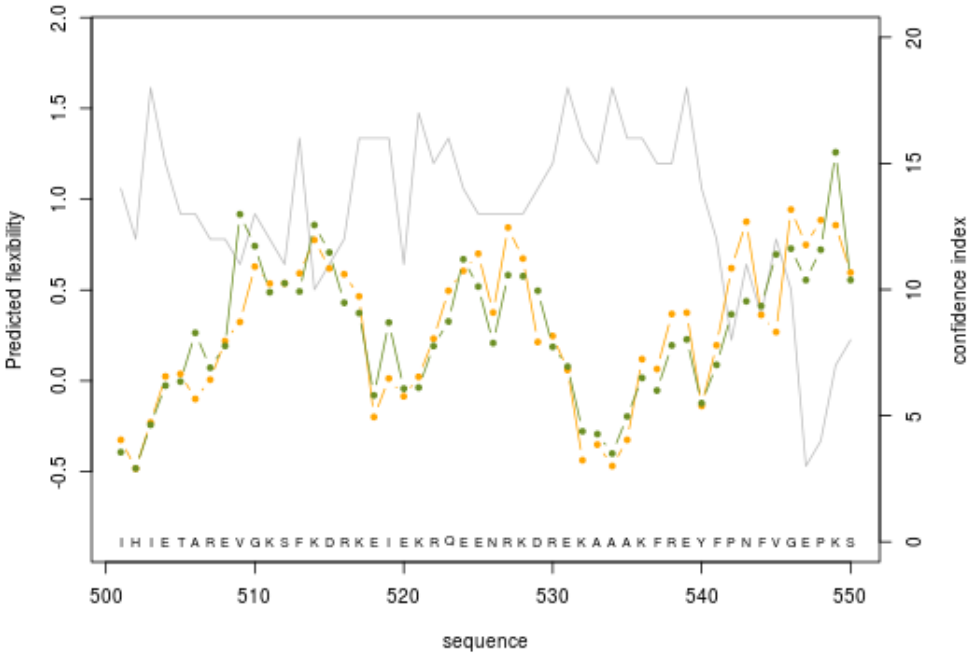

Predicted flexibility
confidence index
sequence
VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDT
250
260
270
280
290
300

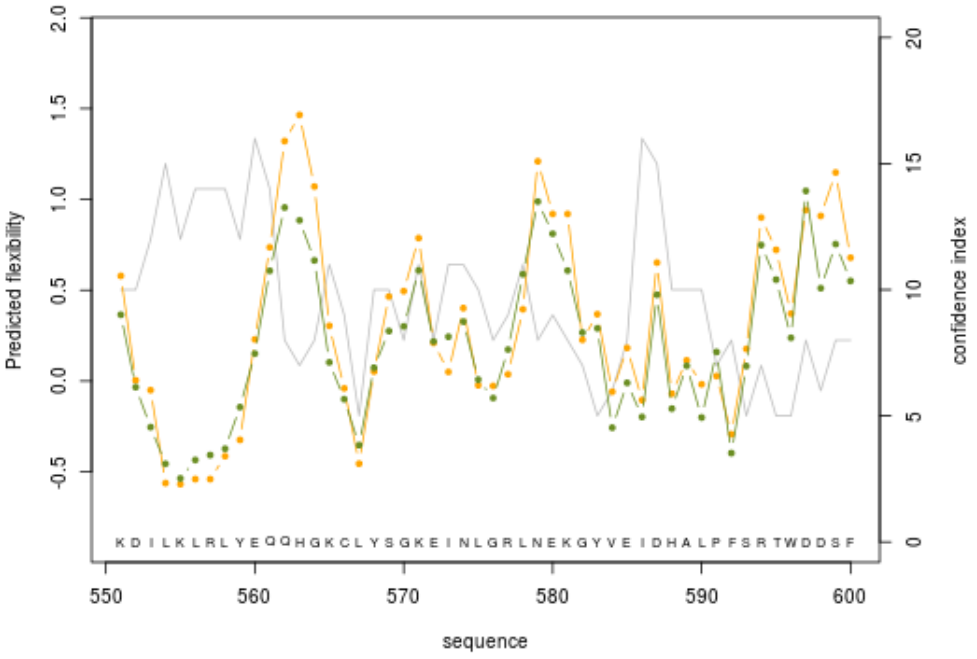

Predicted flexibility
confidence index
sequence
E R A T L M D E P Y R K S K L T Y A Q A R K L L G L E D T A F F K G L R Y G K D N A E A S T L M E M
300
310
320
330
340
350

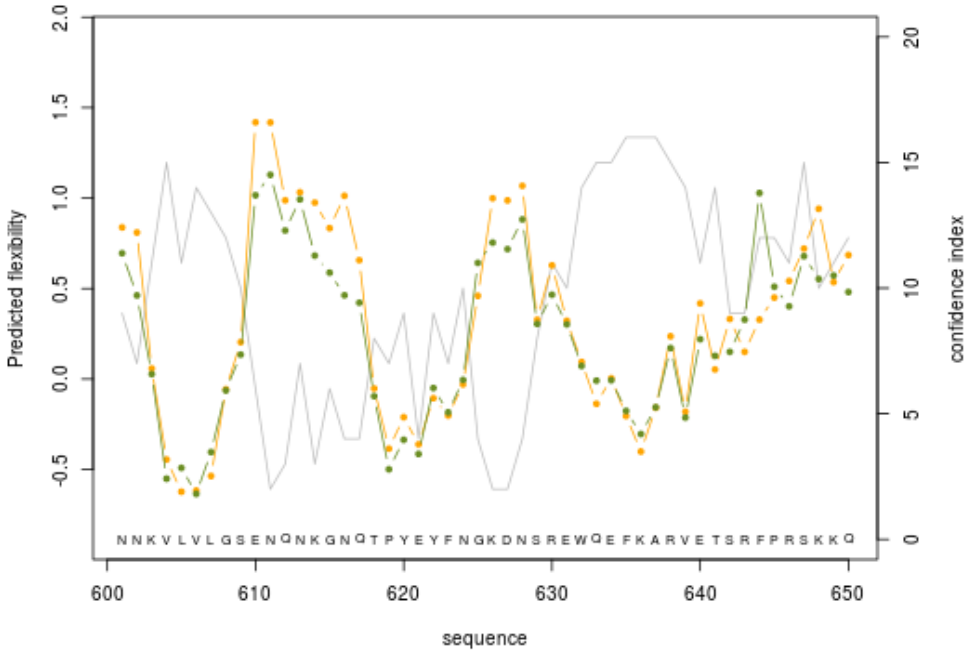

Predicted flexibility
confidence index
sequence
K A Y H A I S R A L E K E G L K D K K S P L N L S P E L Q D E I G T A F S L F K T D E D I T G R L K
350
360
370
380
390
400

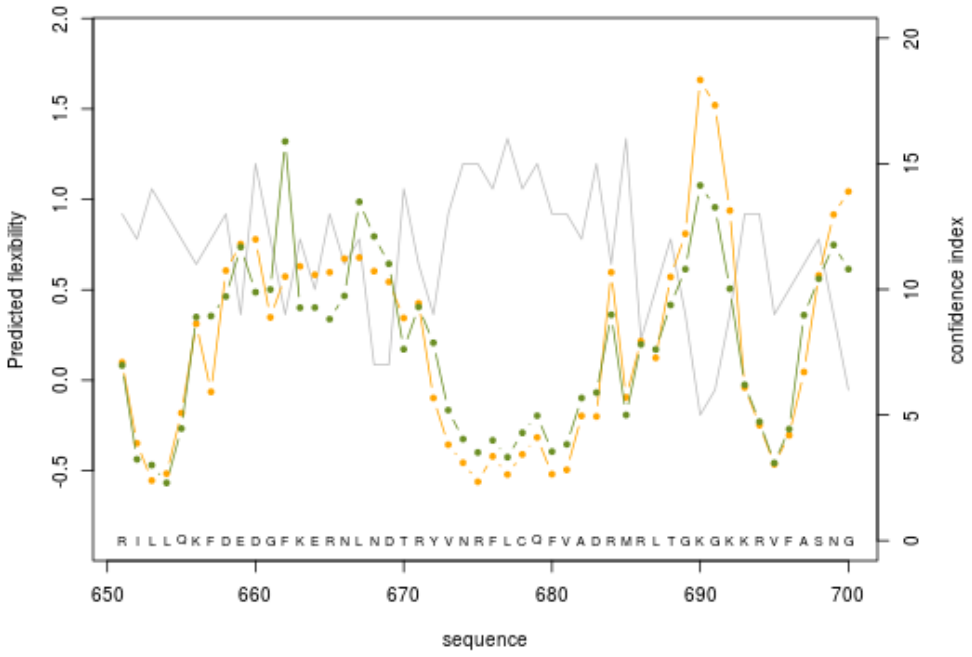

Predicted flexibility
confidence index
sequence
DRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYG
400
410
420
430
440
450

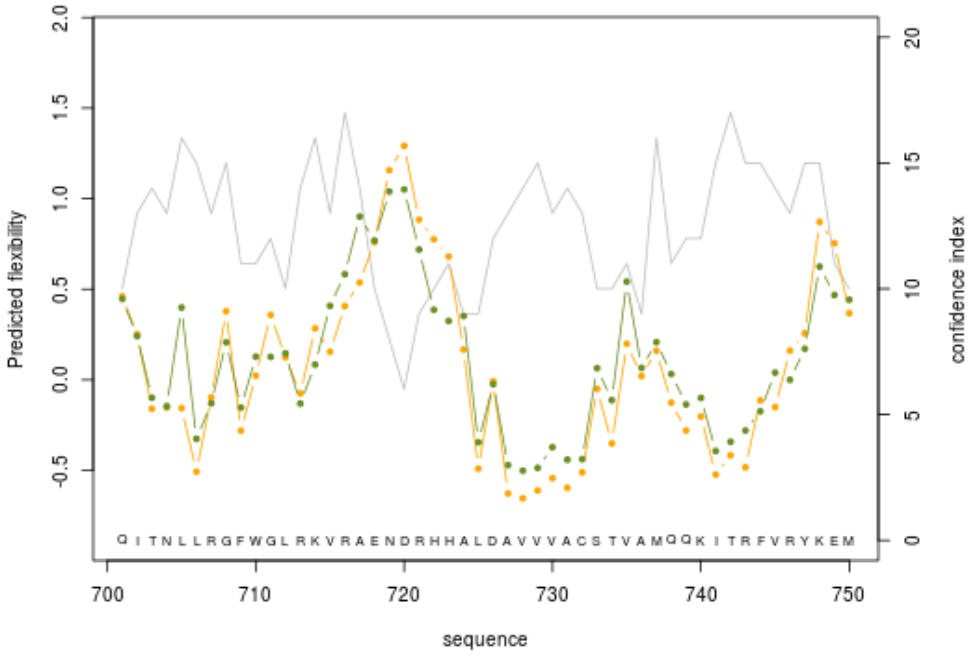

Predicted flexibility
confidence index
sequence
DHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPAR
450
460
470
480
490
500

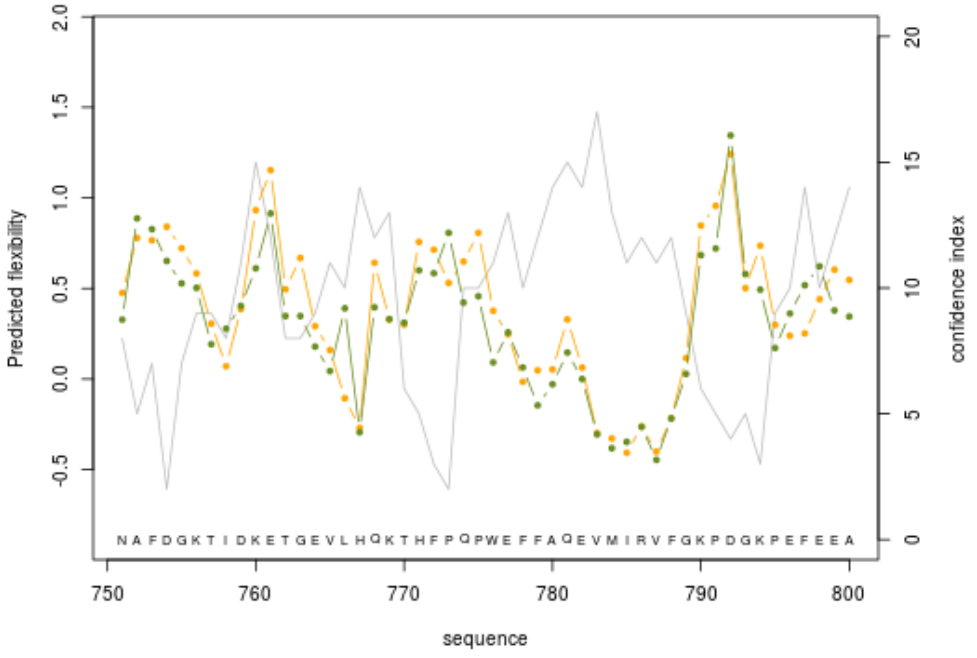

Predicted flexibility
confidence index
sequence
I H I E T A R E V G K S F K D R K E I E K R Q E E N R K D R E K A A A K F R E Y F P N F V G E P K S

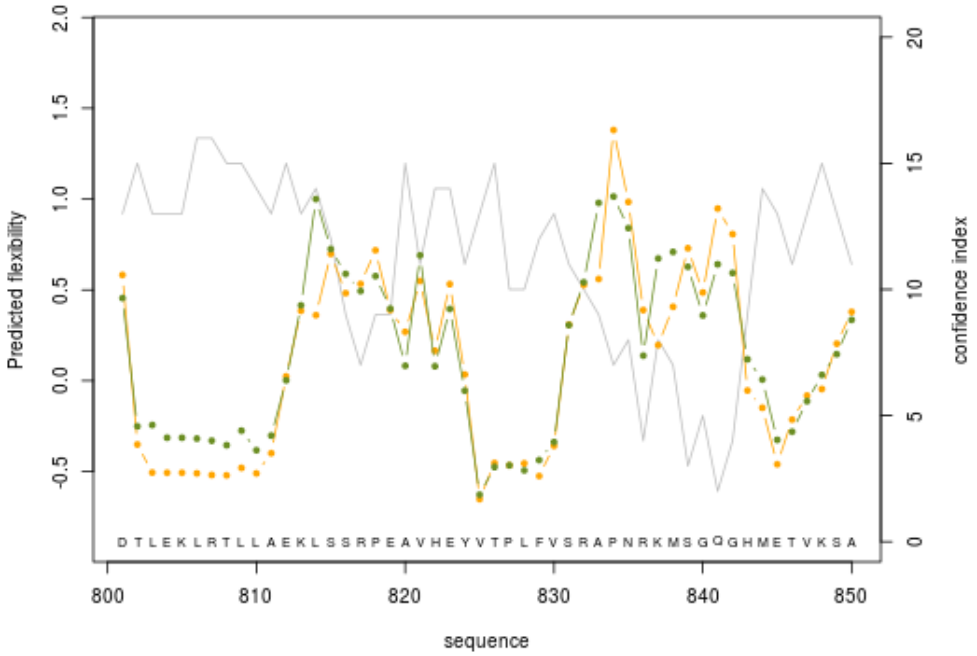

Predicted flexibility
confidence index
sequence
K D I L K L R L Y E Q Q H G K C L Y S G K E I N L G R L N E K G Y V E I D H A L P F S R T W D D S F

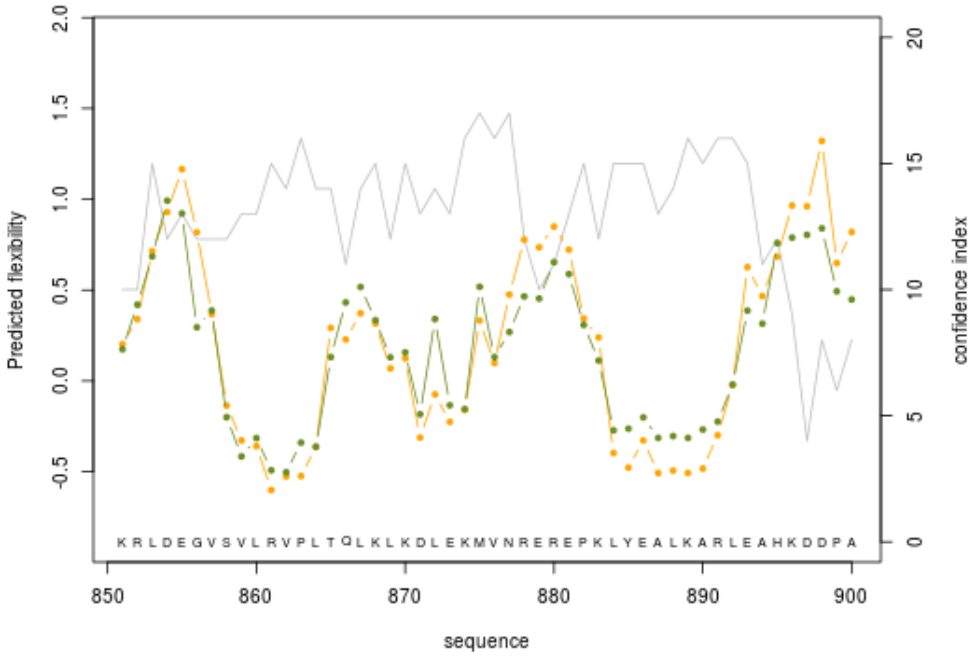

Predicted flexibility
confidence index
sequence
N N K V L V L G S E N Q N K G N Q T P Y E Y F N G K D N S R E W Q E F K A R V E T S R F P R S K K Q

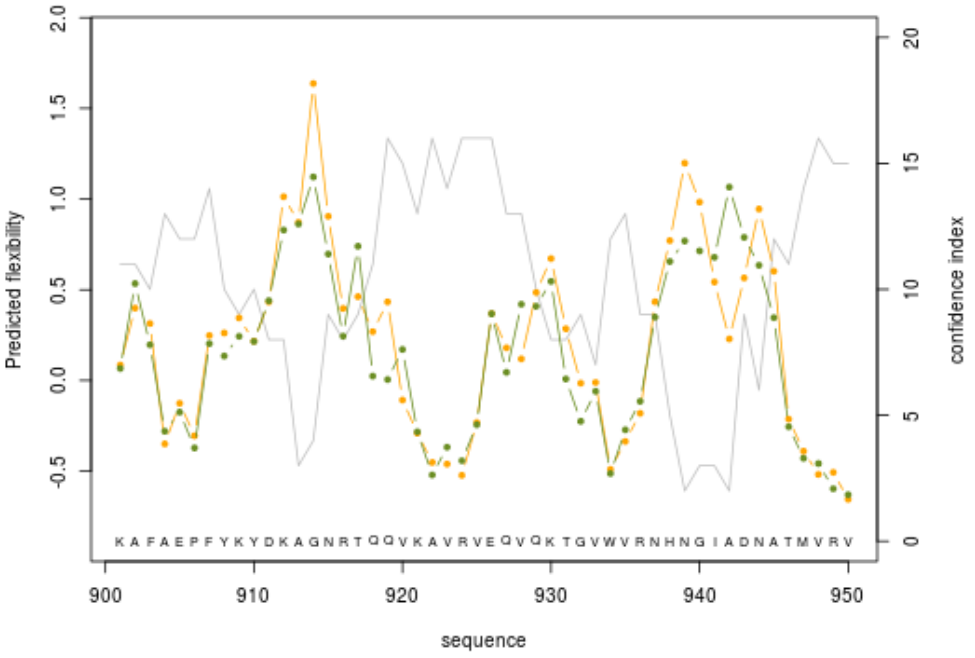

Predicted flexibility
confidence index
sequence
R I L L Q K F D E D G F K E R N L N D T R Y V N R F L C Q F V A D R M R L T G K G K K R V F A S N G

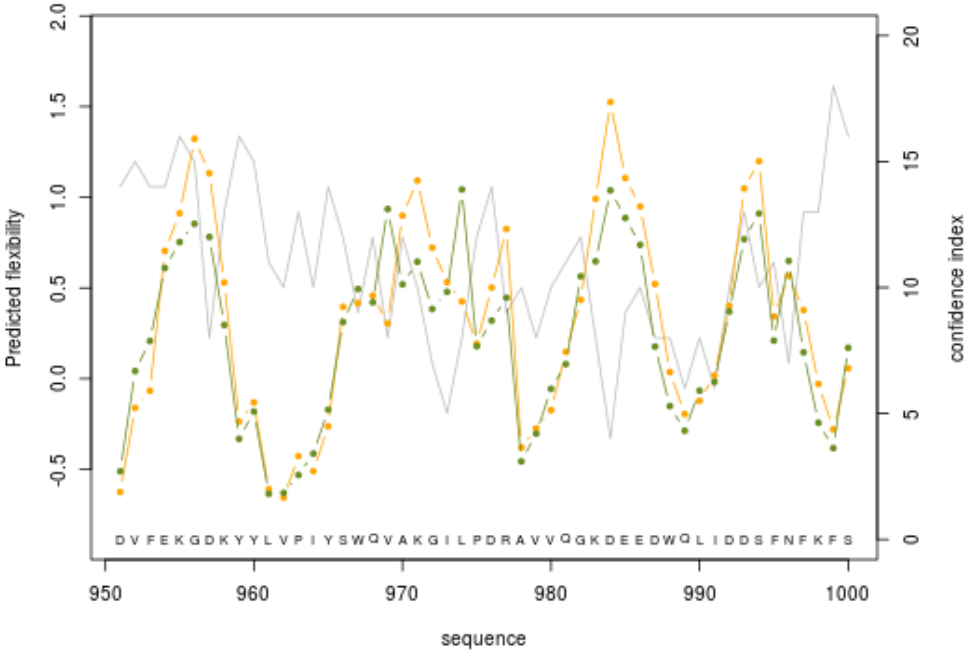

Predicted flexibility
confidence index
sequence
Q I T N L L R G F W G L R K V R A E N D R H H A L D A V V V A C S T V A M Q Q K I T R F V R Y K E M
700
710
720
730
740
750

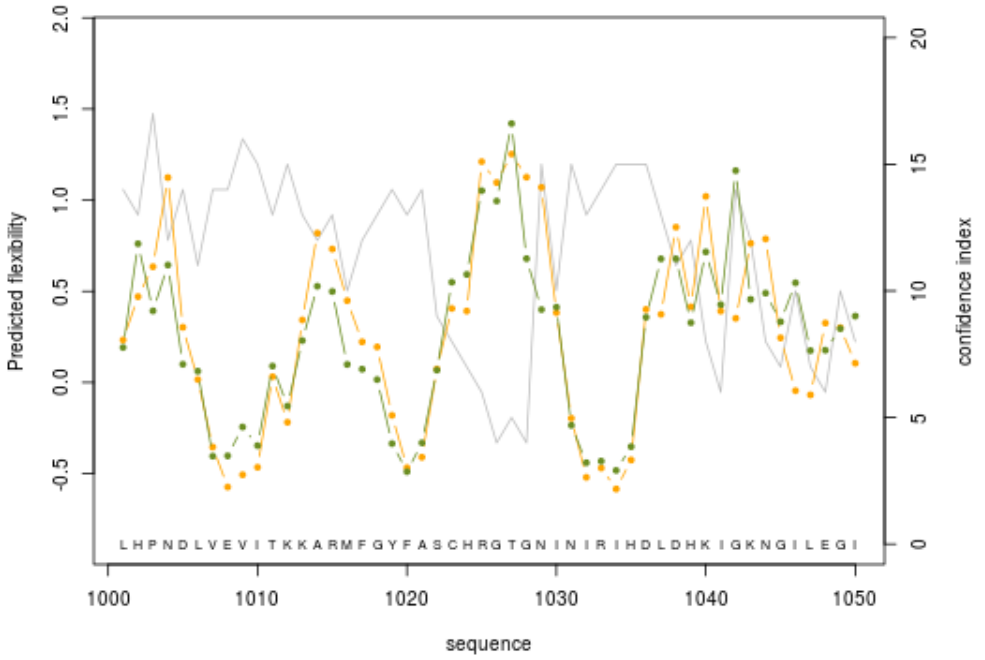

Predicted flexibility
confidence index
sequence
N A F D G K T I D K E T G E V L H Q K T H F P Q P W E F F A Q E V M I R V F G K P D G K P E F E E A
750
760
770
780
790
800

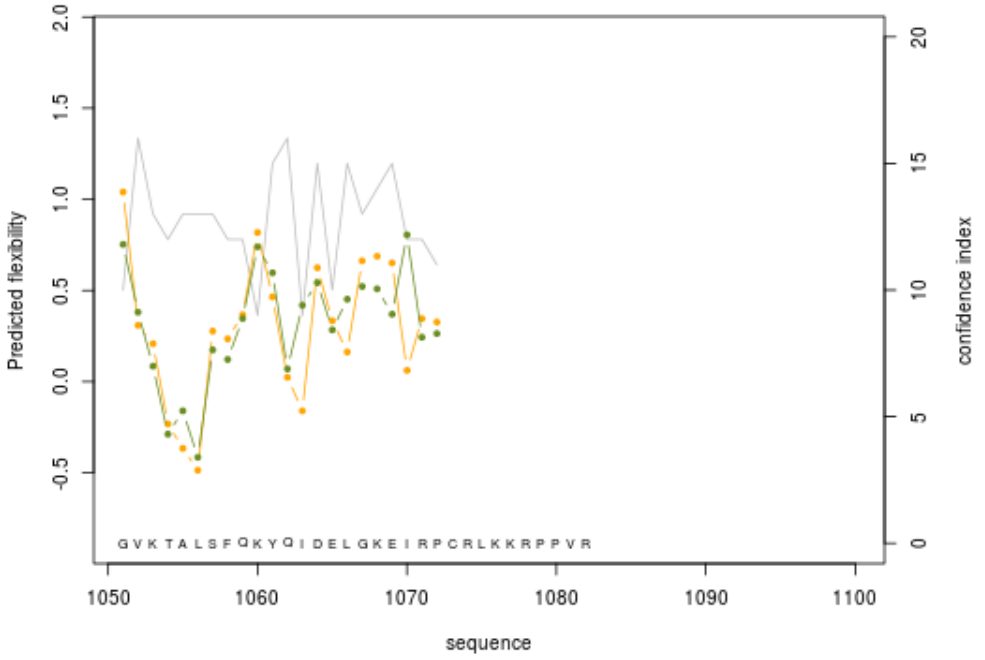

Predicted flexibility
confidence index
sequence
D T L E K L R T L L A E K L S S R P E A V H E Y V T P L F V S R A P N R K M S G Q G H M E T V K S A
800
810
820
830
840
850


Predicted flexibility
confidence index
sequence
K R L D E G V S V L R V P L T Q L K L K D L E K M V N R E R E P K L Y E A L K A R L E A H K D D P A
850
860
870
880
890
900

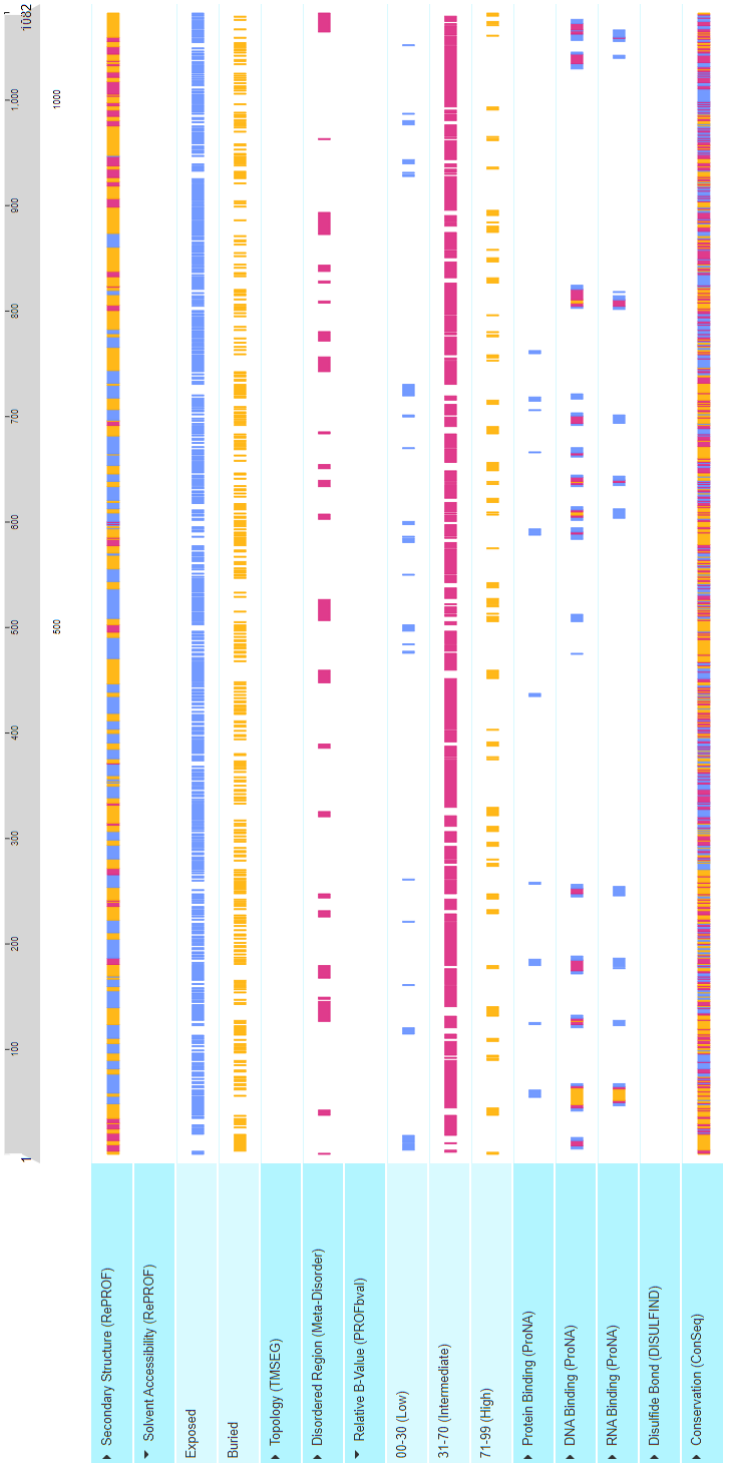

Predicted flexibility
confidence index
sequence
K A F A E P F Y K Y D K A G N R T Q Q V K A V R V E Q V Q K T G V W V R N H N G I A D N A T M V R V
900
910
920
930
940
950


Predicted flexibility
confidence index
sequence
D V F E K G D K Y Y L V P I Y S W Q V A K G I L P D R A V V Q G K D E E D W Q L I D D S F N F K F S
950
960
970
980
990
1000


Predicted flexibility
confidence index
sequence
L H P N D L V E V I T K K A R M F G Y F A S C H R G T G N I N I R I H D L D H K I G K N G I L E G I
1000
1010
1020
1030
1040
1050


Predicted flexibility
confidence index
sequence
G V K T A L S F Q K Y Q I D E L G K E I R P C R L K K R P P V R
1050
1060
1070
1080
1090
1100

Supplementary Fig. 6 PredyFlexy analysis of the rigid site residues present in the Nme1Cas9 protein. Graphs exhibit the predicted flexibility of each amino acid along the protein sequence. Dots depict individual measure points. Green and orange lines represent B-factor and RMSF prediction respectively deduded from flexibility class prediction. Gray line represents the confidence index. Accurate predictions show a high confidence index. Residues with predicted flexibility equal to or lower than 0.5 were considered rigid sites.



Supplementary Fig. 7 PredictProtein analysis of the rigid site residues present in the ThermoCas9 protein. Low relative B-value (0-30) represents rigid site residues, while high relative B-value (71-99) represents flexible site residues. For each amino acid, the solvent accessibility is also predicted (exposed or buried).

----> Slowest mode 1: PDB file

Rigid Part No	Residues	Score
1	A:1-50	0.43
2	A:51-455	0.99
3	A:456-1082	0.94

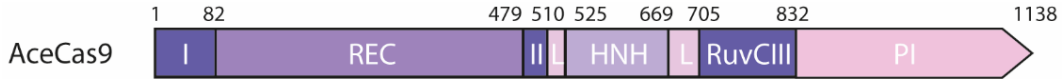
Hinge residues: 50A 455A

----> Slowest mode 2: PDB file

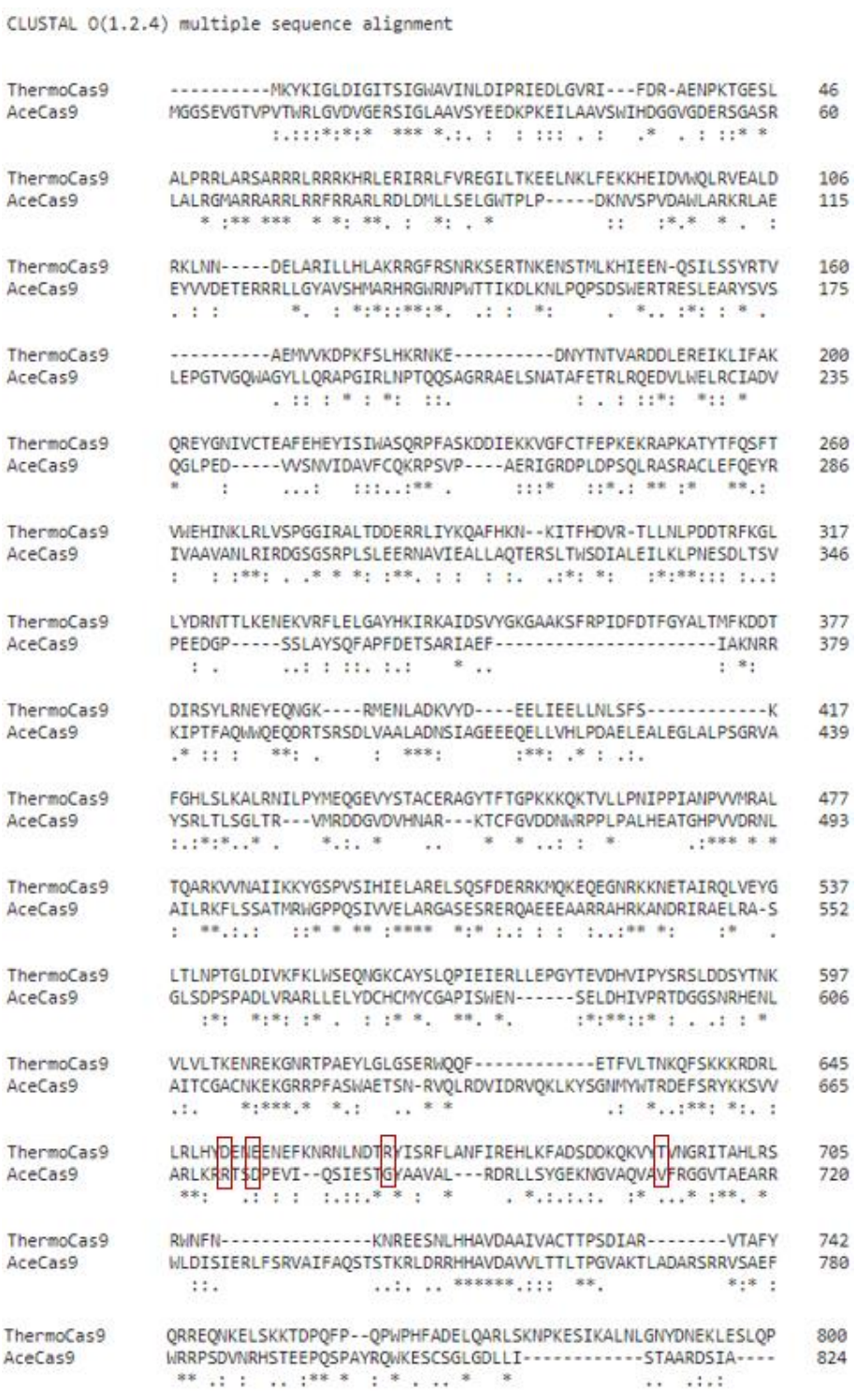
Rigid Part No	Residues	Score
1	A:1-56,447-806	0.96
2	A:57-446	0.99
3	A:807-1082	0.95

Hinge residues: 56A 446A 806A

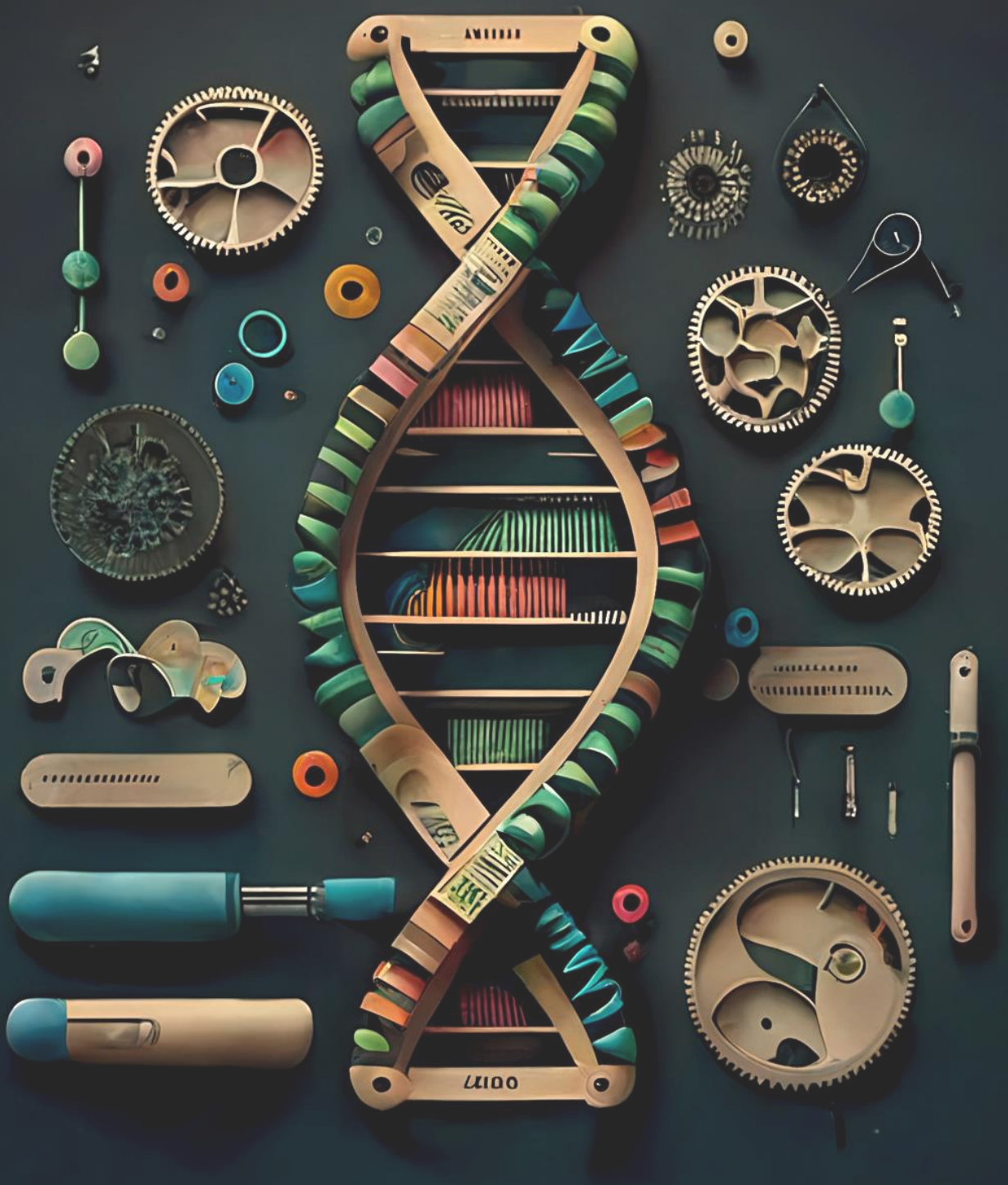
Supplementary Fig. 8 HingeProt analysis of the hinge site residues present in the ThermoCas9 protein.



Supplementary Fig. 9 Domain architecture of the AceCas9 protein.


CLUSTAL O(1.2.4) multiple sequence alignment
ThermoCas9 ----------MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRI---FDR-AENPKTGESL 46
AceCas9 MGGSEVGTVPVTWRLGVDVGERSIGLAAVSYEEDKPKEILAAVSWIHDGGVGDERSGASR 60
ThermoCas9 ALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALD 106
AceCas9 LALRGMARRARRLRRFRRARLRDLDMLLSELGWTPLP-----DKNVSPVDAWLARKRLAE 115
ThermoCas9 RKLNN-----DELARILLHLAKRRGFRSNRKSERTNKENSTMLKHIEEN-QSILSSYRTV 160
AceCas9 EYVVDETERRRLLGYAVSHMARHRGWRNPWTTIKDLKNLPQPSDSWERTRESLEARYSVS 175
ThermoCas9 ----------AEMVVKDPKFSLHKRNKE----------DNYTNTVARDDLEREIKLIFAK 200
AceCas9 LEPGTVGQWAGYLLQRAPGIRLNPTQQSAGRRAELSNATAFETRLRQEDVLWELRCIADV 235
ThermoCas9 QREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFT 260
AceCas9 QGLPED-----VVSNVIDAVFCQKRPSVP----AERIGRDPLDPSQLRASRACLEFQEYR 286
ThermoCas9 VWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKN--KITFHDVR-TLLNLPDDTRFKGL 317
AceCas9 IVAAVANLRIRDGSGSRPLSLEERNAVIEALLAQTERSLTWSDIALEILKLPNESDLTSV 346
ThermoCas9 LYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDT 377
AceCas9 PEEDGP-----SSLAYSQFAPFDETSARIAEF----------IAKNRR 379
ThermoCas9 DIRSYLRNEYEQNGK----RMENLADKVYD----EELIEELLNLSFS-----------K 417
AceCas9 KIPTFAQWWQEQDRTSRSDLVAALADNSIAGEEEQELLVHLPDAELEALEGLALPSGRVA 439
ThermoCas9 FGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRAL 477
AceCas9 YSRLTLSGLTR---VMRDDGVDVHNAR---KTCFGVDDNWRPPLPALHEATGHPVVDRNL 493
ThermoCas9 TQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYG 537
AceCas9 AILRKFLSSATMRWGPPQSIVVELARGASESRERQAEEEAARRAHRKANDRIRAELRA-S 552
ThermoCas9 LTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNK 597
AceCas9 GLSDPSPADLVRARLLELYDCHCMYCGAPISWEN------SELDHIVPRTDGGSNRHENL 606
ThermoCas9 VLVLTKENREKGNRTPAEYLGLGSERWQQF----------ETFVLTNKQFSKKKRDRL 645
AceCas9 AITCGACNKEKGRRPFASWAETSN-RVQLRDVIDRVQKLKYSGNMYWTRDEFSRYKKSVV 665
ThermoCas9 LRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRS 705
AceCas9 ARLKRRTSDPEVI--QSIESTGYAAVAL---RDRLLSYGEKNGVAQVAVFRGGVTAEARR 720
ThermoCas9 RWNFN----------KNREESNLHHAVDAAIVACTTPSDIAR--------VTAFY 742
AceCas9 WLDISIERLFSRVAIFAQSTSTKRLDRRHHAVDAVVLTTLTPGVAKTLADARSRRVSAEF 780
ThermoCas9 QRREQNKELSKKTDPQFP--QPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQP 800
AceCas9 WRRPSDVNRHSTEEPQSPAYRQWKESCSGLGDLLI----------STAARDSIA---- 824

Supplementary Fig. 10 ClustalO alignment of the ThermoCas9 with AceCas9 protein sequence for the identification of hinge loop residues crucial for high catalytic activity.



Chapter 8

Summary & General Discussion

Despoina Trasanidou

Summary

Languishing in the quagmire of food shortage and devastating diseases, humanity has always been pursuing novel ways to improve the quality of life. As described in **Chapter 1**, the millennia-old practices of domestication and selective breeding of agricultural plants and animals were first subjected to the magnifying glass of scientists around 240 years ago[1]. Studying the evolution and genetics of wild fauna and flora, the restless spirit of humankind quickly led to the first attempts of direct but random genetic intervention, which were followed by the discovery and manipulation of the molecule of life (DNA)[2-8]. Around 30 years ago, the first genetically modified organisms were commercialized, and, shortly after, the sequencing technology gave rise to the development of primary tools for site−directed genomic modification[9-17]. However, these tools were inefficient, required selectable markers, and were leaving genomic scars[18-21]. Hence, efficient and programmable endonucleases (meganucleases, Zinc Finger Nucleases, and TALE Nucleases) entered the scene, though necessitating laborious, expensive, and time−consuming protein engineering[22-23]. The discovery, characterization and repurposing of CRISPR-Cas systems during the last two decades have revolutionised genetic engineering, providing simple, cost-effective, and quick methods for site-specific genome editing and transcriptional regulation[24-29]. CRISPR-Cas is a highly diverse set of prokaryotic adaptive immune systems against viruses, plasmids, and conjugative elements. Upon recognition and cleavage of foreign genetic material, a short DNA fragment is integrated into the CRISPR array (adaptation), forming immunological memory. The CRISPR array is transcribed into a long pre-crRNA molecule that is processed into short, mature crRNAs (expression) that guide Cas nucleases to target and cleave the DNA or RNA of cognate invaders[30-35]. CRISPR-Cas systems are classified into two classes, six types and several subtypes[36]. Due to their widespread presence and the eternal evolutionary arms race between cellular host organisms and their MGE invaders, CRISPR-Cas systems exhibit an impressive plethora and variety, part of which has been harnessed for genetic engineering purposes. As such, genome editing is performed using DNA nucleases, DNA base-editors, prime-editors, and Cas-related transposases or recombinases[37-44]. Transcriptional regulation takes place through Cas-related transcriptional effector domains, RNA nucleases, and RNA base-editors[45-48]. The excessive activity of CRISPR-Cas tools is controlled via ligand-inducible expression, complementary oligonucleotides, and virus-encoded anti-CRISPR proteins[49-51]. CRISPR-Cas technology currently holds the reins of genetic engineering, being constantly optimised for safe and

efficient applications in gene therapy, crop improvement, and industrial strain development.

One of the latest advances of CRISPR-Cas technology is base-editing, which enables precise genome and transcriptome editing without the need of DSDBs and donor DNA templates. Base-editors (BEs) are synthetic fusions of a partially or completely inactivated Cas nuclease and one or more deaminase enzyme(s), that use a designed crRNA guide to find a specific target site for DSDB-less editing[37,38]. **Chapter 2** introduces the evolutionary divergence of cytidine and adenosine deaminases (CDs and ADs) as well as their classification, molecular mechanisms, and biological roles. Prokaryotic CDs are categorized into the CDD/CDA-like and the BTCDAs families, while eukaryotic CDs are grouped into AID/APOBEC, the non-AID/APOBEC, and the UCDARs families[52-55]. Similarly, prokaryotic ADs are classified into ADAs and TadA, whilst eukaryotic ADs are divided into ADAs, ADARs, ADATs, ADADs, and UADARs[55-60]. Most deaminases are essential for survival and play a major role in normal organ function and development. Dysregulation of these enzymes may lead to a wide plethora of diseases, including different forms of cancer. Hence, many eukaryotic deaminases have been thoroughly characterised during the last decades[52-60]. Moreover, BTCDA, AID/APOBEC, TadA, and ADAR deaminases have recently been used in base-editing applications[37,61-63]. Five distinct groups of BEs have been developed to date, including those converting C•G-to-T•A (cytidine base-editors, CBEs), A•T-to-G•C (adenine base-editors, ABEs), C•G-to-G•C (cytidine-to-guanine base-editors, CGBEs), C•G-to-A•T (cytidine-to-adenine base-editors, CBEs), and simultaneously C•G-to-T•A and A•T-to-G•C (dual base-editors, DBEs) [37,62,64,65]. Exploitation of alternative deaminase and/or Cas protein variants may expand the targeting scope, optimize the precision and product purity as well as facilitate the delivery of BEs**.**

Interestingly, some predicted ADs, except from their AD domain, additionally bear a CARF-domain (CARF-AD proteins)[66]. **Chapter 3** describes that CARF-ADs are found exclusively in prokaryotes, and associate with type III CRISPR-Cas systems as well as with enzymes involved in purine recycling. Specifically, a CARF-AD enzyme from the thermophilic bacterium *Methylacidiphilum fumariolicum* (mfCARF-AD**)** is predicted to not only sense but also degrade cOAs produced after activation of its neighbouring type III-B CRISPR-Cas system. However, experimental analyses revealed that mfCARF-AD does not show deaminase activity on free adenosines, free adenines, AMP, and ATP. Possible deaminase activity on cOAs, RNA, DNA, dATP, NAD, NADP, FAD, or CoenzymeA (CoA) cannot be excluded. Moreover, mfCARF-AD is predominantly found in an oligomeric state, either in

the presence or the absence of cOA molecules. The exact biological role and the substrate specificity of mfCARF-AD remains elusive.

The vast majority of base-editors and of other Cas effectors, such as nucleases, prime-editors, and transcriptional regulators, present undesired off-target effects, usually due to their low specificity and/or excessive activity[67]. During the last decade, natural inhibitors of Cas proteins, known as anti-CRISPRs, have been detected in viral genomes, prophage regions, and plasmids, as introduced in **Chapter 4**. These tiny proteins (typically 50–330 amino acids) show highly variable sequences, structures, and mechanisms[68]. Therefore, they are classified based on the type of CRISPR-Cas systems that they block. Nevertheless, anti-CRISPR genes usually co-occur with genes that code for a protein that uses a helix-turn-helix (HTH) motif to control the expression of the anti-CRISPR protein. Several of these anti-CRISPR-associated genes (aca's) have been identified today and have been used for homology-based discovery of new anti-CRISPRs ('guilt-by-association approach')[69]. Alternatively, presence of anti-CRISPR gene(s) in prophage regions is revealed by spacers in the CRISPR-array that target the host genome ('self-targeting spacers')[70]. The last method of anti-CRISPR gene identification is the transformation of metagenomic libraries in an anti-CRISPR-selection strain, usually followed by synthetic genetic circuit-based selection for CRISPR-Cas inhibition[71]. Despite the quick obstruction of CRISPR-Cas immunity, collaboration and/or a certain concentration of anti-CRISPRs is often required for successful host immunosuppression[72]. On the other hand, prophage-encoded anti-CRISPRs may be beneficial for the prokaryotic host by enhancing horizontal gene transfer and thus host fitness in the population[73]. Most characterized anti-CRISPRs hamper the interference stage of CRISPR-Cas immunity. Class 1 anti-CRISPRs mainly block binding of the Cascade complex to its target DNA, through steric occlusion or by DNA mimicry. Alternatively, some also prevent the recruitment of the Cas3 nuclease to the Cascade:dsDNA complex[74-83]. Class 2 anti-CRISPRs interrupt Cas9 binding to the target DNA, via DNA mimicry and steric occlusion, by DNA and sgRNA mimicry, or through inactivation of Cas9 by causing its dimerization. Sometimes, they block Cas cleavage activity, by interacting directly with the protein (catalytic sites of Cas9, Cas13b) or indirectly with associated DNA and/or RNA molecules[70,84-86]. This impressive arsenal of inhibitory mechanisms have been exploited for spatiotemporal and conditional control of (d)Cas proteins in genome editing applications, among others[70,87,88]. The continuous discovery and structural/ functional characterization of anti-CRISPR proteins as well as the development of novel anti-CRISPR applications may further expand the CRISPR-Cas technologies.

Chapter 5 delves into the *in vivo* characterization of an anti-CRISPR protein (AcrIIC1) from *Neisseria meningitidis*[84] for the development of novel anti-CRISPR applications. AcrIIC1 robustly blocks the DNA cleavage activities of the thermoactive type II-C Cas9 variants from *Geobacillus thermodenitrificans* T12 (ThermoCas9)[89] and *Geobacillus stearothermophilus* (GeoCas9)[90] in *E. coli*. As such, it successfully lowers or completely hampers ThermoCas9- and GeoCas9-based genome editing. Although AcrIIC1 prevents DNA cleavage, it still allows the *in vivo* DNA binding of ThermoCas9 and GeoCas9 to their targets. Hence, the AcrIIC1:Cas9 complexes can be used for transcriptional regulation in *E. coli*, presenting efficiencies comparable to those of their dCas9 counterparts. Due to their stable DNA binding, AcrIIC1:Cas9 complexes were further harnessed for base-editing purposes, by fusing the active ThermoCas9 or GeoCas9 protein with a cytidine deaminase enzyme from *Petromyzon marinus* (PmCDA1)[38]. Compared to their dCas9 base-editors (dThermoTarget-AID; dGeoTarget-AID), the resulting anti-CRISPR base-editors (AcrThermoTarget-AID; AcrGeoTarget-AID) have the advantage of active Cas9-mediated counter-selection of unedited cells by relieving the anti-CRISPR-mediated inhibition of DNA cleavage activity of Cas9. This results in enrichment for clones with clean mutations. Both the dCas9 and AcrIIC1:Cas9 base-editors exhibit high efficiencies and large editing windows (reaching up to 30 base pairs), preferably editing the PAM distal or the PAM proximal end of the protospacer, respectively. Overall, these type II-C editors mediate efficient base-editing in wide activity windows and multiple positions simultaneously, facilitating gene inactivation in bacteria in the absence of donor DNA.

After proving that ThermoCas9 is active *in vivo* at 37°C in the absence of AcrIIC1 and that it can be used for targeting, silencing, genome editing and base-editing in *E. coli*, its activity was studied in human cells. **Chapter 6** shows that ThermoCas9 is able to introduce DSDBs in human genomes at 37°C, inducing NHEJ-mediated indel formation to disrupt targeted genes for the generation of knock-out cell lines. The cleavage efficiencies of ThermoCas9 are comparable to those of the traditionally used SpyCas9 nuclease. In addition, fusion of nThermoCas9 with a cytidine deaminase from rat (rAPOBEC1)[37] has resulted in the ThermoBE4 system that enables multiple, single nucleotide modifications in a three times larger window of activity compared to the corresponding SpyCas9 base-editor (BE4), facilitating gene mutagenesis. Altogether, ThermoCas9 can be used as an alternative platform for efficient genome- and base-editing applications in human cells, expanding in parallel the targeting scope.

Although ThermoCas9 is relatively stable and precise at moderate temperatures, its catalytic activity could be further improved for genome editing applications at mesophilic

bacteria, human cells, and plants. **Chapter 7** describes initial steps of random and rational engineering of ThermoCas9 for enhanced cleavage activity at temperatures between 20-37°C. Using an *in vivo* selection, toxin-targeting system in *E. coli*, ThermoCas9 exhibited efficient activity at the whole temperature range when combined with optimal spacer length and perfect PAM, while no activity was detected in the presence of truncated spacers or non-perfect PAMs. This characteristic can be harnessed to reduce the background level of survival and facilitates the selection of only the optimised ThermoCas9 variants. A ThermoCas9 library was generated through *in vitro* random mutagenesis (error-prone PCR) for *in vivo* selection using a truncated sgRNA. Surviving colonies that express ThermoCas9 variants with enhanced cleavage activity at the different temperatures will be selected and screened in the near future. In addition, *in silico* structural analyses revealed residues of ThermoCas9 that could be targets for rational engineering to confer lower structural rigidity and higher flexibility to the protein and thus allow for higher catalytic activity at mesophilic temperatures. All in all, this chapter provides the first steps towards the random and/or rational engineering of ThermoCas9 for optimal activity at low temperatures and thus for wide applicability in biotechnology and medicine.

In a nutshell, this thesis summarizes the current advances of CRISPR-Cas technology in the field of genetic engineering. Moreover, it describes the characterisation as well as the exploitation of Cas nucleases, anti-CRISPR proteins, and deaminase enzymes with special traits (thermostability, unique inhibition mechanism, CARF deaminase, respectively) for the development of novel genetic engineering tools with application in bacterial and human cells.

General Discussion

The ground-breaking discovery of CRISPR-Cas systems constitutes a milestone in the field of genetic engineering. During the last years, extensive characterization of the impressive diversity of CRISPR-Cas systems led to the establishment of the multifaced CRISPR-Cas technology that involves distinct Cas nuclease variants combined with different types of enzymes/proteins (deaminases, reverse transcriptases, transposases, recombinases and anti-CRISPRs)[37-51]. This final Chapter describes the modern catalogue of Class 2 CRISPR-Cas DNA nucleases and the special characteristics of the type II-C Cas9 orthologue from *Geobacillus thermodenitrificans* (ThermoCas9). We discuss the current genetic engineering advancements achieved through the exploitation of this Cas nuclease, and we address novel ideas for the expansion of its toolbox. Last, we delineate the updated global legislation on the safe application of genetic engineering techniques in several fields of human life.

Class 2 CRISPR–Cas DNA nucleases: Why ThermoCas9?

To date, numerous RNA-guided DNA nucleases of Class 2 CRISPR-Cas systems have been characterised and exploited for genome editing and transcriptional regulation. These naturally occurring or engineered Cas nucleases exhibit distinct characteristics, such as size, genotype, PAM specificity, cleavage sites, efficiency, fidelity, stability, and toxicity (Table 1)[91-185]. Nowadays, the most widely used Cas nuclease is the type II-A Cas9 from *Streptococcus pyogenes* (SpyCas9) because of its robust dsDNA cleavage activity[91]. However, SpyCas9 exhibits a high number of non-specific interactions with its target DNA, allowing for cleavage of mismatched off-target sites[92-94]. Moreover, it requires a 5'-NGG-3' PAM for target dsDNA binding, limiting the range of targetable sites[91]. In addition, SpyCas9 presents low stability in human plasma and, due to its large size (> 1,350 amino acids, >4,000 nucleotides), it cannot be packaged with its sgRNA into adeno-associated virus (AAV) vectors for *in vivo* gene therapy. As a consequence, electroporation of the SpyCas9/sgRNA ribonucleoprotein (RNP) complex is frequently used for delivery to human cells[90,95]. Low protein stability is also observed at temperatures higher than 42°C, restricting its application only to mesophilic organisms[96]. Last, SpyCas9 may cleave ssRNA for transcriptional repression, in the presence of either a short DNA oligonucleotide that bears the PAM sequence (a PAMmer) or of a repetitive RNA molecule[97-100]. To overcome the above-mentioned limitations, several Cas nuclease variants with higher fidelity,

alternative PAM recognition, higher stability, smaller size, or lack of toxicity have been harnessed.

1. High fidelity

For applications that require high precision editing without off-target modifications, high accuracy Cas nuclease variants are entailed. Positively charged or polar amino acid groups in the REC3 domain are often substituted in order to reduce the off-target activity of SpyCas9. Consequently, the interaction of SpyCas9 with the RNA/DNA heteroduplex is weakened, restricting re-orientation of the REC2 domain and thus preventing efficient docking of the HNH catalytic domain at the active site (target DNA cleavage)[93,101]. Slower dsDNA cleavage leads to lower kcat, increasing fidelity[94,102,103]. Following this strategy, various engineered SpyCas9 variants with improved accuracy have been generated[92,93,101,104-113] (Table 1). In addition, naturally occurring Cas9 (mainly from type II-C CRISPR-Cas systems), Cas12a, Cas12b, or Cas12f orthologues as well as engineered variants thereof present remarkably lower off-target activity compared to the type II-A SpyCas9[90,114-134] (Table 1). Indicatively, in **Chapter 5** we show that the type II-C ThermoCas9 tolerates less spacer-protospacer mismatches than SpyCas9 *in vivo* at 37°C[89]. These engineered variants and natural Cas orthologues enable precise genetic manipulation, minimising off-target editing.

2. Alternative PAM specificity

In cases that the target site does not contain the 5'-NGG-3' PAM, Cas nucleases that recognize distinct PAMs are necessary. As such, amino acid substitutions have been introduced in the PAM-interacting (PI) domain of SpyCas9 or, alternatively, the PI domain of SpyCas9 has been replaced by that of another orthologue, generating variable PAM-altered SpyCas9 variants[105,109,112,113,135-142] (Table 1). Moreover, several Cas9, Cas12a, Cas12b and Cas12f orthologues naturally recognize non-NGG PAMs, broadening the spectrum of available target sites[90,114,117-134,140,143-163,164-170] (Table 1). Some of them have also been engineered to cover an even wider range of PAMs[126,128-132,145,164-171] (Table 1). Notably, ThermoCas9 and few other type II-C Cas9 orthologues recognize the largest PAMs discovered to date, additionally contributing to higher fidelity (however this comes with a prize of reduced availability of potential target sites). The access to various Cas

nuclease variants nowadays has significantly expanded the targeting scope of genetic engineering applications.

3. High stability

Proper folding and high stability of Cas nucleases is crucial for efficient RNP delivery as well as for high catalytic activity at elevated temperatures. Contrary to most characterized Cas nucleases, few type II-C, type V-B and type V-F orthologues are active and stable at a wide temperature range (Table 1). The moderately thermostable Cas9 from *Pasteurella pneumotropica* (PpCas9), Cas12b from *Alicyclobacillus acidoterrestris* (AacCas12b) or *Alicylobacillus acidiphilus* (AaCas12b), and Cas12f1 from *Acidibacillus sulfuroxidans* (AsCas12f1) are active between 20 and 60°C[133,158,172]. However, the highly thermostable proteins from *Geobacillus thermodenitrificans* (ThermoCas9), *Geobacillus stearothermophilus* (GeoCas9), *Acidothermus cellulolyticus* (AceCas9), and *Geobacillus sp.* LC300 (CaldoCas9) are active between 20 and 70°C. Exceptionally, the Cas12f orthologue from *Syntrophomonas palmitatica* (SpaCas12f1) is active between 37 and 80°C. All of these proteins have been applied in human, plant or *E. coli* cells, while ThermoCas9, GeoCas9, and CaldoCas9 have additionally enabled efficient engineering of (extreme) thermophiles[89,90,133,147,172-178]. In **Chapter 5**, we show that ThermoCas9 and GeoCas9 can be used for gene deletion, transcriptional control, and base-editing in *E. coli*. Moreover, in **Chapter 7** we demonstrate that ThermoCas9 is active *in vivo* at low temperatures (20–37°C) under optimal conditions (full length guide, optimal PAM), and in **Chapter 6** we prove that it can be used for gene disruption and base-editing in human cells. ThermoCas9 may also present a longer lifetime in human plasma compared to the mesophilic SpyCas9, as previously observed for GeoCas9 and AaCas12b[90,133,172]. Hence, these highly stable Cas nucleases may increase the efficiency of RNP delivery in eukaryotic cells and/or enable genetic engineering of industrially relevant thermophilic bacteria.

4. Compact size

All-in-one AAV delivery is the most popular choice for *in vivo* gene therapy and is applicable only to proteins smaller than 1,100 amino acids. Several compact Cas nucleases have been identified and characterized to date, including type II-C (like ThermoCas9), type V-B, type V-F, type V-I, and type V-J orthologues (Table 1) [89,90,120,123,124,126,134,146,147,157-

[163,171,178-181]. Thanks to their small size, these proteins can be successfully be packaged with their guide RNA in a single AAV vector, facilitating delivery for therapeutic genome editing.

5. Lack of toxicity

Safe application of CRISPR nucleases implies their interaction solely with the target substrate, avoiding unspecific interference with cellular nucleic acids. Similar to SpyCas9, the type II-B Cas9 orthologue from *Francisella novicida* (FnoCas9) as well as the type II-C orthologues from *Neisseria meningitidis* (Nme1Cas9) and *Campylobacter jejuni* (CjeCas9) nick and/or cleave dsDNA in a PAM- and RNA-independent manner, especially in the absence of guide RNA (Table 1)[96,182]. RNA-guided CjeCas9 and Nme1Cas9 additionally cleave cellular ssDNA and ssRNA molecules in a PAM- and RNA-independent manner[103,183,184]. Moreover, trans-acting non-specific ssDNA degradation is a fundamental feature of type V nucleases, including LbCas12a, FnCas12a, AsCas12a, AaCas12b, Un1Cas12f1, SpaCas12f1, AsCas12f1, and Cas12j3 (Table 1)[134,158,179-181,185,186]. The promiscuity of these Cas orthologues renders them less suitable for genetic engineering applications. In contrast, ThermoCas9 and other orthologues solely create DSDBs, without the ability to cleave ssDNA[89]. In this way, undesired interference with unrelated cellular processes is prevented.

Overall, ThermoCas9, among others, fulfils all the above mentioned criteria (high fidelity, high stability, compact size, lack of toxicity), rendering it a suitable alternative platform for genetic engineering applications. However, its weaker catalytic activity compared to SpyCas9 at low temperatures limits its applicability. Hence, in **Chapter 7** we describe some first steps towards the development of a ThermoCas9 variant with enhanced catalytic activity at mesophilic temperatures. Similarly, several other CRISPR endonucleases, such as MAD7, Un2Cas12f1 and Mi1Cas12f2, present interesting traits that could be exploited in the genetic engineering field[180,187].

Table 1. Class 2 CRISPR-Cas nucleases that have been applied for genetic engineering purposes.

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
Type II-A					
SpyCas9	1,368	Wild-type	NGG	-3/-3	+ High activity - Low fidelity - Large size
SpyCas9 DE	1,368	D1135E	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
HiFi Cas9	1,368	R691A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 + Higher activity than SpyCas9-HF1, eSpyCas9(1.1), and HypaCas9 + Similar or lower activity compared to SpyCas9 - Large size
SpyCas9 KA	1,368	K855A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
Sniper-Cas9	1,368	F539S, M763I, K890N	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9, xCas9(3.7), eSpCas9(1.1), Cas9-HF1 + Similar activity to SpyCas9 + Higher activity than SpyCas9-HF1, evoCas9, and HypaCas9 - Large size
SpyCas9-HF1	1,368	N497A, R661A, Q695A, Q926A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 + Similar activity to SpyCas9 - Large size
SpyCas9-HF2	1,368	N497A, R661A, Q695A, Q926A, D1135E	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
SpyCas9-HF3	1,368	N497A, R661A, Q695A, Q926A, D1135E, L169A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
SpyCas9-HF4	1,368	N497A, R661A, Q695A, Q926A, D1135E, Y450A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9, SpyCas9-HF2, SpyCas9-HF3 - Lower activity than SpyCas9 - Large size
HypaCas9	1,368	N692A, M694A, Q695A, H698A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 + Similar activity to SpyCas9 + Similar or higher activity than SpyCas9-HF1, eSpyCas9(1.1) - Large size

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
eSpyCas9(1.0)	1,368	K810A, K1003A, R1060A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 + Similar activity to SpyCas9 - Large size
eSpyCas9(1.1)	1,368	K848A, K1003A, R1060A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 + Similar activity to SpyCas9 - Large size
evoCas9	1,368	M495V, Y515N, K526E, R661L	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9, SpyCas9-HF1, eSpyCas9(1.1) + Similar or lower activity compared to SpyCas9 - Large size
HeFSpyCas9	1,368	SpyCas9-HF1 combined with eSpyCas9(1.1)	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9, SpyCas9-HF1, and eSpyCas9(1.1) - Lower activity than SpyCas9 - Large size
HeFm1SpCas9	1,368	SpyCas9-HF1 combined with eSpyCas9(1.1) but lacking N497A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9, SpyCas9-HF1, and eSpyCas9(1.1) - Lower activity than SpyCas9 - Large size
HeFm2SpCas9	1,368	SpyCas9-HF1 combined with eSpyCas9(1.1) but lacking K1003A	NGG	-3/-3	Engineered SpyCas9 variant with: + Higher fidelity than SpyCas9 + Similar activity to SpyCas9; higher than SpyCas9-HF1 or eSpyCas9(1.1) - Lower fidelity than SpyCas9-HF1, and eSpyCas9(1.1) - Large size
xCas9	1,368	A262T, R324L, S409I, E480K, E543D, M694I, E1219V	NG, GAA, GAT	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
xCas9(3.6) *[codon-optimized xCas9]*	1,368	A262T, R324L, S409I, E480K, E543D, M694I, E1219V	NG, GAA	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
xCas9(3.7) *[codon-optimized xCas9]*	1,368	A262T, R324L, S409I, E480K, E543D, M694I, E1219V	NGW	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Large size
SpyCas9-NG	1,368	L1111R, D1135V, G1218R, E1219F, A1332R, R1335V,	NGN, NANG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher fidelity than SpyCas9 + Higher activity than xCas9 variants - Lower activity than SpyCas9

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
		T1337R			- Large size
SpG	1,368	D1135L, S1136W, G1218K, E1219Q R1335Q, T1337R	NGN	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + High activity + High fidelity - Large size
SpyCas9-NGA	1,368	SpyCas9-NG and R1335Q	NG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than SpyCas9(NGX)-NGC, and SpyCas9NGX(-NGA) - Large size
SpyCas9-NGC	1,368	SpyCas9-NG and R1335E	NG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Large size
SpyCas9-NGX	1,368	xCas9(3.7) combined with SpyCas9-NG	NG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Large size
SpyCas9-NGX-NGC	1,368	xCas9(3.7) combined with SpyCas9-NG and R1335E	NG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Large size
SpRY	1,368	SpG + A61R, L1111R, 1317R, A1322R, 1333P	NR/Y	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Lower fidelity than SpyCas9 - Large size
SpyCas9-NRRH	1,368	V1139A,D1180 G, Q1221H, A1320V, R1333K, A10T, I322V, S409I, E427G, R654L, R753G, R1114G, D1135N, 1219V	NRRH	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than SpyCas9 - Similar to higher fidelity than SpyCas9 - Large size
SpyCas9-NRTH	1,368	D1180G, 1218S, Q1221H, 1249S, E1253K, 1321S, D1332G, R1335L, A10T, I322V, S409I, E427G, R654L, R753G, R1114G, D1135N, E1219V	NRTH	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than SpyCas9 - Similar to higher fidelity than SpyCas9 - Large size
SpyCas9-NRCH	1,368	D1332N, R1335Q, T1337N, S1338T, H1349R, A10T, I322V, S409I, E427G, R654L, R753G, R1114G,	NRCH	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than SpyCas9 - Similar to higher fidelity than SpyCas9 - Large size

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
		D1135N,E1219V			
SpyCas9-VQR	1,368	D1135V, R1335Q, T1337R	NGAN	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Lower activity than SpyCas9 - Similar fidelity to SpyCas9 - Large size
SpyCas9-VRER	1,368	D1135V, G1218R, R1335E, T1337R	NGCG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Similar activity to SpyCas9 - Similar fidelity to SpyCas9 - Large size
SpyCas9-VRQR	1,368	D1135V, G1218R, R1335Q, T1337R	NGA	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than SpyCas0-VQR + Higher fidelity than SpyCas9-VRQR - Similar fidelity to SpyCas9 - Large size
SpyCas9-EQR	1,368	D1135E, R1335Q, T1337R	NGNG	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Lower activity than SpyCas9-VQR - Similar fidelity to SpyCas9 - Large size
iSpyCas9	1,360	SpCas9 (1-1099) -SmacCas9 (1,078-1,338), R221K, N394K	NAA	-3/-3	Engineered SpyCas9 variant with altered PAM specificity - Large size
SpyMac	1,353	SpyCas9 with PAM of Smac Cas9	NAAN	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than Smac Cas9 - Large size
iSpyMac	1,353	SpyMac with R221K and N394K	NAAN	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher activity than SpyMac and similar to AsCas12a and LbCas12a - Low fidelity - Large size
HiFi-iSpyMac	1,353	iSpyMac with R691A	NAAN	-3/-3	Engineered SpyCas9 variant with altered PAM specificity + Higher fidelity than iSpyMac - Lower activity than iSpyMac - Large size
SmacCas9	1,338	Wild-type	NAA	-3/-3	Distinct to SpyCas9 PAM specificity - Low efficiency
SauriCas9	1,061	Wild-type	NNGG	-3/-3	Distinct to SpyCas9 PAM specificity + Similar activity to SaCas9 + Compact size - Low fidelity

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
SauriCas9-KKH	1,061	Q788K, Y973K, R1020H	NNAG	-3/-3	Engineered SauriCas9 variant with distinct to SpyCas9 PAM specificity + Higher fidelity than SauriCas9 + Compact size - Lower activity than SauriCas9
ScCas9	1,375	Wild-type	NNG	-3/-3	Distinct to SpyCas9 PAM specificity. - Lower or similar activity to SpyCas9 - Low to moderate fidelity - Large size
ScCas9+	1,375	T1227K	NNG	-3/-3	Engineered ScCas9 variant with distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 - Low to moderate fidelity - Large size
ScCas9++	1,375	ScCas9+ with I367K, G368D, I369K, H371L, T375S, T376G	NNG	-3/-3	Engineered ScCas9 variant with distinct to SpyCas9 PAM specificity + Higher activity than ScCas9 and ScCas9+ + Higher fidelity than SpyCas9 and similar to HiFi Cas9 - Large size
HiFi-Sc++	1,375	ScCas9++ with R701A	NNG	-3/-3	+ Higher activity than ScCas9, ScCas9+, SpCas9-NG + Higher fidelity than SpyCas9 and HiFi Cas9 - Large size
SaCas9	1,053	Wild-type	NNGRRT	-3/-3	Distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 + Compact size - Low fidelity but higher than SpyCas9
SaKKHCas9	1,053	E782K, N968K, R1015H	NNNRRT	-3/-3	Engineered SaCas9 variant with: + Distinct PAM specificity + Similar to SpyCas9 activity + Compact size - Similar fidelity to SaCas9
eSaCas9	1,053	R499A, Q500K, K572A	NNGRRT	-3/-3	Distinct to SpyCas9 PAM specificity + Higher fidelity than SaCas9 + Compact size - Lower activity than SaCas9
eSa-SauriCas9	1,061	SaCas9 with PI domain of SauriCas9	NNGG	-3/-3	Distinct to SpyCas9 PAM specificity + Higher fidelity than SauriCas9 - Lower activity than SauriCas9
St1Cas9-LMD9	1,122	Wild-type	NNAGAAW	-3/-3	Alternative to SpyCas9 PAM specificity

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
					+ Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
St1Cas9-LMG18311	1,122	St1Cas9 with WED and PI domain of LMG18311 Cas9	NNGYAA(A)	-3/-3	Engineered St1Cas9 variant with distinct to SpyCas9 PAM specificity. + Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
St1Cas9-CNRZ1066	1,122	St1Cas9 with WED and PI domain of CNRZ1066 Cas9	NNACAA(W)	-3/-3	Engineered St1Cas9 variant with distinct to SpyCas9 PAM specificity. + Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
St1Cas9-TH1477	1,122	St1Cas9 with WED and PI domain of TH1477 Cas9	NNGAAA	-3/-3	Engineered St1Cas9 variant with distinct to SpyCas9 PAM specificity. + Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
St1Cas9-MTH17CL396	1,122	St1Cas9 with WED and PI domain of MTH17CL396 Cas9	NNAAAA	-3/-3	Engineered St1Cas9 variant with distinct to SpyCas9 PAM specificity. + Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
St3Cas9	1,409	Wild-type	NGGNG	-3/-3	Distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
FrCas9	1,372	Wild-type	NNTA	-3/-3	Distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 + Higher fidelity than SpyCas9 - Large size
Type II-B					
FnoCas9	1,629	Wild-type	NGG	-7/-3	Distinct to SpyCas9 PAM specificity. + Much higher fidelity than SpyCas9 - Lower activity than SpyCas9 - Very large size - Toxic!
FnoCas9-RHA	1,629	E1369R, E1449H, R1556A	YG	-7/-3	Engineered FnCas9 variant with distinct PAM specificity - Lower activity than SpyCas9 - Very large size - Toxic!
Type II-C					
Nme1Cas9	1,082	Wild-type	NNNNGATT	-3/-3	Distinct to SpyCas9 PAM specificity.

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
					+ Higher fidelity than SpyCas9 + Compact size - Lower activity than SpyCas9 and SauCas9 - Toxic!
Nme2Cas9	1,082	Wild-type	NNNNCC	-3/-3	Distinct to SpyCas9 PAM specificity. + Higher fidelity than SpyCas9 + Compact size - Lower activity than SpyCas9
CjeCas9	984	Wild-type	NNNNRYAC, NNNVRYM	-3/-3	Distinct to SpyCas9 PAM specificity. + Higher fidelity than SpyCas9 and SauCas9 + Compact size + Similar activity to SaCas9 - Toxic!
enCjeCas9	984	L58Y, D900K	N3VRYAC, N3VAYTC, N3VAYGC	-3/-3	Engineered CjeCas9 variant with distinct to SpyCas9 PAM specificity. + Higher efficiency than CjeCas9 + Compact size - Lower fidelity than CjeCas9
BlatCas9	1,092	Wild-type	NNNNCNAA	-3/-3	Distinct to SpyCas9 PAM specificity. + Similar efficiency to SpyCas9 + Compact size - Low fidelity but higher than SpyCas9
PpCas9	1,055	Wild-type	NNNNRTT	-3/-3	Distinct to SpyCas9 PAM specificity. + Compact size + High stability - Low activity
ThermoCas9	1,082	Wild-type	NNNNCNAA, NNNNCMCA	-3/-3	Distinct to SpyCas9 PAM specificity. + Higher fidelity than SpyCas9 + Higher stability than SpyCas9 + Compact size + High activity at high temperatures - Lower activity than SpyCas9 at low temperatures
GeoCas9	1,087	Wild-type	NNNNCRAA	-3/-3	Distinct to SpyCas9 PAM specificity. + Higher fidelity than SpyCas9 + Compact size + Higher stability than SpyCas9 + High activity at high temperatures - Lower activity than SpyCas9 at low temperatures

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
CaldoCas9	1,087	Wild-type	NNNNGNMA	-3/-3	Distinct to SpyCas9 PAM specificity. + High activity at high temperatures + High stability + Compact size
AceCas9	1,138	Wild-type	NNNCC	-3/-3	Distinct to SpyCas9 PAM specificity. + High fidelity + High activity at high temperatures + High stability - Lower activity than SpyCas9 at low temperatures - Relatively large size
AceCECas9(G685A)	1,138	G685A	NNNCC	-3/-3	Engineered AceCas9 variant with distinct to SpyCas9 PAM specificity. + High fidelity + Higher activity than AceCas9 at mesophilic temperatures - Relatively large size
AceCECas9(V709A)	1,138	V709A	NNNCC	-3/-3	Engineered AceCas9 variant with distinct to SpyCas9 PAM specificity. + Higher activity than AceCas9 at mesophilic temperatures - Lower fidelity than AceCas9 and AceCECas9(G685A) - Relatively large size
LbCas12a	1,228	Wild-type	TTTV	+19/+23	Distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 and higher than FnCas12, AsCas12, and MbCas12a + High activity at low temperatures + Higher fidelity than SpyCas9 - Large size - Toxic!
LbCas12a-RR	1,228	G532R, K595R	TYCV	+19/+23	Engineered LbCas12a variant with altered PAM specificity + High activity (low temperatures) + Higher activity than FnCas12a-RR, AsCas12a-RR, and MbCas12a-RR - Large size
Type V-A					
LbCas12a-RVR	1,228	G532R, K538V, Y542R	TATV	+19/+23	Engineered LbCas12a variant with altered PAM specificity + High activity (especially at low temperatures)

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
					- Large size
LbCas12a-RVRR	1,228	LbCas12a-RVR with K595R	TNTN	+19/+23	Engineered LbCas12a variant with + High activity (especially at low temperatures) + Higher activity than FnCas12a-RVRR and AsCas12a-RVRR + Similar activity to LbCas12a - Lower fidelity than LbCas12a - Large size
impLbCas12a	1,228	LbCas12a-RVRR with D156R	NTTV	+19/+23	Engineered LbCas12a variant with + Higher activity than LbCas12a-RVRR - Large size
FnCas12a	1,300	Wild-type	NTTN, KYTV	+19/+23	Distinct to SpyCas9 PAM specificity + High activity + High fidelity - Large size - Toxic!
FnCas12a-RR	1,300	N607R, K671R	TYCV	+19/+23	Engineered FnCas12a variant with altered PAM specificity - Very low activity - Large size
FnCas12a-RVR	1,300	N607R, K613V, N617R	TATV	+19/+23	Engineered FnCas12a variant with altered PAM specificity - Very low activity - Large size
FnCas12a-RVRR	1,300	FnCas12a-RVR with K687R	TNTN	+19/+23	Engineered FnCas12a variant with altered PAM specificity - Low activity - Large size
AsCas12a	1,308	Wild-type	TTTV	+19/+23	Distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 + Higher activity at higher temperatures than LbCas12a variants + Higher fidelity than SpyCas9 - Large size - Toxic!
AsCas12a-RR	1,308	S542R, K607R	TYCV, CCCC, TTTV	+19/+23	Engineered AsCas12a variant with altered PAM specificity + Higher activity than AsCas12a + Higher activity at higher temperatures than LbCas12a variants - Similar fidelity to AsCas12a - Large size
AsCas12a-RRA	1,308	AsCas12a-RR with K949A	TYCV, CCCC	+19/+23	Engineered AsCas12a variant with altered PAM specificity

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
					+ Similar or higher activity than AsCas12a + Higher fidelity than AsCas12a-RR + Higher activity at higher temperatures than LbCas12a variants - Large size
AsCas12a-RVR	1,308	S542R, K548V, N552R	TATV, TTTV	+19/+23	Engineered AsCas12a variant with altered PAM specificity + Similar or higher activity than AsCas12a + Higher activity at higher temperatures than LbCas12a variants - Similar fidelity to AsCas12a - Large size
AsCas12a-RVRA	1,308	AsCas12a-RVR with K949A	TATV	+19/+23	Engineered AsCas12a variant with altered PAM specificity + Similar or higher activity than AsCas12a + Higher fidelity than AsCas12a-RVR + Higher activity at higher temperatures than LbCas12a variants - Large size
AsCas12a-RVRR	1,308	AsCas12a-RVR with K607R	TNTN	+19/+23	Engineered AsCas12a variant with altered PAM specificity + Higher activity at higher temperatures than LbCas12a variants - Low activity - Large size
enAsCas12a	1,308	E174R, S542R, K548R	TTYN, VTTV, TRTV	+19/+23	Engineered LbCas12a variant with altered PAM specificity + Higher activity than AsCas12a-RVRR - Large size
MbCas12a	1,373	Wild-type	NTTN	+19/+23	Distinct to SpyCas9 PAM specificity + Similar activity to SpyCas9 - Lower activity than LbCas12a and AsCas12a - Large size
MbCas12a-RR	1,373	N576R, K637R	TYCV	+19/+23	Engineered LbCas12a variant with altered PAM specificity - Very low activity - Large size
MbCas12a-RVR	1,373	N576R, K582V, N586R	TATV	+19/+23	Engineered LbCas12a variant with altered PAM specificity - Large size

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
MbCas12a-RVRR	1,373	MbCas12a-RVR with	TNTN	+19/+23	Engineered LbCas12a variant with altered PAM specificity - Large size
Type V-B					
AaCas12b	1,129	Wild-type	TTN	+17/+23	Distinct to SpyCas9 PAM specificity. + High stability + High fidelity (higher than SpyCas9) + Higher activity than AkCas12b and AsCas12a, and similar to FnCas12a - Large size - Toxic!
AacCas12b	1,129	Wild-type	TTN	+17/+24	Distinct to SpyCas9 PAM specificity. + High stability + High fidelity - Low activity at 37°C - Large size
AkCas12b	1,147	Wild-type	TTTN	Unknown	Distinct to SpyCas9 PAM specificity. - Nicks non-target DNA strand instead of creating DSDB (especially at low temperatures) - Large size
BhCas12b	1,109	Wild-type	TTTA	Unknown	Distinct to SpyCas9 PAM specificity. - Nicks non-target DNA strand instead of creating DSDB (especially at low temperatures) - Large size
BhCas12b-v4	1,108	K846R, S893R, E837G	ATTN	Unknown	Engineered BhCas12b variant with improved dsDNA cleavage activity + Lower ssDNA nicking activity than BhCas12b + Higher dsDNA cleavage activity than BhCas12b (at low temperatures) +Higher fidelity than SpyCas9 - Large size
BthCas12b	1,108	Wild-type	ATTC	+16/+23	Distinct to SpyCas9 PAM specificity. - Large size
AmCas12b	1,146	Wild-type	TTN	Unknown	Distinct to SpyCas9 PAM specificity. - Low activity - Large size
Bs3Cas12b	1,112	Wild-type	TTN	Unknown	Distinct to SpyCas9 PAM specificity. + Higher activity than AmCas12b, AkCas12b, and LsCas12b

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
					- Large size
LsCas12b	1,090	Wild-type	TTN	Unknown	Distinct to SpyCas9 PAM specificity. + Higher activity than AsCas12a, and similar to FnCas12a + Compact size
TcCas12b	1,142	Wild-type	TTN	Unknown	Distinct to SpyCas9 PAM specificity. + Similar activity to AaCas12b - Large size
Type V-E					
DpbCasX (DpbCas12e)	986	Wild-type	TTCN	+12,+13,+14/ +22,+23,+24, +25	Distinct to SpyCas9 PAM specificity. - Low efficiency + Compact size
PlmCasX (PlmCas12e)	978	Wild-type	Unknown	Unknown	Distinct to SpyCas9 PAM specificity. + Higher efficiency than DpbCasX, comparable to SpyCas9 and LbCas12a + Compact size
Type V-F					
AsCas12f1	422	Wild-type	TTN	24/+22	Distinct to SpyCas9 PAM specificity + Similar fidelity to SpaCas12f1and Un1Cas12f1 + High stability + Hyper-compact size - Very low activity - Toxic!
Un1Cas12f1 (Cas14a1)	529	Wild-type	TTTR	+22/+24	Distinct to SpyCas9 PAM specificity + (When combined with optimized guide RNA) similar activity to SpyCas9 and higher than LbCas12a + Similar fidelity to AsCas12a, SpaCas12f1and AsCas12f1 + Hyper-compact size - Toxic!
CasMINI-V3.1	529	Un1Cas12f1 with D143R, T147R, and E151A	TTTR	(Between +20 and +24)	Engineered Un1Cas12f1 variant with higher efficiency (when combined with optimized guide RNA).
SpaCas12f1	497	Wild-type	RTTY	+24/+22	Distinct to SpyCas9 PAM specificity + Similar activity to FnCas12a + Similar fidelity to AsCas12f1 and Un1Cas12f1 + Hyper-compact size + High stability - Toxic!

Cas DNA nuclease	Size (aa)	Genotype	PAM (5'->3')	Cleavage sites (non-target/ target strands)	Characteristics
Type V-I					
Cas12i2	1,054	Wild-type	TTY	+31/+24	Distinct to SpyCas9 PAM specificity + Compact size - Low activity
Cas12i2-ABR-001 (Cas12i2 v1)	1,054	D581R, I926R, V1030G	TTY	+31/+24	Distinct to SpyCas9 PAM specificity + Higher activity than Cas12i2 + High fidelity + Compact size
Type V-J					
Cas12j3 (CasΦ)	700-800	Wild-type	TTTN	+23/+12,+13, +14	Distinct to SpyCas9 PAM specificity + Hyper-compact size - Toxic!
CasΦ−2	700-800	Wild-type	TBN	+13/+22	Distinct to SpyCas9 PAM specificity + Hyper-compact size
CasΦ−3	700-800	Wild-type	VTTN	Unknown	Distinct to SpyCas9 PAM specificity + Hyper-compact size
AsCas12j-2	757	Wild-type	TTN	Unknown	Distinct to SpyCas9 PAM specificity + Higher activity than FnCas12a

Expanding the genetic engineering toolbox using ThermoCas9

In 2017, researchers in the Van der Oost lab characterised a type II−C CRISPR−Cas system found in the chromosome of the facultative thermophilic bacterium *G. thermodenitrificans* T12[89,188]. The Cas9 endonuclease of this system (ThermoCas9) was shown to be active *in vitro* between 20 to 70°C (optimal 50–60°C), contrary to the mesophilic SpyCas9 that is functional only below 44°C[89]. This unique characteristic rendered ThermoCas9 the first readily applicable Cas9 variant for genetic engineering of both mesophilic and thermophilic organisms. Indeed, ThermoCas9 has been successfully applied for gene deletion in the mesophilic bacterium *Pseudomonas putida* as well as for gene deletion and/or gene silencing in several thermophilic bacteria (*Bacillus smithii*, *Hungateiclostridium thermocellum*, *Thermus thermophilus*)[89,176,177]. To further extend its potential, we used ThermoCas9 for efficient gene deletion, gene silencing, and gene disruption (C•G-to-T•A base-editing) in *Escherichia coli*, as described in **Chapter 5**. We discovered that a small anti-CRISPR protein (AcrIIC1) from the prophage region in the genome of *N. meningitidis* can trap ThermoCas9 *in vivo* in a DNA-bound state to robustly inhibit its dsDNA cleavage activity. Leveraging this special feature, we developed the first

reported anti-CRISPR base-editing tools. In addition, we proved that ThermoCas9 can be used as an alternative platform for genome- and base-editing (C•G-to-T•A conversion) in human cells (**Chapter 6**). However, due to its thermophilic nature, the activity of ThermoCas9 at moderate temperatures (20–37°C) was lower than that of the highly efficient SpyCas9 orthologue. Hence, we made some initial steps towards the optimization of the ThermoCas9 catalytic efficiency at low temperatures (20–37°C), through random and rational protein engineering (**Chapter 7**). Altogether, ThermoCas9 provides a novel genetic engineering platform for both prokaryotes (mesophilic or thermophilic) and eukaryotes.

To further expand the ThermoCas9 portfolio and enhance its input in diverse biotechnological applications, we hereby provide some suggestions on future research directions.

1. ThermoCas9 structure

First, it is crucial to analyse the molecular mechanisms of ThermoCas9 regarding target dsDNA binding and cleavage in order not only to understand, but also to use it as a basis to optimise the enzyme's catalytic performance. In **Chapter 7**, we predicted the three-dimensional (3D) structure of ThermoCas9 using the bioinformatic tools Phyre2 and Alphafold. However, these predictions are mainly based on the structure of the close homologue from *Neisseria meningitidis* (Nme1Cas9), which presents only 42% aa identity to ThermoCas9, interacts with a shorter sgRNA, and is mesophilic[189]. Moreover, Alphafold is not able to predict different conformational states and interactions with nucleic acids, co-factors, metals, and ligands[190]. Hence, there is an imperative need to experimentally identify and characterize the 3D structure of ThermoCas9. Nowadays, single-particle cryo-electron microscopy (cryo-EM) is extensively used for the elucidation of macromolecular structures (≥64 kDa) at near-atomic resolution, circumventing the crystallization bottleneck of the traditionally used technique of X-ray crystallography[191]. Cryo-EM provides insight into conformational and energy landscapes, and it has been widely applied for several CRISPR-Cas nucleases, such as Nme1Cas9 (kDa) and SpyCas9 (kDa)[189,191,192]. ThermoCas9 has a size that is identical to that of Nme1Cas9 (1082 amino acids), rendering it suitable for cryo-EM analysis. The preparation for this analysis requires production and purification of the ThermoCas9 protein, its sgRNA, and the target dsDNA. We have previously purified a catalytically inactive variant of ThermoCas9 (dThermoCas9)

harboring mutations in the RuvC (D8A) and HNH (H582A) domains[193]. However, the final product additionally contained four ThermoCas9 degradation products of smaller size (not shown), verified by Mass Spectrometry. Therefore, for the cryo-EM analysis we recommend co-purification of ThermoCas9 with the sgRNA or use of an excessive amount of protease inhibitors to enhance protein homogeneity. Importantly, the catalytically active ThermoCas9 variant should be selected for purification, as this would allow to capture the protein structure in a catalytically active state. In case that the sgRNA is not co-purified with the protein, high sgRNA yields of high purity can be achieved using our previously optimised *in vitro* transcription protocol[193]. Resolution of the ThermoCas9 structure would shed light on the details that govern its biological mechanism, and would facilitate structure-based protein engineering for increased structural flexibility and thus enhanced catalytic activity at moderate temperatures.

2. ThermoCas9 base-editors: CBEs, ABEs, CGBEs, CABEs and DBEs

In **Chapter 5**, we developed a ThermoCas9-associated cytidine base-editor (CBEs) in *E. coli*, called dThermoTarget-AID, by fusing the *Petromyzon marinus* cytidine deaminase (PmCDA1) gene and the bacteriophage PBS2 uracil DNA glycosylase inhibitor (UGI) gene to the C-terminus of the *dthermocas9* gene. The dThermoTarget-AID system preferentially edits cytidines at the PAM-distal end of the protospacer, as previously observed for the SpyCas9 base-editors[194]. However, dThermoTarget-AID exhibits a six-times larger editing window (30 bp instead of 5 bp) and is able to edit up to eight cytidine bases simultaneously, facilitating gene inactivation in bacteria in the absence of donor DNA. In addition, we replaced the *dthermocas9* gene in the dThermoTarget-AID system with its wild-type variant and we co-expressed the AcrIIC1 protein (AcrThermoTarget-AID) to block the nuclease activity of the ThermoCas9-PmCDA1-UGI fusion protein, still allowing binding to the target DNA for cytidine deamination. Release of AcrIIC1 expression enables counter-selection of non-edited loci, increasing the number of clean mutants and the preference for editing at the PAM-proximal protospacer region. Moreover, we developed a ThermoCas9-associated CBE for human cells, by fusing the D8A nickase (*nthermocas9*) gene at the N-terminus with the rat Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 1 (rapobec1) gene, and at the C-terminus with two copies of the *ugi* gene followed by the SV40 NLS signal (**Chapter 6**). Similar to dThermoTarget-AID, the ThermoBE4 system presents a wide editing window (22 bp) and edits multiple (up to seven) cytidine bases simultaneously, preferentially at the PAM-distal end of the

protospacer. Collectively, the dThermoTarget-AID, the AcrThermoTarget-AID, and the ThermoBE4 CBEs provide alternative approaches for targeted gene mutagenesis in prokaryotes and eukaryotes. For applications that strictly require a single-nucleotide modification, nThermoCas9 can be combined with the naturally occurring PpAPOBEC1 deaminase to form ThermoBE4-PpAPOBEC1, which is expected to have narrow editing window (3 bp), minimal Cas9-independent DNA or RNA off-target editing, and similar efficiency to ThermoBE4, based on previous data on the SpyCas9-associated BE4-PpAPOBEC1 system[195]. Alternatively, the engineered rAPOBEC1 (YE1, YE2, YEE, EE, FE1, YFE, AALN) or APOBEC3A (A3A182, A3A186, hA3A(Y130F)186) variants that show equally narrow editing windows could be applied[196-199].

As described in **Chapter 2**, nowadays different types of single nucleotide conversions have been implemented through the development of ABEs (A•T-to-G•C), CGBEs (C•G-to-G•C), CABEs (C•G-to-A•T), and DBEs (C•G-to-T•A and A•T-to-G•C simultaneously). However, only few Cas orthologues have been adapted to these base-editing systems, limiting the targeting scope. In this line, we propose fusion of the *nthermocas9* gene at its N-terminus with the most efficient engineered adenosine deaminase variant reported to date (e.g. the *tadA9* gene[200]). The ThermoABE9 editor is predicted to present reduced by-stander edits and minimal Cas9-independent DNA and RNA off-target editing, as previously observed for its cognate SpyCas9-based ABE9 editor[200,201]. However, in case that a very narrow editing window is of highest importance, nThermoCas9 could be adapted similarly to the recently established ABE7.10(F148A) or the ABE7.10(E50A,V106W) systems[202]. For the generation of a ThermoCas9-based CGBE, we suggest fusion of nThermoCas9 with the rAPOBEC1 and the base excision repair protein rXRCC1 (ThermoACX-rXRCC1 editor) to achieve maximum editing efficiency and product purity as well as narrow editing window, as previously reported for nSpyCas9[203]. For the construction of a ThermoCas9-based CABE, nThermoCas9 can be linked to a uracil-DNA glycosylase (UNG) and the activation-induced cytidine deaminase (AID) to form AID-nThermoCas9-UNG, which is expected to provide high editing efficiency and specificity as well as narrow editing window, as previously reported for nSpyCas9[204]. Moreover, a ThermoCas9-based DBE can also be developed to further facilitate gene mutagenesis applications, introducing not only C•G-to-T•A mutations like the already existing ThermoCas9 CBEs (**Chapter 5**) but also A•T-to-G•C at the same time. As such, nThermoCas9 can be adapted either to the CABE-RY or to the A&C-BEmax systems, as both present large editing windows for both deamination types and low off-target activity[205,206]. In summary, all types of transition and transversion modifications can be achieved by adjusting nThermoCas9 to the current base-editing

architectures, broadening in parallel their targeting scope at alternative non-NGG PAM sites.

Despite its widespread application in both prokaryotes and eukaryotes, the base-editing technology has never been implemented in thermophilic organisms due to the absence of a thermostable base-editing system. To enable base-editing in thermophiles, nThermoCas9 or dThermoCas9 can be fused to a thermostable deaminase that acts on DNA substrates. Several thermostable cytidine deaminases have been characterised during the last decades, such as those derived from *Nocardioides* sp., *Streptomyces hygroscopicus*, *Streptomyces viridiviolaceus*, *Penicillium sizowi*, *Penicillium politans*, *E. coli*, *Bacillus caldolyticus* (CDD; CDA_{Bcald}), and *G. stearothermophilus*. However, all of them act on free nucleotides or nucleosides instead of DNA, hampering their use for base-editing purposes[207-215]. In addition, some proteins with a predicted adenosine deaminase domain and a predicted CARF-domain (CARF-ADs) have been detected in thermophiles[216]. Nevertheless, none of them has been experimentally characterised. Therefore, in **Chapter 3**, we focused on the characterization of a CARF-AD enzyme from the thermophilic bacterium *Methylacidiphilum fumariolicum* SolV. Revealing its substrate specificity and its deaminase activity will elucidate its biological role. In case deaminase activity on DNA substrates is detected, this would make it a potential candidate for developing robust, thermostable ABEs.

3. ThermoCas9 prime-editors

The current prime-editing technology is exclusively based on a single Cas protein (nSpyCas9) and a single reverse transcriptase (RT) enzyme (wild-type or engineered Moloney murine leukaemia virus, M-MLV)[217-220]. To further broaden the targeting scope, to optimise the editing efficiency, and to minimise undesired off-target activity, prime-editors composed of alternative Cas and/or RT orthologues should be developed. Similar to base-editing, prime-editing has never been applied in thermophiles. Therefore, the generation of a robust and/or wide-temperature range prime-editor would enable implementation in not only mesophilic but also thermophilic organisms. In this context, the H582A nickase ThermoCas9 (nThermoCas9) can be linked to the engineered M-MLV2 variant[217,218] or another thermostable RT enzyme with favourable characteristics. Although several thermostable RT enzymes have been characterised to date, most of them are not suitable for prime-editing applications because of their large size, DNA

endonuclease activity (intron mobility), and RNase H activity (removal of RNA in RNA-DNA heteroduplex intermediate to allow retrohoming), low RT efficiency, and inability to use RNAs with strong secondary structure for RNA transcription[221-235].

4. Off-target activity of ThermoCas9 nuclease, base-editors and prime-editors

In **Chapter 5**, we showed that the type II-C ThermoCas9, at least under the tested conditions at 37°C, tolerates fewer spacer-protospacer mismatches in *E. coli* compared to previous data for the type II-A SpyCas9[236]. This observation is in agreement with our previous *in vitro* study[89]. The enhanced fidelity of ThermoCas9 at 37°C may be explained by its natural ability to recognize longer PAMs (4 bp) and protospacer sequences (23 bp) as well as to sub-optimal kinetics of dsDNA binding and cleavage (compact REC lobe)[94,89,102]. In this way, ThermoCas9 and other thermophilic type II-C orthologues, such as GeoCas9 and CaldoCas9, minimise editing at undesired sites, preventing toxic off-target effects[96,122,124]. However, mesophilic type II-C orthologues like Nme1Cas9 and CjeCas9 additionally cleave ssDNA and ssRNA in a PAM- and RNA-independent manner[103,182-184], questioning their safe application in human cells. In contrast, ThermoCas9 does not possess ssDNA activity, rendering it a more secure genetic engineering option[89]. However, ThermoCas9 should be further examined for nonspecific activity on ssRNA substrates as well as for genome-wide, *in vivo* off-target cleavage.

Biased off-target detection can be analysed through computational prediction of possible off-target sites with high sequence homology to the guide RNA, followed by screening of these potentially edited genomic sites using mismatch-sensitive endonucleases, Sanger sequencing, or targeted deep sequencing. Several computational off-target prediction tools have been developed to date, such as CasOT, Cas-OFFinder, FlashFry, CrisFlash, MIT, CFD, CRISTA, Elevation, and DeepCRISPR[92,237-252]. However, the accuracy of *in silico* prediction may vary depending on the ‘‘off-by’’ threshold settings that are selected and on the computational off-target prediction tool. Low ‘‘off-by’’ thresholds and/or use of tools that disregard off-target sites with gaps, bulges, or non-canonical PAMs can result in missing important off-target sites[242-256]. Importantly, the major bottleneck of this approach is that most computational tools have been developed to predict off-target sites exclusively for the SpyCas9 orthologue and a 20 nt spacer[242-256]. Moreover, access to most of these tools is costly. Therefore, we suggest unbiased genome-wide off-target detection

and quantification. Several *in vitro* (Digenome-seq, SITE-seq, CIRCLE-seq, CHANGE-seq) and *in vivo*/ cell-based (End-Seq, DSBCapture, BLESS, BLISS, IDLV capture, Chip-seq, GUIDE-seq, DISCOVER-seq, HTGTS, LAM–HTGTS, GOTI, VIVO) methods for unbiased detection have been developed to date[144,253-267]. GUIDE-Seq is the most popular method nowadays, as it is an *in vivo* approach (clinically relevant), it is highly sensitive (0.1%), it does not require high sequencing depth, it presents low false-positive rates, it utilizes open-access computational pipelines, it provides a user-friendly wet-lab protocol, it is applicable to various cell lines, and it is cost-effective[257,261]. Altogether, we propose that unbiased genome-wide detection of ThermoCas9-mediated off-targets should be performed using the GUIDE-Seq method.

DNA base-editors exhibit three different types of off-target editing: (a) Cas-dependent DNA off-targets; (b) Cas-independent DNA off-targets; and (c) Cas-independent RNA off-targets. Cas-dependent DNA off-targets are generally less abundant than those generated by Cas9 nucleases, as base-editors require not only target DNA binding but also R-loop accessibility by the deaminase, presence of a cytidine base (for CBEs, CGBEs, CABEs, and DBEs) or adenine base (for ABEs or DBEs) within the editing window, and recognition of a deaminase-specific nucleotide context neighbouring to the target base[268-270]. The importance of Cas-mediated R-loop formation for the generation of off-targets renders CIRCLE-seq one of the most preferable screening approaches[258,264]. Cas-independent DNA off-targets are caused by prolonged expression of deaminases that randomly act on transiently accessible nucleotides across the genome[263,271,272]. The random localization and the very low frequency rates of these events (lower than spontaneous somatic cell mutation rates) render their detection challenging[263]. However, several *in vitro* (kinetics assay, enzymatic assay) as well as bacterial (rifampin resistance assay, thymidine kinase toxicity assay) and mammalian (orthogonal R-loop assay, ssDNA deamination assay) cell-based assays have been recently developed for this purpose[195,198]. Finally, Cas-independent RNA off-targets can also be transiently generated during overexpression of deaminases that naturally act on both DNA and RNA substrates[202,273-275]. Similar to the Cas-independent DNA off-targets, this type of events presents random localization and low frequency rates, and they can be detected through RNA sequencing[202,274-276]. Collectively, we provide alternative off-target screening methods for base-editors compared to Cas nucleases, due to inherent differences in the biological mechanisms of the systems.

Prime-editors present very low DNA and RNA off-target editing because of the strict requirement for three consecutive hybridization events. Specifically, complementarity is

needed between (a) the pegRNA spacer with the genome, (b) the pegRNA PBS with the nicked genomic DNA primer, and (c) the synthesised 3' DNA flap with the genome (PE editors) or with the second 3' DNA flap (twinPE editors). Screening for DNA off-target prime-editing is usually performed via Whole Genome Sequencing (WGS), Guide-Seq, CHANGE-seq or nDigenome-seq, while RNA off-target prime-editing is examined through RNA sequencing[218-220,277-281]. Nickase-based Digenome-seq (nDigenome-seq) is a recently developed unbiased *in vitro* assay that identifies Cas9-induced single-strand breaks and is ideal for use in prime-editing applications[277]. Overall, we propose nDigenome-seq and RNA sequencing for screening of DNA and RNA off-target prime-editing, respectively.

5. Relaxed PAM for ThermoCas9

CRISPR-Cas DNA nucleases recognize distinct PAMs regarding sequence and length (Table 1). Generally, Cas9 orthologues prefer C/G-rich PAMs, while Cas12 proteins prefer A/T-rich PAMs (Table 1). The length of the PAM may vary from 1 nt (e.g. 5'-NNG-3' for ScCas9) up to 5 nt (e.g. 5'-NNAGAAW-3' for St1Cas9-LMD9)[118,143]. Although long PAMs reduce the number of possible off-target sites and thus enhance fidelity, they also limit the number of available on-target sites limiting the targeting scope. ThermoCas9 recognizes a long (4 nt; 5'-NNNNCNAA-3' or 5'-NNNNCMCA-3') PAM, restricting the range of potential target sites[89]. Notably, it associates with various PAMs at high temperatures, while its PAM requirement is more strict at low temperatures[89]. To expand the targeting scope of ThermoCas9 at low temperatures, its PAM requirement can be relaxed to 1-2 nt through replacement or random/rational engineering of its PI domain.

6. Activity of ThermoCas9-based tools in clinically relevant human cell lines

To date, the activity of ThermoCas9 has only been tested in bacteria and HEK293T cells[89,176,177] (**Chapter 5**, **6**). In order to make the step towards the exploitation of ThermoCas9 in the field of gene therapy, its genome-, base-, and prime-editing efficiency should also be studied in clinically relevant human cell lines, such as autologous hematopoietic stem and progenitor cells (HSPCs) and T cells[282,283].

7. *In vivo* delivery of ThermoCas9 and its sgRNA

The therapeutic efficiency and safety of the ThermoCas9 toolbox is greatly influenced by the method that will be selected for delivery to human cells. Nowadays, most clinical trials implement *ex vivo* CRISPR delivery (genetic engineering of cells outside the patient and re-introduction to the patient), while only few have successfully applied *in vivo* CRISPR delivery (genetic engineering of cells directly to the patient)[284]. Although *ex vivo* CRISPR delivery provides higher safety, technical feasibility, and screening of successfully edited cells, this method is limited by the requirement for cell types that are able to survive and multiply outside of the patient after extensive *in vitro* culturing and genetic manipulation. As such, *ex vivo* CRISPR delivery can be advantageous specifically for cancer immunotherapy, hematological disorders and regenerative medicine, while most other diseases (such as Duchenne muscular dystrophy, DMD) necessitate *in vivo* CRISPR delivery[282-284].

The ThermoCas9/sgRNA editing machinery (in the form of plasmid DNA, or mRNA/sgRNA, or Cas9 protein/sgRNA) can be delivered *in vivo* mainly through tissue-specific injections or intravenous injections. In the first case, edited cells are typically distributed unevenly in the target tissue, leading to poor therapeutic outcomes[285]. In the second case, the CRISPR components are transferred to the target tissue through the circulatory system and CRISPR cargos are quickly degraded by endogenous proteases, nucleases, opsonins, phagocytes, and lysosomal enzymes before reaching the target tissue[284-287]. Moreover, CRISPR cargos cannot easily cross the cell membrane, due to their hydrophilic nature, anionic charge, and large size. In addition, they induce innate immune response (through pattern recognition receptors) as well as adaptive immunity, especially in the case of SpyCas9 or SauCas9 proteins as they derive from human pathogens[288,289]. Hence, the CRISPR components are usually packaged into vesicles for protection against degradation, efficient delivery, and prevention of immune responses. These vesicles are either viral or non-viral vectors. Although viral vectors provide highly efficient *in vivo* delivery, they have raised concerns regarding toxicity and immunogenicity[290-295]. Specifically, lentiviruses (LVs, packaging capacity 8 kb) integrate their genome into host cells, resulting in overexpression of the CRISPR components and thus in extensive off-target editing, insertional mutagenesis and persistent immune reactions[296]. The widely used adeno-associated viruses (AAVs, packaging capacity 4.7 kb) provide episomal CRISPR expression, avoiding persistent genomic integration and off-target editing[297,298]. However, they may trigger cell-mediated and humoral immune responses, questioning their safe use[299-302].

Exploiting alternative options, several non-viral vectors of controlled size, shape and charge have been developed and applied for efficient and safe *in vivo* gene editing in diverse types of tissues (eyes, brain, adipocytes, liver, muscles) and malignancies (superficial, internal)[303-309]. The most common non-viral vectors are based on lipid, inorganic, polymeric, and hybrid nanoparticles, or on cell-penetrating peptides. Overall, viral vectors generally provide high efficiency but low safety, while non-viral vectors operate in the opposite way. Hence, we suggest *in vivo* delivery of ThermoCas9 and its sgRNA (as nucleic acids or RNP) through lipid-based nanoparticles, as these not only protect the CRISPR cargo from degradation but also facilitate endocytosis into the intracellular compartments and present enhanced efficiencies[310-312]. Alternatively, especially in the case of *ex vivo* delivery, the Induced Transduction by Osmocytosis and Propanebetaine (iTOP) technology that harnesses a natural cellular uptake process (macropinocytosis) is proposed[313,314]. The iTOP technology presents several advantages over conventional delivery methods (microinjection, electroporation and lipofection) because of its high delivery efficiency in diverse (even difficult-to-transduce) human cell types as well as because of its ability to preserve cell viability[313,314].

Regulation and ethical considerations of CRISPR-Cas technology

During the last decade, the CRISPR-Cas technology has convinced the scientific community that it has the potential to revolutionise future developments in biotechnology (microorganisms), in agriculture (plants), and in gene therapy (humans). However, before entering the golden age of gene therapy and crop improvement, multiple ethical, societal and political concerns are constantly being raised, creating the need for multidisciplinary debates that also should engage the general public. Here, we discuss the current advances in the regulation of the CRISPR-Cas technology and we pinpoint some related ethical considerations.

In the field of human medicine and reproduction, gene editing can be applied in either somatic or germline cells. Gene editing of somatic cells is not heritable and thus it is allowed for basic research and treatment of serious diseases. However, it is banned for improvement of natural human traits (eugenics) and for embryos beyond 14 days of their development[315,316]. On the other hand, germline cell gene editing passes the genetic modification to the next generation(s) and, therefore, it is prohibited and highly

discouraged by the UNESCO Declaration on the Human Genome and Human Rights as well as the Council of Europe Convention on Human Rights and Biomedicine[317-320]. Nevertheless, in 2018 a Chinese scientist claimed that they had created genetically modified babies with immunity against human immunodeficiency virus (HIV). This violation of ethical and regulatory norms led them to jail and raised awareness in both the scientific community and the society[321]. Until the issue of off-target editing is resolved, the CRISPR-Cas technology cannot be applied in germline cells due to unpredictable future consequences not only in the babies and their descendants but also in the human race[322].

Regarding ecology, the CRISPR-mediated gene drive technology may speed up the transmission of a genetic recombination event from an organism (for example mosquito) to its entire population, surpassing the typical Mendelian inheritance rate[323-325]. As such, gene drives have high potential for elimination of deadly diseases, like malaria, and for control or alteration of certain animals and plant pests[326-329]. Once released, however, the impact on the environmental ecology is irreversible. As such, it is crucial that robust off-switch/kill-switch control systems are developed, especially for unexpected or emergency cases[330-333]. Currently, there is no regulation regarding the gene drive technology, leaving it vulnerable to conflicts among scientists and politicians[334-336].

In the agricultural sector, gene editing of plants is regulated based on either the process (European Union, New Zealand) or the product (United States, Canada, Argentina, Japan, Australia)[337-343]. Regulation based on the process involves genetic engineering without or with introduction of foreign genetic material. In the first case, the CRISPR-Cas technology triggers site-directed mutagenesis to disrupt the target gene, while in the second case a transgene is provided[344]. Notably, products generated using site-directed mutagenesis are considered as non-genetically modified organisms (non-GMOs) in the aforementioned non-EU countries, while they are considered as GMOs by the European Union because the genetic material is modified in a 'non-natural way'[343,345]. However, at least in case of relatively small genetic changes, this modification could potentially be the outcome of natural evolution, and presents higher specificity and efficacy compared to traditionally used chemical or physical mutagenesis techniques[346-350]. Moreover, the CRISPR cargo is not delivered in the genome when expressed transiently (RNP or RNA delivery), or it can be removed from the progeny via genetic segregation if needed (DNA integration). Consequently, site-directed mutagenesis is considered safe by some countries[350,351]. Regulation based on the product refers to the generation of cisgenic or transgenic crops. Cisgenic crops do not contain foreign genetic material, while transgenic crops have received exogenous genetic information. Cisgenic crops are considered as non-GMO by

several non-EU countries, while they are considered as GMOs by the European Union if they have been created in a 'non-natural way'[346]. As a consequence, the inconsistency of GMO legislation among different countries has created a gap in the global market of plant biotechnology, posing bureaucratic burden and financial costs for marketing and labelling of GMO products especially within European Union and New Zealand[352,353]. Hence, there is an imperative need to resolve these bipolar regulatory criteria and align the conception of GMO definition among law-makers, scientists and society. In 2023, the European Union will announce future plans on this matter.

Altogether, it is crucial to pinpoint the correct bottlenecks of safer implementation of the CRISPR-Cas technology and motivate fruitful discussions among universities, governments, and societies to articulate an updated and secure ethical and legal framework. After decades of a polarised discussion between groups for or against gene editing in general and of crops in particular, it may be important to discuss examples of editing that not only are safe in terms of health and ecology, but that also will benefit society.

References

1. Joubert, D. M. (1977). Animal breeding--from Bakewell to Hammond: 1725-1964. Journal of the South African Veterinary Association, 48(4), 233-236.
2. Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *Die Entdeckung der Doppelhelix*, 97.
3. Watson, J. D., & Crick, F. H. (1953). A structure for deoxyribose nucleic acid.
4. Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiological reviews*, *32*(4), 403-430.
5. Little, J. W., Zimmerman, S. B., Oshinsky, C. K., & Gellert, M. (1967). Enzymatic joining of DNA strands, II. An enzyme-adenylate intermediate in the dpn-dependent DNA ligase reaction. *Proceedings of the National Academy of Sciences*, *58*(5), 2004-2011.
6. Smith, H. O., & Welcox, K. W. (1970). A restriction enzyme from Hemophilus influenzae: I. Purification and general properties. *Journal of molecular biology*, *51*(2), 379-391.
7. Mandel, M., & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *Journal of molecular biology*, *53*(1), 159-162.
8. Jackson, D. A., Symons, R. H., & Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proceedings of the National Academy of Sciences*, *69*(10), 2904-2909.
9. Cohen, S. N., & Chang, A. C. (1973). Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants. *Proceedings of the National Academy of Sciences*, *70*(5), 1293-1297.
10. Jaenisch, R., & Mintz, B. (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proceedings of the national academy of sciences*, *71*(4), 1250-1254.
11. Bevan, M. W., Flavell, R. B., & Chilton, M. D. (1983). A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, *304*(5922), 184-187.
12. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.

13. Mullis, K. B. (1994). *The polymerase chain reaction* (Vol. 41, No. 5). Springer science & business media.
14. Neumann, E., Schaefer-Ridder, M., Wang, Y., & Hofschneider, P. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. *The EMBO journal*, *1*(7), 841-845.
15. Klein, T. M., Wolf, E. D., Wu, R., & Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, *327*(6117), 70-73.
16. Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., & Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, *330*(6148), 576-578.
17. Thomas, K. R., & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, *51*(3), 503-512.
18. Sauer, B., & Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proceedings of the National Academy of Sciences*, *85*(14), 5166-5170.
19. Sadowski, P. D. (1995). The Flp Recombinase of th 2-µm Plasmid of Saccharomyces cerevisiae. *Progress in nucleic acid research and molecular biology*, *51*, 53-91.
20. Murphy, K. C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. *Journal of bacteriology*, *180*(8), 2063-2071.
21. Muyrers, J. P., Zhang, Y., Benes, V., Testa, G., Rientjes, J. M., & Stewart, A. F. (2004). ET Recombination. In *Bacterial Artificial Chromosomes* (pp. 107-121). Humana Press.
22. Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, *93*(3), 1156-1160.
23. Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., ... & Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, *186*(2), 757-761.
24. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, *31*(3), 233-239.
25. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *elife*, *2*, e00471.
26. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.
27. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823-826.
28. Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D. L., Wei, P., ... & Zhu, J. K. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell research*, *23*(10), 1229-1232.
29. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173-1183.
30. Van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., & Brouns, S. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in biochemical sciences*, *34*(8), 401-407.
31. Deveau, H., Garneau, J. E., & Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annual review of microbiology*, *64*, 475-493.
32. Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, *327*(5962), 167-170.
33. Karginov, F. V., & Hannon, G. J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. *Molecular cell*, *37*(1), 7-19.
34. Koonin, E. V., & Makarova, K. S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 biology reports*, *1*.
35. Sorek, R., Kunin, V., & Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, *6*(3), 181-186.
36. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., ... & Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67-83.
37. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424.
38. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., ... & Kondo, A. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, *353*(6305), aaf8729.
39. Jasin, M., & Rothstein, R. (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology*, *5*(11), a012740.
40. Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., ... & She, Q. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research*, *44*(4), e34-e34.

41. Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39), E2579-E2586.
42. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., ... & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149-157.
43. Klompe, S. E., Vo, P. L., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219-225.
44. Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H., & Liu, D. R. (2016). A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic acids research*, *44*(20), 9758-9770.
45. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173-1183.
46. Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., ... & Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*(2), 442-451.
47. Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., ... & Koonin, E. V. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*, *60*(3), 385-397.
48. Abudayyeh, O. O., Gootenberg, J. S., Franklin, B., Koob, J., Kellner, M. J., Ladha, A., ... & Zhang, F. (2019). A cytosine deaminase for programmable single-base RNA editing. *Science*, *365*(6451), 382-386.
49. Gao, Y., Xiong, X., Wong, S., Charles, E. J., Lim, W. A., & Qi, L. S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nature methods*, *13*(12), 1043-1049.
50. Li, B., Zeng, C., Li, W., Zhang, X., Luo, X., Zhao, W., ... & Dong, Y. (2018). Synthetic oligonucleotides inhibit CRISPR-Cpf1-mediated genome editing. *Cell reports*, *25*(12), 3262-3272.
51. Shin, J., Jiang, F., Liu, J. J., Bray, N. L., Rauch, B. J., Baik, S. H., ... & Doudna, J. A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. *Science advances*, *3*(7), e1701620.
52. Frances, A., & Cordelier, P. (2020). The emerging role of cytidine deaminase in human diseases: a new opportunity for therapy?. Molecular Therapy, 28(2), 357-366.
53. Iyer, L. M., Zhang, D., Rogozin, I. B., & Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic acids research, 39(22), 9473-9497.
54. Nikkilä, J., Kumar, R., Campbell, J., Brandsma, I., Pemberton, H. N., Wallberg, F., ... & Lord, C. J. (2017). Elevated APOBEC3B expression drives a kataegic-like mutation signature and replication stress-related therapeutic vulnerabilities in p53-defective cells. British journal of cancer, 117(1), 113-123.
55. Teichert, I. (2020). Fungal RNA editing: who, when, and why?. Applied Microbiology and Biotechnology, 104(13), 5689-5695.
56. Skaldin, M., Tuittila, M., Zavialov, A. V., & Zavialov, A. V. (2018). Secreted bacterial adenosine deaminase is an evolutionary precursor of adenosine deaminase growth factor. Molecular Biology and Evolution, 35(12), 2851-2861.
57. Bar-Yaacov, D., Mordret, E., Towers, R., Biniashvili, T., Soyris, C., Schwartz, S., ... & Pilpel, Y. (2017). RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. Genome research, 27(10), 1696-1703.
58. Moriwaki, Y., Yamamoto, T., & Higashino, K. (1999). Enzymes involved in purine metabolism-a review of histochemical localization and functional implications. Histology and histopathology, 14(4), 1321-1340.
59. Wolf, J., Gerber, A. P., & Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. The EMBO journal, 21(14), 3841-3851.
60. Mannion, N., Arieti, F., Gallo, A., Keegan, L. P., & O'Connell, M. A. (2015). New insights into the biological role of mammalian ADARs; the RNA editing proteins. Biomolecules, 5(4), 2338-2362.
61. Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., ... & Liu, D. R. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature, 583(7817), 631-637.
62. Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. Nature, 551(7681), 464-471.
63. Merkle, T., Merz, S., Reautschnig, P., Blaha, A., Li, Q., Vogel, P., ... & Stafforst, T. (2019). Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. Nature biotechnology, 37(2), 133-138.
64. Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., ... & Zhang, X. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. Nature biotechnology, 39(1), 35-40.
65. Sakata, R. C., Ishiguro, S., Mori, H., Tanaka, M., Tatsuno, K., Ueda, H., ... & Yachie, N. (2020). Base editors for simultaneous introduction of C-to-T and A-to-G mutations. Nature biotechnology, 38(7), 865-869.

66. Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č., & Koonin, E. V. (2020). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic acids research*, *48*(16), 8828-8847.
67. Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids*, *4*, e264.
68. Marino, N. D., Zhang, J. Y., Borges, A. L., Sousa, A. A., Leon, L. M., Rauch, B. J., ... & Bondy-Denomy, J. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, *362*(6411), 240-242.
69. Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., ... & Davidson, A. R. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature microbiology*, *1*(8), 1-6.
70. Rauch, B. J., Silvis, M. R., Hultquist, J. F., Waters, C. S., McGregor, M. J., Krogan, N. J., & Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell*, *168*(1-2), 150-158.
71. Uribe, R. V., van der Helm, E., Misiakou, M. A., Lee, S. W., Kol, S., & Sommer, M. O. (2019). Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. *Cell host & microbe*, *25*(2), 233-241.
72. Borges, A. L., Zhang, J. Y., Rollins, M. F., Osuna, B. A., Wiedenheft, B., & Bondy-Denomy, J. (2018). Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity. *Cell*, *174*(4), 917-925.
73. Stanley, S. Y., & Maxwell, K. L. (2018). Phage-encoded anti-CRISPR defenses. *Annual review of genetics*, *52*, 445-464.
74. Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M. F., Hidalgo-Reyes, Y., ... & Davidson, A. R. (2015). Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*, *526*(7571), 136-139.
75. Pawluk, A., Bondy-Denomy, J., Cheung, V. H., Maxwell, K. L., & Davidson, A. R. (2014). A new group of phage anti-CRISPR genes inhibits the type IE CRISPR-Cas system of Pseudomonas aeruginosa. *MBio*, *5*(2), e00896-14.
76. Maxwell, K. L., Garcia, B., Bondy-Denomy, J., Bona, D., Hidalgo-Reyes, Y., & Davidson, A. R. (2016). The solution structure of an anti-CRISPR protein. *Nature communications*, *7*(1), 1-5.
77. Wang, J., Ma, J., Cheng, Z., Meng, X., You, L., Wang, M., ... & Wang, Y. (2016). A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. *Cell research*, *26*(10), 1165-1168.
78. Van Erp, P. B., Jackson, R. N., Carter, J., Golden, S. M., Bailey, S., & Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic acids research*, *43*(17), 8381-8391.
79. Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., ... & Wiedenheft, B. (2017). Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell*, *169*(1), 47-57.
80. Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., ... & Subramaniam, S. (2017). Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR-Cas surveillance complex. *Cell*, *171*(2), 414-426.
81. Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., ... & Gao, G. F. (2017). Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. *Cell research*, *27*(7), 853-864.
82. Wang, X., Yao, D., Xu, J. G., Li, A., Xu, J., Fu, P., ... & Zhu, Y. (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nature structural & molecular biology*, *23*(9), 868-870.
83. Pawluk, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T. F., Davidson, A. R., & Maxwell, K. L. (2017). Disabling a type IE CRISPR-Cas nuclease with a bacteriophage-encoded anti-CRISPR protein. *MBio*, *8*(6), e01751-17.
84. Harrington, L. B., Doxzen, K. W., Ma, E., Liu, J. J., Knott, G. J., Edraki, A., ... & Doudna, J. A. (2017). A broad-spectrum inhibitor of CRISPR-Cas9. *Cell*, *170*(6), 1224-1233.
85. Zhu, Y., Gao, A., Zhan, Q., Wang, Y., Feng, H., Liu, S., ... & Gao, P. (2019). Diverse mechanisms of CRISPR-Cas9 inhibition by type IIC anti-CRISPR proteins. *Molecular cell*, *74*(2), 296-309.
86. Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H. E., ... & Sontheimer, E. J. (2018). Potent Cas9 inhibition in bacterial and human cells by AcrIIC4 and AcrIIC5 anti-CRISPR proteins. *MBio*, *9*(6), e02321-18.
87. Smargon, A. A., Cox, D. B., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., ... & Zhang, F. (2017). Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Molecular cell*, *65*(4), 618-630.
88. Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., ... & Davidson, A. R. (2016). Naturally occurring off-switches for CRISPR-Cas9. *Cell*, *167*(7), 1829-1838.
89. Mougiakos, I., Mohanraju, P., Bosma, E. F., Vrouwe, V., Finger Bou, M., Naduthodi, M. I., ... & Van Der Oost, J. (2017). Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature communications*, *8*(1), 1-11.
90. Harrington, L. B., Paez-Espino, D., Staahl, B. T., Chen, J. S., Ma, E., Kyrpides, N. C., & Doudna, J. A. (2017). A thermostable Cas9 with increased lifetime in human plasma. *Nature communications*, *8*(1), 1-8.
91. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science*, *337*(6096), 816-821.
92. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nuclease with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490-495.

93. Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84-88.
94. Bisaria, N., Jarmoskaite, I., & Herschlag, D. (2017). Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell systems*, *4*(1), 21-29.
95. Wu, Z., Yang, H., & Colosi, P. (2010). Effect of genome size on AAV vector packaging. *Molecular Therapy*, *18*(1), 80-86.
96. Sundaresan, R., Parameshwaran, H. P., Yogesha, S. D., Keilbarth, M. W., & Rajan, R. (2017). RNA-independent DNA cleavage activities of Cas9 and Cas12a. *Cell reports*, *21*(13), 3728-3739.
97. O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M., & Doudna, J. A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*, *516*(7530), 263-266.
98. Nelles, D. A., Fang, M. Y., O'Connell, M. R., Xu, J. L., Markmiller, S. J., Doudna, J. A., & Yeo, G. W. (2016). Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell*, *165*(2), 488-496.
99. Batra, R., Nelles, D. A., Pirie, E., Blue, S. M., Marina, R. J., Wang, H., ... & Yeo, G. W. (2017). Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. *Cell*, *170*(5), 899-912.
100. Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A., & Doudna, J. A. (2018). RNA-dependent RNA targeting by CRISPR-Cas9. *elife*, *7*, e32724.
101. Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., ... & Doudna, J. A. (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*, *550*(7676), 407-410.
102. Mir, A., Edraki, A., Lee, J., & Sontheimer, E. J. (2018). Type II-C CRISPR-Cas9 biology, mechanism, and application. *ACS chemical biology*, *13*(2), 357-365.
103. Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K., & Doudna, J. A. (2015). Single-stranded DNA cleavage by divergent CRISPR-Cas9 enzymes. *Molecular cell*, *60*(3), 398-407.
104. Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., ... & Toki, S. (2019). Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. *Nature plants*, *5*(1), 14-17.
105. Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., ... & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, *556*(7699), 57-63.
106. Kim, S., Bae, T., Hwang, J., & Kim, J. S. (2017). Rescue of high-specificity Cas9 variants using sgRNAs with matched 5'nucleotides. *Genome biology*, *18*(1), 1-6.
107. Vakulskas, C. A., Dever, D. P., Rettig, G. R., Turk, R., Jacobi, A. M., Collingwood, M. A., ... & Behlke, M. A. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nature medicine*, *24*(8), 1216-1224.
108. Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y. H., ... & Kim, J. S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications*, *9*(1), 1-10.
109. Tsai, S. Q., Nguyen, N., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets. *Nature*, *529*(7587), 490-495.
110. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., ... & Cereseto, A. (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nature biotechnology*, *36*(3), 265-271.
111. Kulcsár, P. I., Tálas, A., Huszár, K., Ligeti, Z., Tóth, E., Weinhardt, N., ... & Welker, E. (2017). Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome biology*, *18*(1), 1-17.
112. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., ... & Nureki, O. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science, 361(6408), 1259-1262.
113. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., ... & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481-485..
114. Friedland, A. E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., ... & Bumcrot, D. (2015). Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome biology*, *16*(1), 1-10.
115. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., & Joung, J. K. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature biotechnology*, *33*(12), 1293-1298.
116. Raitskin, O., Schudoma, C., West, A., & Patron, N. J. (2019). Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: an expanded toolkit for precision genome engineering. *PLoS One*, *14*(2), e0211598.
117. Agudelo, D., Carter, S., Velimirovic, M., Duringer, A., Rivest, J. F., Levesque, S., ... & Doyon, Y. (2020). Versatile and robust genome editing with Streptococcus thermophilus CRISPR1-Cas9. *Genome research*, *30*(1), 107-117.
118. Müller, M., Lee, C. M., Gasiunas, G., Davis, T. H., Cradick, T. J., Siksnys, V., ... & Mussolino, C. (2016). Streptococcus thermophilus CRISPR-Cas9 systems enable specific editing of the human genome. *Molecular Therapy*, *24*(3), 636-644.

119. Acharya, S., Mishra, A., Paul, D., Ansari, A. H., Azhar, M., Kumar, M., ... & Chakraborty, D. (2019). Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proceedings of the National Academy of Sciences*, *116*(42), 20959-20968.
120. Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L. F., Sontheimer, E. J., & Thomson, J. A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proceedings of the National Academy of Sciences*, *110*(39), 15644-15649.
121. Wang, Y., Liu, K. I., Sutrisnoh, N. A. B., Srinivasan, H., Zhang, J., Li, J., ... & Tan, M. H. (2018). Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells. *Genome biology*, *19*(1), 1-16.
122. Amrani, N., Gao, X. D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., ... & Sontheimer, E. J. (2018). NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome Biology*, *19*(1), 1-25.
123. Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C. Q., ... & Sontheimer, E. J. (2019). A compact, high-accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. *Molecular cell*, *73*(4), 714-726.
124. Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., ... & Kim, J. S. (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nature communications*, *8*(1), 1-12.
125. Hand, T. H., Roth, M. O., Smith, C. L., Shiel, E., Klein, K. N., Gilbert, D. M., & Li, H. (2021). Catalytically Enhanced Cas9 through Directed Protein Evolution. *The CRISPR Journal*, *4*(2), 223-232.
126. Hu, Z., Wang, S., Zhang, C., Gao, N., Li, M., Wang, D., ... & Wang, Y. (2020). A compact Cas9 ortholog from Staphylococcus Auricularis (SauriCas9) expands the DNA targeting scope. *PLoS biology*, *18*(3), e3000686.
127. Cui, Z., Tian, R., Huang, Z., Jin, Z., Li, L., Liu, J., ... & Hu, Z. (2022). FrCas9 is a CRISPR/Cas9 system with high editing efficiency and fidelity. *Nature communications*, *13*(1), 1-12.
128. Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S. H., & Kim, J. S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature biotechnology*, *34*(8), 863-868.
129. Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., ... & Joung, J. K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature biotechnology*, *34*(8), 869-874.
130. Tóth, E., Czene, B. C., Kulcsár, P. I., Krausz, S. L., Tálas, A., Nyeste, A., ... & Welker, E. (2018). Mb-and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. *Nucleic acids research*, *46*(19), 10272-10285.
131. Tu, M., Lin, L., Cheng, Y., He, X., Sun, H., Xie, H., ... & Gu, F. (2017). A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic acids research*.
132. Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., ... & Zhang, F. (2019). Engineering of CRISPR-Cas12b for human genome editing. *Nature communications*, *10*(1), 1-8.
133. Teng, F., Cui, T., Feng, G., Guo, L., Xu, K., Gao, Q., ... & Li, W. (2018). Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell discovery*, *4*(1), 1-15.
134. Kim, D. Y., Lee, J. M., Moon, S. B., Chin, H. J., Park, S., Lim, Y., ... & Kim, Y. S. (2022). Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nature biotechnology*, *40*(1), 94-102.
135. Hu, X., Wang, C., Fu, Y., Liu, Q., Jiao, X., & Wang, K. (2016). Expanding the range of CRISPR/Cas9 genome editing in rice. *Molecular plant*, *9*(6), 943-945.
136. Wang, J., Meng, X., Hu, X., Sun, T., Li, J., Wang, K., & Yu, H. (2019). xCas9 expands the scope of genome editing with reduced efficiency in rice. *Plant biotechnology journal*, *17*(4), 709.
137. Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., ... & Liu, D. R. (2020). Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature biotechnology*, *38*(4), 471-481.
138. Xu, L., Zhang, C., Li, H., Wang, P., Gao, Y., Mokadam, N. A., ... & Han, R. (2021). Efficient precise in vivo base editing in adult dystrophic mice. *Nature communications*, *12*(1), 1-14.
139. Liang, F., Zhang, Y., Li, L., Yang, Y., Fei, J. F., Liu, Y., & Qin, W. (2022). SpG and SpRY variants expand the CRISPR toolbox for genome editing in zebrafish. *Nature communications*, *13*(1), 1-10.
140. Chatterjee, P., Lee, J., Nip, L., Koseki, S. R., Tysinger, E., Sontheimer, E. J., ... & Jakimo, N. (2020). A Cas9 with PAM recognition for adenine dinucleotides. *Nature communications*, *11*(1), 1-6.
141. Zhong, Z., Sretenovic, S., Ren, Q., Yang, L., Bao, Y., Qi, C., ... & Zhang, Y. (2019). Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. *Molecular Plant*, *12*(7), 1027-1036.
142. Ge, Z., Zheng, L., Zhao, Y., Jiang, J., Zhang, E. J., Liu, T., ... & Qu, L. J. (2019). Engineered xCas9 and SpCas9-NG variants broaden PAM recognition sites to generate mutations in Arabidopsis plants. *Plant Biotechnology Journal*, *17*(10), 1865.
143. Chatterjee, P., Jakimo, N., & Jacobson, J. M. (2018). Minimal PAM specificity of a highly similar SpCas9 ortholog. *Science advances*, *4*(10), eaau0766.

144. Ran, F. A. C. L., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., ... & Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, *520*(7546), 186-191.
145. Hirano, H., Gootenberg, J. S., Horii, T., Abudayyeh, O. O., Kimura, M., Hsu, P. D., ... & Nureki, O. (2016). Structure and engineering of Francisella novicida Cas9. *Cell*, *164*(5), 950-961.
146. Gao, N., Zhang, C., Hu, Z., Li, M., Wei, J., Wang, Y., & Liu, H. (2020). Characterization of Brevibacillus laterosporus Cas9 (BlatCas9) for mammalian genome editing. *Frontiers in cell and developmental biology*, *8*, 583164.
147. Fedorova, I., Vasileva, A., Selkova, P., Abramova, M., Arseniev, A., Pobegalov, G., ... & Severinov, K. (2020). PpCas9 from Pasteurella pneumotropica—a compact Type II-C Cas9 ortholog active in human cells. *Nucleic acids research*, *48*(21), 12297-12309.
148. Hirano, S., Abudayyeh, O. O., Gootenberg, J. S., Horii, T., Ishitani, R., Hatada, I., ... & Nureki, O. (2019). Structural basis for the promiscuous PAM recognition by Corynebacterium diphtheriae Cas9. *Nature communications*, *10*(1), 1-11.
149. Steinert, J., Schiml, S., Fauser, F., & Puchta, H. (2015). Highly efficient heritable plant genome engineering using Cas9 orthologues from Streptococcus thermophilus and Staphylococcus aureus. *The Plant Journal*, *84*(6), 1295-1305.
150. Kaya, H., Mikami, M., Endo, A., Endo, M., & Toki, S. (2016). Highly specific targeted mutagenesis in plants using Staphylococcus aureus Cas9. *Scientific reports*, *6*(1), 1-9.
151. Qin, R., Li, J., Li, H., Zhang, Y., Liu, X., Miao, Y., ... & Wei, P. (2019). Developing a highly efficient and wildly adaptive CRISPR-SaCas9 toolset for plant genome editing. *Plant Biotechnology Journal*, *17*(4), 706.
152. Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., & Church, G. M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods*, *10*(11), 1116-1121.
153. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., ... & Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759-771.
154. Zhong, Z., Zhang, Y., You, Q., Tang, X., Ren, Q., Liu, S., ... & Qi, Y. (2018). Plant genome editing using FnCpf1 and LbCpf1 nucleases at redefined and altered PAM sites. *Molecular plant*, *11*(7), 999-1002.
155. Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. *Cell*, *167*(7), 1814-1828.
156. Teng, F., Cui, T., Gao, Q., Guo, L., Zhou, Q., & Li, W. (2019). Artificial sgRNAs engineered for genome editing with new Cas12b orthologs. *Cell discovery*, *5*(1), 1-4.
157. Takeda, S. N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., ... & Nureki, O. (2021). Structure of the miniature type VF CRISPR-Cas effector enzyme. *Molecular Cell*, *81*(3), 558-570.
158. Bigelyte, G., Young, J. K., Karvelis, T., Budre, K., Zedaveinyte, R., Djukanovic, V., ... & Siksnys, V. (2021). Miniature type VF CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nature communications*, *12*(1), 1-8.
159. Xu, X., Chemparathy, A., Zeng, L., Kempton, H. R., Shang, S., Nakamura, M., & Qi, L. S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Molecular Cell*, *81*(20), 4333-4345.
160. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., ... & Ji, Q. (2021). Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nature Chemical Biology*, *17*(11), 1132-1138.
161. McGaw, C., Garrity, A. J., Munoz, G. Z., Haswell, J. R., Sengupta, S., Keston-Smith, E., ... & Chong, S. (2022). Engineered Cas12i2 is a versatile high-efficiency platform for therapeutic genome editing. *Nature communications*, *13*(1), 1-11.
162. Tan, L. L., Heng, E., Zulkarnain, N., Leong, C. Y., Ng, V., Yang, L. K., ... & Wong, F. T. (2021). Application of Cas12j for Streptomyces editing and cluster activation. *bioRxiv*.
163. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C. A., Li, Z., Cress, B. F., ... & Doudna, J. A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science*, *369*(6501), 333-337.
164. Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Topkar, V. V., Zheng, Z., & Joung, J. K. (2015). Broadening Staphylococcus aureus Cas9 targeting range by modifying PAM recognition. *Nature biotechnology*, *33*(12), 1293.
165. Liu, T., Zeng, D., Zheng, Z., Lin, Z., Xue, Y., Li, T., ... & Zhu, Q. (2021). The ScCas9++ variant expands the CRISPR toolbox for genome editing in plants. *Journal of Integrative Plant Biology*, *63*(9), 1611-1619.
166. Chatterjee, P., Jakimo, N., Lee, J., Amrani, N., Rodriguez, T., Koseki, S. R., ... & Jacobson, J. (2020). An engineered ScCas9 with broad PAM range and high specificity and activity. *Nature Biotechnology*, *38*(10), 1154-1158.
167. Gao, L., Cox, D. B., Yan, W. X., Manteiga, J. C., Schneider, M. W., Yamano, T., ... & Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. *Nature biotechnology*, *35*(8), 789-792.
168. Li, S., Zhang, X., Wang, W., Guo, X., Wu, Z., Du, W., ... & Xia, L. (2018). Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. *Molecular Plant*, *11*(7), 995-998.
169. Nishimasu, H., Yamano, T., Gao, L., Zhang, F., Ishitani, R., & Nureki, O. (2017). Structural basis for the altered PAM recognition by engineered CRISPR-Cpf1. *Molecular cell*, *67*(1), 139-147.
170. Tóth, E., Varga, E., Kulcsár, P. I., Kocsis-Jutka, V., Krausz, S. L., Nyeste, A., ... & Welker, E. (2020). Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. *Nucleic acids research*, *48*(7), 3722-3733.

171. Nakagawa, R., Ishiguro, S., Okazaki, S., Mori, H., Tanaka, M., Aburatani, H., ... & Nureki, O. (2022). Engineered Campylobacter jejuni Cas9 variant with enhanced activity and broader targeting range. *Communications biology*, *5*(1), 1-8.
172. Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., ... & Koonin, E. V. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*, *60*(3), 385-397.
173. Tsui, T. K. M., Hand, T. H., Duboy, E. C., & Li, H. (2017). The impact of DNA topology and guide length on target selection by a cytosine-specific Cas9. *ACS synthetic biology*, *6*(6), 1103-1113.
174. Adalsteinsson, B. T., Kristjansdottir, T., Merre, W., Helleux, A., Dusaucy, J., Tourigny, M., ... & Hreggvidsson, G. O. (2021). Efficient genome editing of an extreme thermophile, Thermus thermophilus, using a thermostable Cas9 variant. *Scientific reports*, *11*(1), 1-15.
175. Walker, J. E., Lanahan, A. A., Zheng, T., Toruno, C., Lynd, L. R., Cameron, J. C., ... & Eckert, C. A. (2020). Development of both type I–B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium Clostridium thermocellum. *Metabolic engineering communications*, *10*, e00116.
176. Gallo, G., Mougiakos, I., Bianco, M., Carbonaro, M., Carpentieri, A., Illiano, A., ... & Fiorentino, G. (2021). A hyperthermoactive-Cas9 editing tool reveals the role of a unique arsenite methyltransferase in the arsenic resistance system of Thermus thermophilus HB27. *Mbio*, *12*(6), e02813-21.
177. Ganguly, J., Martin-Pascual, M., & van Kranenburg, R. (2020). CRISPR interference (CRISPRi) as transcriptional repression tool for Hungateiclostridium thermocellum DSM 1313. *Microbial biotechnology*, *13*(2), 339-349.
178. Wang, Y., Wang, Y., Pan, D., Yu, H., Zhang, Y., Chen, W., ... & Ji, Q. (2022). Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease. *Cell Reports*, *40*(13), 111418.
179. Harrington, L. B., Burstein, D., Chen, J. S., Paez-Espino, D., Ma, E., Witte, I. P., ... & Doudna, J. A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*, *362*(6416), 839-842.
180. Karvelis, T., Bigelyte, G., Young, J. K., Hou, Z., Zedaveinyte, R., Budre, K., ... & Siksnys, V. (2020). PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic acids research*, *48*(9), 5016-5023.
181. Carabias, A., Fuglsang, A., Temperini, P., Pape, T., Sofos, N., Stella, S., ... & Montoya, G. (2021). Structure of the mini-RNA-guided endonuclease CRISPR-Cas12j3. *Nature Communications*, *12*(1), 1-12.
182. Saha, C., Mohanraju, P., Stubbs, A., Dugar, G., Hoogstrate, Y., Kremers, G. J., ... & Louwen, R. (2020). Guide-free Cas9 from pathogenic Campylobacter jejuni bacteria causes severe damage to DNA. *Science advances*, *6*(25), eaaz4849.
183. Dugar, G., Leenay, R. T., Eisenbart, S. K., Bischler, T., Aul, B. U., Beisel, C. L., & Sharma, C. M. (2018). CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the Campylobacter jejuni Cas9. *Molecular cell*, *69*(5), 893-905.
184. Rousseau, B. A., Hou, Z., Gramelspacher, M. J., & Zhang, Y. (2018). Programmable RNA cleavage and recognition by a natural CRISPR-Cas9 system from Neisseria meningitidis. *Molecular cell*, *69*(5), 906-914.
185. Swarts, D. C., & Jinek, M. (2019). Mechanistic Insights into the cis-and trans-Acting DNase Activities of Cas12a. *Molecular cell*, *73*(3), 589-600.
186. Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., & Doudna, J. A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, *360*(6387), 436-439.
187. Jarczynska, Z. D., Rendsvig, J. K., Pagels, N., Viana, V. R., Nødvig, C. S., Kirchner, F. H., ... & Mortensen, U. H. (2021). DIVERSIFY: A fungal multispecies gene expression platform. *ACS Synthetic Biology*, *10*(3), 579-588.
188. Daas, M. J., Vriesendorp, B., van de Weijer, A. H., van der Oost, J., & van Kranenburg, R. (2018). Complete genome sequence of Geobacillus thermodenitrificans T12, a potential host for biotechnological applications. *Current microbiology*, *75*, 49-56.
189. Sun, W., Yang, J., Cheng, Z., Amrani, N., Liu, C., Wang, K., ... & Wang, Y. (2019). Structures of Neisseria meningitidis Cas9 complexes in catalytically poised and anti-CRISPR-inhibited states. *Molecular cell*, *76*(6), 938-952.
190. Perrakis, A., & Sixma, T. K. (2021). AI revolutions in biology: The joys and perils of AlphaFold. *EMBO reports*, *22*(11), e54046.
191. Shoemaker, S. C., & Ando, N. (2018). X-rays in the cryo-electron microscopy era: structural biology's dynamic future. *Biochemistry*, *57*(3), 277-285.
192. Zhu, X., Clarke, R., Puppala, A. K., Chittori, S., Merk, A., Merrill, B. J., ... & Subramaniam, S. (2019). Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nature structural & molecular biology*, *26*(8), 679-685.
193. Trasanidou, D. (2017). Development of Novel Tools for Genome Editing in Thermophiles. *Master of Science Major Thesis at Wageningen University & Research*, 28-31.
194. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., & Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. *Nature microbiology*, *3*(4), 423-429.

195. Yu, Y., Leete, T. C., Born, D. A., Young, L., Barrera, L. A., Lee, S. J., ... & Gaudelli, N. M. (2020). Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nature communications*, *11*(1), 1-10.
196. Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., ... & Li, Z. (2020). Efficient base editing with high precision in rabbits using YFE-BE4max. *Cell death & disease*, *11*(1), 1-11.
197. Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature biotechnology*, *35*(4), 371-376.
198. Doman, J. L., Raguram, A., Newby, G. A., & Liu, D. R. (2020). Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nature biotechnology*, *38*(5), 620-628.
199. Tan, J., Zhang, F., Karcher, D., & Bock, R. (2020). Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nature communications*, *11*(1), 1-11.
200. Yan, D., Ren, B., Liu, L., Yan, F., Li, S., Wang, G., ... & Zhou, H. (2021). High-efficiency and multiplex adenine base editing in plants using new TadA variants. *Molecular Plant*, *14*(5), 722-731.
201. Chen, L., Zhang, S., Xue, N., Hong, M., Zhang, X., Zhang, D., ... & Li, D. (2022). Engineering a precise adenine base editor with minimal bystander editing. *Nature Chemical Biology*, 1-10.
202. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., ... & Yang, H. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature*, *571*(7764), 275-278.
203. Chen, L., Park, J. E., Paa, P., Rajakumar, P. D., Prekop, H. T., Chew, Y. T., ... & Chew, W. L. (2021). Programmable C: G to G: C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nature communications*, *12*(1), 1-7.
204. Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., ... & Zhang, X. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nature biotechnology*, *39*(1), 35-40.
205. Tao, W., Liu, Q., Huang, S., Wang, X., Qu, S., Guo, J., ... & Huang, X. (2021). CABE-RY: A PAM-flexible dual-mutation base editor for reliable modeling of multi-nucleotide variants. *Molecular Therapy-Nucleic Acids*, *26*, 114-121.
206. Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., ... & Li, D. (2020). Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nature biotechnology*, *38*(7), 856-860.
207. Sakai, T., Daikai, T., Monma, H., & Maeda, H. (2002). Isolation from Nocardioides sp. Strain CT16, Purification, and Characterization of a Deoxycytidine Deaminase Extremely Thermostable in the Presence of D,L-Dithiothreitol. *Bioscience, biotechnology, and biochemistry*, *66*(8), 1646-1651.
208. Ali, N. H. & Mohamed, L. A. (2008). Nucleoside Degradation in Some Streptomyces Strains. *Asian Journal of Biochemistry*, 1(3), 1-10.
209. Elawamry, Z.A. & T.A. Elzainy, 1985. Deamination of cytidine and cytosine Streptomyces viridiviolaceus. Eygpt. J. Bot., 28: 171-176.
210. Bezborodov, A.M., T.A. Reshetilova & I.A. Golubenko, 1974. Isolation of intracellular cytidine deaminase of Penicillium sizowi and investigation of some of its properties. Mickrobiologiya, 34: 600-605.
211. Elshafei, A.M., N.H. Ali and L.A. Mohamed, 2005. Cytidine deaminase activity in *Penicillium politans* NRC-510. J. Basic Microbiol., 45: 335-343.
212. Katsuragi, T., Sakai, T., Matsumoto, K. Y., & Tonomura, K. (1986). Cytosine deaminase from Escherichia coli-production, purification, and some characteristics. *Agricultural and biological chemistry*, *50*(7), 1721-1730.
213. Woo, J. H., Heo, N. J., Ghim, S. Y., Kim, J. G., & Song, B. H. (2002). Purification and characterization of thermostable cytidine deaminase encoded by the Bacillus caldolyticus cdd gene. *Enzyme and microbial technology*, *30*(2), 153-160.
214. Cambi, A., Vincenzetti, S., De Sanctis, G., Neuhard, J., Natalini, P., & Vita, A. (2001). Cytidine deaminase from two extremophilic bacteria: cloning, expression and comparison of their structural stability. *Protein Engineering*, *14*(10), 807-813.
215. Park, Y. M., Phi, Q. T., Song, B. H., & Ghim, S. Y. (2009). Thermostability of chimeric cytidine deaminase variants produced by DNA shuffling. *Journal of microbiology and biotechnology*, *19*(12), 1536-1541.
216. Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č., & Koonin, E. V. (2020). Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic acids research*, *48*(16), 8828-8847.
217. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T., & Lee, S. Y. (2021). A versatile genetic engineering toolkit for E. coli based on CRISPR-prime editing. *Nature communications*, *12*(1), 1-11.
218. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., ... & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, *576*(7785), 149-157.
219. Anzalone, A. V., Gao, X. D., Podracky, C. J., Nelson, A. T., Koblan, L. W., Raguram, A., ... & Liu, D. R. (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nature Biotechnology*, *40*(5), 731-740.

220. Nelson, J. W., Randolph, P. B., Shen, S. P., Everette, K. A., Chen, P. J., Anzalone, A. V., ... & Liu, D. R. (2022). Engineered pegRNAs improve prime editing efficiency. *Nature biotechnology*, *40*(3), 402-410.
221. Brooks, E. M., Sheflin, L. G., & Spaulding, S. W. (1995). Secondary structure in the 3'UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR. *Biotechniques*, *19*(5), 806-12.
222. Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M., & Gerard, G. F. (1988). Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic acids research*, *16*(1), 265-277.
223. Tian, L., Kim, M. S., Li, H., Wang, J., & Yang, W. (2018). Structure of HIV-1 reverse transcriptase cleaving RNA in an RNA/DNA hybrid. *Proceedings of the National Academy of Sciences*, *115*(3), 507-512.
224. Lambowitz, A. M., & Belfort, M. (2015). Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiology spectrum*, *3*(1), 3-1.
225. Qu, G., Kaushal, P. S., Wang, J., Shigematsu, H., Piazza, C. L., Agrawal, R. K., ... & Wang, H. W. (2016). Structure of a group II intron in complex with its reverse transcriptase. *Nature structural & molecular biology*, *23*(6), 549-557.
226. Martín-Alonso, S., Frutos-Beltrán, E., & Menéndez-Arias, L. (2021). Reverse transcriptase: from transcriptomics to genome editing. *Trends in Biotechnology*, *39*(2), 194-210.
227. Toro, N., Jiménez-Zurdo, J. I., & García-Rodríguez, F. M. (2007). Bacterial group II introns: not just splicing. *FEMS microbiology reviews*, *31*(3), 342-358.
228. Dreier, J., Stormer, M., & Kleesiek, K. (2005). Use of bacteriophage MS2 as an internal control in viral reverse transcription-PCR assays. *Journal of clinical microbiology*, *43*(9), 4551-4557.
229. Frolova, L. Y., Metelyev, V. G., Ratmanova, K. I., Smirnov, V. D., Shabarova, Z. A., Prokofyev, M. A., ... & Kisselev, L. L. (1977). Reverse transcription of phage RNA and its fragment directed by synthetic heteropolymeric primers. *Nucleic Acids Research*, *4*(7), 2145-2160.
230. Weise, L. I., Heymann, M., Mayr, V., & Mutschler, H. (2019). Cell-free expression of RNA encoded genes using MS2 replicase. *Nucleic acids research*, *47*(20), 10956-10967.
231. Belfort, M., & Lambowitz, A. M. (2019). Group II intron RNPs and reverse transcriptases: from retroelements to research tools. *Cold Spring Harbor Perspectives in Biology*, *11*(4), a032375.
232. Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., ... & Lambowitz, A. M. (2013). Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *Rna*, *19*(7), 958-970.
233. Shandilya, H., Griffiths, K., Flynn, E. K., Astatke, M., Shih, P. J., Lee, J. E., ... & Bergquist, P. L. (2004). Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles*, *8*(3), 243-251.
234. Nowak, E., Miller, J. T., Bona, M. K., Studnicka, J., Szczepanowski, R. H., Jurkowski, J., ... & Nowotny, M. (2014). Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nature structural & molecular biology*, *21*(4), 389-396.
235. Rausch, J. W., Miller, J. T., & Le Grice, S. F. (2017). Reverse transcription in the Saccharomyces cerevisiae long-terminal repeat retrotransposon Ty3. *Viruses*, *9*(3), 44.
236. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., & Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research*, *41*(15), 7429-7437.
237. Guschin, D. Y., Waite, A. J., Katibah, G. E., Miller, J. C., Holmes, M. C., & Rebar, E. J. (2010). A rapid and general assay for monitoring endogenous gene modification. In *Engineered zinc finger proteins* (pp. 247-256). Humana Press, Totowa, NJ.
238. Ran, F. A. F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, *8*(11), 2281-2308.
239. Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, *31*(9), 822-826.
240. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F., Konermann, S., Agarwala, V., ... & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, *31*(9), 827-832.
241. Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, *31*(9), 839-843.
242. Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., ... & Church, G. M. (2013). Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, *31*(9), 833-838.
243. Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., ... & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, *34*(2), 184-191.

244. Singh, R., Kuscu, C., Quinlan, A., Qi, Y., & Adli, M. (2015). Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic acids research*, *43*(18), e118-e118.
245. Yang, L., Grishin, D., Wang, G., Aach, J., Zhang, C. Z., Chari, R., ... & Church, G. (2014). Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nature communications*, *5*(1), 1-6.
246. Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, *30*(10), 1473-1475.
247. Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., & Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, *30*(8), 1180-1182.
248. Jacquin, A. L., Odom, D. T., & Lukk, M. (2019). Crisflash: open-source software to generate CRISPR guide RNAs against genomes annotated with individual variation. *Bioinformatics*, *35*(17), 3146-3147.
249. McKenna, A., & Shendure, J. (2018). FlashFry: a fast and flexible tool for large-scale CRISPR target design. *BMC biology*, *16*(1), 1-6.
250. Abadi, S., Yan, W. X., Amar, D., & Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS computational biology*, *13*(10), e1005807.
251. Listgarten, J., Weinstein, M., Kleinstiver, B. P., Sousa, A. A., Joung, J. K., Crawford, J., ... & Fusi, N. (2018). Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature biomedical engineering*, *2*(1), 38-47.
252. Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., ... & Liu, Q. (2018). DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome biology*, *19*(1), 1-18.
253. Frock, R. L., Hu, J., Meyers, R. M., Ho, Y. J., Kii, E., & Alt, F. W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature biotechnology*, *33*(2), 179-186.
254. Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., ... & Joung, J. K. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, *33*(2), 187-197.
255. Wang, X., Yang, Y., Zhang, Y., Lv, X., Wei, L., Yu, H., & Liu, B. (2015). Photocurrent properties of KTa0. 65Nb0. 35O3 crystal grown by Czochralski method. *Optical Materials*, *46*, 175-178.
256. Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., ... & Kim, J. S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature methods*, *12*(3), 237-243.
257. Wienert, B., Wyman, S. K., Yeh, C. D., Conklin, B. R., & Corn, J. E. (2020). CRISPR off-target detection with DISCOVER-seq. *Nature protocols*, *15*(5), 1775-1799.
258. Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. V., Aryee, M. J., & Joung, J. K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nature methods*, *14*(6), 607-614.
259. Yan, W. X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M. W., Kallas, T., ... & Crosetto, N. (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nature communications*, *8*(1), 1-9.
260. Wienert, B., Wyman, S. K., Richardson, C. D., Yeh, C. D., Akcakaya, P., Porritt, M. J., ... & Corn, J. E. (2019). Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science*, *364*(6437), 286-289.
261. Sullivan, N. T., Allen, A. G., Atkins, A. J., Chung, C. H., Dampier, W., Nonnemacher, M. R., & Wigdahl, B. (2020). Designing safer CRISPR/Cas9 therapeutics for HIV: Defining factors that regulate and technologies used to detect off-target editing. *Frontiers in Microbiology*, *11*, 1872.
262. Cameron, P., Fuller, C. K., Donohoue, P. D., Jones, B. N., Thompson, M. S., Carter, M. M., ... & May, A. P. (2017). Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nature methods*, *14*(6), 600-606.
263. Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., ... & Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, *364*(6437), 289-292.
264. Akcakaya, P., Bobbin, M. L., Guo, J. A., Malagon-Lopez, J., Clement, K., Garcia, S. P., ... & Joung, J. K. (2018). In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature*, *561*(7723), 416-419.
265. Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B. P., & Nussenzweig, A. (2016). DNA breaks and end resection measured genome-wide by end sequencing. *Molecular cell*, *63*(5), 898-911.
266. Lensing, S. V., Marsico, G., Hänsel-Hertsch, R., Lam, E. Y., Tannahill, D., & Balasubramanian, S. (2016). DSBCapture: in situ capture and sequencing of DNA breaks. *Nature methods*, *13*(10), 855-857.
267. Lazzarotto, C. R., Malinin, N. L., Li, Y., Zhang, R., Yang, Y., Lee, G., ... & Tsai, S. Q. (2020). CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nature biotechnology*, *38*(11), 1317-1327.
268. Kim, D., Lim, K., Kim, S. T., Yoon, S. H., Kim, K., Ryu, S. M., & Kim, J. S. (2017). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nature biotechnology*, *35*(5), 475-480.
269. Liang, P., Xie, X., Zhi, S., Sun, H., Zhang, X., Chen, Y., ... & Songyang, Z. (2019). Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nature communications*, *10*(1), 1-9.

270. Kim, D., Kim, D. E., Lee, G., Cho, S. I., & Kim, J. S. (2019). Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nature biotechnology*, *37*(4), 430-435.
271. Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., ... & Gao, C. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science*, *364*(6437), 292-295.
272. McGrath, E., Shin, H., Zhang, L., Phue, J. N., Wu, W. W., Shen, R. F., ... & Ye, Z. (2019). Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nature communications*, *10*(1), 1-9.
273. Grünewald, J., Zhou, R., Iyer, S., Lareau, C. A., Garcia, S. P., Aryee, M. J., & Joung, J. K. (2019). CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nature biotechnology*, *37*(9), 1041-1048.
274. Grünewald, J., Zhou, R., Garcia, S. P., Iyer, S., Lareau, C. A., Aryee, M. J., & Joung, J. K. (2019). Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*, *569*(7756), 433-437.
275. Rees, H. A., Wilson, C., Doman, J. L., & Liu, D. R. (2019). Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Science advances*, *5*(5), eaax5717.
276. Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., ... & Liu, D. R. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature biotechnology*, *38*(7), 883-891.
277. Kim, D. Y., Moon, S. B., Ko, J. H., Kim, Y. S., & Kim, D. (2020). Unbiased investigation of specificities of prime editing systems in human cells. *Nucleic acids research*, *48*(18), 10576-10589.
278. Jin, S., Lin, Q., Luo, Y., Zhu, Z., Liu, G., Li, Y., ... & Gao, C. (2021). Genome-wide specificity of prime editors in plants. *Nature Biotechnology*, *39*(10), 1292-1299.
279. Schene, I. F., Joore, I. P., Oka, R., Mokry, M., van Vugt, A. H., van Boxtel, R., ... & Fuchs, S. A. (2020). Prime editing for functional repair in patient-derived disease models. *Nature communications*, *11*(1), 1-8.
280. Gao, P., Lyu, Q., Ghanam, A. R., Lazzarotto, C. R., Newby, G. A., Zhang, W., ... & Miano, J. M. (2021). Prime editing in mice reveals the essentiality of a single base in driving tissue-specific gene expression. *Genome biology*, *22*(1), 1-21.
281. Lin, Q., Jin, S., Zong, Y., Yu, H., Zhu, Z., Liu, G., ... & Gao, C. (2021). High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nature Biotechnology*, *39*(8), 923-927.
282. Yu, K. R., Natanson, H., & Dunbar, C. E. (2016). Gene editing of human hematopoietic stem and progenitor cells: promise and potential hurdles. *Human Gene Therapy*, *27*(10), 729-740.
283. Baylis, F., & McLeod, M. (2017). First-in-human phase 1 CRISPR gene editing cancer trials: are we ready?. *Current gene therapy*, *17*(4), 309-319.
284. Uddin, F., Rudin, C. M., & Sen, T. (2020). CRISPR gene therapy: applications, limitations, and implications for the future. *Frontiers in oncology*, *10*, 1387.
285. Tong, S., Moyo, B., Lee, C. M., Leong, K., & Bao, G. (2019). Engineered materials for in vivo delivery of genome-editing machinery. *Nature Reviews Materials*, *4*(11), 726-737.
286. Ablain, J., & Zon, L. I. (2016). Tissue-specific gene targeting using CRISPR/Cas9. In *Methods in cell biology* (Vol. 135, pp. 189-202). Academic Press.
287. Komarova, Y., & Malik, A. B. (2010). Regulation of endothelial permeability via paracellular and transcellular transport pathways. *Annual review of physiology*, *72*, 463-493.
288. Sakurai, H., Kawabata, K., Sakurai, F., Nakagawa, S., & Mizuguchi, H. (2008). Innate immune response induced by gene delivery vectors. *International journal of pharmaceutics*, *354*(1-2), 9-15.
289. Wagner, D. L., Amini, L., Wendering, D. J., Burkhardt, L. M., Akyüz, L., Reinke, P., ... & Schmueck-Henneresse, M. (2019). High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nature medicine*, *25*(2), 242-248.
290. Nelson, C. E., Hakim, C. H., Ousterout, D. G., Thakore, P. I., Moreb, E. A., Rivera, R. M. C., ... & Gersbach, C. A. (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*, *351*(6271), 403-407.
291. Wang, X., Raghavan, A., Chen, T., Qiao, L., Zhang, Y., Ding, Q., & Musunuru, K. (2016). CRISPR-Cas9 targeting of PCSK9 in human hepatocytes in vivo—brief report. *Arteriosclerosis, thrombosis, and vascular biology*, *36*(5), 783-786.
292. Song, X., Liu, C., Wang, N., Huang, H., He, S., Gong, C., & Wei, Y. (2021). Delivery of CRISPR/Cas systems for cancer gene therapy and immunotherapy. *Advanced drug delivery reviews*, *168*, 158-180.
293. Xu, C. L., Ruan, M. Z., Mahajan, V. B., & Tsang, S. H. (2019). Viral delivery systems for CRISPR. *Viruses*, *11*(1), 28.
294. Yang, Y., Wang, L., Bell, P., McMenamin, D., He, Z., White, J., ... & Wilson, J. M. (2016). A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nature biotechnology*, *34*(3), 334-338.
295. Ryø, L. B., Thomsen, E. A., & Mikkelsen, J. G. (2019). Production and validation of lentiviral vectors for CRISPR/Cas9 delivery. In *CRISPR Gene Editing* (pp. 93-109). Humana Press, New York, NY.

296. Rittiner, J. E., Moncalvo, M., Chiba-Falek, O., & Kantor, B. (2020). Gene-editing technologies paired with viral vectors for translational research into neurodegenerative diseases. *Frontiers in Molecular Neuroscience*, *13*, 148.
297. Goswami, R., Subramanian, G., Silayeva, L., Newkirk, I., Doctor, D., Chawla, K., ... & Betapudi, V. (2019). Gene therapy leaves a vicious cycle. *Frontiers in oncology*, 297.
298. Chen, X., & Gonçalves, M. A. (2016). Engineered viruses as genome editing devices. *Molecular Therapy*, *24*(3), 447-457.
299. Chew, W. L., Tabebordbar, M., Cheng, J. K., Mali, P., Wu, E. Y., Ng, A. H., ... & Church, G. M. (2016). A multifunctional AAV–CRISPR–Cas9 and its host response. *Nature methods*, *13*(10), 868-874.
300. Li, A., Tanner, M. R., Lee, C. M., Hurley, A. E., De Giorgi, M., Jarrett, K. E., ... & Lagor, W. R. (2020). AAV-CRISPR gene editing is negated by pre-existing immunity to Cas9. *Molecular Therapy*, *28*(6), 1432-1441.
301. Ronzitti, G., Gross, D. A., & Mingozzi, F. (2020). Human immune responses to adeno-associated virus (AAV) vectors. *Frontiers in Immunology*, *11*, 670.
302. Xu, G. J., Kula, T., Xu, Q., Li, M. Z., Vernon, S. D., Ndung'u, T., ... & Elledge, S. J. (2015). Comprehensive serological profiling of human populations using a synthetic human virome. *Science*, *348*(6239), aaa0698.
303. Chou, S. J., Yang, P., Ban, Q., Yang, Y. P., Wang, M. L., Chien, C. S., ... & Chiou, S. H. (2020). Dual Supramolecular Nanoparticle Vectors Enable CRISPR/Cas9-Mediated Knockin of Retinoschisin 1 Gene—A Potential Nonviral Therapeutic Solution for X-Linked Juvenile Retinoschisis. *Advanced Science*, *7*(10), 1903432.
304. Chung, J. Y., Ain, Q. U., Song, Y., Yong, S. B., & Kim, Y. H. (2019). Targeted delivery of CRISPR interference system against Fabp4 to white adipocytes ameliorates obesity, inflammation, hepatic steatosis, and insulin resistance. *Genome research*, *29*(9), 1442-1452.
305. Lee, K., Conboy, M., Park, H. M., Jiang, F., Kim, H. J., Dewitt, M. A., ... & Murthy, N. (2017). Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nature biomedical engineering*, *1*(11), 889-901.
306. Lee, B., Lee, K., Panda, S., Gonzales-Rojas, R., Chong, A., Bugay, V., ... & Lee, H. Y. (2018). Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nature Biomedical Engineering*, *2*(7), 497-507.
307. Cheng, Q., Wei, T., Farbiak, L., Johnson, L. T., Dilliard, S. A., & Siegwart, D. J. (2020). Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR–Cas gene editing. *Nature nanotechnology*, *15*(4), 313-320.
308. Wang, P., Zhang, L., Zheng, W., Cong, L., Guo, Z., Xie, Y., ... & Jiang, X. (2018). Thermo-triggered release of CRISPR-Cas9 system by lipid-encapsulated gold nanoparticles for tumor therapy. *Angewandte Chemie International Edition*, *57*(6), 1491-1496.
309. Rosenblum, D., Gutkin, A., Kedmi, R., Ramishetti, S., Veiga, N., Jacobi, A. M., ... & Peer, D. (2020). CRISPR-Cas9 genome editing using targeted lipid nanoparticles for cancer therapy. *Science advances*, *6*(47), eabc9450.
310. Rouatbi, N., McGlynn, T., & Al-Jamal, K. T. (2022). Pre-clinical non-viral vectors exploited for in vivo CRISPR/Cas9 gene editing: an overview. *Biomaterials Science*.
311. Wang, M., Zuris, J. A., Meng, F., Rees, H., Sun, S., Deng, P., ... & Xu, Q. (2016). Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proceedings of the National Academy of Sciences*, *113*(11), 2868-2873.
312. Campbell, L. A., Coke, L. M., Richie, C. T., Fortuno, L. V., Park, A. Y., & Harvey, B. K. (2019). Gesicle-mediated delivery of CRISPR/Cas9 ribonucleoprotein complex for inactivating the HIV provirus. *Molecular Therapy*, *27*(1), 151-163.
313. D'Astolfo, D. S., Pagliero, R. J., Pras, A., Karthaus, W. R., Clevers, H., Prasad, V., ... & Geijsen, N. (2015). Efficient intracellular delivery of native proteins. *Cell*, *161*(3), 674-690.
314. Kholosy, W. M., Visscher, M., Ogink, K., Buttstedt, H., Griffin, K., Beier, A., ... & Chatsisvili, A. (2021). Simple, fast and efficient iTOP-mediated delivery of CRISPR/Cas9 RNP in difficult-to-transduce human cells including primary T cells. *Journal of Biotechnology*, *338*, 71-80.
315. National Academies of Sciences, Engineering and Medicine (NASEM) (2017) Human Genome Editing: Science, Ethics, and Governance. Washington, DC: The National Academies Press.
316. Hyun, I., Wilkerson, A., & Johnston, J. (2016). Embryology policy: Revisit the 14-day rule. *Nature*, *533*(7602), 169-171.
317. Lander, E. S., Baylis, F., Zhang, F., Charpentier, E., Berg, P., Bourgain, C., ... & Winnacker, E. L. (2019). Adopt a moratorium on heritable genome editing.
318. Ishii, T. (2017). Germ line genome editing in clinics: the approaches, objectives and global society. *Briefings in functional genomics*, *16*(1), 46-56.
319. Nuffield Council on Bioethics. (2017) Human embryo culture: discussions concerning the statutory time limit for maintaining human embryos in culture in light of some recent scientific developments

320. Dickenson, D., & Darnovsky, M. (2019). Did a permissive scientific culture encourage the 'CRISPR babies' experiment?. *Nature Biotechnology*, *37*(4), 355-357.
321. Ye, Z. J., Zhang, X. Y., Liang, J., & Tang, Y. (2020). The challenges of medical ethics in China: are gene-edited babies enough?. *Science and Engineering Ethics*, *26*(1), 123-125.
322. Ma, Y., Zhang, L., & Qin, C. (2019). The first genetically gene-edited babies: It's "irresponsible and too early". *Animal Models and Experimental Medicine*, *2*(1), 1-4.
323. Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, *348*(6233), 442-444.
324. Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences*, *112*(49), E6736-E6743.
325. Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., ... & Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature biotechnology*, *34*(1), 78-83.
326. Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., ... & Crisanti, A. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, *473*(7346), 212-215.
327. Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J., & Burt, A. (2017). Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proceedings of the National Academy of Sciences*, *114*(2), E255-E264.
328. Moro, D., Byrne, M., Kennedy, M., Campbell, S., & Tizard, M. (2018). Identifying knowledge gaps for gene drive research to control invasive animal species: the next CRISPR step. *Global Ecology and Conservation*, *13*, e00363.
329. Courtier-Orgogozo, V., Morizot, B., & Boëte, C. (2017). Using CRISPR-based gene drive for agriculture pest control. *EMBO reports*, *18*(9), 1481-1481.
330. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). Safeguarding CRISPR-Cas9 gene drives in yeast. *Nature biotechnology*, *33*(12), 1250-1255.
331. Roggenkamp, E., Giersch, R. M., Schrock, M. N., Turnquist, E., Halloran, M., & Finnigan, G. C. (2018). Tuning CRISPR-Cas9 gene drives in Saccharomyces cerevisiae. *G3: Genes, Genomes, Genetics*, *8*(3), 999-1018.
332. Basgall, E. M., Goetting, S. C., Goeckel, M. E., Giersch, R. M., Roggenkamp, E., Schrock, M. N., ... & Finnigan, G. C. (2018). Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae. *Microbiology*, *164*(4), 464.
333. Vella, M. R., Gunning, C. E., Lloyd, A. L., & Gould, F. (2017). Evaluating strategies for reversing CRISPR-Cas9 gene drives. *Scientific reports*, *7*(1), 1-8.
334. Beech, C. (2014). Regulatory experience and challenges for the release of GM insects. *Journal für Verbraucherschutz und Lebensmittelsicherheit*, *9*(1), 71-76.
335. Shapiro, S. A. (2007). OMB and the Politicization of Risk Assessment. *Environmental Law*, 1083-1106.
336. Bolleyer, N., & Börzel, T. A. (2010). Non-hierarchical policy coordination in multilevel systems. *European Political Science Review*, *2*(2), 157-185.
337. Jouanin, A., Boyd, L., Visser, R. G., & Smulders, M. J. (2018). Development of wheat with hypoimmunogenic gluten obstructed by the gene editing policy in Europe. *Frontiers in Plant Science*, *9*, 1523.
338. Fritsche, S., Poovaiah, C., MacRae, E., & Thorlby, G. (2018). A New Zealand perspective on the application and regulation of gene editing. *Frontiers in Plant Science*, *9*, 1323.
339. Kleter, G. A., Kuiper, H. A., & Kok, E. J. (2019). Gene-edited crops: towards a harmonized safety assessment. *Trends in biotechnology*, *37*(5), 443-447.
340. Schuttelaar (2015) The regulatory status of new breeding techniques in countries outside the European union. The Hague (NL).
341. Pfeiffer, M., Quétier, F., & Ricroch, A. (2018). Genome editing in agricultural biotechnology. In *Advances in Botanical Research* (Vol. 86, pp. 245-286). Academic Press.
342. Davison, J., & Ammann, K. (2017). New GMO regulations for old: Determining a new future for EU crop biotechnology. *GM crops & food*, *8*(1), 13-34.
343. Ishii, T., & Araki, M. (2017). A future scenario of the global regulatory landscape regarding genome-edited crops. *GM crops & food*, *8*(1), 44-56.
344. Waltz, E. (2018). With a free pass, CRISPR-edited plants reach market in record time. *Nature biotechnology*, *36*(1), 6-8.
345. Araki, M., & Ishii, T. (2015). Towards social acceptance of plant breeding by genome editing. *Trends in plant science*, *20*(3), 145-149.
346. European Court of Justice (ECJ). (2018a) PRESS RELEASE No 111/18, Luxembourg, 25 July 2018.

347. Zilberman, D., Holland, T. G., & Trilnick, I. (2018). Agricultural GMOs—what we know and where scientists disagree. *Sustainability*, *10*(5), 1514.
348. Bartkowski, B., Theesfeld, I., Pirscher, F., & Timaeus, J. (2018). Snipping around for food: Economic, ethical and policy implications of CRISPR/Cas genome editing. *Geoforum*, *96*, 172-180.
349. Globus, R., & Qimron, U. (2018). A technological and regulatory outlook on CRISPR crop editing. *Journal of cellular biochemistry*, *119*(2), 1291-1298.
350. Huang, S., Weigel, D., Beachy, R. N., & Li, J. (2016). A proposed regulatory framework for genome-edited crops. *Nature genetics*, *48*(2), 109-111.
351. Li, G., Liu, Y. G., & Chen, Y. (2019). Genome-editing technologies: the gap between application and policy. *Sci China Life Sci*, *62*, 1534-1538.
352. Schulman, A. H., Oksman-Caldentey, K. M., & Teeri, T. H. (2020). European Court of Justice delivers no justice to Europe on genome-edited crops. *Plant biotechnology journal*, *18*(1), 8.
353. Medvedieva, M. O., & Blume, Y. B. (2018). Legal regulation of plant genome editing with the CRISPR/Cas9 technology as an example. *Cytology and Genetics*, *52*(3), 204-212.

About the author

Despoina Trasanidou (17-03-1993) was born and raised in Greece. During her BSc studies in Food Science and Nutrition at the University of the Aegean (2011 – 2015), she developed a green process for the extraction of antioxidants and pigments from food wastes and she performed Internships at the Hellenic Food Authority (EFET) and the dairy company MEVGAL. After graduating honored with the 1st Rank Student Award, she received Financial Scholarships from the Erasmus Replacement and the ETH Zurich to perform two Internships at the ETH Zurich of Switzerland. There, she studied the impact of breast-feeding to the gut microbiome of infants and she characterized and engineered a CRISPR-Cas system from *Listeria* for phage engineering. The University of the Aegean acquainted her with the Award of Excellence for international activity & contribution to the international reputation of the University.

In 2016, Despoina received a 2-year Financial Scholarship from the Alexander S. Onassis Foundation and moved to the Netherlands to study her MSc in Cellular and Molecular Biotechnology at Wageningen University. During her MSc Major and Minor theses at the Laboratory of Microbiology of Wageningen University, she developed novel tools to change the DNA of thermophilic and mesophilic bacteria. In 2018, she started her PhD studies at the same laboratory under the supervision of Prof. John van der Oost and Dr. Raymond Staals, being recipient of a 4-year Financial Scholarship from the Alexander S. Onassis Foundation. During her PhD, she developed novel genetic engineering tools for bacteria and human cells. Most of this work has been described in this thesis. In addition, she participated in the international Genetically Engineered Machine (iGEM) competitions of 2019 and 2021. The iGEM team of 2019 received the 2nd place worldwide for the graduate division and two additional Awards.

List of publications

1. **Trasanidou, D.**, Barendse, P.*, Bouzetos, E.*, de Haan, L., Bouwmeester, H., Staals, R. H. J., Mougiakos, I.**, & van der Oost, J.**. Efficient genome and base editing in human cells using ThermoCas9. *Manuscript accepted for publication (The CRISPR Journal)*.

2. **Trasanidou, D.**, Potocnik*, A., Barendse*, P., Mohanraju, P., Bouzetos, E., Karpouzis, E., Desmet, A., van Kranenburg, R., van der Oost, J., Staals, R. H. J.**, & Mougiakos, I.**. *In vivo* characterization of the AcrIIC1 anti-CRISPR protein for Cas9-based genome engineering. *Manuscript under revision (Nature Communications)*.

3. **Trasanidou, D.**, Geros, A. S., Mohanraju, P., Nieuwenweg, A. C., Nobrega, F. L. **, & Staals, R. H. J**. (2019). Keeping crispr in check: diverse mechanisms of phage-encoded anti-crisprs. *FEMS microbiology letters*, 366(9), fnz098.

4. **Trasanidou, D.**, van der Oost, J.**, & Staals, R. H. J.**. Base-editing: from biology to technology. *Manuscript in preparation*.

5. Bouzetos*, E., **Trasanidou*, D.**, Kobus, Y., Villegas Warren, R., Staals, R. H. J., & van der Oost, J.**. Rational and random engineering of ThermoCas9 for optimal activity at moderate temperatures. *Manuscript in preparation*.

*Equal contribution

**Corresponding authors

Patent Applications

Mougiakos, I., **Trasanidou, D.**, Van der Oost, J. (2021). Novel base editing tools for bacteria and human cells. Invention ATV119.

Co-author affiliation

Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

John van der Oost, Raymond H. J. Staals, Victor Pool, Jurre Steens, Carl Salazar, Prarthana Mohanraju, Anna Cornelia Nieuwenweg, Ana Potocnik, Patrick Barendse, Evgenios Bouzetos, Efthymios Karpouzis, Amber Desmet, Richard van Kranenburg, Ioannis Mougiakos, Youri Kobus, Ricardo Villegas Warren

Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands.

Ana Sousa Geros, Franklin L. Nobrega

Corbion, Arkelsedijk 46, 4206 AC Gorinchem, The Netherlands.

Richard van Kranenburg

SNIPR Biome, Lersø Parkallé 44, 2100 Copenhagen, Denmark.

Ioannis Mougiakos

Laboratory of Toxicology, Department of Agrotechnology and Food Sciences, Wageningen University and Research, Stippeneng 4, Wageningen 6708 WE, The Netherlands.

Laura de Haan, Hans Bouwmeester

Completed training activities

Discipline specific activities

Meetings & conferences

- CRISPR Conference, Quebec, Canada (2019)
- 5th Applied Synthetic Biology in Europe, Online (2020)
- Netherlands Biotechnology Congress (NBC)-20: BioTECH Talks, Online (2020)
- Precision Genome Editing without Double-Strand Breaks, Online (2020)
- CRISPR Conference, Online (2021)
- 3rd International Conference on Base Editing - Enzymes and Applications, Online (2022)
- CRISPR Conference, Boston, United States (2022)

General courses

- VLAG PhD week, ?, The Netherlands (2018)
- Philosophy and Ethics of Food Science and Technology, Wageningen, The Netherlands (2020)
- Reviewing a Scientific Manuscript, Online (2020)
- Scientific Publishing, Online (2020)
- Presenting with Impact, Wageningen, The Netherlands (2020)
- Project and Time Management, Wageningen, The Netherlands (2020)
- Searching and Organizing Literature, Online (2020)
- Career Perspectives, Wageningen, The Netherlands (2021)

Other activities

- Preparation of research proposal, Wageningen, The Netherlands (2018)
- PhD study tour, Boston and New York, United States (2019)
- Weekly BacGen meetings, Wageningen, The Netherlands (2018 - 2022)
- Monthly MIB meetings, Wageningen, The Netherlands (2018 - 2022)

- PhD study tour, California, United States (2022)
- Organization of PhD study tour, Wageningen, The Netherlands (2020 - 2022)

Acknowledgements

Promoters

John, I will never forget the day I knocked your office door asking for a PhD position and you saying with a smile 'Let's do it! What project would you like to do?'! That moment has been deeply engraved in my mind, as I had been following for years your research wishing that one day I become part of your team. I am very grateful for the opportunity you gave me to join your group for my MSc Major and Minor theses as well as for my PhD. You have offered me endless support as well as freedom, from the very first moment. You are a very welcoming person, always there to hear my ideas, thoughts, concerns, and feelings. You are an exemplary supervisor and a very inspiring person! I enjoy so much being with you, discussing science and random staff. I will always remember how much you supported me also for my post-PhD steps, offering me a PostDoc position, helping me with the application to the VLAG grant, supporting my PostDoc visit to Florida State University and St. Jude Children's Research Hospital as well as assisting me find my next job. I have deep respect and appreciation for you both on scientific and personal level. It is my honor to have learned so many things next to you and to have been part of your amazing lab family!

Raymond, my Greek-speaking, chess puzzle solving, type III enthusiastic mentor/co-promoter! Meeting you was always so much fun and informative! You taught me so many things about fundamental research but also about how the science worlds works. I have many memories with you from the CRISPR Conferences in Quebec and Boston, the PhD trip in California as well as our iGEM meetings. You are a great leader, brilliant mind, innovative and enthusiastic scientist as well as an open-minded, pleasant, kind-hearted person with lots of respect and caring for other people. I am deeply grateful for all the support, advices and directions you provided me, I wish one day I manage to become a mentor like you!

PhD thesis committee

I would like to thank the thesis committee members **Niels Geijsen**, **Michiel Kleerebezem**, **Daan Swarts** and **Elleke Bosma** for reading and evaluating my PhD thesis as well as for travelling to Wageningen from other city or even other country.

Collaborators

Ioannis, συγκάτοικος, φίλος, MSc supervisor, PhD co-mentor και ένας από τους λόγους που κάποιος διαβάζει αυτή τη Διατριβή τώρα. Δεν έχω λόγια να εκφράσω την ευγνομοσύνη μου για όλα αυτά που μου έμαθες όλα αυτά τα χρόνια, την υποστήριξη που μου πρόσφερες ειδικά στα πρώτα βήματα του Διδακτορικού μου. Ήσουν πάντα ο άνθρωπος που πήγαινα στις πιο δύσκολες στιγμές στο εργαστήριο. Ο άνθρωπος που μου έδινε τις καλύτερες συμβουλές την κατάλληλη χρονική στιγμή. Ο άνθρωπος που μου δίδαξε πόσο υπέροχη αλλά και δύσκολη μπορεί να είναι η έρευνα. Είχαμε και τις καλές και τις κακές μας στιγμές αλλά αυτό είναι πάντα μέρος αληθινών ανθρωπίνων σχέσεων. Εκτιμώ βαθύτατα την εντυπωσιακή επιστημονική σου πορεία, είσαι ένας από τους πιο εργατικούς, δημιουργικούς και εύστροφους επιστήμονες που ξέρω. Είσαι επίσης πολύ caring, διαφανής και ευχάριστος άνθρωπος, με υψηλό αίσθημα δικαιοσύνης. Εύχομαι τα καλύτερα στην προσωπική και επαγγελματική σου ζωή στην Copenhagen. Είμαι πολύ περίφανη για σένα!

Prarthana, my be-loved office-mate, MSc co-supervisor and PhD collaborator. You were always a raw model for me: hard-working and highly skilled scientist, with profound knowledge on the field but also a very kind, pleasant person with a lovely smile and a beautiful personality. It was a great pleasure to share not only the same office but also several projects with you. Every time I purify a protein I remember the days and nights we spent together at the FPLC with you teaching me everything in detail. Thank you for everything you taught me and for all the support you offered me during my MSc and PhD studies. Your contribution to this thesis is very important. I wish you have a brilliant career and life with your husband. Lots of love to you.

My students **Phin**, **Patrick**, **Ana**, **Amber**, **Efthymios**, **Laura**, **Kevin**, **Victor** and **Evgenios**, thank you for all the great moments, discussions and scientific adventures we had together. You worked very hard during your thesis and I would wholeheartedly like to thank you for your enormous contribution to the research described in this book. **Phin**,

Patrick and **Ana** our collaboration has been transformed into a beautiful friendship with many memories to share. Ana drawing flowers on the head of a random bold man at the Amsterdam Centraal train station before the pride, Patrick and Phin with colorful clothes messing around, we visiting Phin at Zwolle to swim at the sea 1st January and to play board games, many lunches, dinners, trash movies/music, biking in nature… **Amber** Highly motivated and always with positive energy, able to easily collaborate with others and make the best out of her research. I wish you have a happy life in your home country, Belgium. **Efthymios** 'αγαπημένε μου μαθητή', κύριε Μαρκορά, έφερες τον Κισλόφσκι και απαραίτητη ποιότητα στην καθημερινότητα μου αχαχαχαχα Είμαι πολύ περίφανη για τα επιτεύγματά σου (Ουτρέχτη, Ελβετία, Γερμανία) και εύχομαι να δούμε μια μέρα μαζί την τριλογία (αν αντέξω) αχαχαχα. **Laurita**, mi querida Spanish student, always bringing happiness in the room. Enthusiastic, motivated and innovative researcher as well as very kind-hearted person with high emotional intelligence and caring for others. Miss you Laurita. **Kevin**, coming every day from Utrecht full of motivation and ground-breaking ideas! Highly skillful, ready for any career path, solid character, very honest and direct, great scientist and beautiful personality. I still can't forget your body-building photo and your stories about this experience especially regarding your special diet! I wish you have a happy life with your girlfriend in Utrecht. **Victor**, the craziest guy in the lab, biking endless hours, eating tones of peanut butter, super social, highly skilled, extremely independent, very direct and honest. Despite my limited time in the lab due to thesis writing, you managed to do miracles with your project, never stopping surprising me! You are a great scientist! All the best with your DSM experience, I am very proud of you!

My collaborators **Jurre**, **Carl**, **Ana Sousa**, **Carina**, **Franklin**, **Richard**, **Laura de Haan**, **Hans**, **Youri** and **Ricardo**, thank you for working together on these exciting projects! **Jurre**, I am very happy to have shared with you not only the type III project adventure but also the CRISPR Conference and the PhD trip experiences. You are a very inspiring person and I highly admire you! **Carl**, always with a smile and open to talk science. You are a very kind and kind-hearted person, I wish you all the best with your PhD journey. **Ana Sousa**, thank you for the nice collaboration. **Carina**, you showed me for first time the Doppio experience that you like so much hahaha I loved the pizza dinners we had together with Suzan as well as the photos and stories of your cute cats, your new house, your cross-fit experience… You have a very interesting personality that I never want to stop learning about! I loved sharing my bed with you in the CRISPR Conference hahaha and organizing the PhD trip to California next to you. Kisses! **Franklin**, thank you for the nice collaboration and the nice talks at the CRISPR Conference in Quebec. **Richard**, thank you for the opportunity to learn

from you about engineering thermophiles. You are a very nice person and all your students are always very happy with your talent to motivate them and advise them in the best way. **Laura de Haan**, thank you for teaching us how to culture human cells and for welcoming us in the Laboratory of Toxicology. **Hans**, thank you for the nice collaboration. **Youri**, nice to see you growing scientifically. All the best! **Ricardoooo**, I wrote for you in the next section ahahaha.

iGEM team-mates

My iGEM students **Cleo**, **Alex** and **Sophie**, it was a great pleasure to live with you the adventurous iGEM experience! I am very proud of all of you. Cleo, it was really nice to meet you and I am sure you will have a lot of success in your life. You are a very nice person and a great scientist! **Alex**, very motivated, pleasant and skillful scientist as well as a golden personality. I was very happy to work with you on the anti-CRISPR circuit, and to see you growing scientifically throughout the years. **Sophie**, I wish you reach your goals and dreams in your life.

My iGEM collaborators **Enrique**, **Jurre**, **Thomas**, **Carina**, **Joep**, **Maria**, **Wen**, **Lyon**, **Lorenzo**, **Max**, **Eric**, **Luis**, **James**, **Niek**, **Costas**, **Robert**, **Raymond**, and **Jasper**, thank you for sharing the adventurous iGEM experiences with me.

Office-mates

My old office-mates **Prarthana**, **Sjoerd**, **Wen**, **Jorrit**, **Yifan**, **Maartje** and **Belen**, thank you for the nice moments we had in the office. **Prarthana** you have a special section above and in my heart. **Sjoerd**, being with you was such a special experience. You taught me soooooo many things, from science to music and even ancient Greek language! I admire you really a lot, I loved having (non-stop hahaha) discussions with you. You are a brilliant scientist and orchestra director! You are a very nice person and I really missed you when you left. You left but you will always have a special place in my heart. **Wen**, the most positive, motivating, social, super-girl of the planet! I loved our discussions in the office and was really nice to share so many scientific interests with someone in the lab. You have a unique personality and lots of scientific/presenting talent! **Jorrit**, the tallest office-mate I've ever had, always helpful and nice. **Yifan**, thank you very much for teaching me all the RNA purification techniques. You are a very nice person. Wish you all the best! **Maartje**,

the wonderful mum that practically showed me that is possible to combine science with family. Such a pleasure to talk with you every day. **Belen**, you have a special place at the Friends part.

My new office-mates **Ricardo**, **Thomas**, **Isabelle** and **Laureen**, I have so much fun with you guys both in the lab and in our dinners. **Ricardooooo**, mi Colombiano favorito! I love hanging out with you. You have a golden heart and a beautiful way of thinking. I really enjoyed sharing with you the CRISPR Conference experience at Boston but also the PhD trip at California. I will never forget how much you supported me psychologically in my hardest moments in the lab. Your kind caring and understanding was therapeutic. You deserve a lot as a person and as a scientist. A big hug for you! **Thomas**, the kindest person I know In my life and with the most peaceful aura. Always open to discuss and hang out, you bring such a calmness in my heart every time I see you. Very happy to sit next to you! **Isabelle**, you are an exemplary scientist and person. Very smart, always to the point, deeply knowledgeable, highly skilled, very social, very caring. I really like it when you laugh, you are such a real and inspiring person! **Laureen**, every time I see you in the office I can't keep the pain in my stomach from laughing so hard. I love your sense of humor and your beautiful personality. Looking forward to meet you soon!

PhD trip committee-mates

My PhD trip committee team **Carina**, **Anastasia**, **Maria**, **Lyon**, **Peter**, **Max**, **Lorenzo** and **Valentina**, it was a great pleasure to organize with you our road trip to California!

Lab-mates

My BacGen colleagues **Ricardo**, **Thomas**, **Isabelle**, **Laureen**, **Belen**, **Lorenzo**, **Caglar**, **Vittorio**, **Miguel**, **Panos**, **Afonso**, **Rob**, **Carl**, **Anneleen**, **Christian**, **Jurre**, **Costas**, **Prarthana**, **Ioannis**, **Joep**, **Jorrit**, **Eric**, **Max**, **Costas**, **Suzan**, **Wen**, **Sjoerd**, **Maartje**, **Evgenios**, **Carina**, **Thijs**, **Yifan**, **Mihris**, **Joyshree** and **Jeroen**, thank you for all the fun times we had in the lab and the offices.

Friends

My international friends **Maria**, **Lyon**, **Enrique**, **Yajia**, **Daniel**, **Belen**, **Lorenzo** and **Pilar**, my life in Wageningen would be completely different without you! Such amazing trips to Belgium and Terschelling, Fiesta Macumbas, Carnivals, dinners, parties, PhD trips, iGEMs…! **Maria**, the 1st person I met the 1st day of my MSc studies at Wageningen University. And still together! <3 So many moments during our MSc classes and team projects, endless hours in the library with Dani bringing you period pads ahahaha Then, together again at our MSc theses, sharing our experiences. Then again together in our PhD but this time at different labs. You are such a positive, inspiring, clever, hard-working and brilliant scientist but also my lovely 6-year friend that is always there both at good and bad times. Thank you for all the support. I love you. **Lyon**, or better Lyon Bruuuuuuinsma hahaha. How lucky I feel to have met you years ago and grow next to you. You are my favorite dance-mate and you have a very special place in my heart. I try to find words to express my feelings about you and what comes is pure love. You have such a golden heart, smart brain, high sense of humor and a big huge for everyone. You deserve the best in your personal and professional life. I admire you extremely. **Enrique**, I first met you during my MSc studies through Maria and I am grateful to be friends the last years. You are such a pure person, always supporting and advising me in the hard moments. You are very special for me. Love you. **Yajia**, I feel very happy every time I hang out with you. You are such a kind-hearted and caring person. Thank you for all the beautiful dinners at your place and muchos besos a Blueeee! **Belen**, querida amiga with the beautiful smile. Always taking care of others. Always kind, bringing calmness and happiness in the atmosphere. Always there for me in the difficult moments. You are an office-mate, lab-mate but most importantly a trustworthy friend. **Lorenzo**, chichino, I 1st met you during our MSc theses at BacGen together with Alberto. It was the 1st time I'd ever listened to Italian rap music and tasted real, slowly cooked Italian dish! Still remember your crazy graduation day that I was trying with Lyon to speak Italian to all these people hahahaha. I can't also forget the handball match we gave in the WUR tournament hahaha. Keep posting suggestions for movies, I am your biggest Insta fan! **Daniel**, I 1st met you at the library when you bought period pads for Maria hahahaha It was really nice to see how your flirt was converted into true love… I am really proud of your recent graduation and everything you have achieved both in personal and professional level. Muchos besos! **Pilarita**, you were the 2nd person I met the 1st day of our MSc studies. We also experienced together our MSc theses at BacGen. My Haarweg neighbor but also my favorite Mexican friend, you are such a positive and cute person and I wish all the best in your life!

My Greek friends in the Netherlands **Vivi**, **Menia**, **Nancy, Anastasia**, **Kassiani**, **Johnys**, **Alex**, **Foteini**, **Costas**, **Stamatis**, **Prokopis**, **Xristaras**, και **Giannis**, thank you for making my days at Wagenignen so special. **Vivouli** μου, συγκάτοικε, κολλητή, soul-mate. Ήρθες ξαφνικά στη ζωή μου, με μια απίστευτη χημεία μεταξύ μας ήδη από την συνέντευξη. Αυτή η γνωριμία εξελίχθηκε σε μία δυνατή φιλία που το μόνο σίγουρο είναι πως θα κρατήσει πολλά χρόνια ακόμη. Είσαι πάντα εκεί στις πιο δύσκολες στιγμές αλλά και στις πιο ευχάριστες. Ζήσαμε τόσα μαζί, από μεθυσμένα βραδινά χορευτικά μέχρι pole dancing με skates, butt challenges, olaplex stories, asian dinners, εξόδους με Πάνο και Ελίζα, βαθιές συζητήσεις, αλληλοστήριξη στην PhD εμπειρία, αραλίκια με τα κορίτσια στην Dolder, beauty nights, νοσοκομεία, αστυνομίες αχαχαχαχα Σ'αγαπώ! **Menia** μου, αδυναμία μου και πλέον συγκάτοικε. Ήσουν πάντα Dolder στην καρδιά και πλέον και πρακτικά. Είμαι πολύ χαρούμενη που θα έχω την ευκαιρία να ζήσω μαζί σου έστω και αυτούς τους λίγους μήνες. Απ'την 1η στιγμή μου είχες κάνει το κλικ αλλά πάντα ντρεπόμουν να σε προσεγγίσω. Χάρηκα πολύ μοιραστήκαμε κάποιες στιγμές στο PhD trip. Η Dolder family μας έφερε πιο κοντά και τα καλύτερα έρχονται! **Nancy** μου, γλυκούλα μου. Γειτόνισσα του Haarweg, olaplex co-ambassador και νυχού μου. Σε θαυμάζω πολύ για την προσωπική και επαγγελματική σου πορεία. Ευχαριστώ που είσα πάντα εκεί για μένα, έχεις μια ξεχωριστή θέση στην καρδιά μου. **Anastasia** μου, διπλή γειτόνισσα, χαίρομαι πάρα πολύ που έχω την ευκαιρία να σε γνωρίσω καλύτερα. Είσαι πολύ καλός άνθρωπος και πολύ όμορφη παρέα. Μου αρέσουν πολύ τα αράγματα σπίτι σου κ σπίτι μου και σε ευχαριστώ πολύ για τη στήριξη που μου πρόσφερες όταν τη χρειάστηκα. **Kassiani** μου, περάσαμε πολύ όμορφα στην California αλλά και στις Dolder nights. Χαίρομαι πολύ που σε γνώρισα! **Johnαρα** μου, 4 χρόνια γνωριμίας πλέον. Σε ευχαριστώ πολύ για όλες τις όμορφες στιγμές που περάσαμε μαζί, όπως αυτές στο Βέλγιο, στο Leiden και στο Amsterdam. Σε θαυμάζω πολύ για τις γνώσεις και την προσωπικότητά σου και εύχομαι να κρατήσουμε αυτή την επαφή τα ερχόμενα χρόνια. **Alex**, είσαι ο πιο κοινωνικός και περιπετειώδης άνθρωπος που γνωρίζω. Έχεις πολύ όμορφη προσωπικότητα και σε ευχαριστώ για τις όμορφες στιγμές στο Amsterdam και τη Γερμανία. Εύχομαι να κρατήσουμε αυτή την όμορφη επαφή. **Costas**, ο καλύτερος σύμβουλος σε όλο το PhD μου, πάντα δίπλα μου στην προσπάθειά μου να γίνω καλύτερη επιστήμονας. Δεν θα ξεχάσω ποτέ πόσο με βοήθησες να φτιάξω την παρουσίαση για το Broad Institute στο PhD trip. Φοβερό μυαλό και προσωπικότητα. Με ό,τι και να καταπιαστεί πάντα επιτυχημένος. Η έρευνα χρειάζεται ανθρώπους σαν κι εσένα. Keep going! **Stamatis**, ευχαριστώ για τις όμορφες συζητήσεις μας στο Wageningen και στην Αθήνα. See you soon! **Prokopis**, φοβερό μυαλό, δυνατός επιστήμονας και πολύ όμορφη προσωπικότητα. Τιμή μου που γνώρισα έναν άνθρωπο σαν κι εσένα. Χαίρομαι που

μοιραστήκαμε όμορφες στιγμές στο PhD trip. **Xristaras**, τα λόγια είναι περιττά για τον Χρηστάρα της καρδιάς μας. Πέρασα υπέροχα μαζί σου, τον Προκόπη και τη Μένια στο PhD trip. Είσαι πολύ special άτομο και έχεις χρυσή καρδιά! Ό,τι καλύτερο Χρηστάρα μου μέσα απ την καρδιά μου. **Giannis Kostopoulos**, πατέρας πλέον, ευχαριστώ για τις όμορφες συζητήσεις μας τον πρώτο καιρό του PhD μου.

My housemates **Ioannis**, **Foteini**, **Aslihan** and **Vivi**, thank you for all the beautiful moments we had in Dolder 48D.

My friends in Greece **Ελενίτσα**, **Μαρία**, **Μιχάλης**, **Κατερίνα**, **Ελενάρα**, **Χριστίνα**, **Τζίνα**, **Γεωργία**, **Ηρώ**, **Κλόντια**, **Παναγιώτη** και **Άννα**., σας ευχαριστώ για τις υπέροχες στιγμές που περνάμε μαζί. **Ελενίτσα**, 24 χρόνια κολλητές, μια ολόκληρη ζωή. Είσαι αδερφή μου και η οικογένειά σου 2η μου οικογένεια. Μαζί μεγαλώσαμε στην αυλή του σπιτιού σου παίζοντας πίου πίου, τοίχο, bubble bubble, χορεύοντας Βανδή, τρώγοντας καλαμπόκι και τσουρέκι απ΄τα χεράκια της κυρα Ρούλας. Μαζί πάντα στα εύκολα στα δύσκολα. Σε καμαρώνω πολύ για όλα αυτά που είσαι και που έχεις καταφέρει στη ζωή σου. Σ'αγαπώ πολύ. Φιλάκια στο Γιωργάκη! <3 **Μαρία**, η Μαρία της γειτονιάς και 2η αδερφή μου. Αυτή που λατρεύω να αράζω πίνοντας ελληνικό καφέ στη Χρύσα. Κάθε συζήτηση μαζί σου μου δίνει τεράστια δύναμη και χαρά. Ανυπομονώ να σε ξαναδώ! Σ'αγαπώ πολύ. **Μιχάλη** μου αγαπημένε, ιδρυτικό μέλος της 5άδας αχαχαχα Λατρεύω τις μουσικές μας βραδιές, τις νυχτερινές μας βόλτες, τις συζητήσεις μας, τις τρέλες μας. Είσαι αδερφός μου, σ'αγαπώ πολύ. **Κατερινούλι** μου, η αδερφή ψυχή μου. Σε γνώρισα πριν 10 χρόνια στη Λήμνο και μου έκανες τρομερό κλικ! Και κοίτα σήμερα, αχώριστες με άπειρα κοινά μεταξύ μας. Έχεις το ταλέντο να καταλαβαίνεις πάντα τι σκέφτομαι και τι θέλω να πω. Η χημεία μας είναι τρομερή. Είσαι κολλητή και αναπόσπαστο κομμάτι της ζωής μου. **Ελενάρα**, ο πιο ανάποδος άνθρωπος του πλανήτη με τις πιο όμορφες μπούκλες και φιόγκους αχαχαχα 14 χρόνια φίλες, γνωρίζουμε η μία την άλλη απ΄όλες τις μεριές. Μαζί στα εύκολα και στα δύσκολα. Είμαι πολύ περίφανη για σένα και τις επιτυχίες σου. Ανυπομονώ να σε δω! **Χριστίνα** μου, Μποτιτσέλισσα και άλλα πολλά που δεν μου επιτρέπεται να γράψω εδώ, ξέρεις εσύ! αχαχα Γνωριστήκαμε στο Λύκειο και κρατάει χρόνια αυτή η κολόνια. Είσαι ο πιο καλόψυχος άνθρωπος που ξέρω, πάντα εκεί να προσφέρει βοήθεια και αγάπη στις αδύναμες κοινωνικές ομάδες. Αξιοθαύμαστη προσωπικότητα. **Τζινάκι** μου γλυκό, travel buddy μου, τόσα roadtrips μαζί, τόσες συζητήσεις... Αγαπώ τον ενθουσιασμό σου και τη δίψα σου για ζωή, την αθωότητα και την καλοσύνη σου. Είσαι υπέροχος άνθρωπος. Τα καλύτερα έρχονται! **Γεωργία** μου, από την 1η στιγμή με αγκάλιασες και με έκανες να νιώθω ευπρόσδεκτη στην παρέα. Έχεις φοβερή ποιότητα σαν άνθρωπος και καλοσύνη. Σε θαυμάζω πολύ για αυτό που είσαι. Θα είσαι η πιο όμορφη νυφούλα! **Ηρώ** μου, πόσο

χαίρομαι που ήρθες Ολλανδία, εδώ μαζί μας! Θαρραλαία και αισιόδοξη πάντα, με ένα πλατύ πανέμορφο χαμόγελο. Ανυπομονώ να ζήσουμε περισσότερες στιγμές μαζί! **Κλόντια**, βραζιλιάνα μου, πάντα στην τρίχα και με περιπέτειες να αφηγηθείς. Σε ευχαριστώ για τις όμορφες στιγμές μας στη Λήμνο και την Αθήνα όλα αυτά τα χρόνια. **Παναγιώτη** μου, ευχαριστώ πολύ για όλη τη βοήθειά που μου πρόσφερες στο διδακτορικό μου. Ανυπομονώ να τα ξαναανταμώσουμε! **Άννα** μου, ήσουν η αγαπημένη μου συμμαθήτρια στο Γυμνάσιο και το αγαπημένο μου δίδυμο στο θρανίο. Έχεις εξελιχθεί σε μία φοβερή γυναίκα γεμάτη εμπειρίες και ενδιαφέροντα χόμπι. Πόσο χαίρομαι που ξαναβρεθήκαμε! **Γιάννη** μου, ο καλύτερος μηχανόβιος της Αθήνας με πάντα όμορφους προορισμούς! Σε γνώρισα στο Δημοτικό και χαίρομαι βαθύτατα που ξανα ανταμώσαμε! Ανυπομονώ να για τον επόμενο παραλιακό μας καφέ! **Μιχαλάρα**, διατροφολόγε μου, τι όμορφες βόλτες κάναμε στην Κηφισσιά αλλά και αραλίκια στο Venlo. Εύχομαι ό,τι καλύτερο στη ζωή σου!

Family

Ένα ΤΕΡΑΣΤΙΟ ευχαριστώ στην οικογένειά μου. **Μανούλα**, είσαι η καλύτερη μαμά του κόσμου! Σ'αγαπώ πάρα πολύ! Χωρίς εσένα δεν θα είχα καταφέρει ούτε να σπουδάσω ούτε να βρίσκομαι σε αυτή τη χώρα σήμερα. Αυτό το βιβλίο είναι ένας κάρπος των κόπων σου όλα αυτά τα χρόνια να μας μεγαλώσεις και να μας προσφέρεις τα πάντα, συναισθηματικά και πρακτικά. Ευχαριστώ για όλα μανούλα! **Μπαμπάκα**, από τότε που θυμάμαι τον εαυτό μου είμαι πάντα μέσα στην γαλήνια αγκαλιά σου. Όταν σε βλέπω νιώθω απέραντη ηρεμία. Σε ευχαριστώ πολύ για όλες τις θυσίες και τους κόπους σου για μένα όλα αυτά τα χρόνια. Μου λείπεις, σ'αγαπώ πολύ! **Μπάμπη** μου, αδερφούλη μου, η αγάπη μου για σένα είναι απέραντη... Δεν έχω λέξεις να περιγράψω πόσο σημαντικός είσαι στη ζωή μου. Σ'αγαπάω πάρα πολύ! **Νικολάκη** μου, αδερφούλη μου, είσαι το μεγαλύτερο δώρο της ζωής μου. Σ'αγαπάω πάρα πολύ! **Κυρ Νίκο**, χαίρομαι πάρα πολύ που είσαι μέλος αυτής της οικογένειας. Ευχαριστώ για όλα όσα έχεις κάνει για μένα. Σ'αγαπώ! **Θεία Δέσποινα**, η ήρεμη δύναμη της οικογένειας! Έχεις μία μοναδική θέση στην καρδιά μου. **Θείο Μήτσο**, είσαι ο καλύτερος θείος του κόσμου! Πάντα λεβέντης και η καλύτερη παρέα! Πάντα δίπλα μου σε ό,τι χρειαστώ. **Μορφούλα**, Φρούλι μου, μεγαλώσαμε σαν αδερφές και τώρα είσαι μανούλα δύο υπέροχων κοριτσιών. Είμαι σίγουρη πως είσαι μια υπέροχη μαμά. Σε έχω πάντα μέσα στην καρδιά μου. **Νίκο**, πάντα εκεί να συμμαζεύεις τις βλακίες μου αχαχα Ο άνθρωπος εμπιστοσύνης μου, από μικρή ηλικία. Σ'αγαπώ πολύ. **Πωλίνι** μου, φασαία της φαμίλιας αχαχαχα Είσαι η μικρή μου

αδερφούλα, μου λείπεις και σ'αγαπάω πολύ! **Γιαγιά Φούλα** και **γιαγιά Λευκοθέα**, σας αγαπώ πολύ και σας έχω πάντα μέσα στο μυαλό μου. **Παππού Νίκο** και **παππού Χαράλαμπε**, εύχομαι να με βλέπετε και να χαμογελάτε από εκεί ψηλά... Μου λείπετε πολύ.

Φιλάνθη μου αγαπημένη, με υποδέχθηκες με μια ζεστή αγκαλιά και ένα πλατύ χαμόγελο. Είσαι ένας πολύ γνήσιος και αληθινός άνθρωπος, που προσφέρει απλόχερα φροντίδα και αγάπη. Απολαμβάνω πολύ τις συζητήσεις μας, τα τραπέζια που ετοιμάζεις, τα επιτραπέζια και τις βόλτες στο θερινό σινεμά. Δεν έχω λόγια να εκφράσω πόσο ευπρόσδεκτη με κάνεις να νιώθω. Θαυμάζω πολύ την ωριμότητα, την ανεξαρτησία και την θετικότητά σου! Σε ευχαριστώ πολύ για όλα. Τα καλύτερα έρχονται...! **Κυρ Κώστα** μου, ευχαριστώ πολύ για τις όμορφες στιγμές στις ταβέρνες και στο αμάξι. Με κάνεις πάντα να νιώθω τόσο άνετα και όμορφα. Και είσαι πάντα εκεί σε ό,τι ιατρικό χρειαστώ. Θα χαρώ πολύ να σε ξαναδώ σύντομα. **Άλκη** μου, χαίρομαι τόσο πολύ που σε έχω γνωρίσει. Μου αρέσει πολύ όταν παίζουμε επιτραπέζια με τα παιδιά και περνάμε στιγμές μαζί. Στο ιατρείο όμως παίζουν δυσκολίες αχαχα Ευχαριστώ πολύ για όλα! **Δήμητρά** μου γλυκιά, είσαι αξιολάτρευτος άνθρωπος και με πολύ ενδιαφέρουσα προσωπικότητα που όσο ανακαλύπτω τόσο μου αρέσει! Είδες δεν σε είπα κι εγώ Δέσποινα αχαχαχ Δεν θα ξεχάσω την όμορφή μας συζήτηση στο Μοναστηράκι εκείνη τη νύχτα. **Κα Μαρία** μου, ευχαριστώ για όλα τα γεύματα και τις συζητήσεις στο σπίτι σου. Μου αρέσει πολύ ο δυναμισμός και ο τρόπος σκέψης σου. Ανυπομονώ να σε ξαναδώ! **Κ. Τάσο**, ευχαριστώ για τα όμορφα γεύματα και τις επιχειρηματικές συζητήσεις. Έχω μάθει πολλά από εσένα και πάντα μου προκαλείς έμπνευση! **Φίλιππέ** μου, τι όμορφα που περνάμε κάθε φορά είτε Ελλάδα είτε Ολλανδία! Μου αρέσει πολύ να μιλάω μαζί σου, έχεις πάντα πολύ ενδιαφέρουσες απόψεις! **Λενάκι** μου, γκαβιλάνα No.1 αχαχα Φοβερή εκείνη η βραδιά στο Μοναστηράκι. Περνάω πάντα υπέροχα μαζί σου. Είσαι πολύ γλυκιά και εύστροφη! **Μπάμπη** μου, χαίρομαι πολύ που σε γνώρισα και περνάμε τόσο όμορφα όταν είμαστε όλοι μαζί. **Ελένη** μου, γκαβιλάνα No.2 αχαχα Πάντα χαμογελαστή και ευχάριστη, σε ευχαριστώ για τις όμορφες στιγμές! **Δέσποινα** και γλυκύτατη **γιαγιά Αγγέλα**, χάρηκα πολύ που σας γνώρισα και εύχομαι να τα ξαναανταμώσουμε σύντομα.

Φώτη μου, ζωή μου. Ευχαριστώ πολύ για την στήριξη που μου πρόσφερες όλα αυτά τα χρόνια. Είσαι πάντα εκεί για μένα στις χαρές και στις λύπες. Αυτό το μακρύ PhD ταξίδι το ζήσαμε μαζί σχεδόν από την αρχή μέχρι το τέλος. Έτσι, αυτό το βιβλίο είναι αφιερωμένο σε εσένα... Τα ομορφότερα ταξίδια μας έρχονται σύντομα! Σ'αγαπώ πολύ.

The research described in this thesis was financially supported by the Alexander S. Onassis Foundation grant (F ZM 083-2/2018-2019), the Dutch Research Council (NWO Spinoza grant SPI 93-537, and NWO Gravitation grant 024.003.019), and the European Research Council (ERC-AdG-834279).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Despoina Trasanidou and Panagiotis Achilladelis.