Needles in a Haystack

Pursuing directed evolution of CRISPR-Cas proteins for improved genome editing

Evgenios Bouzetos

Propositions

1. Protein evolution in artificial microcompartments is an excellent example of the reductionist approach.
   (this thesis)
2. Spacer sequences are the prokaryotic analogues of antibodies, and CRISPR arrays the prokaryotic analogues of memory T-cells.
   (this thesis)
3. The principles of natural selection are applicable to human-made inventions.
4. Modern technology and medicine have lowered selection pressure on human species.
5. Society's structure is dictated by a blend of the Darwinian rules and social norms.
6. Social media promote an asocial lifestyle.

Propositions belonging to the thesis entitled

Needles in a haystack:
Pursuing directed evolution of CRISPR-Cas proteins for improved genome editing.

Evgenios Bouzetos
Wageningen 4th October 2024

Needles in a haystack

Pursuing directed evolution of CRISPR-Cas proteins for improved genome editing

Evgenios Bouzetos

Thesis Committee

Promotor

Prof. Dr John van der Oost

Personal chair at the laboratory of Microbiology

Wageningen University and Research

Co-promotor

Dr Raymond H.J. Staals

Associate Professor at the Laboratory of Microbiology

Wageningen University & Research

Other Members

Dr. Johannes Hohlbein, Wageningen University & Research

Dr. Ioannis Mougiakos, SNIPR Biome

Dr. Prathana Mohanraju, Leiden University Medical Centre

Prof. Dr Arjan de Visser, Wageningen University & Research

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

Needles in a haystack

Pursuing directed evolution of CRISPR-Cas proteins for improved genome editing

Evgenios Bouzetos

Thesis

Submitted in fulfilment of the requirement for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr. C.Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 4 October 2024
at 13:00 p.m. in the Omnia Auditorium

Evgenios Bouzetos

Needles in a haystack: Pursuing directed evolution of CRISPR-Cas proteins for improved genome editing

Pages: 160

PhD thesis, Wageningen University, Wageningen, The Netherlands (2024)

With references, with summary in English

DOI: 10.18174/671676

Table of Contents

Chapter 1

General Introduction and Thesis Outline

From philosophy to biology

The first biological studies, in which a systematic approach was used, are the works of Aristotle (384-322 BC) and his student Theophrastus (371-281 BC). Lacking any previous experimental knowledge, Aristotle and Theophrastus set out to an "ὁτι: what" investigation on the differences and similarities of animal and plant properties. In the known book History of Animals, Aristotle argues that the documentation of these properties is necessary before starting the research of "διὁτι: why" and finally connecting cause with effect (1). As such he introduces causality in the study of biological phenomena.

Purely biological studies were not performed again until the late Renaissance. The invention of the microscope was an important step. Robert Hook (1635-1703) reports the earliest observation of living matter that is invisible with the naked eye and introduces the term "cell". Antoni van Leeuwenhoek (1632-1723), who is considered the father of Microbiology, succeeded in generating significantly stronger lenses that allowed him to observe and describe single cell microorganisms. The "quantum" of biology has been discovered.

Charles Darwin (1809-1882) in his book On the Origin of Species (1859) is the first to formulate the principles of evolution and natural selection. His work explains how evolution through the loss or gain of heredity traits in a species population depends on the environmental conditions that the population encounters (2). Around the same time, the principles of heredity were discovered by Gregor Mendel (1822 -1884). His experiments on plant breeding revealed that certain properties (e.g. the colour of flowers) can be independently inherited and that each trait has two alleles, one inherited from each parent.

Still, the molecular basis of heredity was a mystery. First, Frederick Griffith (1877-1941) showed that there is a component that can be transformed in bacteria and change their pathogenic properties. Avery, MacLeod and McCarty performed a series of exclusion experiments and found that the

chemical properties of the heredity molecule best match those of deoxyribonucleic acid (DNA) (3). Due to slight amounts of protein contamination, however, they could not completely rule out that protein could be the heredity molecule. Final confirmation that the DNA is the heredity molecule came from the Hershy and Chase Experiment. They independently tracked DNA and protein molecules by probing them with radioactive phosphorus or sulphur, respectively. They exposed bacterial cells to bacterial viruses (phages) which had either their DNA or their capsid protein radioactively labelled. They observed that bacterial cells infected by DNA-labelled phages emitted higher levels of radiation, conclusively proving that DNA is the heredity molecule (4).

It was not long after that the structure of DNA was resolved as a double strand helix by James Watson and Franscis Crick, a model that was based on Rosalind Franklin's X-crystallography results (5).

Finally, the central dogma of biology was formulated by Francis Crick in 1970. The central dogma describes the flow of the biological information from DNA to RNA to protein (6). Through the central dogma, the biological traits that Aristotle and Theophrastus documented were linked with the heredity and evolution principles that Mendel and Darwin unlocked.

Prokaryotic cell defence and CRISPR-Cas systems

Prokaryotic cells (bacterial & archaeal) and parasitic genetic elements (MGEs) such as transposons, plasmids and phages fight each other for the right to heredity, following the principles of Darwinian evolution. Their arms race has resulted in great diversity of defence mechanisms and resembles the relationship of that of the prey with the predator. The bacterial defence points reflect the invader's infection cycle: i) the bacterial cell can adapt to reject association with the phage by accumulating mutations on its surface proteins, ii) the cell can counter-attack the phage by cleaving its genome, iii) it ca1n prevent phage propagation by suppressing the replication of the

phage's genome, or iv) the cell sacrifices itself through suicide to protect the rest of its population, either by regulating toxin/antitoxin systems or by triggering programmed cell death (7,8).

Two decades ago, a defence system based on Clustered Regularly Interspaced Palindromic Repeat sequences (CRISPR) and CRISPR-associated *(cas)* genes was discovered. CRISPR-Cas immunity is encoded by two main components: i) an array of palindromic repeats that are interspaced by unique sequences called spacers and ii) an operon of CRISPR-associated genes. The CRISPR-array stores the genetic information of invading genomes as short spacer sequences. It is transcribed into long precursor CRISPR-RNAs (pre-crRNAs) which is subsequently processed into mature crRNAs that consist of a single spacer sequence and one repeat copy. Mature crRNAs will form ribonucleoprotein complexes (RNPs) with certain Cas proteins called effectors. Active RNPs will look to establish sufficient base complementarity between a portion of the crRNA (spacer) and the target sequence (protospacer) which is found in the invading nucleic acids. Establishment of base complementarity between spacer/protospacer triggers nucleolytic activity of the Cas-protein which introduces cuts in the target sequence. Genomic breaks restrict the propagation of the invading element and provide immunity to the host.

Cas proteins are involved in every stage of CRISPR-Cas immunity, from the acquisition of new spacer sequences to the maturation of crRNAs and interference with target nucleic acids. Because new spacer sequences can always be integrated to the CRISPR-array, CRISPR-Cas systems confer adaptive immunity, setting them apart from other procaryotic defence systems such as restriction modification systems that offer innate immunity.

Classification of CRISPR-Cas systems

CRISPR-Cas systems are divided into two classes: i) Class 1 systems that recruit multi-protein Cas effector complexes which recognize and interfere with foreign sequences, ii) Class 2 systems employ a single and multidomain Cas protein as their effector complex (**Figure 1.1**). The two classes are further subdivided into types according to the domain architecture of the effector complex. The diversity reflects in the nature of the target nucleic acid (DNA or RNA), the requirement of a secondary, trans-activating crRNA module (tracrRNA) and the type and number of nucleolytic domains (9).

Each class subdivides into three types: i) Class 1 (Types I, III, IV) and ii) Class 2 Types (II, V, VI). A new candidate Class 1 type (Type-VII) has recently proposed as the newest member of the CRISPR-Cas family (10).

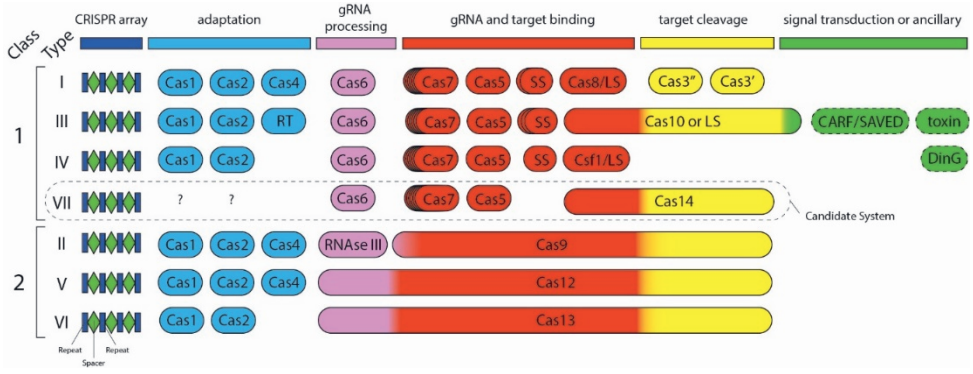


Figure 1.1 Classification and modules of CRISPR-Cas.
Coloured bars represent proteins. Colours represent the different protein functions. LS (large subunit) and SS (small subunit) of Cas11. Dashed lines indicate that some subtypes lack the shown factors. Type VII is proposed as candidate CRISPR-Cas system. Adapted from (11) by taking into consideration (10).

Molecular mechanisms of guided interference in type II and type V CRISPR-Cas systems

Due to their extensive use in genome editing applications and our in-depth knowledge, the molecular mechanism of guided DNA-interference of type II and type V CRISPR-Cas systems is described here.

Type II systems belong to Class 2 and they employ a single multidomain Cas protein as effector complex called Cas9. Cas9 is guided to the target sequence by a dual-RNA guide (12). That consists of a crRNA and a tracrRNA which pair through repeat/anti-repeat sequences, respectively. The tracrRNA is also required for pre-crRNA maturation (13). crRNA and tracrRNA can be genetically fused together to form a chimeric single guide RNA (sgRNA), reducing the system's components into two members. Cas9 will associate with a sgRNA, mainly through the repeat/anti-repeat duplex and stem loop 1, to form a ribonucleoprotein complex (RNP) (14). RNP complexes will scan DNA for the presence of a short sequence motif at the 3'-end of the target sequence called protospacer adjacent motif (PAM). The PAM motif is crucial for discrimination between self and non-self and rescues the cell from self-targeting at the spacer sequence located in the CRISPR-array (15). The commonly used *Streptococcus pyogenes* Cas9 (SpCas9) recognises a 5'-NGG-3' PAM. Upon PAM recognition, Cas9 will initiate local melting of the double strand DNA and the single RNA strand will invade (displacement of the non-target strand) to form an RNA-DNA heteroduplex (14). The length of the RNA-DNA heteroduplex usually is in the range of 20-30 nucleotides, and upon its formation Cas9 is activated for DNA cleavage. Cas9 uses two nuclease domains (HNH and RuvC) that cleave either the target (TS) or the non-target strand (NTS), respectively. DNA cleavage generally occurs 3 nucleotides upstream of the PAM motif for SpCas9 and results in blunt end DNA double-strand breaks (DDSBs).

Type V systems also belong to Class 2 and employ a single protein as effector complex which is named Cas12. Type V systems of the A-subtype (Cas12a) have extensively been applied for genome editing purposes. In

contrast with Cas9, Cas12a does not require a tracrRNA and can process its own pre-crRNA, facilitating the targeting of multiple protospacers (16,17). Cas12a associates with the crRNA through the pseudoknot structure in the palindromic repeat portion of the crRNA. The RNP complex will search for T-rich PAM sequences located in 5'-end of the target sequence (18). Upon PAM-recognition, the double strand DNA is locally melted and RNA-DNA heteroduplex formation takes place. Cas12a recruits a single nucleolytic domain (RuvC) for cleavage of both DNA-strands. Crystal structures combined with biochemical analyses of Cas12ain an R-loop conformation, have indicated that the cleavage of the displaced NTS strand happens first, followed by cleavage of the TS (19,20). Double strand cleavage by Cas12a results in staggered ends.

Adaptation into a gene editing toolbox

SpCas9 being the workhorse, CRISPR-Cas systems have been found application as a heterologous system in a variety of species. When DDSBs are introduced in a cell's genome, the DNA-damage triggers the response repair mechanism of the cell. DNA double strand cleavage is repairable through two main pathways: i) the non-homologous end joining (NHEJ) pathway, and ii) the homology directed repair (HDR) pathway (21). Through NHEJ repair the two DNA-ends are trimmed and resynthesized before ligated to each other. NHEJ often results in the insertion or deletion (indel) of nucleotides, thereby causing a shift in the reading frame of a protein-coding gene, resulting in its a functional knock out (22). Homology directed repair, on the other hand, is based on the presence of a repair template that shares a certain degree of homology with the DDSBs surrounding sequences. Although repair through a homology template allows for precise editing, the HDR pathway is restricted to the G2 and S phase of the cell cycle and suffers from relatively low editing efficiency. (23,24).

By introducing mutations to the catalytic domains of a Cas-effector, its nucleolytic activity is abolished but the activated RNP complex can still bind with target sequences. If guided to bind to a gene or a promoter sequence, it represses the transcriptional process (25). In contrast, a "dead" Cas-effector (dCas) can be genetically fused with transcriptional activators such as VP64 to upregulate transcription of a target gene when guided to the corresponding promoter sequence (26).

An exciting development for the field of CRISPR-Cas genome editing was the genetic fusion of Cas-effectors with nucleotide deaminating enzymes (27,28). Cytosine or adenosine deaminases are guided to the target sequence using the RNP-complex as vehicle. The nucleotides of the NTS are exposed to deamination during R-loop formation and function as deamination substrates, resulting to C->T or A->G substitutions depending on the deaminating enzyme (cytosine or adenosine deaminase, respectively). Even if revolutionary, base editors are not capable for every possible nucleotide substitution and cannot introduce indels. The solution to this issue came with the design of a Cas9-based prime editor: a fusion of a nickase variant of Cas9 (nCas9) with a reverse transcriptase (RT). The sgRNA is elongated with a primer binding site to the nicked NTS. The pegRNA/NTS heteroduplex serves as primer binding site for the RT domain, which extends the NTS by synthesizing a complementary DNA by using part of the pegRNA with the desired edit. As nCas9 only cleaves the NTS, the synthesized cDNA serves as repair template to repair the nick and to introduce the edits through homologous recombination (29).

Optimization of the gene editing toolbox by Cas-protein engineering through directed evolution

Even if the currently known set of natural Cas effectors is diverse, the CRISPR-Cas toolbox still fails in certain applications, leaving room for

further optimization. Well known constraints are the high off-target effect (low specificity) (30), requirement of specific PAM-motifs (31), unpredictable performance of guides (32), and low nuclease activity under certain physiological conditions (33,34).

Following the principles of Darwinian evolution, natural Cas-effectors can be evolved in a laboratory setting to acquire the desired traits (**Figure 1.2**). A native Cas-effector may be subjected to random mutagenesis by error-prone PCR, resulting in a library containing diverse versions of the original Cas-effector. Subsequently the gene library is expressed to obtain the corresponding protein library in a manner that retains the linkage of gene to protein (genotype to phenotype). Next the protein library is subjected to an appropriate selection or screening procedure, resulting in the enrichment of members of which the desired properties have been improved. Linkage of genotype to phenotype allows isolation of the corresponding gene variants, that may serve as the starting point of a new evolutionary cycle (**Figure 1.2**).

In many cases, directed evolution of Cas-effectors has resulted in variations that demonstrate one of the desired properties: i) shorter or different PAM motif (35), ii) increased target specificity (36), iii) enhanced catalytic activity (37), and iv) stronger binding affinity (38). Commonly, Cas-effectors are evolved in combination with in vivo selection or screening systems (39). However, in vivo selection schemes can often suffer from selection pressure escapees. Alternatively, directed evolution can be performed in an in vitro laboratory setting (40).

This thesis focuses on: i) the development of a novel in vitro directed evolution platform, ii) on the optimization of CRISPR-Cas properties by in vivo directed evolution, and iii) on the expansion of the current CRISPR-Cas toolbox.

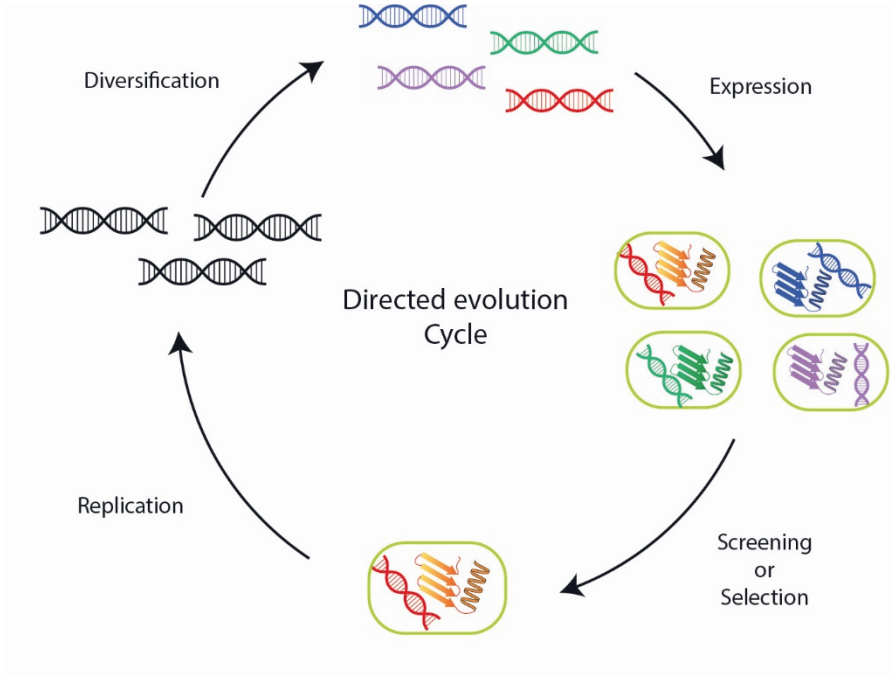


Figure 1.2. Illustration of directed evolution cycle.
The gene of interest is diversified into a population of different versions of itself. The gene library is expressed into proteins while retaining phenotype to genotype linkage. The library is screened for variants that demonstrate the desired property and variants are selected from the population due to an evolutionary advantage. The chosen variants can serve as the starting point of a new evolutionary cycle.

Thesis Outline:

Chapter 1 | Introduction

gives a general introduction to the history of the biological sciences, by noting foundational developments in the field. The principles described, have led to a revolution in molecular biology, including the recent discovery of CRISPR-Cas systems. These systems are a product of the continuous arms race between prokaryotic cells and their parasitic genetic elements such as viruses. They stand out from other prokaryotic immune systems because they confer adaptive cell immunity. Their RNA guided nature of the CRISPR-Cas system has allowed for the unprecedented flexibility to adjust the Cas effectors' targeting specificity, which has resulted in the development of a diverse gene editing toolbox. The relevant details of the molecular mechanisms of the most commonly applied CRISPR-Cas systems (Cas9 and Cas12) is described. Finally, the potential further optimization of natural CRISPR-Cas systems through directed evolution is portrayed.

Chapter 2 | (R)evolution on a chip: Synthetic microcompartments in directed evolution

The second chapter reviews the application of artificial microcompartments in directed evolution schemes. Artificial microcompartments can function as the physical barrier that links genotype to phenotype, hence serving as a viable alternative to cellular compartmentalization. Special focus is given to microfluidics assisted in vitro compartmentalization (μIVC). Scientific principles and technical aspects are discussed through given examples, while following the steps of a directed evolution cycle. The latest advances in μIVC are discussed and propositions for further expansion of μIVC are made. The high customizability of μIVC is emphasized and therefore a (R)-evolution on a chip is foreseen.

Chapter 3 | Towards a microfluidics-assisted in vitro compartmentalization platform for the directed evolution of CRISPR-associated nucleases:

In Chapter 3 we describe our pursuit to build a μIVC platform for the evolution of CRISPR-Cas nucleases. We started by devising fluorescent based schemes to link CRISPR-Cas activity to a readable fluorescence signal. In parallel, we worked on the production of artificial compartments by microfluidics chips and we demonstrate fluorescent activated sorting of water in oil in water droplets by a common cell sorter. Next, we demonstrate successful sorting of mixed droplet populations engulfing Cas effector proteins with different DNA-interference capabilities. Finally, we discuss the necessary parts left for a fully fledged μIVC platform for the directed evolution of Cas nucleases.

Chapter 4 | Evolving a thermophilic type II-C Cas9 with increased catalytic activity in mesophilic temperatures.

ThermoCas9 is a Type II-C effector protein that has been successfully harnessed for genome editing of prokaryotic microorganisms and human cell cultures. However, because of its thermophilic nature, ThermoCas9 has low catalytic activity at 20 ^{o}C, which hampers its application for genome editing of plants and human cells. In Chapter 3 we launch a directed evolution campaign to evolve ThermoCas9 variants with increased catalytic activity at lower temperatures. We make use of an in vivo, positive selection scheme applied in *Escherichia coli*. Initially, we set up stringent selection conditions that allow for selection of ThermoCas9 protein variants with enhanced catalytic activity. We next applied selective pressure on a random gene library and discovered enrichment for mutated amino acids in the RuvC domain and the LII hinge domain. Finally, we purified the enriched protein variants and performed in vitro characterization of their cleaving properties in relation to temperature.

Chapter 5 | Base editing by the miniature Cas12f effector

Even if Type II effectors have undoubtfully proven their wide applicability as genome editing tools in multiple species, their large size remains a restrictive factor when it comes to in vivo gene delivery approaches. In chapter 5 we develop a base editor that meets the needs for simpler and easier in vivo gene delivery. We demonstrate cytosine and adenine base editing in target plasmid substrates in *Escherichia coli*. Interestingly, we observed a unique feature of efficient base editing on the DNA target strand. We also test for adenine base editing in a human cell line but with no base substitution being detected.

Chapter 6: Summary and General Discussion

A general summary of the research described in the different chapters of this thesis are provided. We discuss final experimental results of the thesis and underscore next steps that would promote and facilitate relevant scientific. Finally, a general discussion is included on state of the art genome editing tools as well as on the future of CRISPR-based genome editing. The thesis concludes with certain bioethical considerations concerning the application of genome editing outside of academic context.

Chapter 2

(R)-evolution on a chip: Synthetic microcompartments in directed evolution

Evgenios Bouzetos[*], Ketan Ashok Ganar[*], Enrico Mastrobattista, Siddharth Deshpande[**] and John van der Oost[**]

* authors contributed equally

** corresponding authors

Adapted from publication:

Bouzetos, et al., (R)evolution-on-a-chip. Trends Biotechnol. 2022 Jan;40(1):60-76.

Abstract:

Billions of years of Darwinian evolution has led to the emergence of highly sophisticated and diverse life forms on Earth. Inspired by natural evolution, similar principles have been adopted in laboratory evolution for the fast optimization of genes and proteins for specific applications. In this review, we highlight state-of-the-art laboratory evolution strategies for protein engineering, with a special emphasis on in vitro strategies. We further describe how recent progress in microfluidic technology has allowed the generation and manipulation of artificial compartments for high-throughput laboratory evolution experiments. Expectations for the future are high: we foresee a revolution on-a-chip.

Laboratory evolution in artificial compartments

Laboratory Evolution is based on consecutive cycles of differentiation and selection. To trace a desired phenotype back to its genetic origins, phenotype-to genotype linkage is a key requirement. It can be achieved by physically linking the gene and gene-encoded product (e.g., DNA display (41), mRNA display (42), ribosome display(43) or by compartmentalizing the gene and gene-encoded product within the same physical space. Unicellular microorganisms (e.g., bacteria or yeast; (44,45) or viral particles (phage-display by bacteriophages M13 and lambda;(46,47)) can serve as *in vivo* micro-compartments that separate each gene variant of the population pool. However, despite several successful studies (48–50) *in vivo* approaches can suffer from transformation bottlenecks, host genome mutations, and in case of a fluorescent reporter, it needs to be cell-constrained (51). Alternatively, compartmentalization can be obtained *in vitro* from artificial compartments (52). For example, microtiter well-plates are commonly used for this purpose, but even in combination with sophisticated robotics systems, this system is low-throughput, typically allowing the screening of libraries consisting of 10^4-10^5 members (51,53). Moreover, microtiter plates have already reached their physical barrier because assays with volumes less than 1 µl are problematic due to capillarity and evaporation (54).

A more promising solution is provided by *In vitro* Compartmentalization (IVC)(55). In IVC, libraries of gene variants are generated, after which single gene copies are engulfed into (sub)picolitre-sized artificial compartments such as water-in-oil droplets. Gene expression inside artificial compartment is catalyzed by an *In Vitro* Transcription and Translation machinery (56). A subset of enzyme variants successfully catalyzes the desired substrate-to-product conversion after which enrichment for the product (and the associated gene variants) is performed (57). The process of replication-variation-selection is repeated, until enzyme variants reach the desired level of performance are obtained. The advantage of IVC is that billions of such droplets can be produced using

simple and rather crude techniques allowing effective selection of huge libraries (10^8-10^{11}) (58). A drawback of bulk droplet generation, however, is the inevitable variation in container size. This polydispersity issue complicates accurate quantification of enzymic activity, often resulting in unreliable screening outcomes (59). Another challenge concerns the chemical synthesis required for generation a physical link between the encoding gene and the enzyme substrate (60). As a result, IVC applications have been mainly restricted to DNA-interacting proteins (52). Recent progress in microfluidic technology (**see Box 1**) has allowed the production of highly monodisperse droplets, widely extending the application scope of IVC and giving birth to microfluidics-assisted IVC (μIVC) (61) .

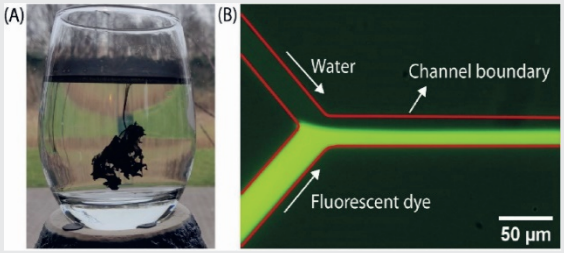


Figure I. (A) An ink drop mixing in water in a turbulent manner. (B) Demonstration of laminar flow in a lab-on-a- chip device by co-flowing two streams of water (with and without a fluorescent dye) parallel to each other without mixing

Box 1. Microfluidics

Since sample quantity is often a limiting resource in biological experiments, reducing sample volumes is highly favorable. Microfluidics enables handling small fluid volumes (in the microliter range or less) and makes use of the predictable behavior of fluids at low-enough flow rates within microchannels of desired geometries. Microfluidic technology has revolutionized biological analysis by downscaling the laboratory-based systems into, lab-on-a-chip devices. These miniaturized systems allow for a substantial reduction of required reagents as well as the analysis time. For example, a typical microtiter plate (96 wells) generally requires 100 µl of sample per well, whereas a typical water-in-oil droplet of 20 µm diameter produced during on-chip experiments corresponds to a volume of merely ~4 pl (i.e., more than seven orders of magnitude smaller).

How do fluids flowing at smaller length scales and low-enough flow velocities differ from fluid flows that we experience in daily life, such as stirring your morning black coffee? The main difference is a predictable laminar flow against an unpredictable turbulent flow, depending on whether the viscous forces or the inertial forces dominate the system. To determine which of these forces get an upper hand, one can calculate the Reynolds number (*Re*) for a system, which is defined as the ratio of inertial forces to viscous forces (Equation I).

$$Re = \frac{\rho u L}{\eta} \quad [I]$$

Here, ρ is the fluid density, u is the velocity, L is the characteristic linear dimension of the system, and η is the dynamic viscosity. Typically, for Re <2000, viscous forces dominate, resulting in laminar flow (62). Thus, while mixing of an ink drop in water is obviously a turbulent phenomenon (Figure IA), two water streams meeting each other in a microfluidic device will simply flow parallel to each other (although they will mix in a diffusive manner further downstream) (Figure IB). Calculating Re for a typical microfluidic system makes this clearer. For a water stream flowing through a microfluidic channel ($\eta \approx 10^{-3}$ Pa.s, $\rho \approx 103$ kg/m^3, $u \approx 1$ mm/s, $L \approx 100$ µm), $Re \approx 0.1$, confirming a laminar flow. The laminar nature of the fluid flow, combined with competing effects of the deformation of the interface by a local shear stress and the resistance to this deformation by interfacial forces, leads to production of µIVC containers (single/double emulsions, liposomes, etc.) in a highly controlled manner (63).

In this review, we focus on μIVC as a platform to perform laboratory evolution. As μIVC has started moving from simple **enrichments assays** to actual directed evolution campaigns, we consider reviewing this approach highly timely. The structure of our review follows the structure of a classic laboratory evolution scheme (**Figure 2.1**): (i) gene differentiation, (ii) gene expression, and (iv) selection and high-throughput screening strategies. However, we are not addressing the subject of gene differentiation as it has been already extensively covered (51). Additionally, noteworthy examples of μIVC campaigns are discussed throughout the review. Finally, we present a perspective on future developments in the field.

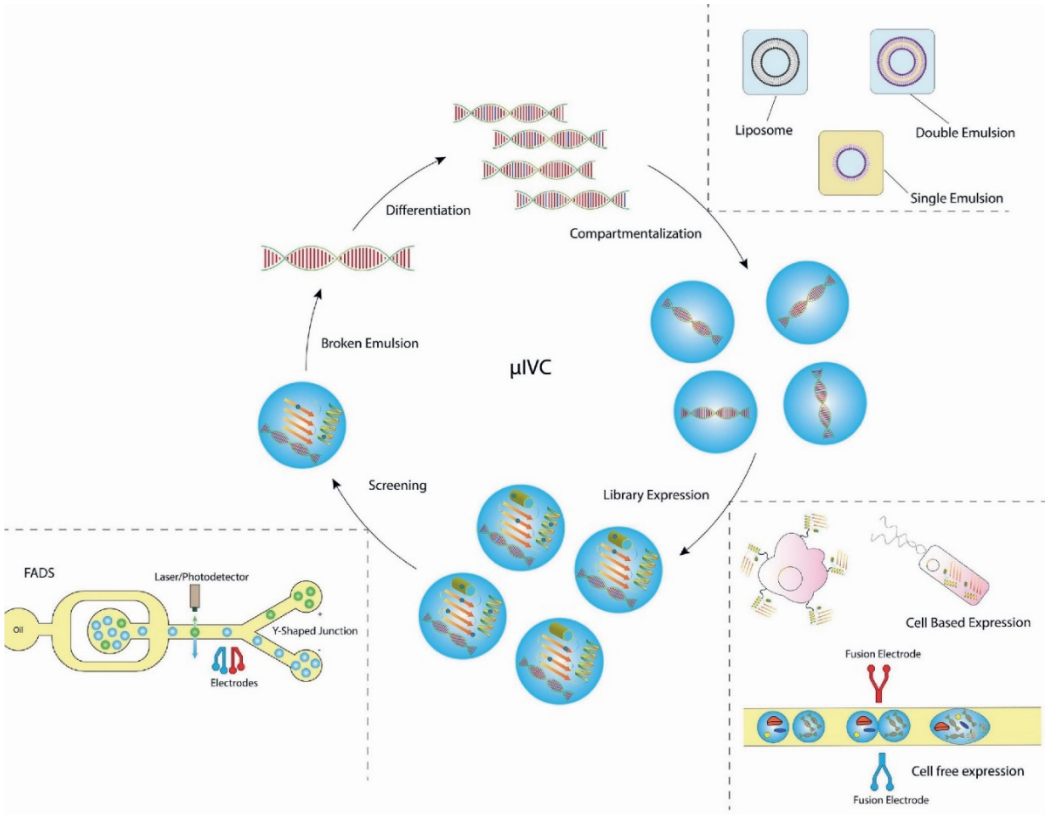


Figure 2.1. A starter gene is randomly differentiated into a library of variants. A single gene variant is compartmentalized into artificial compartments (single/double emulsions, liposomes, etc.). Gene variants are cell-based or cell-free expressed. Clonal amplification of gene variants usually precedes cell-free expression and in vitro transcription and translation (IVTT) components are provided by droplet fusion. High-throughput screening of droplets is based on fluorescence-activated droplet sorting (FADS). Active gene variants that are selected by the screening process serve as the starter genes for the next round of laboratory evolution. Abbreviation: μIVC, microfluidics-assisted in vitro compartmentalization.

Microcompartments for laboratory evolution

Evolution of genes in a laboratory setting strictly requires a genotype-to phenotype association. This is normally achieved by ensuring a single gene variant per compartment (**see Box 2**). Although cells function as natural compartments, their use is restricted by transformation bottlenecks, interference of host-cell proteins with target protein and toxic-product encoding genes. In the following section we discuss the most common artificial compartments currently used in μIVC: single emulsions, double emulsions, and liposomes, how they are generated, their advantages and disadvantages and their applicability.

Single emulsions: water-in-oil confinements

Emulsion-based compartments can serve as an excellent alternative to cells, of which, single emulsions (hereby simply referred as emulsions) are most frequently used. Depending on the volume ratio of water to oil and on the surfactant type, either oil-in-water or water-in-oil emulsions are formed, with the droplet size ranging from a few nanometers to hundreds of micrometers (67). Since oil-in-water emulsions are unsuitable to be used as compartments for the described biochemical reactions, emulsions in the context of laboratory evolution basically refer to water-in-oil emulsions (**Table 1**). To prevent emulsion droplets from fusing, surfactants are routinely added to the oil and sometimes also to the aqueous phase (61).

Emulsions were initially generated in large numbers using straightforward bulk approaches, *i.e.*, by mixing aqueous buffer with oil in presence of surfactants. However, this poorly controlled process generally resulted in highly polydisperse emulsion populations, seriously affecting the uniform encapsulation of reagents. Despite the polydispersity, numerous evolutionary campaigns have been successful (68–71). Nonetheless, polydispersity tends to hamper downstream screening, as small volumetric deviations can substantially affect the concentration of enzyme-derived products (72). Incremental evolution campaigns will significantly benefit from uniform volumes as well as constant reagent concentrations and stoichiometry, highlighting the importance of monodisperse emulsions.

Box 2. Encapsulation of a single gene copy per compartment

Independently of the compartment type, the genotype to phenotype link is essential for a successful laboratory evolution scheme. It is achieved by compartmentalizing a single gene variant per compartment. The gene variant can be encoded by 'naked' DNA templates or by single cells transformed with a gene library. In both cases, encapsulation of a gene variant in an artificial compartment is a stochastic process and obeys Poisson's distribution, where the probability of finding a droplet with k encoding templates follows Equation [II], with λ being the average number of encoding templates per droplet (64,65).

$$P(k,\lambda) = \frac{e^{-\lambda}}{k!}\,\lambda^{k} \quad [II]$$

Thus, in order to ensure the monoclonal nature of each compartment and prevent the simultaneous sorting of an active variant with an inactive variant, most of the compartments are better left empty. For example, at λ = 0.3, 74% of the droplets will be left empty, 22% will have one template, and only 3% will have two (66). Once inside the compartment, cells or an IVTT system will drive gene expression.

Highly monodisperse water-in-oil droplets are generated using **on-chip microfluidics**. For example, droplets, 45 μm in diameter, have been produced at rate of 1-10 KHz (72). Production rate can be substantially increased by operating multiple parallel production channels in single

microfluidic device, or by using a serial droplet-splitting technique, subsequently resulting in smaller droplets. Frequency up to 1.3 MHz was achieved by running multiple microfluidic devices in parallel (73). Mechanical splitting of large droplets into smaller vesicles by introducing splitter array can also increase the throughput, preferably >1000 droplets/hour (74). The high throughput nature, the excellent encapsulation, and the feasibility of subsequent gene expression within emulsions make this approach a highly suitable platform for laboratory evolution. Yet, one of the major limitations of these emulsions is their incompatibility for downstream screening using fluorescence-assisted cell sorting (FACS). This is due to the external oil phase which is not suitable for commercial cell-sorting machines.

Double emulsions: water-in-oil-in-water confinements

This limitation of single emulsions can be resolved by using double emulsions. A typically used double emulsion for biological applications is a surfactant-stabilized water-in-oil-in-water emulsion, i.e., an aqueous droplet containing the gene-of-interest along with the IVTT machinery, surrounded by an oil shell, and dispersed in the outer aqueous phase (**Table 1**). The presence of this external aqueous phase makes double emulsions compatible with commercially available droplet-sorting techniques like FACS. Double emulsions can be made in bulk by further emulsifying water-in-oil single emulsions, with the same key disadvantage of polydispersity. This limitation can be overcome using a microfluidic approach to obtain monodisperse double emulsions with efficient encapsulation. While double emulsions have their own merit owing to its compatibility with downstream sorting processes, the on-chip production rate is relatively low but good enough to be used for laboratory evolution campaigns (75). Indeed, double emulsions produced on chip have has been successfully applied to screen and evolve soluble proteins (76,77).

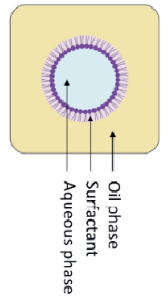
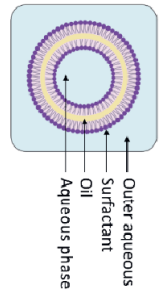
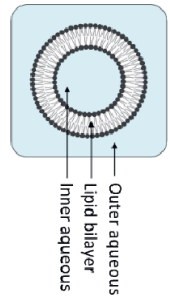
Liposomes: semi-permeable membranous confinements

An alternative to emulsions is liposomes, whose boundary is composed of a continuous lipid bilayer. The lipid bilayer in turn is composed of phospholipids, which are surfactant-like amphiphilic molecules with a hydrophilic head and two hydrophobic chains, allowing them to self-assemble in aqueous environments to form such three-dimensional containers (**Table 1**). Liposomes have gained tremendous interest in the context of building synthetic cells (78,79), and they are suitable compartment candidates, especially for membrane proteins. Their applicability has already been demonstrated by engineering an α-hemolysin variant, a membrane-bound toxin from Staphylococcus aureus, with enhanced pore-forming activity (80).

Liposomes can be produced in various sizes (tens of nm to hundreds of μm in diameter) and different lamellarities (uni-lamellar or multilamellar – having a single or multiple lipid bilayer(s), respectively). For IVC applications, uni-lamellar liposomes of few micrometers in diameter, also known as Giant Uni-lamellar Vesicles (GUVs), are preferred partly due to the detection limit of FACS measurements. GUVs can be generated via bulk techniques such as thin film hydration, rehydration of freeze-dried empty liposome (FDEL), and inverted emulsion transfer (81). However, similar to emulsions, GUVs produced by bulk techniques suffer from polydispersity and an additional size-based sorting step in FACS is required. Two other drawbacks originating from bulk liposome production are the formation of multiple compartments (liposome(s) within a liposome) and multi-lamellarity. Multi-compartment liposomes tend to trap components between different sub-compartments, causing inaccurate evaluation of enzymic activity. They also tend to yield low protein expression, which is undesirable for laboratory evolution (82). Fortunately, several effective on-chip production techniques have been developed over the years, which could prove to be very well suitable for evolution experiments. Octanol-assisted Liposome Assembly (OLA) is a process akin to bubble-blowing, where the initially produced double emulsions give rise to uni-lamellar liposomes

within minutes (83,84). A similar technique was developed using glass capillary-based approach and cell-free, GFP expression was carried out inside the liposomes(85). Spatz and coworkers recently presented another approach, starting with copolymer-stabilized droplets to generate liposomes, termed droplet-stabilized GUVs (86). When combined with a droplet splitting technique, an impressive production rate of $>10^6$ GUVs/min was reported (87). Overall, on-chip GUV production is a promising approach, especially in terms of sample mono-dispersity and encapsulation efficiency.

Table: Types and characteristics of artificial microcompartments

Compartment	Compartment Schematic	Advantages	Limitations
Single emulsions		1) Highly established approach 2) High generation frequency rate $>10^7$ droplets/hour 3) Robust compartment	1) Incompatible with FACS 2) Incompatible with membrane proteins
Double emulsions		1) Compatible with FACS 2) Established approach, $>10^6$ droplets/hour 3)Robust compartment	1) Incompatible with membrane proteins 2)Production through two-step process
Liposomes		1) Compatible with membrane proteins 2) Compatible with FACS 3) Sensitive compartment	1) Low established approach 2) Low production rate

Coupling microfluidics to protein expression

Cell based expression

Cell-based expression is the most straightforward way to achieve protein production. To make use in a microfluidic platform, single microbial cells are encapsulated in water-in-oil droplets where substrate-to-product catalysis takes place. However, as protein expression occurs intracellularly and the substrate for the enzyme is provided extracellularly, enzyme-substrate interaction must be ensured for a successful activity assay. Hydrophobic substrates can freely diffuse through cell membrane and interact with target-enzyme. However, diffusing substrates are not available for every target enzyme. **Cell-surface display**, enzyme secretion, and cell lysis are common alternatives to achieve enzyme-substrate interaction.

Surface display was applied by *Agresti et all.,* in the first demonstration of a directed evolution campaign using μIVC (88). Yeast cells were transformed with a randomly generated variants of horseradish peroxidase (HRP) that were fused with the membrane anchoring peptide (Aga2) which allowed surface display. Each yeast cell was displaying ~10000 copies of a single HRP mutant in its surface (88). Single cells were compartmentalized into emulsions along with a non-fluorescent substrate (**Figure 3.2A**) and active HRP variants converted the substrate to a fluorescent product, allowing bright droplet sorting. Surface display was recently shown with *E. coli* cells too (89). By exploiting the autotransporter protein system the target enzyme takes the place of the passenger and it is surface displayed by fusion with the β-barrel through the linker domain. The advantage of using surface display is that after droplet sorting, live cells can be retrieved and re-cultured.

Yarrowia lipolytica is a non-model yeast species that made a case in μIVC for its biotechnological significance and secretion abilities (90). Efficient secretion of heterologous proteins in *Y. lipolytica* is directed by signal

sequences fused upstream of the protein of interest (91). To screen for thermostable mutants of endo-β-1,4-xylanase, *Y. lipolytica* cells were transformed with two randomly generated xylanase gene libraries generated with high and low mutation rates. Eight clones were found to have up to 10-fold higher residual activity than the wild type, following a heat shock at 90ºC (91). However, small libraries sizes (<1000 clones) could indicate a transformation bottleneck.

When enzyme substrates can freely diffuse through cell membrane enzyme-substrate interaction is ensured without any additional prerequisites. Amplex Ultra Red (AUR) was used as freely diffused substrate in an enrichment assay for CotA laccase which is a periplasmic enzyme from *Bacillus subtilis*(92). Single *E. coli* cells expressing either CotA laccase or a frameshifted ΔCotA variant were compartmentalized into single emulsion (92). AUR was later pico-injected (**see Box 3**) into the droplets, allowing fluorescence-based droplet sorting reaching enrichment up to 437-fold. Like hydrophobic substrates, secondary metabolites can also diffuse through cell membrane. In another enrichment test, *E. coli* cells producing either D-lactate or L-lactate were mixed in a mimic library along with reagents for L-lactate reaction assay (64). Enrichment for L-lactate producing cells reached up to 5800-fold (64) .

Even though living-cells drive efficiently protein expression, cellular barriers can obstruct enzyme-substrate interactions. To overcome this issue, compartmentalization can be followed by cell lysis. This allows the proteinaceous content of the cell to get released into the synthetic compartment, greatly facilitating the interaction of the target enzyme with its substrate. In case of evolution of a thermostable enzyme, *E. coli cells* can be lysed by exposure to near-boiling temperatures. Heat lysis was successfully applied for engineering and enriching xeno, nucleic, acid-incorporating, DNA polymerases (77,93). However, as high temperatures tend to compromise the functionality of most mesophilic enzymes, chemical lysis is the preferred method. Chemolytic agents and *E. coli* cells are mixed on-chip just before they are compartmentalized into droplets to ensure cell

lysis after encapsulation (**Figure 3.2B**). This approach has been successfully tested by enrichment assays for arylsulfatase enzymes reaching 2500-fold enrichment (76). Chemical lysis has been successfully applied for the laboratory evolution of an esterase with altered

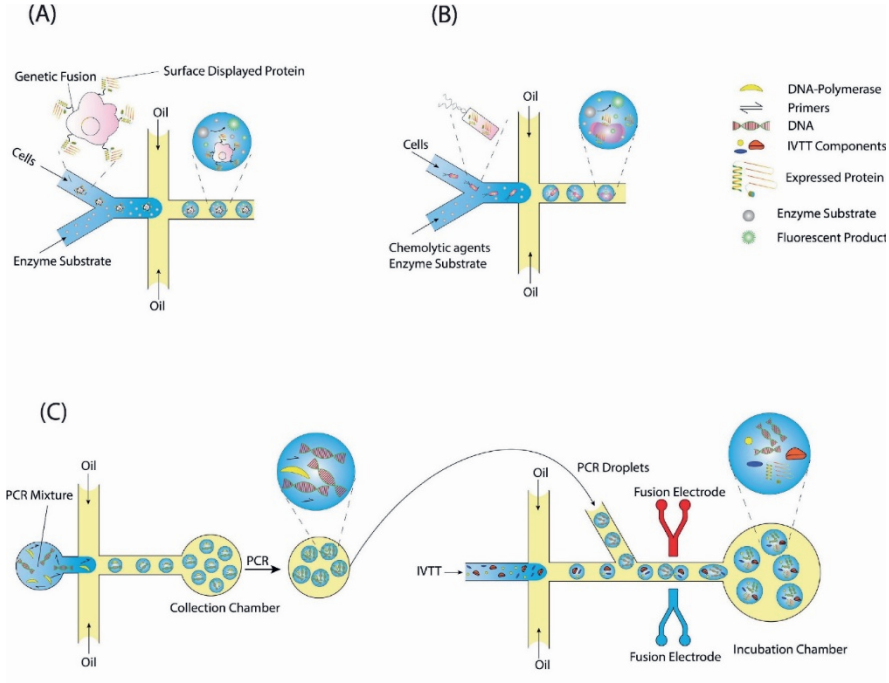


Figure 3. On chip protein expression and enzymatic activity.
(A) The target protein is programmed to be displayed on the microbial cell surface through a genetic fusion with a membrane-anchoring peptide. Each cell displays thousands of copies of the same protein variant. Single cells are compartmentalized into water-in-oil droplets along with appropriate enzyme substrate. Upon encapsulation, active protein variants catalyse the conversion of the substrate to its fluorescent product. **(B)** Target protein expression happens intracellularly. Single cells are compartmentalized into water-in-oil droplets along with cellulolytic agents and the enzyme substrate. After encapsulation, cells are lysed and their proteinaceous content is released inside the compartment. Enzymes encounter the substrate and active variants catalyse its conversion to a fluorescent product. **(C)** Single DNA templates are encapsulated along with a PCR mixture (top). After DNA amplification, the DNA-containing droplets are transferred to a new microfluidic chip (bottom), where it is combined with compartments that contain in vitro transcription and translation (IVTT) components for gene expression. The two types of droplets are fused through electro-coalescence.

enantioselectivity (94), for remodeling the active site of cyclohexylamine oxidase (95) and for the laboratory evolution of an artificial aldolase (96).

Cell-free expression

While cell-based protein production is certainly a valuable approach to perform μIVC, it does not allow for the expression of toxic genes or the introduction of non-natural amino acids, and the biological background can impair screening accuracy. An interesting alternative is cell-free expression by IVTT. IVTT consists of either purified transcription/translation components expressed in *E. coli* (Protein synthesis using Recombinant Elements (PURE)) (97), or by cell lysates provided with additional supplements (98). As with cells, cell-free expression is obtained by single gene copies engulfed into artificial compartments (65). Importantly, in vitro expression of single genes should result in a detectable phenotype. For example, the expression of the green fluorescent protein (GFP) has been detected inside droplets by single DNA templates (99). Moreover, when penicillin-acylase activity was coupled to the expression of a GFP-reporter, it was also detectable from single gene copies encoding for the acylase enzyme and the GFP (100). Such fluorescence-based detection of enzymatic activity encoded by single DNA templates is, however, not always possible. For example, β-galactosidase and FeFe hydrogenase activity could not be detected when encoded by single DNA molecules (65,101). To enhance *in vitro* expression in such cases, a clonal population of a gene inside artificial compartments can be achieved by PCR amplification or rolling cycle amplification (65,101). Nonetheless, DNA amplification (replication) and cell-free expression are incompatible processes because the sensitive IVTT components will be inactivated at the elevated PCR temperatures and rolling cycle amplification is inhibited by IVTT components Therefore, DNA amplification is first performed and components for the cell-free expression and the screening assay are added by droplet fusion (**Figure 2C**) (102) or pico-injection (103) (**Box 3**).

A completed and cell-free, protein directed evolution campaign had long eluded as only enrichment assays were making their case (100,102). However, recently a subtilisin-like protease was evolved to a 5-fold more

active variant completely cell-free (103) . In that case single plasmids were compartmentalized into single emulsion, and they were clonally amplified through rolling cycle amplification. IVTT reagent and enzyme substrate were provided by two consecutive steps of picoinjection (103). It seems that high controllability over the consecutive steps (clonal amplification, protein expression, enzyme catalysis) is the key for a cell-free directed evolution campaign by μIVC.

Apart from optimizing enzymic activity, cell-free expression has been applied for ribozyme engineering. Ribozymes can be expressed by simpler In Vitro Transcription (IVT) machineries because any translation step is unnecessary. X-motif is a ribozyme that catalyzes RNA-cleavage and was initially engineered by **Sequential Evolution of Ligands by Exponential Enrichment (SELEX)** (104). After 9 rounds of μIVC, a new variant of X-motif demonstrated 28-fold increased catalytic activity (105). Spinach is another ribozyme which binds the DFHBI dye acquiring fluorescent properties (106). Randomly generated variants of Spinach were compartmentalized into single emulsions and were clonally amplified by droplet PCR (107). IVT mixture and DHBFI were later provide by droplet fusion. The most optimal isolated variant was named ISpinach and showed increased fluorescence, enhanced thermostability and reduced salt sensitivity (107).

Box 3. Compartment content modification

Droplet modification can be defined as the process of manipulation of droplet content by merging two or more droplets. In droplet-based laboratory evolution, it is a common practice to first express the gene of interest in the host droplet and then add the substrate required for enzyme catalysis, which is present in another droplet. This is often achieved by fusing the substrate-containing droplet to the host droplet. The governing principle to induce droplet fusion is to temporarily destabilize the oil–water interface. Destabilization through electro-coalescence (i.e., electric field-induced droplet merging) (108) has gained most ground in the μIVC field (Figure IA).

Building upon electro-coalescence, the idea of pico-injection was developed. A second microfluidic device allows reagent delivery to existing droplets by applying an electric field as they pass a pressurized channel containing reagents (Figure IB) (109) [. With this design, fluid delivery was attained at high precision, up to sub-picoliter volumes. Pico-injection-based droplet modification has tremendously helped droplet modification in laboratory evolution experiments. Wang and colleagues utilized pico-injection to deliver AUR assay reagents with enzymes (oxidase and HRP) into pre-existing droplets to detect metabolites of interest (64) [. Subsequently, new versions of pico-injectors have been developed, eliminating the requirement for electrodes by using dissolved electrolytes in solution (110) .

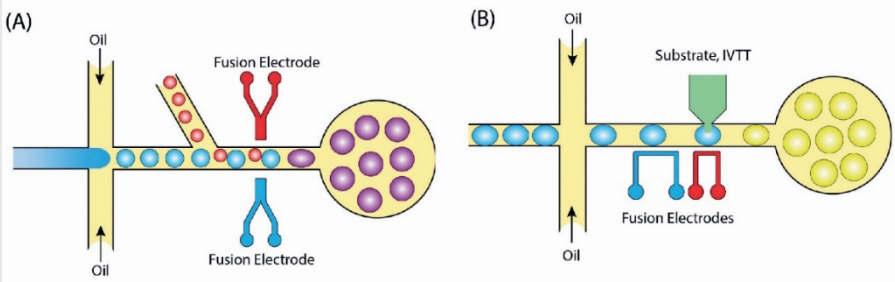


Figure I. Droplet content modification.
(A) In droplet fusion, droplets are fused by temporal destabilization through electric current application. (B) In pico-injection, reagent delivery [substrate, in vitro transcription and translation (IVTT)] is applied through an additional microfluidic channel through temporal destabilization by electric current.

Selection and High Throughput Screening

A crucial part of a laboratory evolution experiment is selection or screening for the gene variants with the desired characteristics. When selection pressure is applied in a population, only members with the required key mutations will be able to survive and/or reproduce. On the other hand, screening is a testing procedure that checks individually each genetic variant for the desired activity. In selection schemes, there is no need for highly monodisperse emulsions because there is no quantification of an assay reaction product. Therefore, selection schemes are well-suited for IVC studies.

Selection Pressure in IVC

Selection within artificial compartments is well-suited for DNA-interacting proteins because the protein-encoding gene and its target substrate can easily be combined on a single DNA molecule, linking the genotype and the phenotype. For example, in the first demonstration of IVC, an enrichment assay was performed between methyltransferase encoding genes and non-methyltransferase encoding genes (55). Successful DNA-methylation protected DNA by subsequent restriction digestion. Therefore, only intact DNA-molecules could be rescued by PCR (55) .

A different approach is followed with DNA-polymerases. DNA-polymerases can drive polymerase chain reactions inside single emulsion leading to the overrepresentation of active genes in the gene library. Compartmentalized Self Replication (CSR) has been applied successfully to engineer DNA-polymerases capable of unnatural nucleotide incorporation (111). Building upon CSR it was shown that gene variants that drive the expression of DNA polymerase more efficiently, will be preferentially amplified during the subsequent compartmentalized in vitro PCR step. Compartmentalized partner replication (CPR) has been successfully applied for the laboratory evolution of T7 RNA-polymerase that drives DNA-polymerase transcription from an alternative T7 promoter and of an aminoacyl-tRNA

synthetase:tRNA pair able to incorporate an unnatural amino acid, restoring DNA-polymerase expression (70).

Screening with Fluorescent Activated Cell Sorters (FACS)

All the aforementioned selection schemes have been only applied to DNA-interacting proteins because the encoding DNA is coupled to a DNA-motif that acts as a substrate, ensuring the genotype-to-phenotype link. Hence, in case of proteins that do not act on DNA, different screening/selection strategies must be developed. The first example of an IVC approach with an alternative, DNA-independent selection was the evolution of a phosphotriesterase and was based on microbead display (60). The role of the microbead is to act as the physical link between the genotype and the phenotype. To do so, magnetic microbeads were coated with streptavidin and the gene library was biotinylated by PCR (60). Due to strong streptavidin- biotin affinity, microbead-gene complexes were formed. Moreover, a biotinylated enzyme substrate was also attached to the microbead (60). Single microbead-gene-substrate complexes were compartmentalized into droplets, where gene expression took place (60) . After incubation to allow for protein expression and substrate-to-product conversion, the emulsion was resolved. The microbeads were then exposed to an anti-product antibody, and as a result, only those microbeads carrying functional gene variants were selected (60). Microbead display is suitably combined with FACS. In this case, microbeads are coated with a fluorogenic substrate that turns to a fluorescent product. Fluorescent microbeads are then sorted by FACS. This approach is gaining some ground and has been positively valuated by enrichment assays for an oxygen-resistant [FeFe] hydrogenase and horseradish peroxidase activity (101,112). It has also been applied for screening of a random library of formate dehydrogenase variants and on engineering a protease with post-translational modification specificity (113,114).

Even if useful, microbead display still requires complex biochemistry schemes. Hence, one would like to have artificial compartments that can be sorted by FACS, without the use of microbeads. As indicated above, double emulsions are synthetic compartments that can be sorted by FACS (115) and microfluidic technology has made possible the generation of highly monodisperse compartments that allow for high screening accuracy. This was convincingly demonstrated by screening for wild-type arylsulfatase variants from a mimic library containing a low activity arylsulfatase variant in overwhelming numbers. Enrichment was found to reach up to 100000-fold (76). The approach has also been applied successfully to engineer a manganese-independent polymerase that can incorporate non-natural nucleic acid building blocks (77). FACS is expected to be compatible with different artificial compartments if the continuous phase remains aqueous. For instance, a library of aminoacyl-tRNA synthetases (ARSs) was co-compartmentalized in liposomes with a GFP gene carrying an amber codon (116). Successful incorporation of an unnatural amino acid by ARSs restored GFP expression, allowing for sorting of liposomes containing active ARS variants by FACS (116). Alternatively, by adding agarose and alginate in the aqueous solution of water-in-oil droplets and cooling to 4^{o}C, a solid 'gel-shell' bead is formed (117). By exchanging the oil phase to water, the gel-shell bead remains stable and can be sorted by a normal FACS facility. After two rounds of sorting, enrichment >100000-fold has been obtained for gel-shell beads that encapsulated active phosphotriesterase variants (117) .

On-chip, high throughput droplet sorting

Microfluidic IVC (μIVC) owes its rapid progression to technological advancements that allowed the performance of all the necessary steps to perform laboratory evolution on-chip. FACS is not suitable for sorting the commonest of artificial compartments: the water-in-oil emulsions. Even alternative compartments such liposomes are more complicated to generate

and must be transferred to the FACS facility. Fluorescent-activated droplet sorting (FADS) is a system to sort water-in-oil droplets on-chip by **dielectrophoresis**. If fluorescent and non-fluorescent droplets are to be sorted, they are exposed to a laser. In case of detected fluorescence, a sensor triggers the generation of a non-uniform electric field that polarizes the fluorescent droplet causing its deflection towards the positive arm of a Y-shaped junction (**Figure 3A**). Droplets without positive signal spontaneously flow in the negative arm because of the lower hydraulic pressure of the Y-shaped junction. Apart from fluorescence discrimination, the system was found suitable for other types of sorting, for example of a model library of *E. coli* cells expressing β-galactosidase (118). Based on the enzymatic conversion of a chromogenic substrate (ONP-Gal, releasing a yellow color after cleavage) sorting was performed at ~300Hz (300 droplets per second) with false positives rates less than 1 in 10000 (118).

The first application of FADS was to screen a library of HRP variants for enhanced activity (88). Single yeast cells that expressed HRP on their cell surface were compartmentalized with a fluorogenic substrate in water-in-oil droplets (88) . Droplets were sorted at ~2000 Hz for up to 3 hours and a 10-fold faster HRP variant was found (88). FADS has been successfully modified to screen for enantioselective enzymes using a dual-channel screening platform (94). The goal was to improve the enantioselectivity of an esterase towards (S)-profens.

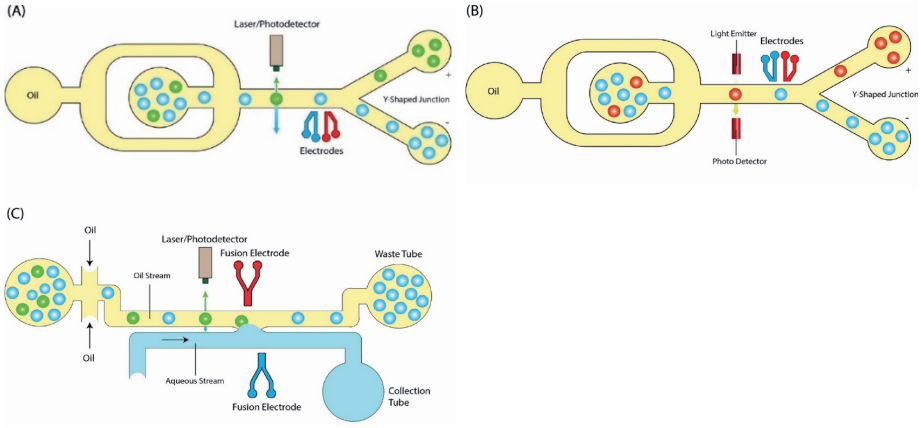


Figure 4. On chip, high-throughput screening.
(A) Fluorescent-activated droplet sorting (FADS). Water-in-oil droplets are initially spaced out by an oil phase. Each droplet is interrogated for fluorescence. When fluorescence above a set threshold is detected, a pair of electrodes send nonuniform pulses of electric current that polarize the fluorescent droplet. The polarized droplet is then deflected to the positive arm of a Y-shaped junction. Nonfluorescent droplets passively flow towards the negative arm due to smaller hydraulic pressure. **(B)** Absorbance-activated droplet sorting (AADS). Water-in-oil droplets are once again spaced out with an oil phase. Then they are interrogated for alterations in absorbance of their content. Positive hits are polarized by electrodes and, similar to FADS, are deflected to the positive arm of a Y-shaped junction. Negative-hit droplets passively flow to the negative arm of the junction. **(C)** Fluorescent-activated electrocoalescence (FAE). FAE is a screening system that collects only the content of the droplets. Water-in-oil droplets flow parallel to an aqueous stream. Detection of fluorescence above a set threshold triggers a voltage that causes the droplets to coalesce with the aqueous stream. The droplet is dissolved and its contents are collected. Droplets with a

Although the aforementioned cases of μIVC-based sorting have shown success, there are more variations that are worth mentioning. The first one is an approach in which rather than collecting individual droplets, only their content is retrieved. This technique, introduced as Fluorescent Activated Electro-coalescence (FAE), consists of an aqueous and the droplet containing, oil stream that co-flow between two electrodes (**Figure 3C**) (102). When fluorescence is detected, an electric pulse is applied, resulting in the coalescence, i.e., merging of the aqueous content of the passing

droplet with the aqueous stream below. The droplet disintegrates and its contents are collected. Yet another sorting approach that uses a magnetic field has been developed to separate droplets encapsulating different number of microalgal cells, reflecting different growth rates between cell strains (119). All generated droplets contain iron oxide nanoparticles and are attracted by the same magnetic force. According to Newton's second law of motion, differences in mass/cell density of the microdroplets results in different acceleration and hence distinct lateral displacement of each droplet, allowing sorting (119) . A third case is also label-free and uses electrochemical principles. Electrical current passing through droplets can be quantified by a boron-doped diamond electrode, accurately reflecting the NADH concentration (120). Droplets above a given threshold are then sorted dielectrophoretically using the Y-shaped junctions. Isocitrate dehydrogenase reduces NAD^+ to NADH, and after two rounds of laboratory evolution, 5 variants showing higher activity than the wild type were isolated (120) . Finally, it was shown that droplet content can be analyzed by mass spectrometry in combination with electron spray ionization (ESI-MS) (121,122). However, in these cases the droplet was destroyed in the process, thus the beneficial genotype could not be tracked down. Mass activated droplet sorting (MADS) is based on an ingenious chip design and a programmable sorting algorithm (123). Droplets injected for sorting are first spitted into two parts. One-part flows through a short line to the mass spectrometry inlet for subsequent analysis and the other part flows through a long delay line that concludes to a Y-shaped sorting device. While the second part travels through the delay line, content analysis of the first part takes place and the sorting algorithm decides, above a given threshold, if the second droplet part corresponds to a positive hit. MADS has been tested over a range of enzyme-substrate concentrations and sorting accuracy was found at 98% (123) . Even if promising, MADS performed at 0.7 Hz which falls greatly behind FADS (123) . However, the universality coming with its label-free nature leaves expectations for further development.

Table 1. Enzymes evolved by μIVC

Enzyme	Compartment	Screening	Target Property	Reference
Alpha hemolysin	liposome	FACS	Membrane pore formation	(80)
Amicoancyl-tRNA synthetase	Liposome	FACS	Incorporation of unnatural aminoacids	(116)
X-motif ribozyme	Single emulsion	FACS	Enhanced catalytic activity	(105)
Spinach ribozyme	Single emulsion	FADS	Fluorescence, thermostability, salt tolerance	(107)
Phelylalanine dehydrogenase	Single emulsion	AADS	Enhanced catalytic activity	(124)
DNA-polymerase	Double emulsion	FACS	Manganese independence	(77)
Aldolase	Single emulsion	FADS	Stereoselectivity, wider substrate compatibility	(96)
Esterase	Single emulsion	FADS	enantioselectivity	(94)
Oxidase	Single emulsion	FADS	Non-natural substrate catalysis	(95)
Protease	Single emulsion	FADS	Increased catalytic activity	(103)

Abbreviations: epPCR: error-prone PCR, STEP: staggered extension process

Conclusions and Future Perspectives

While microfluidics has not yet been a true game changer in the field of life sciences as it was expected to be (62), the strength of droplet microfluidics has been clearly recognized and aptly utilized to progress the field of laboratory evolution. One major disadvantage in the current state of μIVC is that in most cases microfluidic modules are generated by specialized laboratories. Standardization and commercialization of microfluidics devices

will allow for more consistent assays and give higher accessibility to non-specialists. While the simplest, 'hand-made' version of IVC is an appropriate method for selection of DNA-interacting proteins, μIVC has allowed for high-throughput screening of different enzyme classes. The level of control of microfluidics-based compartment generation allows for standardized, accurate, downstream screening of different compartment types. In addition, its compatibility with both cell-based and cell-free expression systems offers flexibility in terms of gene expression, allowing screening of protein features under desired conditions. Finally, on-chip, fluorescence-activated droplet screening has resulted in obtaining a variety of enzymes with desired properties, meaning that μIVC is now moving from the development stage to real life applications.

It is anticipated that more developments to occur as μIVC gains more space as a laboratory evolution scheme. Starting from screening, FADS has become the norm but there is a well-placed desire for label-free screening methods. AADS has already come into play but droplet-MS could be the next big novelty. However, throughput must be increased substantially to compare with the current screening ratios already achieved by FADS which can reach up to 30KHz (125). Because water-in-oil emulsions may cause problems in sorting devices (FACS), on-chip sorting methods (FADS, AADS) and alternative containers (double emulsions, liposomes) have been sought for. In addition, commercial systems have recently become available that use disposable cartridges for sorting (https://www.miltenyibiotec.com/NL-en/products/macs-flow-cytometry/cell-sorter.html, https://on-chipbio.com/product-onchip_sort/). For example, the On-Chip sorter is discussed that successfully sorted single emulsions engulfing GFP expressing *E. coli* cells (https://on-chipbio.com/emulsion-sorting/).

High-throughput, accurate, on-chip sorting of huge libraries is the holy grail of laboratory evolution. As every μIVC module is generated in-house, the approach is highly characterized by controllability and flexibility. However, this fact comes in hand with significant complexity. Outstanding questions (see Outstanding Questions) indicate specific points that needs addressing

to expand and simplify μIVC. As described in this review, the combination of evolution principles with the emerging microfluidics technology allows for unprecedented possibilities with respect to high-throughput screening and selection to obtain biocatalysts (proteins, ribozymes) with desired optimal features. As Darwin concludes in his famous book: "There is grandeur in this view of life, (...) from so simple a beginning endless forms most beautiful and wonderful have been and are being evolved" (2). This is not only true for biological creatures, but also for proteins, nucleic acids, as well as inorganic compounds. Hence, expectations for the future are high, we foresee a revolution-on-a-chip.

Chapter 3

Cell-free, screening of CRISPR-Cas activity by microfluidics assisted in vitro compartmentalization (μIVC)

Evgenios Bouzetos[*], Ketan Ashok Ganar[*], Siddharth Deshpande[**] and John van der Oost[**]

* authors contributed equally
** corresponding authors

Manuscript in submission

Abstract

Natural CRISPR-Cas systems are responsible for anti-viral immunity of prokaryotic cells. Some CRISPR-associated (Cas) nucleases have been repurposed as powerful genome editing tools. To further increase the applicability of natural Cas nucleases, protein engineering through rational design and/or random mutagenesis has resulted in a variety of synthetic Cas nuclease variants with improved performance. Libraries of random Cas protein variants are generally screened using in vivo selection strategies which, however, can suffer from transformation bottlenecks and selection pressure escapees. In vitro compartmentalization (IVC) is an interesting alternative in which cell-free gene expression is executed in artificial microcompartments. IVC has been successfully applied previously for optimization of the activity and specificity of endonucleases. Here, we set the foundations for a microfluidic-based, cell-free, high-throughput screening approach tailored towards the optimization and characterization of CRISPR-Cas systems. We show guided DNA-interference through protein fluorescence and ensure linkage to the genotype by compartmentalization into micron-sized water-in-oil-in-water double emulsions generated by on-chip microfluidics. Screening of compartments is demonstrated in high-throughput fashion using a common cell sorter. We demonstrate tracing of the expected genotypes from a model compartment population according to their relative fluorescence. Finally, we discuss the next steps necessary for the adaptation into a fully-fledged, high-throughput screening method.

Introduction

CRISPR-Cas systems allow for an adaptive immune response in prokaryotic cells that contributes to the defence against infections of viruses and other mobile genetic element. CRISPR-Cas systems consist of two main components, namely the Cas-nuclease protein and a guide RNA (gRNA) that together form a ribonucleoprotein (RNP) complex (126). This RNP complex proceeds to interrogate large double stranded DNA sequences to find a 20-30nt long target sequence (also called protospacer) for base complementarity with the variable part of the gRNA (12,127). Prerequisite for initiating formation of the heteroduplex between the gRNA and the target DNA strand, is the presence of a 2-4 nt long sequence, named protospacer adjacent motif (PAM) that, depending on the CRISPR-Cas system, is located at the 5' or 3'-end of the protospacer (128,129). After formation of an R-loop structure (a guide-RNA/target-DNA heteroduplex and a displaced non-target strand), specific cleavage of both DNA strands is catalysed by the nuclease domain(s) of the Cas-effector (12,127). The ease of generating synthetic target-specific guide sequences as well as the functional expression of the corresponding RNP complexes in a wide variety of host organisms have established CRISPR-Cas systems a versatile genome engineering platform, with applications ranging from biotechnology to gene therapy (130).

The ease of programmable target sequence cleavage by CRISPR-Cas systems has had a huge impact in the field of genetic engineering (130,131). Yet, naturally occurring Cas-effector variants often fail in certain applications, leaving room for innovation and improvement of the Cas effectors. Some well-known constraints of native CRISPR-Cas systems are their low specificity (high off-target interference), requirement of specific PAM motifs, and low nuclease activity under certain physiological conditions (10). Protein engineering of Cas-effectors has produced a variety of synthetic variants with new properties such as altered PAM compatibility, improved specificity and/or increased activity (132–134) .

Engineering of Cas effectors can be performed either rationally when structure-function relations are well understood, or randomly in case relevant information is lacking (135,136). Random mutagenesis is performed usually in combination with *in vivo* selection systems in microorganisms such as *Escherichia coli*, yeast or mammalian cells (49,132–134,137). However, these *in vivo* selection systems can suffer from low gene expression, transformation bottlenecks and other host-specific factors such as selection pressure escapees (51,138).

In vitro compartmentalization (IVC) eliminates the use of living cells in favour of cell-free gene expression inside synthetic compartments (55,139,140). IVC has been successfully applied for selection of endonucleases and other nucleic acid-interacting protein variants (68,111,141,142). In the case of DNA-cleaving enzymes, the required genotype-to-phenotype link is ensured by including the target sequence in the nucleic acid fragment that also encodes the protein of interest. When an active enzyme cleaves its target dsDNA sequence, resulting in sticky ends, compatible adaptors can be ligated to the nuclease product. The adaptor sequence subsequently serves as a primer binding site for a PCR reaction, amplifying only the genotypes (gene variants) that encoded enzymes with the desired phenotype (activity and/or specificity) (69,143).

DNA targeting CRISPR-Cas systems can have multiple cleavage patterns and therefore different cohesive ends can potentially be produced by a single Cas-effector, thus complicating the design of adaptor sequences (14,144). Also deactivated Cas-effectors can still interfere with target sequences although they do not generate dsDNA breaks (145–147). Hence, to circumvent the need of adaptors, we set out to adapt the classic IVC approach to a high throughput screening scheme, based on fluorescence level alterations.

To accurately assess the fluorescent intensity between artificial compartments, high emulsion mono-dispersity is important. On-chip microfluidics allow effective manipulation of fluids at the micrometre scale,

allowing the generation of highly monodisperse compartments, such as water-in-oil-in-water double emulsions which are compatible with the aqueous sheath fluid of common cell sorters (148,149). The uniform compartment size and efficient encapsulation of biomolecules within a population makes it easy to couple gene activity with fluorescence intensity (150). The double emulsion nature of the compartments makes them compatible with common cell-sorting equipment and abolishes the need for complex, droplet-screening microfluidics devices (58,77). It should be noted that double emulsions are prone to deformation under shear (151) and can even rupture in an osmotic imbalance environment (152). Thus, downstream calibration is still necessary for the screening procedure by the cell sorter. However, sdDE-FACS (single droplet Double Emulsion-FACS) and extensively optimized instrument settings (laser, flow rate, and drop delays) for commercially available cell sorters has resulted in droplet recovery up to 70% (153).

Here we describe a strategy that is based on substituting living cells with monodisperse double emulsions produced using microfluidic devices, and the subsequent screening for CRISPR-Cas activity according to the alteration in fluorescence levels. We first prototyped and tested genetic circuits to connect CRISPR-Cas activity with protein fluorescence intensity. We then encapsulated the components of the gene circuits into double emulsions and validated the fluorescence-based assay. Finally, as a proof-of-principle, we mixed emulsions of known compositions and single droplets were sorted based on their fluorescence intensity using sdDE-FACs, after which we confirmed enrichment for the expected genotype.

Material and Methods

Microfluidic chip fabrication and droplet generation

The master wafer was prepared using previously described UV lithography and the protocol was adjusted at attain a channel height of 20 µm. To prepare microfluidic devices, polydimethylsiloxane (PDMS; SYLGARDTM184 elastomer) and curing agent were mixed in 10:1 weight ratio. The mixture was poured on the master wafer, degassed using vacuum desiccator and baked at 70°C for 4 hours. The hardened PDMS block was carefully removed and inlet and exit holes were punched using a biopsy punch of 1.2 mm diameter (PT-T983-05, Darwin Microfluidics). The PDMS block was then bonded to PDMS-coated glass slide (1mm thickness, Corning 2947-75x25) and incubated at 80°C for 2 hours. To produce water-in-oil-in-water DEs, the outer aqueous and exit channels were passivated by flowing 5 % w/v polyvinyl alcohol (Sigma-Aldrich P8136) solution. The device was then baked at 120 °C for 15 minutes and stored at room temperature in a dry place. The device was rested for at least 12 hours before utilizing for double emulsion production. To flow the fluid through the microfluidic channels of the device, PTFE tubing (LVF-KTU-15, Darwin microfluidic) was connected to microfluidic reservoir XS (LVF-KPT-4XS, Darwin microfluidics). The fluid flow was controlled using pressure controller OBK1 Mk3+ (Elveflow).

The inner aqueous consisted of myTXTL cell-free extract (supplemented with the corresponding vectors plasmids. The oil phase consisted of HFE-7500 with 2% w/w Fluosurf-C (EU-FSC-V10-2%-HFE7500, Darwin microfluidics). The outer aqueous consisted of PBS pH7.4 with 1% Tween-20.

Microscopy examination of droplets and image analysis

All the microscopic images were acquired using Nikon-Ti2-Eclipse inverted fluorescence microscope equipped with pE-300ultra illumination system. The droplets were visualized using Nikon CFI Plan Fluor 10x Objective. The

fluorescence from green fluorescence protein was observed using 482/35 nm excitation filter, 505 nm dichroic mirror and the emitted light was collected through 536/40 nm emission filter (Semrock). Bright field images were captured at 10 ms exposure while the fluorescent images were captured by varying the exposure between 0.2-1 s using a Prime BSI Express sCMOS camera. The fluorescence intensity of GFP within the droplets was analysed using Fiji Image 1.5.2. For analysis, background-corrected signal intensity was used, calculated as the difference between the fluorescence intensity within the double emulsions and the background intensity (fluorescence signal intensity outside the DEs). The average background-corrected fluorescence intensity difference between two distinct double emulsion populations was determined using a two-tailed significance test indicated.

Propagation and assembly of expression constructs

pCas, pQuide and pTarget plasmid vectors were assembled by standard cloning methods and propagated in *Escherichia coli* DH5a. When transcription was driven by the lambda phage P70a promoter, expression vectors were propagated in in *Eschecichia coli* KL740 cI857+ (Supp. Table 1). *E. coli* was routinely cultured in Luria Bertani medium (LB), 10 g/L tryptone, 10 g/L NaCl, 5 g\L yeast extract supplemented with the appropriate antibiotic. pQuides plasmids included a pair of BbsI restriction sites flanking the spacer sequence to facilitate consequent spacer sequence exchange. Plasmid vector sequence was verified by Sanger Sequencing.

deGFP-6xHis Purification and protein fluorescence calibration curve

deGFP gene was fused with a C-terminal 6xHis-tag in a pET28b vector. The pET28b-deGFP vector was transformed into *E. coli* BL21 (DE3) cells. A single colony of cells was inoculated into LB-media until OD600 = 0.5-0.6. Recombinant deGFP expression was induced by 0.5 mM IPTG, over 16 hours

at 20 ºC. Cell pellet was resuspended in Binding and Wash buffer (50 mM NaH2PO4, 300 mM NaCl, 20mM imidazole, pH=8.0), complemented with protease inhibitors (Roche cOmplete, EDTA-free Protease Inhibitor Cocktail). Cells were lysed by a Q500 Sonicator, at 30% power through consecutive run and pause rounds of 1 s and 2 s, respectively, over 15 minutes. Cell lysate was filtered through 0.22 µm filters and applied to a Histrap HP (GE Lifesciences) nickel column. After sample application, the column was washed with 5x column volumes of binding and wash buffer. Bound protein was eluted with elution buffer (50mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH = 8.0) and fractions of 1ml were collected and analysed by SDS-PAGE gel electrophoresis. Fractions were pulled together and dialyzed in 500 mL 20 mM Tris-HCL, pH = 8.07, overnight at 4 ºC using a dialysis tube (Sigma Aldrich #D0505-100FT). Dialyzed protein solution was applied in anion exchange chromatography, HiTrap Q HP column. Bound protein was eluted into 1ml fractions by applying a gradient of elution buffer (20 mM Tris-HCL, NaCl 1 M, pH = 8.07) up to 50%. Fractions were analysed in SDS-PAGE gel and pulled together accordingly (Supp. Figure 4a). deGFP protein was quantified at 280nm (Ext. coefficient=20525 1/M*cm, MW:26.454 kDa) with three technical replicates. deGFP protein was diluted accordingly in PBS 1X and fluorescence intensity was measured in a biotek synergy Neo2 plate reader (Ex = 485 nm, Em = 528 nm)

Protein fluorescence kinetic measurements

Cell free TXTL extract was provided under the commercial name myTXTL Sigma 70 Master Mix Kit by Arbor Biosciences. myTXTL reactions were assembled according to manufacturer's instructions. Plasmid vectors were purified using the Zymoclean midiprep kit (#D4200) and further cleaned and concentrated using Zymoclean Clean and Concentrator (#D4013) according to manufacturer's instructions. In short, pCas was used at 3 nM, pTarget at 1 nM and pQUIDE at 1 nM. When expression was driven by linear templates, GamS protein (ArborBiosciences #501024) was added at 10 µM.

Each 12 µL reaction was splitted in two wells of a PP, V-bottom plate 96well plate (greiner BIO-ONE, 651201) for two technical replicates (5.5 µL per replicate). Fluorescence (Ex = 485nm, Em = 528nm) was monitored on a biotek synergy Neo2 plate reader for 16 hours at 30 ^{o}C with point measurements being taken every 3 minutes.

FACS sorting of water-in-oil-water compartments and nucleic acid recovery

Prior sorting the double emulsions were diluted in 1:5 v/v ratio in PBS. The fluorescence-based sorting was performed using Sony sorter SH800 using a 130 µm sorting chip. The sorting conditions were optimized according to previously described study6 along with manufacturer's manual.

1.5 µL of sorted double emulsions were thermocycled using the Q5-HF 2X MM (NEB, #M0492) and the BG22686 and B20724. Concentrated amplicons were sent for Sanger sequenced.

Microbead display

Iron microbeads M-270 (Thermofischer, #65305) were coated with the 5'-, doubly biotinylated BG29694 oligonucleotides (/52-Bio/, IDT) according to the bead's manufacturer's instructions.

The emulsion's oil phase was based on water-saturated mineral oil, supplemented with Span80 4.5%, Tween80 0.4% and Triton X-100 0.05%. The inner aqueous phase Q5 Hifi Fidelity 2X mix (NEB, #M0492), BG29694 (0.05 µM), BG29695 (0.5 µM), 3 ng of /52-Bıo/-P70a_deGFP linear template (λ=2.5-25) and varying amounts of M-270 microbeads (S1=7x10^{6}, S2=13x10^{6}, S3=20x10^{6}) in a total volume of 200 µL.

Water-in-oil emulsions were prepared in cryogenic vials (Corning, CLS430659) by dropwise addition of 200 µL of aqueous phase while stirring

at 1100rpm with a 8X3 mm magnetic bar over 1.5 min and additional stirring for 5 min.

Emulsions were splitted into 50-60 μL parts and transferred to PCR eppendorves. PCR cycling was carried out with the following conditions: 98 ^{o}C / 30 sec, 35 cycles of 98 ^{o}C / 10 sec, 64 ^{o}C / 30 sec and 72 ^{o}C 36 sec and one final step of 72 ^{o}C / 2 minutes.

After completion of PCR the emulsion parts were pooled together and centrifuged at 13.000 rpm for 5 minutes. Oil supernatant was disposed and 800 μL of NX-buffer (100mM NaCl/1% Triton X-100, 10mM Tris-HCl, pH=7.5, and 1mM EDTA) were added. The mixture was homogenized by pipetting and centrifuged for 1.5 minutes at 8,000 rpm. After disposing the supernatant without disturbing the precipitated beads, the procedure was repeated with 600 μL of NX buffer until the solution reached transparency.

When transparency was reached, beads were captured by a magnetic tray and supernatant was removed. Beads were suspended in 200μl of B&W Buffer 1X (Thermofischer, #65305) and transferred to a clean 1.5 mL eppendorf. After magnetic capture of the beads and removal of the B&W Buffer, 12μl of myTXTL cell-free expression reaction mix, supplemented with GamS protein (10 μM) (ArborBiosciences #501024) were added. The samples were incubated overnight at 30 ^{o}C, under mild rotation to avoid bead precipitation. After ~16 hours, beads were magnetically captured and the cell free TXTL mix was transferred in the wells of a PP, V-bottom plate 96well plate (greiner BIO-ONE, 651201). Three end point fluorescence measurements were taken in a Neo2 biotek synergy Neo2 plate reader (Ex=485 nm, Em=528 nm).

Results & Discussion

On-chip encapsulation of cell-free TXTL extract in double emulsions

To generate micron-sized water-in-oil-in-water double emulsions (DEs), we built PDMS-based microfluidic devices with two consecutive flow-focusing junctions connected by a resistance loop (**Figure 1A**; **Supp. Figures S1, S2**). A horizontal channel designated to carry the inner contents of DEs, hereafter referred to as the inner aqueous channel, feeds into the first flow-focusing junction formed by two perpendicular oil channels carrying the continuous oil phase (**Figure 1A**; **Supp. Figure S3**). The exit of the first junction is strategically designed as a single loop which effectively increases the length of the channel providing enough resistance to optimize the pressures in inner aqueous and oil channels during production. The resulting single emulsions (SEs) are then directed to the second junction, where perpendicular channels facilitate the pinching-off of the SEs by a surfactant-containing outer aqueous phase, forming DEs (**Figure 1B**, **Supp. Figure S3**). The generated DEs flow through the exit channel and are collected via the exit hole and stored in a glass vial. Prior to the production of DEs, the microfluidic devices undergo surface passivation to make the exit channel hydrophilic (154) (details in Material and Methods; **Supp Figure S2**).

To assess the encapsulation and retention of small molecules, 10 µM of fluorescein isothiocyanate (FITC) fluorescent dye dissolved in PBS was used as the inner aqueous phase. The produced DEs efficiently encapsulated and retained the FITC dye for >24 hours (**Figure 1B**) indicating retention of small molecules. For a typical experiment, the DEs were produced with an average diameter of 37.6 ± 1.8 µm (N = 1875; coefficient of variation = 4.6%) (**Figure 1C**) and with a throughput of around 500 Hz, corresponding to a production rate of $>10^6$ droplets per hour. After characterizing the stability, encapsulation, retention and monodisperse nature of the DEs, we further tested its biocompatibility.

We produced DEs with inner aqueous phase consisting of the cell-free transcription and translation (TXTL) extract, supplemented with enhanced green fluorescent protein Del6-229 (deGFP) expression vectors (155) (**Figure 1D**). TXTL drives efficient cell-free expression from the *E. coli* endogenous RNA polymerase and sigma-70 (σ70) factors. We tested four known σ70 promoters for deGFP expression in the cell-free extract (**Supp. Figure S5**). deGFP concentration was determined based on of calibration curve of deGFP-6xHis fluorescence signal (**Supp. Figure S4**). As expected, we observed detectable but varying protein expression between the σ70 promoters, generally reflecting the promoter strength (45), whereas no deGFP expression was detected driven by the T7 promoter (**Supp. Figure S5B**). Hereafter, we chose the P70a_deGFP construct as our core gene expression cassette.

TXTL is a complex lysate of *E*scherichia *coli of* which the precise osmolarity is not known, the use of 1% Tween-20 surfactant in PBS (pH 7.4) as the outer aqueous phase did result in stable DEs. Upon incubation at 30°C for 16 hours, the DEs showed bright fluorescence intensity (1044 ± 217; mean ± standard deviation; N = 412) indicating successful transcription and translation of the plasmid-encoded GFP protein (**Figure 1E**). Under identical imaging conditions, a negative control (no GFP plasmid but only TXTL) showed negligible fluorescence intensity. Statistical analysis also confirmed significant different between the two samples with p value < 0.001 (**Figure 1F**).

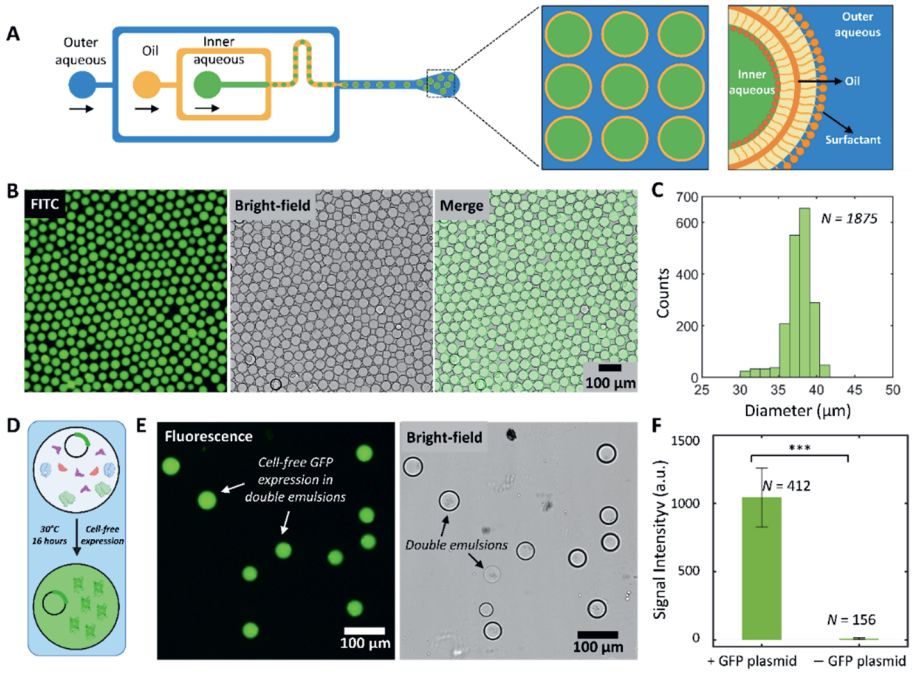


Figure 1. On-chip controlled generation of TXTL-compatible double emulsions. (A) Molecular gene circuit showing guided DNA-interference by CRISPR-Cas systems on a deGFP-encoding DNA template results in decreased fluorescence levels. **(B)** Kinetic measurements of deGFP(µM) concentration in presence of 1nM: pP70a_deGFP, 3nM: pJ23108_Cas and 1nM: pJ23119_gRNA. Three target sites in the deGFP expression cassette are shown. Circles represent the PAM sequence and straight lines the spacer sequence. Error bars represent the SD from three independent measurements triplicates. **(C)** Fold repression of protein expression produced by targeting gRNAs as compared with a non-targeting gRNA. **(D)** Microscopy images of double emulsions encapsulating cell-free extract supplemented with 1nM: pP70a_deGFP, 3nM: pJ23108_SpCas9 and 1nM pJ23119_gRNA 9 or 1nM pJ23119_gRNA NT. **(E)** Quantitative analysis of the average fluorescence intensity within the double emulsions encapsulating the gene circuit without a targeting gRNA (N = 66) showed a significantly higher value than for the double emulsions encapsulating the gene circuit with a targeting sgRNA9 (N = 57, p-value < 0.001).

CRISPR-Cas activity in double emulsions

Building upon the work of Marshall et. al, 2018 and J. Shin & V. Noireaux, 2010 (155,156) , we tested three different CRISPR nucleases for three different positions against the P70a_deGFP expression cassette: the Type II-A SpCas9 nuclease from *Streptococcus pyogenes* (12), the Type II-C ThermoCas9 nuclease from *Geobacillus thermodenitrificans* T12 (157), and the miniature Type V-F AsCas12f1 nuclease from *Acidobacillus sulfuroxidans* (158). All three nucleases were demonstrated to cause robust guided repression of deGFP expression, for all tested target positions (Supp. **Figure S6**).

After confirming robust CRISPR-Cas activity in TXTL, we proceeded to compartmentalize the cell-free expression reaction mix for the SpCas9 effector into DEs (**Figure 2**). As expected, when using a circuit in which Cas9 targeting results in repression of *gfp* expression (**Figure 2A, B**), we observed notably lower fluorescence in compartments engulfing a gRNA guide with a targeting spacer sequence than in compartments engulfing a non-targeting (NT) spacer sequence (**Figure 2C-E**). The time required for gene expression, ribonucleoprotein complex formation and eventually target interference explains the lag phase during the initial stages of the cell-free expression reactions, needed both for the production of GFP and its Cas9-based inhibition. This results in lower but detectable levels of fluorescence even in the case of targeting spacers (**Figure 2C-E**). Background-corrected fluorescence intensity revealed that the DEs containing gene circuit with non-targeting gRNA displayed a signal intensity of approximately 609 ± 111 (N = 66) (**Figure 2E**). In contrast, when targeting guide RNA (gRNA 9) was present, the fluorescence intensity was measured to be 339 ± 70 (N = 57). The two populations were found to be statistically significant (p-value < 0.001, see Materials and Methods for details).

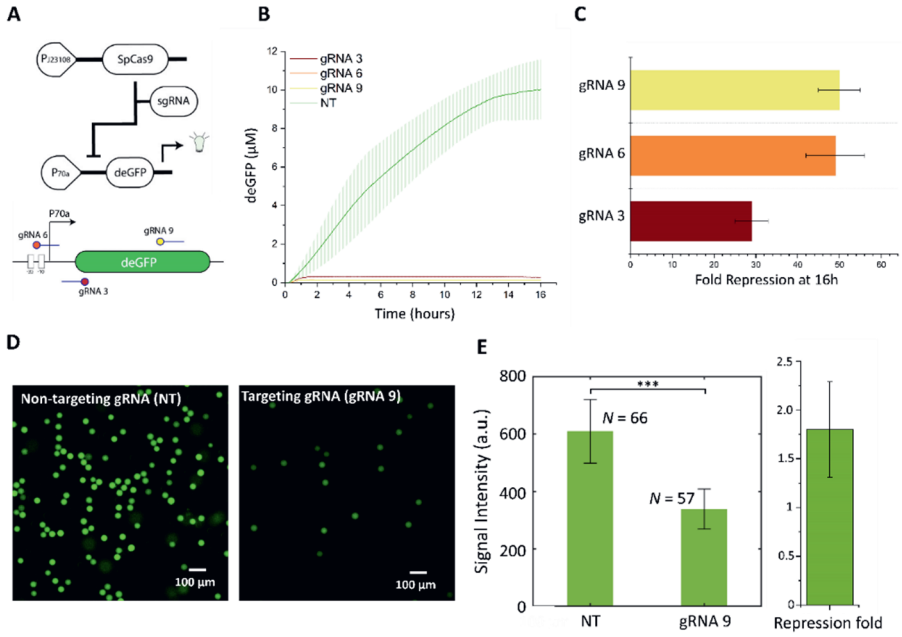


Figure 2. Guided regulation of protein expression by CRISPR-Cas systems in double emulsions. (A) Molecular gene circuit showing guided DNA-interference by CRISPR-Cas systems on a deGFP-encoding DNA template results in decreased fluorescence levels. **(B)** Kinetic measurements of deGFP(µM) concentration in presence of 1nM: pP70a_deGFP, 3nM: pJ23108_Cas and 1nM: pJ23119_gRNA. Three target sites in the deGFP expression cassette are shown. Circles represent the PAM sequence and straight lines the spacer sequence. Error bars represent the SD from three independent measurements triplicates. **(C)** Fold repression of protein expression produced by targeting gRNAs as compared with a non-targeting gRNA. **(D)** Microscopy images of double emulsions encapsulating cell-free extract supplemented with 1nM: pP70a_deGFP, 3nM: pJ23108_SpCas9 and 1nM pJ23119_gRNA 9 or 1nM pJ23119_gRNA NT. (**e**) Quantitative analysis of the average fluorescence intensity within the double emulsions encapsulating the gene circuit without a targeting gRNA (N = 66) showed a significantly higher value than for the double emulsions encapsulating the gene circuit with a targeting sgRNA9 (N = 57, p-value < 0.001).

Next, we sought to investigate the differentiation between a synthetic variant and a wild-type variant. We chose to compare the wild type SpCas9 (wtSpCas9), that requires a 5'-NGG PAM, with the "near PAM-less"

SPRYCas9 (159). Three protospacers, flanked by a 5'-YAC PAM motif, were tested with for wtSpCas9 and SPRYCas9 (**Figure 3A, B, Supp. Figure S7**). With the non-canonical PAM targets, the wtSpCas9 with a targeting spacer did not result in altered fluorescence (**Figure 3C)**. In contrast, SPRYCas9 successfully interfered with the non-canonical PAM targets, in all three cases, resulting in greatly reduced GFP production and fluorescence (**Figure 3C-E**). We then proceeded to create DEs encapsulating either wtSpCas9 or SPRYCas9 along with the non-canonical PAM targeting spacer-3 (**Figure 3D**). In agreement with our previous observations, compartments that encapsulated SPRYCas9-encoding plasmids demonstrated significantly lower fluorescence than those engulfing wtSpCas9 plasmid vectors (**Figure 3E**). The fluorescent intensities in the DEs under identical imaging conditions revealed a fluorescence signal intensity of SpCas9 around 490 ±91 (*N*= 197), whereas the signal intensity of the SpRYCas9 variant was 269 ± 75 (*N* = 359) (**Figure 3E**).

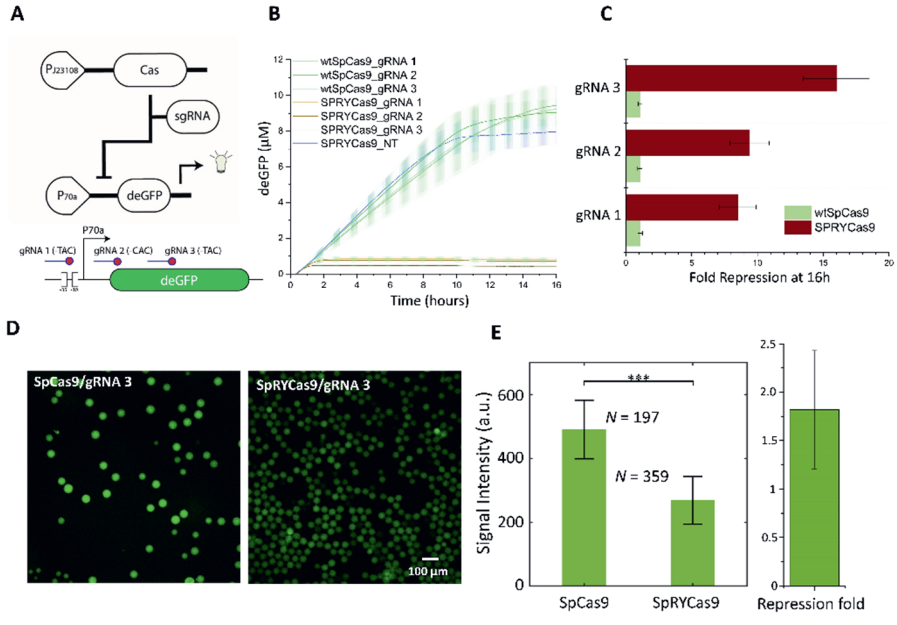


Figure 3. Cas9 variant discrimination in TXTL. (A) Molecular gene circuit showing guided DNA-interference by CRISPR-Cas systems on a deGFP-encoding DNA

template results in decreased fluorescence levels. **(B)** Kinetics measurements of deGFP concentration in presence of 1nM: pP70a_deGFP, 3nM: pJ23108_Cas (wtSpCas9 or SPRYCas9) and 1nM: pJ23119_gRNA. Three target sites for non-canonical PAMs in the deGFP expression cassette are shown. Circles represent the PAM sequence and straight lines the spacer sequence. Error bars represent the SD from three repetitions. **(C)** Fold repression of protein expression produced by targeting gRNAs as compared with a non-targeting gRNA. **(D)** Microscopy images of double emulsions encapsulating cell free extract supplemented with 1nM: pP70a_deGFP, 3nM: pJ23108_SpCas9 or pJ23108_SPRYCas9 and 1nM: pJ23119_sgRNA Sp3 or 1nM: pJ23119_gRNA NT. **(E)** Quantitative analysis of the average fluorescence intensity within the double emulsions encapsulating SpRYCas9 (N = 197) showed a significantly lower value than for the double emulsions encapsulating SpCas9(N = 359, p-value < 0.001).

Linking CRISPR-Cas activity to increased deGFP expression in TXTL

Guided interference versus the deGFP vectors results in repression of deGFP expression and consequently lower fluorescence levels. We resolved to prototype a repression cascade to indirectly connect CRISPR-Cas activity with increased protein expression. We exploited the guided nature of CRISPR-Cas systems and the negative regulation of the P70a promoter by the cI protein (**Figure 4A, Supp. Figure S7**). By guiding CRISPR-Cas interference to the cI expressing vectors, less cI-protein dimers are formed (inhibition of the inhibitor), resulting in enhanced *gfp* expression from the P70a promoter (**Figure 4B**). We tested the three Cas-effector variants (SpCas9, ThermoCas9, AsCas121f) to indirectly drive P70a_deGFP expression by repressing cI-ssrA expression. The ssrA tag was also fused with cI to shorten to its half-life (155) , minimizing the effect of cI protein that was expressed before efficient guided DNA-interference being able to take place. Guided repression of the cI-ssrA expression indeed resulted in higher fluorescence (higher deGFP concentration) for all three tested Cas-effectors when compared to a non-targeting gRNA (**Figure 4C**, **Supp. Figure S8**). We proceeded to compartmentalize the cell-free extract supplemented with SpCas9 and gRNA plasmid vectors. As expected,

compartments encapsulating a cI-targeting gRNA exhibited higher fluorescence than compartments encapsulating a non-targeting gRNA (**Figure 4D**) Microscopically measuring the fluorescence intensity of these DEs after 16 hours, under identical conditions, showed that DEs in presence of non-targeting gRNA showed an average fluorescent intensity of 105 ± 24 (*N* = 723) (**Figure 4E**). On the contrary, average fluorescence intensity of the droplets in presence of targeting gRNA 3 was 423 ± 119 (*N* = 921), a statistically significant increase.

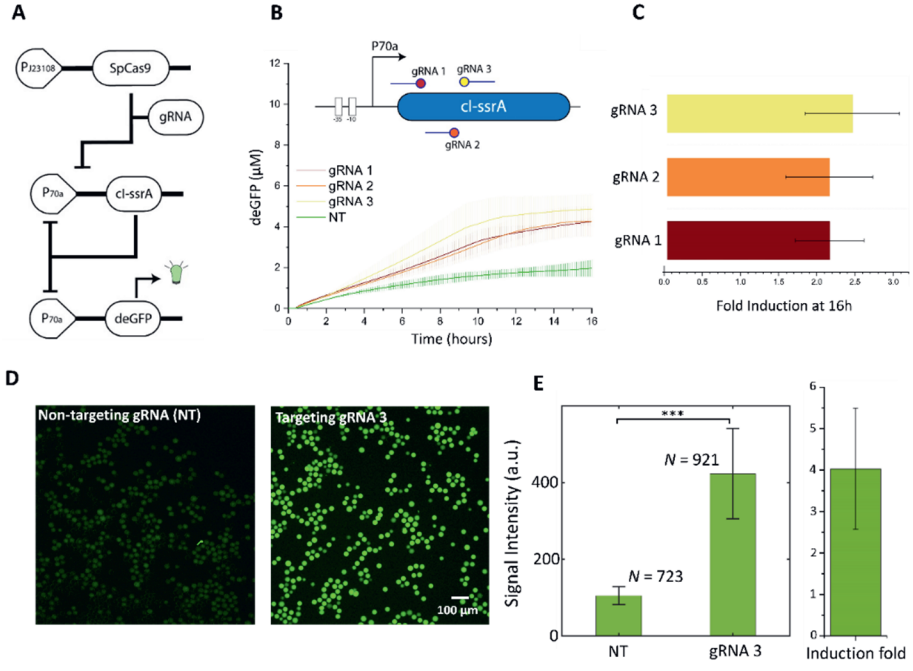


Figure 4. Indirect linkage of CRISPR-Cas activity with increased fluorescence. (A) Molecular gene circuit showing guided DNA-cleavage by CRISPR-Cas systems of the cI_ssrA prevents transcriptional repression of P70a promoter resulting in increased fluorescence levels**. (B)** Kinetics measurements of deGFP concentration in presence of 1nM: pP70a_deGFP, 1nM: pP70a_cI-ssrA, 3nM: pJ23108_Cas and 1nM: pJ23119_gRNA. Three target sites in the cI-ssrA expression cassette are shown. Circles represent the PAM sequence and straight lines the spacer sequence. Error bars represent the SD from three repetitions**. (C)** Fold induction of protein expression produced by targeting gRNAs as compared with a non-targeting gRNA **(D)** Microscopy photographs of double emulsions encapsulating cell free extract supplemented. **(E)** Quantitative analysis of the average fluorescence intensity

within double emulsions encapsulating the gene circuit with SpCas9 (N = 723) showed a significantly lower value than for the double emulsions encapsulating the gene circuit comprising targeting gRNA 3 (N = 921, p-value < 0.001).

Genotype enrichment through droplet fluorescence-activated sorting

In the previous sections, we showed that in each CRISPR-Cas gene circuit the genotype had a significant effect on the phenotype, as confirmed by both qualitative and quantitative analysis. However, an important remaining question is whether we can retrieve the genotype from the DE droplets with the desired phenotype. To test this, we used a fluorescence activated cell sorting (FACS). Since DEs are larger and more deformable than cells (160) substantial optimization of FACS settings was required (for detail, see Material and Methods). To show proof-of-principle retracing of the genotype from droplets with the desired phenotype, we generated two DE populations encapsulating cell-free extract supplemented with wtSpCas9 and either a plasmid encoding a non-targeting guide, or a plasmid encoding targeting the deGFP expression cassette (**Figure 2C**). Separate microfluidic chips for DE generation were used to prevent cross contamination of gRNAs. Both emulsions were incubated overnight for the cell-free expression to take place and afterwards were mixed in a 1:1 volume ratio and fed to FACS apparatus (**Figure 5A**). Size-based analysis of approximately 15,000 events (each event corresponds to a sample passing through the detector) using forward scattering width (FCS-W) and backward scattering area (BSC-A) (**Figure 5B**) showed that approximately 48% (*N*=7318) of the population was in the high FSC-W region (500 < FSC-W < 2000) and high BSC-A region (10^5 < BSC-A < 10^7). We observed a broad distribution of fluorescence intensity ranging from 10^3 to 10^5. However, two distinct peaks of higher (H) and lower (L) fluorescence could be gated and sorted accordingly (**Figure 5C**). Microscopic examination of the sorted DE populations confirmed higher fluorescence intensity of the H-population in comparison with the L-population (**Figure 5D**). Downstream fluorescence intensity analysis on recovered DEs gave a statistically significant (p-value

< 0.001) higher value for Gate-H (79 ± 21; *N* = 248) as compared to that for gate L (63 ± 17; *N*= 233) (**Figure 5E**).

The sorted L compartment populations and the mixed model population were then used for PCR amplification of the gRNA expression cassettes. After verification of correct amplicon size, the amplicons were analysed by Sanger sequencing. Subsequent alignment with the in-silico maps of the gRNA plasmid vectors revealed spacer genotype enrichment according to fluorescent intensity (**Figure 5F**).

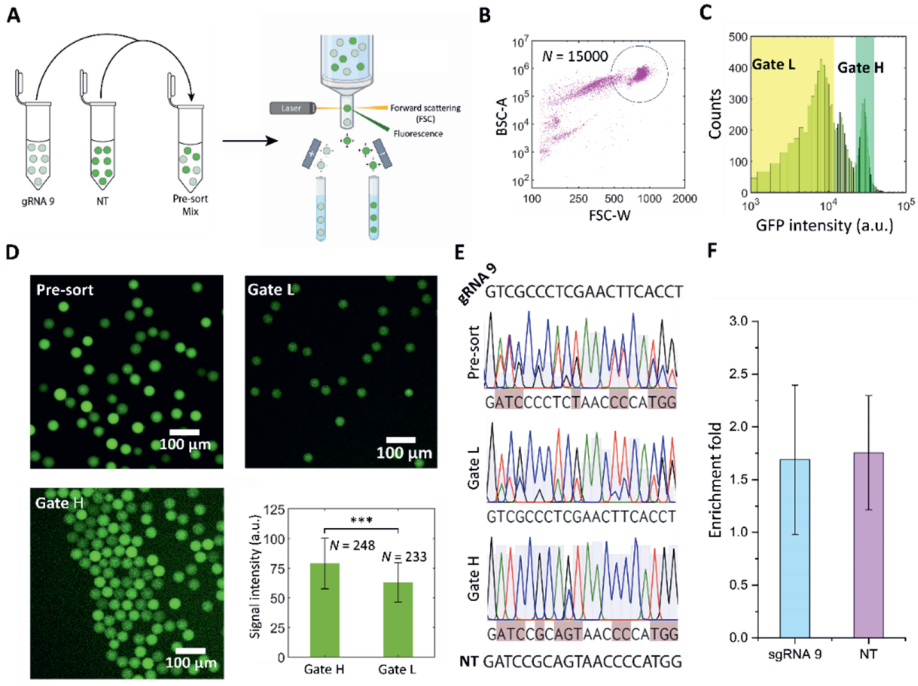


Figure 5. High throughput screening of CRISPR-Cas activity by FACS. (A) Schematic illustrating of sample preparation before fluorescence activated droplet sorting. **(B)** Size-based analysis of the mixed double emulsion population. The input sample comprised of two distinct double emulsions: one with SpCas9, GFP, and targeting gRNA, and second one with gene circuit SpCas9, GFP, and non-targeting gRNA. Approximately 48% (N = 7318) of the total population (N = 15000) exhibited high forward scattering (500 < FSC-W < 2000) and backward scattering (10^5 < BSC-A < 10^7). **(C)** Analysis of the entire population based on GFP-associated fluorescence intensity revealed two distinct peaks between 10^3 and 10^5. Gate-L was designated for the low fluorescence intensity population, while Gate-H was assigned to the high fluorescence intensity population. **(D)** Representative microscopic images of the pre-sorting droplet population and of the recovered double emulsions sorted for Gate-H and Gate-L. Right Bottom: Qualitative analysis of the average signal intensity

measured in the double emulsion droplets in Gate-H (N = 248) and Gate-L (N= 233), show a significant difference (p-value < 0.001). **(E)** Sanger sequencing read alignments versus the targeting gRNA 9. Pre-sort: Alignment of sequenced amplicon of the mixed, pre-sorted sample. Gate L: Alignment of sequenced amplicon of Gate L**-**sorted Gate H: Alignment of sequenced amplicon of Gate H**-**sorted sample. Red colour appears when the most prevalent peak mismatches with query sequence (gRNA 9). **(F)** Enrichment fold of the corresponding gRNA 9 and NT spacer sequences as predicted by the EditR software from the shown Sanger sequencing chromatographs

Conclusions

Here we described the development of an in vitro and high throughput screening system, tailored towards the optimization and characterization of CRISPR-Cas systems. We show the establishment of a phenotype screening that is based on fluorescence level alterations. We ensure linkage of phenotype-to-genotype through in vitro compartmentalization in highly monodisperse, double emulsion droplets, efficiently generated by on chip microfluidics. High-throughput screening is demonstrated by droplet sorting using common cell sorting equipment. Finally, we put all the pieces together by tracing back the expected gRNA genotype from a model, droplet sample according to the relative fluorescence of the artificial compartments.

To fully exploit its potential, the described approach, would be applied into large gene libraries. To ensure individual genotype-to-phenotype linkage among the library members, encapsulation of one library member per droplet would be desired. However, encapsulation of single gene copies suffers from low to undetectable phenotypic readout (65). The issue can be circumvented by clonal PCR amplification of a single gene per droplet, while remaining the droplet structure. Subsequently, cell-free expression will follow where TXTL extract is provided in each individual compartment either through droplet fusion or pico-injection (102,103). Alternatively, a single gene can be covalently attached to magnetic microbeads, that are encapsulated in droplets along with an appropriate PCR mix. Emulsion PCR amplification results in multiple copies of that gene attached to the bead.

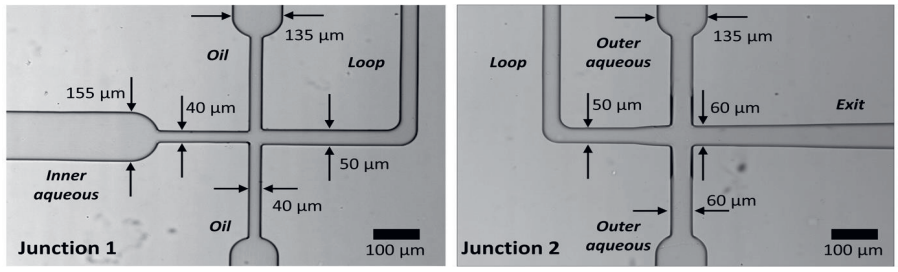
The beads will then be harvested and compartmentalized in droplets along with the TXTL mixture. As a proof of concept, we demonstrated that the clonal amplification of a reporter gene attached to microbeads leads to a detectable phenotypic readout (**Supp. Method S1, Supp. Figure S9**).

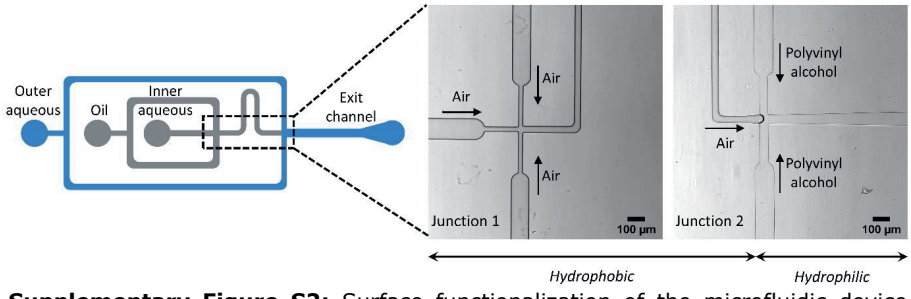
While further studies are still necessary for a fully-fledged method, it is anticipated that the described components of cell-free gene expression and single droplet sorting may serve as a solid basis for future adaptations

Supplementary Material

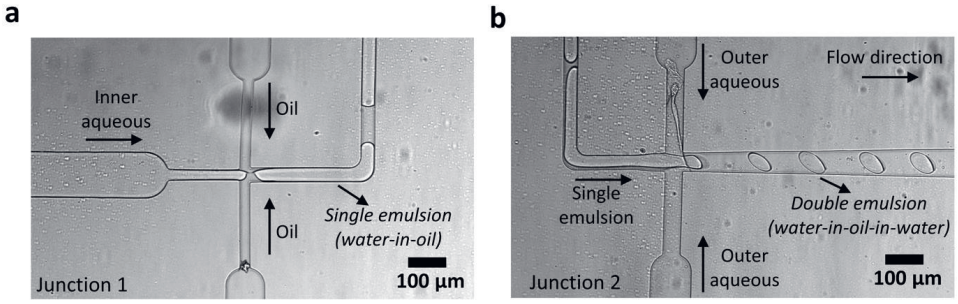
Supplementary Figures

Supplementary Figure S1: Dimensions of the two-junction microfluidic device used to produce double emulsions. Junction 1 consists of a horizontal inner aqueous channel (40 µm in width) intercepted by perpendicular oil channels (40 µm in width) resulting in a loop of slightly larger opening of 50 µm width designed to hold single emulsions. Junction 2 consists of single emulsion- containing horizontal loop (50 µm) intercepted by perpendicular outer aqueous channels (60 µm) leading to cone shaped exit channel. All the channels are approximately 20 µm in height.

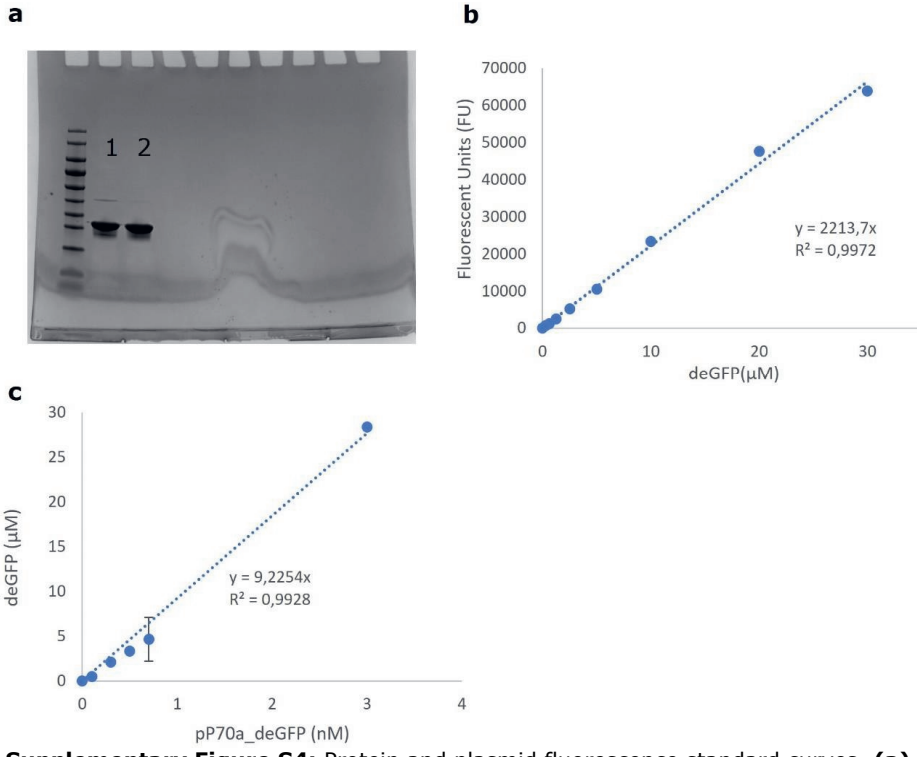




Supplementary Figure S2: Surface functionalization of the microfluidic device. Schematic depicting the hydrophobic channels (grey) and hydrophilic channel (blue) post-surface functionalization. To make the channels after the second junction hydrophilic, 5% w/v polyvinyl alcohol (PVA) solution is flown through the outer aqueous channel towards the exit channel. Positive air pressure is maintained in the inner aqueous and oil channels to prevent the back flow of PVA in the loop.

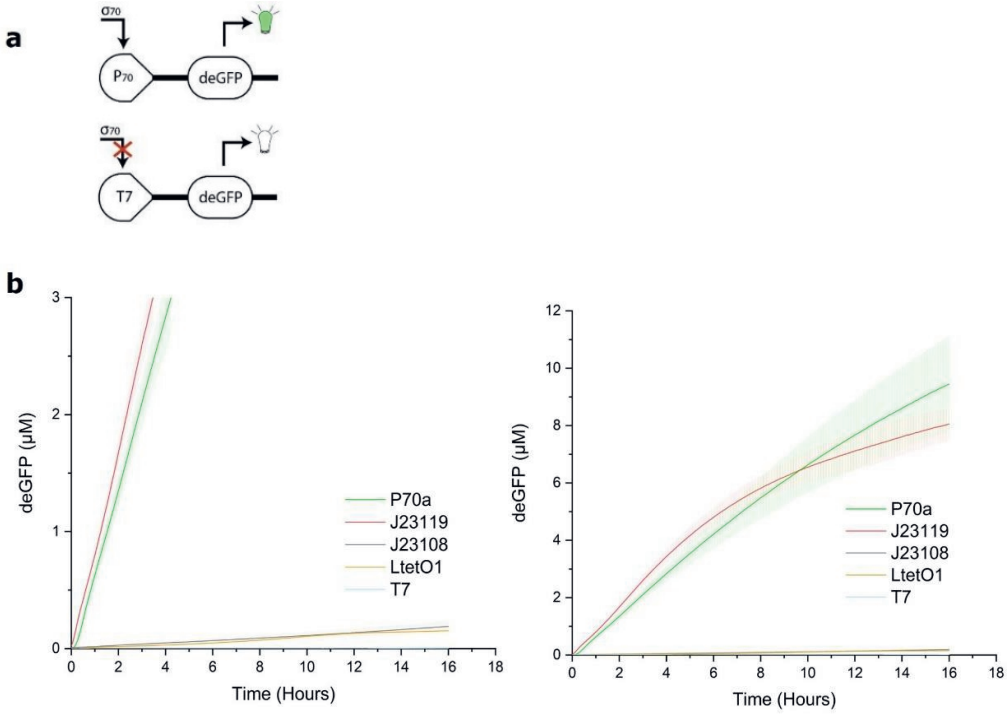


Supplementary Figure S3: Microfluidic production of double emulsions. **(a)** Production of single emulsions at Junction 1. The formed single emulsions travel to the second junction via the loop. **(b)** Production of double emulsions by pinching-off the single emulsions by the outer aqueous phase.

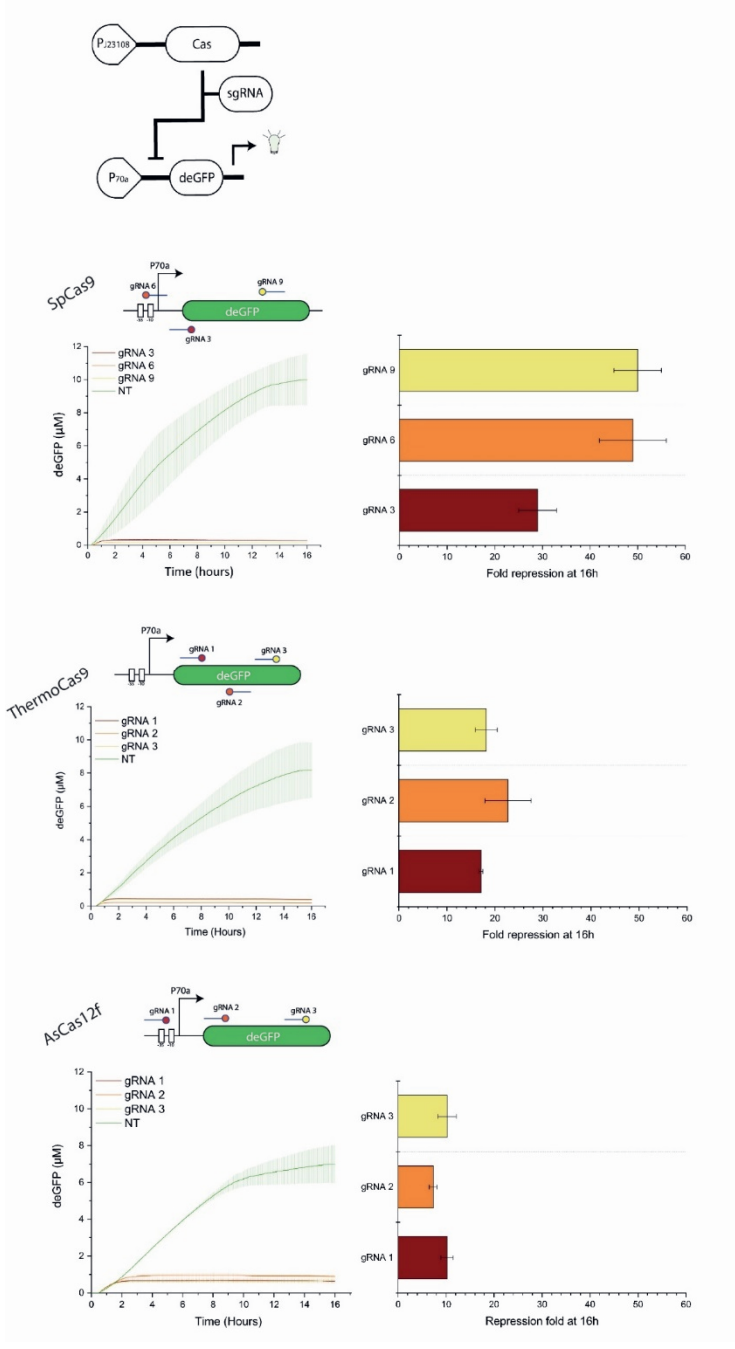


Supplementary Figure S4: Protein and plasmid fluorescence standard curves. **(a)** SDS-PAGE gel loaded with 9.5 µg protein enriched for deGFP-his. Fraction 2 was

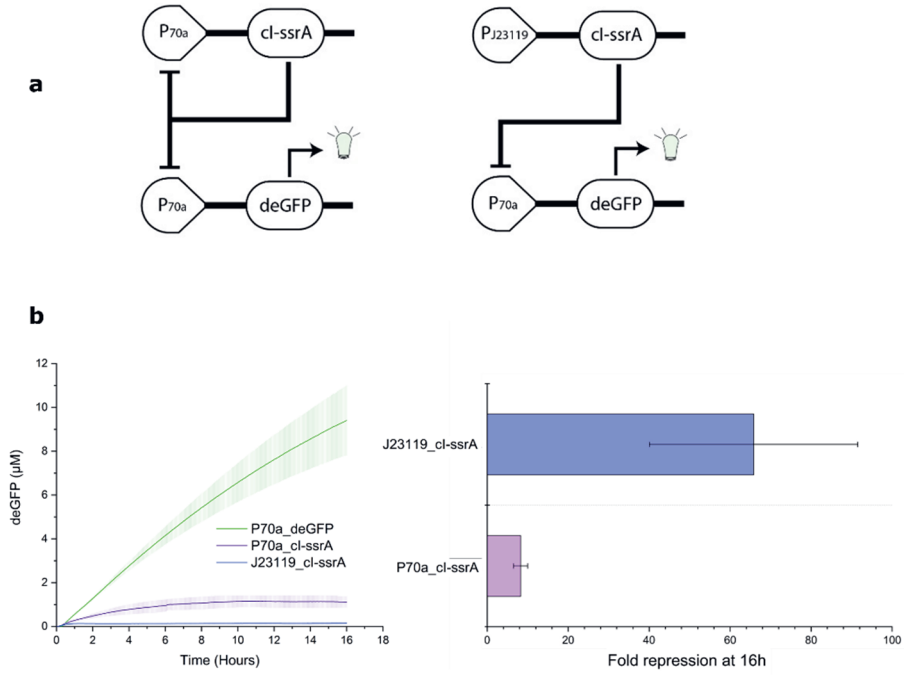
chosen for regression analysis. **(b)** deGFP-His standard curve based on fraction 2. **(c)** Plasmid DNA for P70a_deGFP expression cassette standard curve.



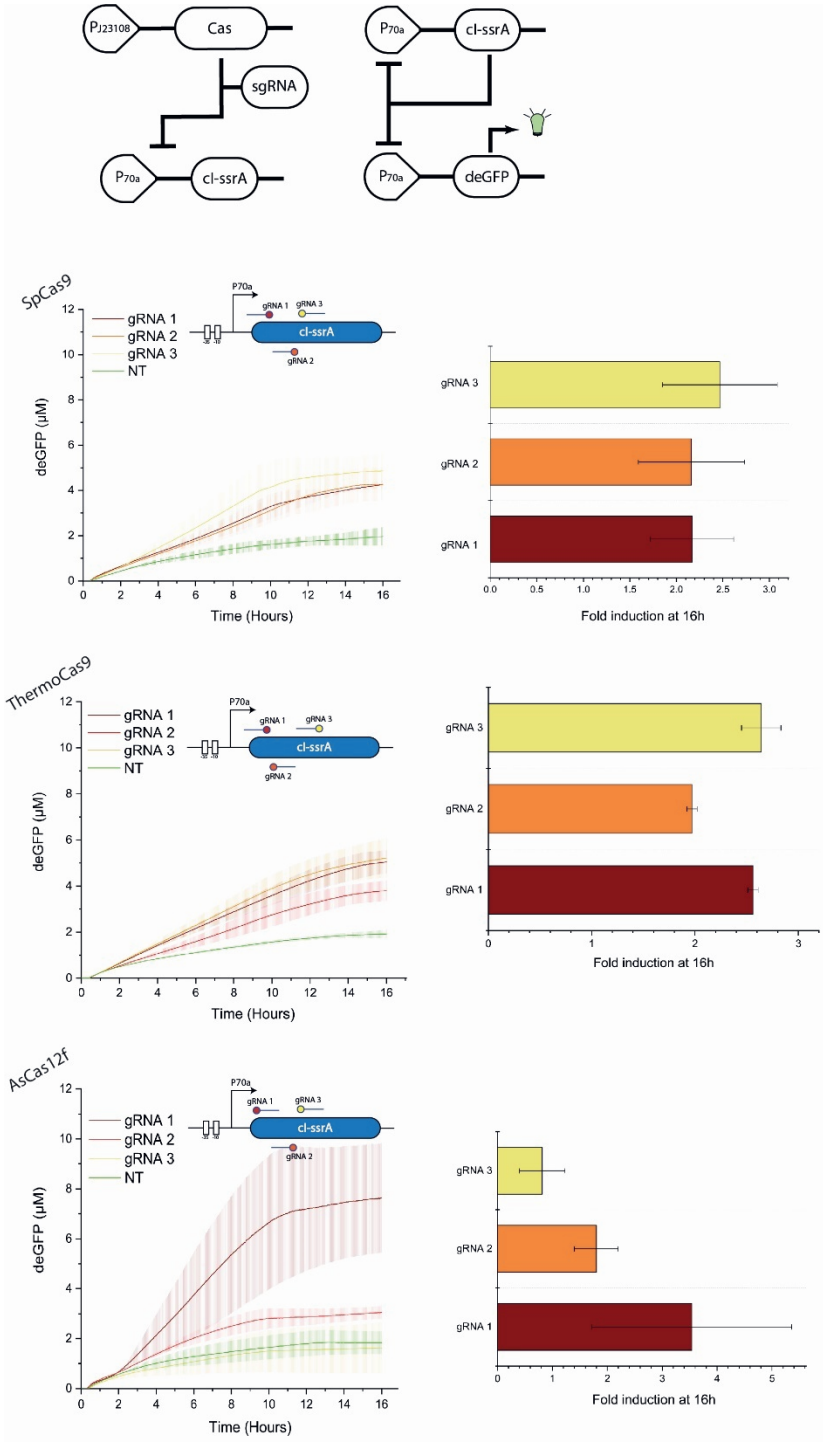
Supplementary Figure S5: Varying promoter strength in TXTL. **(a)** Gene transcription in TXTL is driven only by the E. coli endogenous σ70 factor. **(b)** Kinetic measurements of deGFP concentration in presence of 1 nM: pP70_deGFP or 1 nM: pT7_deGFP as negative control. Error bars represent the SD from three repetitions.


P123108
Cas
sgRNA
P70a
deGFP
SpCas9
P70a
gRNA 6
gRNA 9
deGFP
gRNA 3
gRNA 3
gRNA 6
gRNA 9
NT
deGFP (μM)
Time (hours)
gRNA 9
gRNA 6
gRNA 3
Fold repression at 16h
ThermoCas9
P70a
gRNA 1
gRNA 3
deGFP
gRNA 2
gRNA 1
gRNA 2
gRNA 3
NT
deGFP (μM)
Time (Hours)
gRNA 3
gRNA 2
gRNA 1
Fold repression at 16h
AsCas12f
P70a
gRNA 1
gRNA 2
gRNA 3
deGFP
gRNA 1
gRNA 2
gRNA 3
NT
deGFP (μM)
Time (Hours)
gRNA 3
gRNA 2
gRNA 1
Repression fold at 16h

Supplementary Figure S6: Guided regulation of protein expression by CRISPR-Cas systems in TXTL. (a) Kinetic measurements of deGFP (µM) concentration in presence of 1 nM: pP70a_deGFP, 3 nM: pJ23108_Cas and 1 nM: pJ23119_gRNA. Target sites for three different CRISPR-Cas systems in the deGFP expression cassette were designed. Circles represent the PAM sequence and straight lines the spacer sequence. Error bars represent the SD from three repetitions.



Supplementary Figure S7: Promoter negative regulation in TXTL. (a) cI protein dimers bind to the operator sites of the P70a promoter and obstruct interaction with σ70 factors thus preventing transcription. **(b)** Kinetic measurements of deGFP (µM) concentration in presence of 1 nM: pP70a_deGFP and 1 nM: pP70a_cI-ssrA or 1 nM: pJ23119_cI-ssrA. Error bars represent the SD from three repetitions.


P23108
Cas
sgRNA
P70a
cl-ssrA
P70a
cl-ssrA
P70a
deGFP
SpCas9
gRNA 1
gRNA 2
gRNA 3
NT
deGFP (μM)
Time (Hours)
Fold induction at 16h
ThermoCas9
AsCas12f

Supplementary Figure S8: Repression cascade of protein expression partly based on CRISPR-Cas systems in TXTL. Guided DNA cleavage by CRISPR-Cas systems of the cI_ssrA prevents transcriptional repression of P70a promoter. Target sites for three different CRISPR-Cas systems in the cI-ssrA expression cassette. Circles represent the PAM sequence and straight lines the spacer sequence. Kinetics measurements of deGFP concentration in presence of 1nM: pP70a_deGFP, 1nM: pP70a_cI-ssrA, 3nM: pJ23108_Cas and 1nM: pJ23119_sgRNA. Error bars represent the SD from three repetitions.

Supplementary Methods

***Supplementary Method S1*.** Compartmentalized gene amplification and display of gene copies on iron microbeads.

We previously demonstrated high throughput screening for guided DNA- interference by Cas9 nuclease from a model population of emulsion containing compartments encapsulating a targeting anti-deGFP gRNA coding plasmid or a non-targeting gRNA coding plasmid (Figure 5a). The theoretical λ of gRNA plasmids vectors supplied in 1nM would be ~16,667 assuming a compartment 37.5 µm in diameter (Figure 1c).

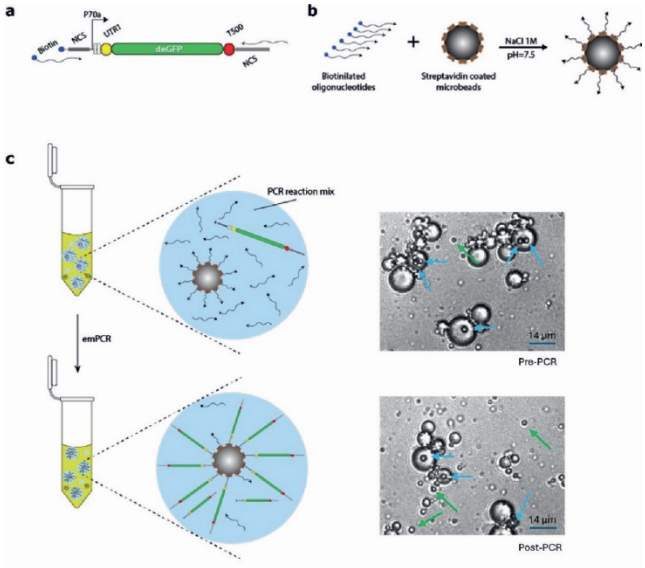
By assuming a compartment 15 µm in diameter, we theoretically estimated the corresponding concentration of a single pJ23108_SpCas9 plasmid copy in such a compartment (λ=1). We proceeded to measure the fluorescence level alteration for the corresponding pJ23108_SpCas9 concentration in 96 multi well plates without observing any significant changes (data not shown). Indeed, low to undetectable levels of protein expression by single gene copies have been reported before (65).

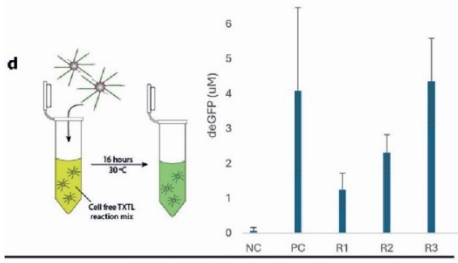
Alternatively, clonal amplification while retaining phenotype-to-genotype linkage can be achieved through microbead display (101,113,161). Single gene copies are compartmentalized into droplets along with magnetic microbeads in a PCR mixture. During gene replication multiple copies will be attached to the microbead through biotin-streptavidin affinity, thus each microbead will represent a gene clonal population. To test its feasibility, we clonally amplified the P70a_deGFP expression cassette in a compartmentalized fashion while retaining genotype to phenotype linkage through microbead display of the amplified nucleic acids. Magnetic microbeads were coated with a forward priming oligonucleotide and they were compartmentalized along P70a_deGFP expression cassette in PCR reaction

We assumed a compartment with diameter of 15 µm and we calculated the theoretical droplet population corresponding to our usual cell-free expression reaction volume (12 µl). Based on the theoretical droplet population we used increasing bead/droplet ratios (R) (R_1=1, R_2=2, R_3=3). The beads were coated with biotinylated, forward-priming, oligonucleotides (Supp Figure 16a, b). Then, they were compartmentalized along with a linear, biotinylated deGFP- encoding dsDNA template (λ_{deGFP}=2.5-25) and PCR reagents into water-in-oil emulsions (**Supp. Figure 10c**). After appropriate

thermocycling, emulsions were broken, and the beads were collected. After thorough washing, the microbeads were incubated in 12 µl cell-free extract and incubated overnight. deGFP concentration was measured and efficient protein expression was observed (Supp. Figure 16d). Making use of a plasmid copy/fluorescence intensity standard curve **(Supp. Figure 4c**) we estimate ~100-150 gene copies per microbead.

The efficiency of amplifying bead displayed gene copies is reversely correlated with the gene copy length and explains the extensive use of microbead display for short oligonucleotides. However, deGFP is an average size protein (225 aa) and ~100-150 copies per microbead are shown to suffice for efficient protein expression. Miniature CRISPR-Cas systems (422-700 aa) have lately gained a lot of attention due to their potential as small and efficient genome editing tools (162)





	NC	PC	R1	R2	R3
biotin-deGFP	+	-	+	+	+
biotin-Fw Oligo	+	-	+	+	+
Rv Oligo	-	-	+	+	+
biotin-Fw Oligo Coated beads	+++	-	+	++	+++
biotin-deGFP coated beads	-	+++	-	-	-
emPCR	+	-	+	+	+

Supplementary Figure S10: Gene copy amplification on magnetic microbeads. **(a)** Schematic representation of the deGFP expression cassette and PCR priming oligonucleotides. **(b)** Coating of iron microbeads with biotinylated forward priming oligos. **(c)** Compartmentalization of oligo-coated microbeads along with PCR mixture in water in oil droplets. **(d)** Incubation of post-PCR microbeads in TXTL and measurements of deGFP protein concentration.

Chapter 4

Catalytically enhanced Cas9 through protein directed evolution

Evgenios Bouzetos[*], Christian Südfeld[*], Roel Boonen[*], Mitchell Roth, Despoina Trasanidou, Kobus Youri, Raymond H.J. Staals, Hong Li, and John van der Oost[**]

* authors contributed equally
** corresponding authors

Manuscript in preparation

Abstract

ThermoCas9 is a robust CRISPR-associated nuclease that has previously been discovered in a thermophilic bacterium *Geobacillus thermodenitrificans*. Because of its elevated temperature optimum (55-60 °C) it has initially been used for genome editing of a thermophilic bacteria. The observation that ThermoCas9 is able to specifically cleave dsDNA at a relatively brought temperature range, also has allowed for successful genome editing of mesophilic bacteria as well as human cells. However, to improve the efficiency of genome editing at lower temperatures, we here describe a random engineering approach to obtain ThermoCas9 variants with better editing performance in the range of 20-37 °C. The directed evolution approach based on an error-prone PCR library, has resulted in different variants with elevated cleavage efficiencies *in vivo* (in *E. coli*), which at least in some cases could be confirmed by *in vitro* analysis. These results provide a strong basis for further optimization of the ThermoCas9 performance at lower temperatures.

Introduction

CRISPR-Cas9 systems are prokaryotic adaptive immune systems that have been repurposed to become programmable genome editing tools. These systems compose a Cas9 nuclease, a CRISPR RNA (crRNA) and a trans-activating RNA (tracrRNA). The tracrRNA and crRNA can be genetically fused to form a single guide RNA (sgRNA) simplifying application. A part of the crRNA generally contains a 20-30 nucleotide long complementary region called spacer sequence that base pairs with the DNA-target called protospacer. The Cas9 nuclease is a multidomain enzyme consisting of: i) the recognition lobe (REC) that interacts with the sgRNA and the RNA-DNA heteroduplex, ii) the PAM-interacting domain (PID) and two nucleolytic domains (RuvC, HNH). Prerequisite for target cleavage is the recognition of a short sequence motif located in the 3'-end of the target sequence called protospacer adjacent motif (PAM). Upon PAM recognition the DNA duplex is locally melted and the spacer region of the sgRNA base pairs with the target strand to form an R-loop structure. Upon R-loop formation the RuvC and HNH domains will cleave the non-target strand and target strand, respectively.

ThermoCas9 is a type II-C enzyme found in the thermophilic bacterium *Geobacillus thermodenitrificans* T12. With a size of 1082 amino acids, the protein is smaller than its type II-A homologues, such as *Streptococcus pyogenes* Cas9 (SpCas9) (31,157). It recognizes a 5'-N_4CRAA PAM motif and cleaves optimally when using a 23 nt long spacer flanked by a 5'-N_4CCAA PAM. ThermoCas9 is a moderately thermophilic enzyme with demonstrated catalytic activity up to 65 ^{o}C. ThermoCas9 has proven as an effective genome editing tool for prokaryotic cells and has successfully been used for editing of mammalian cells (163,164). However, its high temperature optimum is limiting for genome engineering applications in mesophilic organisms, for example crop engineering.

Here, we report a directed evolution campaign using bacterial systems to generate ThermoCas9 variants with increased activity at lower

temperatures. We constructed randomized mutagenesis libraries and selected mutants for improved activity. The used selection scheme (Figure 1A) is based on the *ccdB/ccdA* toxin-antitoxin system which is a plasmid addiction system (165,166) that was initially adapted for the directed evolution of homing endonucleases (138). To this end, the *ccdA* antitoxin gene was omitted from the system and the *ccdB* toxin gene was placed under the tight regulation of the arabinose-inducible pBAD promoter. To facilitate arabinose induction, the system includes the galactosidase permease *lacY*. Arabinose induction causes *ccdB* expression and cell death. Growth phenotype can be rescued by guided cleavage of the *ccdB* encoding plasmid vector.

This selection system has been recently adapted for screening CRISPR-Cas effector variants with altered properties such as PAM-recognition, increased specificity and/or activity (132–134). Here we describe efforts to use the *ccdB* screening in *E. coli* to obtain variants of ThermoCas9 with improved performance at lower temperatures.

Materials and Methods

Propagation and Assembly of pTarget vectors

p11-LacY-wtx1 was generated by Zhao and co-workers (138). The *ccdB*-resistant strain "*E. coli: ccdB-survival*" was cured from p11-lac-wtx1 plasmid through two consecutive inoculations in LB-media supplemented with 10 mM Ara and 1 mM IPTG in absence of antibiotic selection (Supp. figure 1 top left). *ccdB* cloning vectors were propagated and/or assembled in the cured *Escherichia coli ccdB*-resistant strain using LB-media supplemented with 0.4% Glucose (Glu) and the corresponding antibiotic marker (Ampicillin 75µg/ml). ThermoCas9 protospacers were assembled with the p11-lacy-wtx1 vector by XbaI/SphI restriction ligation using the corresponding pre-annealed oligos to create pTarget plasmids (Supp. Table

S1). Electrocompetent *E. coli ccdB*-resistant cells were prepared as described previously (167).

Preparation of E. coli selection strain

Electrocompetent *E. coli* BW25141 cells were electroporated with 1ng pTarget. Cells were recovered in 250µl SOC media [SOB (0.5% yeast extract, 2%(w/v) tryptone, 10mM NaCl, 2.5mM KCl, 20mM MgSO4), 0.4% Glu] for 1 hour at 37 ^{o}C. Equal volumes of the recovery media were plated into non-selective (-Ara: 0.4% Glu, Amp75) and selective media (+Ara: 10mM Ara, 1mM IPTG, Amp75) to verify *ccdB* sensitivity (Supp. Figure 1 top right). After confirming absence of surviving colonies in presence of arabinose, single colonies from the -Ara plate were streaked to +Ara and -Ara agar plates to isolate *ccdB*-sensitive clonal populations. The next day, a *ccdB*-sensitive population was resuspended in 300µl SOC media. 250µl were used to inoculate 250 ml of SOC media and cells were grown at 37^{o}C until OD_{600}=~0.4. Cells were chilled on ice and a sample was plated on -Ara and +Ara agar plates to verify *ccdB*-sensitivity as before (Supp. figure 1 bottom left). Finally, cells were made electro-competent as described previously (167).

ThermoCas9 gene random mutagenesis

Error-prone PCR was performed on the whole *wt-thermoCas9* gene using a random mutagenesis kit (Agilent #200550) according to manufacturer's instructions for an error rate of 1.5 mutations/kbp. The corresponding primers included overhangs for the following Gibson Assembly (Supp. Table S1). The PCR products were column-purified and DpnI-digested to remove the plasmid template. Finally, PCR-products were purified using a gel DNA recovery kit according to manufacturer's instructions (Zymo Research #D4002). The purified linear fragments were used for plasmid assembly through Gibson Assembly. The assembled vector library was transformed

into One ShotTM TOP10 electrocomp™ *E. coli* cells. Cells were recovered in SOC medium for 1 hour at 37 °C and then plated on 245mm^2 agar plates (supplemented with Chloramphenicol at 12.5 µg/ml). After overnight selection at 37 °C, a total of ~8.7 X 10^6 colonies was obtained. 20 individual clones were picked for sequencing before pooling all remaining colonies. We observed high ratio of failed assemblies, introduction of stop codons and deletion in the *thermoCas9* gene, that would result into non-functional gene variations.

Selection of ThermoCas9 mutants by CcdB toxin rescue

To filter out no functional gene variations, *E. coli* BW25141: pTarget_CCAA_23 was electroporated with 100ng of *thermoCas9* random mutagenesis library. Cells were recovered in SOC media for 1hour at 37°C. Recovery media was plated on 245mm^2 selective media (+Ara: 10mM Ara, 1mM IPTG, Cam15) and incubated at 37°C. After 16hours, colonies growing on selective media were collected in 20ml LB medium and the plasmid library was extracted using a midiprep kit (Zymo Research #D4200) according to manufacturer's instructions. *E. coli* BW25141: pTarget_CAAA_23 was electroporated with 100ng of the plasmid library. Equal volumes of recovery media were plated on -Ara (0.2% Glu, Cam15) and +Ara media (10mM Ara, 1mM IPTG, Cam15) and incubated overnight at 37°C. Colonies were manually counted and the survival frequency was calculated as follows: Survival Frequency (%) = (Number of colonies in +Ara media/ Number of colonies in -Ara media) *100. The process was repeated for two consecutive rounds of selection and plasmid purification in *E. coli* BW25141: pTarget_CAAA_23.

ThermoCas9 protein purification

ThermoCas9 variants were propagated in pET28b vectors and transfected to *E. coli* Nico21(DE3) cells for recombinant protein production. *E. coli*

Nico21 (DE3). 15ml of overnight culture were used to inoculate 1.5L of LB medium. Cells were incubated at 37°C until OD_{600}= ~0.6. The cultures were cooled on ice for 30minutes and *ThermoCas9* expression was induced by addition of IPTG at 0.5mM. The culture was incubated at 18°C overnight, harvested, and the cell pellet was stored at -20°C until further processing.

The cell pellet was resuspended in 20ml lysis buffer (0.5M NaCl, 50mM Na_2PO_4, 5mM imidazole) supplemented with one tablet protease inhibitor (2 complete™, mini, EDTA–free Protease Inhibitor Cocktail tablet [Roche]). Cell lysis was performed by using a Sonoplus, Bandelin sonicator (Amplitude 30%, pulse 1 ec on, 2sec off, for 15-20 min) and the cell debris was removed by centrifugation (30,000 x g, 45min) (Sorvall LYNX 4000 Superspeed Centrifuge, Thermo Scientific). The supernatant was filtered (0.45 μm filter; Mdi membrane technologies). NiNTA columns were calibrated with 1.5ml of lysis buffer before application of the cell-free extract and washing with 2.5ml of wash buffer (lysis buffer supplemented with 1mM β-mercaptoethanol). Purified protein was eluted with 2ml elution buffer (50mM NaH_2PO_4, 0.5M NaCl, 0.5M imidazole, pH=8.0). Elution fractions were analyzed by SDS-PAGE electrophoresis and ThermoCas9 protein was concentrated using a 15mL 100,000 MWCO PES centrifugal filter (Sartorius stedim Lab Technology Products) and stored in storage buffer (0.5M NaCl, 50mM Na_2PO_4, 2mM DTT, 5% glycerol) at -80 °C.

In vitro expression and purification of sgRNA

The DNA template for *in vitro* transcription of sgRNA was prepared by annealing two complementary single strand oligonucleotides (Supp. Table S1). The dsDNA was then amplified by PCR and concentrated by ethanol precipitation (168) . 100 ng of DNA was then used as template for sgRNA synthesis using a HiScribeTM T7 Quick High Yield RNA Synthesis Kit (NEB#E2050S) according to the manufacturer's instructions. The *in vitro* transcription reaction was carried out at 37 °C for 4 hours, followed by DNase treatment at 37 °C for 15 minutes.

The sgRNA was resolved in an 8% urea-PAGE gel. The corresponding sgRNA bands were excised and left to dissolve in RNA elution buffer (0.5M CH_3COONA, 10mM $MgCl_2$, 1mM EDTA, 0.1% SDS) overnight at room temperature. The solution was centrifuged (4,700 rpm, 5 min, 4°C) and the supernatant was filtered (0.22µm; Mdi membrane technologies), and the eluted RNA was concentrated (3,000 x g, 20 min) to a final volume of 2 ml in a 10K centrifugal filter (Microcon). The RNA was aliquoted to 500µL fractions and precipitated by addition of 50µl of 2M sodium acetate (pH=5.2) and 1.5mL chilled 100% ethanol, followed by incubation at −20°C for 1 h. After centrifugation (14,000rpm, 40 min, 4°C), the pellet was washed with 200µL chilled 70% ethanol. The RNA pellet was air-dried and dissolved in 100µL MQ. The concentration and the purity of the sgRNA were determined using Nanodrop (NanoDropTM 1000, Thermo Fisher Scientific) and Qubit RNA BR Assay Kit (Thermo Fisher Scientific). Finally, the purity and integrity of the sgRNA was confirmed by gel electrophoresis using an 8% urea-PAGE gel (Supp. Figure 2A).

In vitro cleavage assays

In vitro cleavage assays (IVCAs) were performed with purified ThermoCas9 variants and the *in vitro*-transcribed sgRNA and DNA substrate. Linear DNA substrate was prepared by PCR amplification (Supp. Table 1) and purified (Zymo Research #D4004). ThermoCas9 protein concentration was adjusted according to the band intensities observed by SDS-PAGE gel electrophoresis (Supp. Figure 2B). IVCA reactions contained 160nM of ThermoCas9 protein, 150nM sgRNA and 4nM of substrate DNA in NEB3.1 buffer (NEB #B6003S). The reactions were stopped by adding 6x blue loading dye (NEB #B7021S), and reaction products were analyzed by electrophoresis using 1% agarose gels and SYBR safe (Thermo Fisher Scientific #S33102). DNA band intensities were quantified using ImageJ software. The cleaved fraction (Fc) was calculated and fitted using a non-linear regression model (Equation 1)

with the drc package (of R Statistical Software (v4.3.3; R Core Team 2021) (169).

Equation 1: $$y = \frac{a}{1+e^{-b(x-c)}}$$

Where y: cleaved fraction (Fc) at temperature x, a: right horizontal asymptote, b: growth rate, c: mid-point

Results & Discussion

In vivo selection approach

To enrich for ThermoCas9 mutants with enhanced catalytic activity, we based our selection scheme on the previously described *ccdB* selection system (**Figure 1A**). First, we verified cell-toxicity from the *ccdB* gene by transforming CcdB-sensitive *Escherichia coli*:BW25141 cells with the p11-lacy-wtx1 plasmid (**Supp. Figure 1 top left**). To allow removal of the plasmid by ThermoCas9, we modified this plasmid by adding a 23nt long protospacer sequence "PS1", (**Supp. Table S1**), flanked on the 3' side by a 5'-N_4CCAA-3' PAM motif (pTarget_CCAA_23). To verify that ThermoCas9 could remove this modified plasmid and rescue bacterial growth, *E. coli* BW25141 cells harboring pTarget_CCAA_23 were transformed with pCRISPR vectors encoding for *wt-thermoCas9* and a sgRNA targeting PS1. Guided DNA cleavage of pTarget_CCAA_23 by wtThermoCas9 rescued growth under *ccdB*-induction conditions for three temperature conditions (**Figure 1B**). To enable selection of ThermoCas9 mutants with improved activity, we aimed to create more stringent selection conditions that would not allow wtThermoCas9 to rescue growth. To this end, we constructed a series of pTarget plasmids that carried truncated versions of PS1 in the PAM distal portion. Furthermore, we created pTarget plasmids that carried full length PS1 but flanked by suboptimal PAM motifs. wtThermoCas9 failed to

rescue growth when the protospacer was truncated at 17nt length, or when it was flanked by a suboptimal PAM-motif (**Figure 1C, D**).

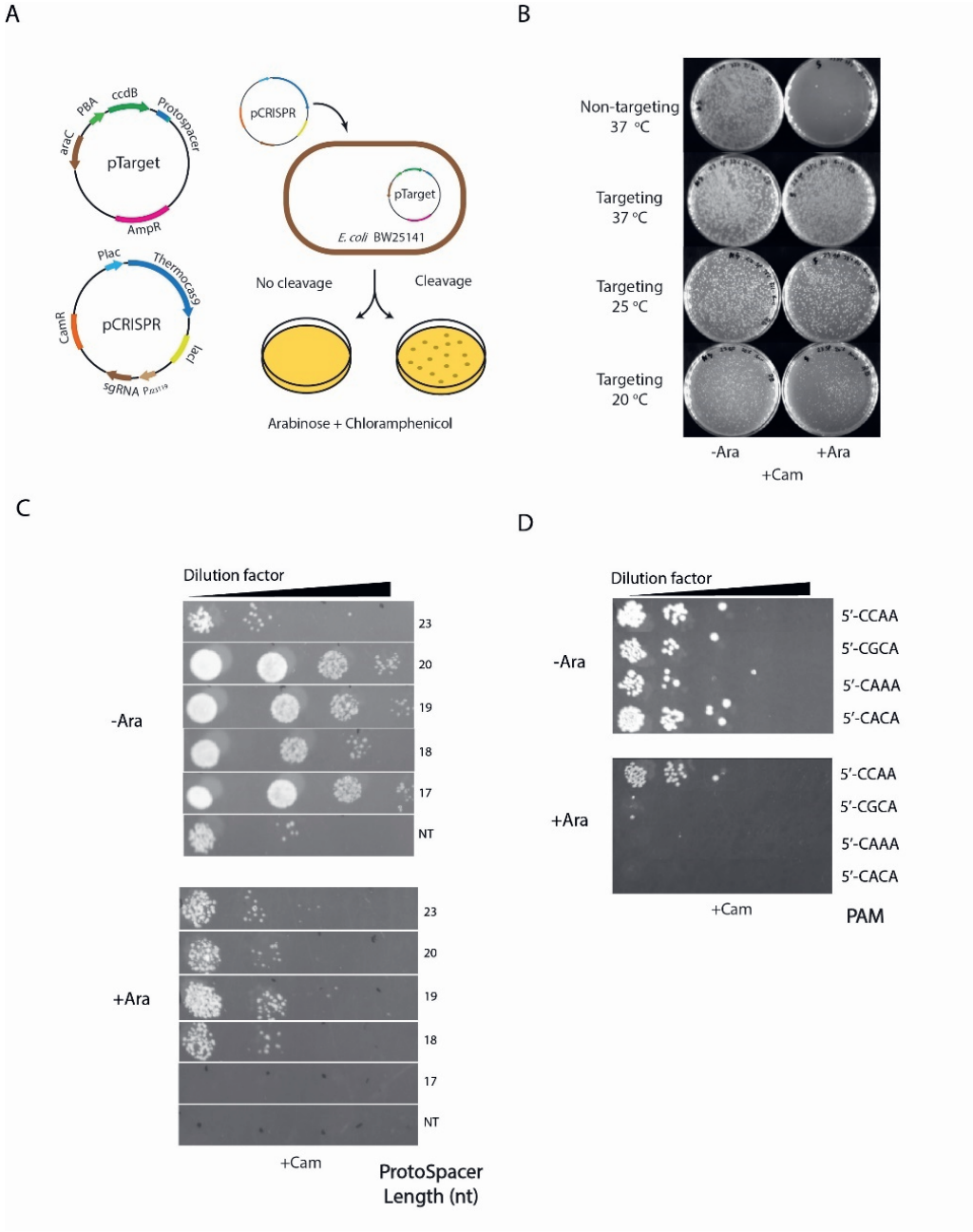


Figure 5. CcdB-based selection conditions for the enrichment of improved ThermoCas9 variants. (A) Schematic representation of positive selection assay. **(B)** *E. coli* Bw25141: pTarget_CCAA_23 transformed with *wtThermoCas9* encoding pCRISPR vector. Transformed cells were recovered and grown in 3 different temperatures (37 °C, 25 °C, 20 °C). **(C)** *E. coli* Bw25141: pTarget_CCAA carrying truncated PS1 versions in the PAM distal portion (20-17 nt) was transformed with *wtThermoCas9* encoding pCRISPR vector and grown at 37 °C. NT: Non-targeting spacer **(D)** *E. coli* Bw25141: pTarget carrying full length PS1 (23nt) but suboptimal PAM motifs transformed with pCRISPR.

Based on these results, we chose to perform the selection of interesting variants from the ThermoCas9 library at 37 °C, using the full length protospacer flanked by the suboptimal 5'-N_4CAAA PAM (pTarget_CAAA_23).

Directed Evolution of ThermoCas9

To unbiasedly examine mutations that could alter the DNA-cleaving properties, we performed error prone PCR over the entire *thermoCas9* gene using a low mutation rate (1.5/kbp) (**Figure 2A**). Roughly 20 clones of the freshly assembled gene library were sequenced to verify correct plasmid assembly and successful mutagenesis. We observed several cases of undesired introduction of STOP codons or indels (data not shown). To remove inactive gene variants, we performed an initial selection cycle using the full length protospacer flanked by the optimal 5'-N_4CCAA PAM. Sequencing of 20 clones from the thus-selected library (LibraryCas9) revealed successful curation of the library from gene variants with STOP codons or indels (data not shown).

We then selected for variants with improved catalytic activity from the curated gene library (LibraryCas9) using variants carrying a full length protospacer flanked by the suboptimal 5'-N_4CAAA PAM motif (pTarget_CAAA_23). Surviving colonies were pooled, and plasmids were extracted and used for a second round of selection under the same

conditions (Libraryx2Cas9). We observed significantly increased survival rates between the second and first selection round and compared with the wtThermoCas9 **(Figure 2B**). Roughly ~8000 colonies were collected from the second selection round and their plasmid content was purified and sequenced by NGS. We observed strong enrichment for amino acid substitutions in the position 6 and 7 (**Figure 2C**). Interestingly, these positions of the RuvC domain are highly conserved among Cas9 genes (**Figure 2D**).

In vivo characterization of ThermoCas9 mutants

We proceeded to test the most abundant amino acid mutations by re-introducing them in the *wt-thermoCas9* gene and subjecting them to the selection pipeline once again. Here, we also included in our tests the E655I, N696I ThermoCas9 mutant. It was derived by subjecting random mutagenesis libraries targeting the LII domain for cleavage of *ccdB*-vectors carrying a 17nt long protospacer (Mitchel Roth, Hong Li unpublished).

All variants that were obtained from the whole-gene library managed to efficiently deplete pTarget_CAAA_23 (**Figure 2E**). The E655G, N696I mutant did not manage to rescue the growth phenotype against the 5'-N_4CAAA PAM (**Figure 2E**). Additionally, all tested mutants retained catalytic activity against the full length protospacer flanked by the 5'-N_4CCAA PAM (**Figure 2E**). Tests over 17 nucleotide-long protospacer were not performed for any of the mutants.

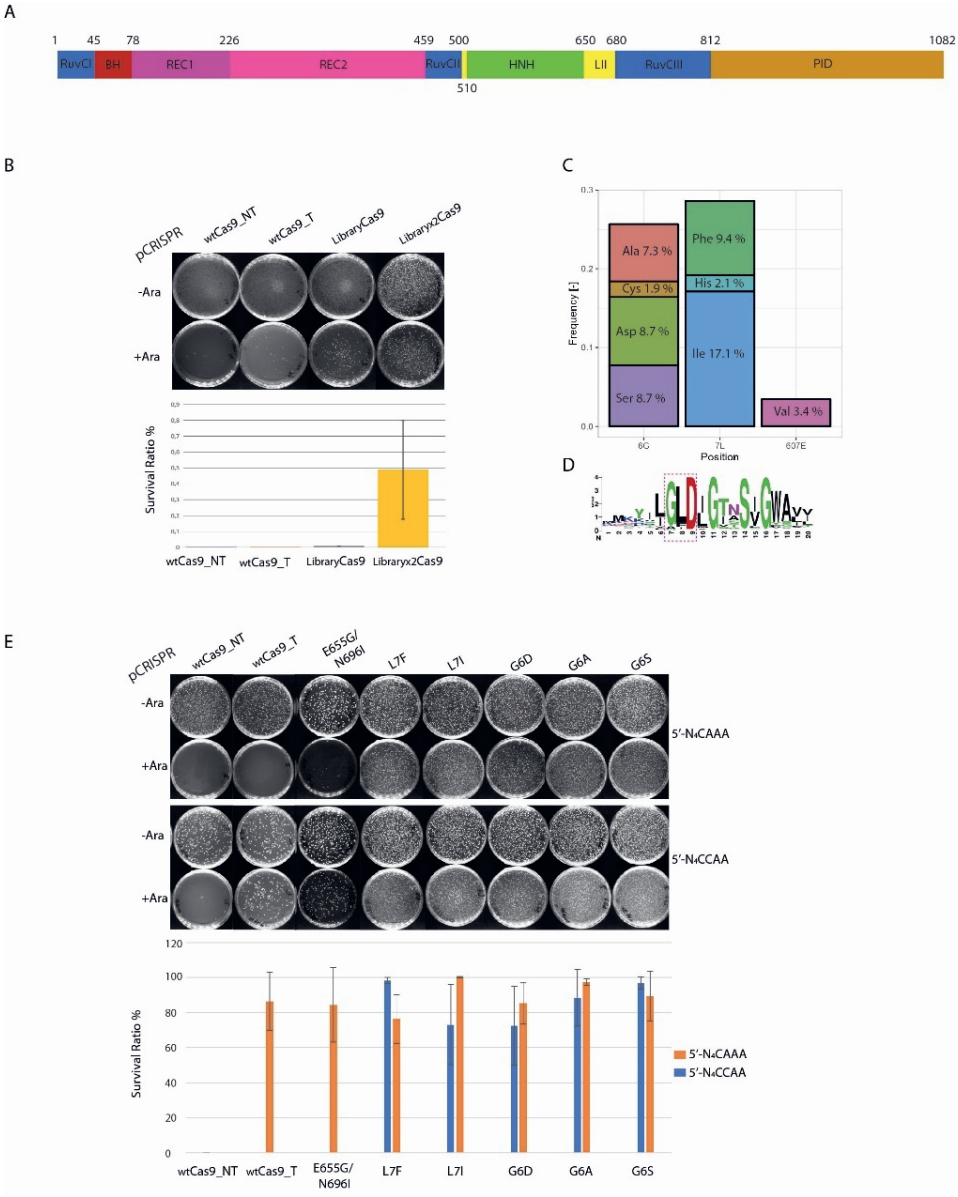


Figure 2. Directed evolution of ThermoCas9 with improved activity. **(A)** Domain organization of thermoCas9. **(B)** top: agar plates of E. coli BW25141: pTarget_CAAA_23 transformed with pCRISPR vectors bottom: bar graph of colony formation ratio between +Ara/-Ara agar plates. T: Targeting Spacer, NT: Non-targeting Spacer. **(C)** Bar graph of enriched aminoacid substitutions as detected by Illumina NGS. **(D)** Sequence-logo of the first 25 aminoacids positions from 3500 Cas9

orthologues (170). **(E)** top: agar plates of E. coli BW25141: pTarget_CAAA_23 or pTarget _CCAA_23 transformed with pCRISPR vectors encoding for most abundant mutants, bottom: bar graph of colony formation ration between +Ara/-Ara agar plates.

In vitro characterization of ThermoCas9 mutants

We continue to characterize the catalytic properties of the most promising mutants in vitro. Therefore, we purified the most interesting protein mutants and performed *in vitro* cleavage assays (IVCAs) against a ~1.9 kbp long linear template carrying a full length PS1 flanked by a 5'-N_4CCAA PAM motif (**Supp. Table S1**). As ThermoCas9 is a moderately thermophilic enzyme with optimal temperature at ~55^{o}C (157), we wanted to see if the selected amino acid substitutions enhanced ThermoCas9 catalytic activity at lower temperatures.

Initially, we performed IVCAs at 20^{o}C, for increased amounts of time. We saw that the first E655G/N696I and second L7F mutants catalysed detectable DNA double strand breaks earlier than wtThermoCas9 (**Figure 3A**). We continued to examine these two variants for a potential shift in temperature optimum. To that end, we performed IVCA at different temperatures. Detectable cleavage was observed approximately at 23^{o}C, 27^{o}C and 28^{o}C for ThermoCas9: i) E665G/N696I, ii) L7F and iii) wtThermoCas9, respectively **(Figure 3B)**. The cleaved fractions were fitted using a non-linear regression model **(Figure 3C)**. Finally, the temperature when half of the linear substrate ([S]/2) was cleaved, was calculated from the regression model **(Figure 3D)**. Tukey's statistical test was used to predict significant change on temperature value when half of the substrate is cleaved ([S]/2), only for the E655G/N696I mutant.

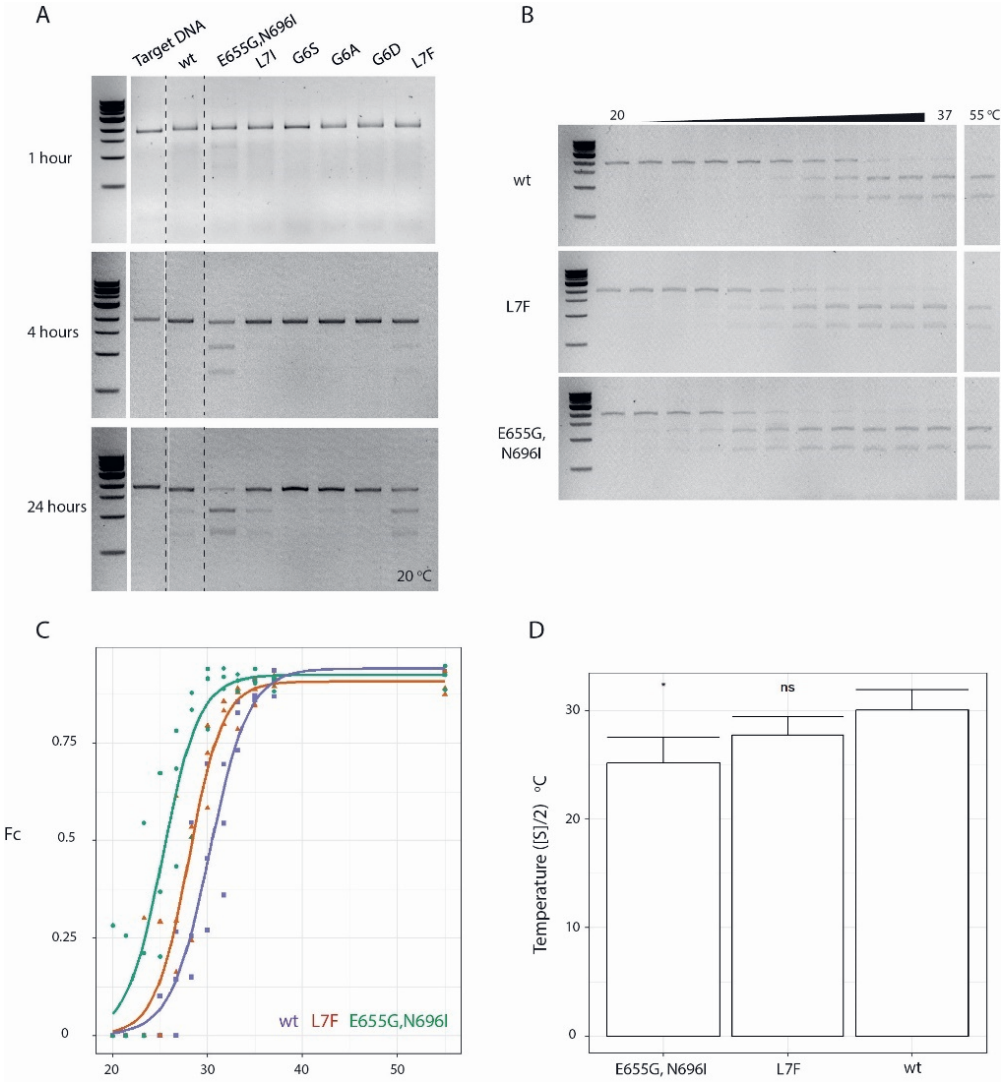


Figure 6. Cleavage assays (*in vitro*) of selected ThermoCas9 variants. (A) In vitro cleavage assay of ThermoCas9 mutants over increased amount of time at 20°C. **(B)** In vitro cleavage assay of L7F and E655G/N696I ThermoCas9 mutants over a range of mesophilic temperatures, incubated for 45minutes. **(C)** In vitro dsDNA activity of wt, L7F and E655G/N696I ThermoCas9 variants over temperature range as determined by agarose gel (n=3). (D) Temperature values when Fc=0.5 for wt, L7F and E655G/N696I ThermoCas9 variants (n=3) $_{(Fc=0.5)}T_{wt}$ =30°C, $_{(Fc=0.5)}T_{L7F}$ =27.7°C, $_{(Fc=0.5)}T_{E655G/N696I}$ =25.15°C.

Conclusions

In this study we focused on ThermoCas9, a Type IIC CRISPR-Cas9 system, with the aim to improve its genome editing activity. Following the principles of directed evolution, we combined random mutagenesis and an *in vivo* selection system. The random mutagenesis library was generated for whole ThermoCas9-encoding gene and was tested against a full length protospacer flanked by suboptimal PAM-motif. We verified the gain of function of the enriched mutations by re-introducing them in the ThermoCas9 gene and test them for cleavage of the suboptimal-PAM, flanked protospacer, in vivo, along with testing for cleavage of the optimal PAM-motif. Following *in vitro* characterization of *in vivo*-selected ThermoCas9 variants, two candidates (E655G/N696I & L7F) demonstrated increased catalytic activity at lower temperatures compared with the wild type protein.

Recently, the tertiary structure of ThermoCas9 alongside the sgRNA targeting a double stranded substrate flanked by a 5'-N_4CCAA PAM was resolved for two functional states: (i) pre-cleavage (open) and (ii) post-cleavage (closed) (Michael Roth & Hong Li, unpublished).

These cryo-EM structures allowed us to examine the interactions of the three mutations with their surroundings. ThermoCas9 catalyses cleavage of the non-target strand through its RuvC domain using a canonical Asp-Glu-Asp (DED) cleavage site (**Figure 4A**). Aspartate 8 (D8), glutamate 500 (E500) and aspartate 723 (D723) form the catalytic triad. In agreement with a recent analysis of the active site of SpyCas9, the histidine 720 (His720) of ThermoCas9 is predicted to form hydrogen bonds with the neighbouring Mg^2 and H_2O, respectively (**Figure 4B**) (34,171).

Residues E655 and N696 of ThermoCas9 are found in the LII-hinge domain. The LII hinge domain undergoes structural rearrangements while ThemorCas9 passes from an "open" pre-catalytic state to a "closed" catalytic state (**Figure 4C**). While in the precatalytic state, E655 is predicted to from hydrogen bonds with neighbouring residues only through

its backbone atoms. However, when catalytic conformation has been established, contacts are predicted between the E655 side chain and the non-target strand backbone of the DNA. E655G substitution abolishes those interactions **(Figure 4D)**.

N696 is predicted to form hydrogen bonds through its side chain with other residues (R661, R698) of the LII domain. As this domain undergoes conformational changes while Cas9 passes from the pre-cleavage to the post-cleavage state, these interactions may stabilize the pre-catalytic conformation of the enzyme **(Figure 4C, E**). N696I cannot form these interactions, which might facilitate the transition from the open to the closed conformation (**Figure 4E**).

The side chain of leucine 7 (L7) is directed in the opposite side from the catalytic site, therefore L7 most likely cannot be directly involved in DNA hydrolysis. Its placement in the RuvC domani leaves it exposed to the aqueous solution and therefore we speculate due to the hydrophobic nature of phenyalanine that L7F substitution promotes the stabilization of the catalytic centre. Phenylalanine is a heavier residue than leucine and it is predicted to result in greater number of clashes with the neighbouring, hydrophobic side chains than leucine does **(Figure 4F)**. However, this approach in rotamer analysis is not recommended by the used software and the observation of increased steric clashes should be taken with a pinch of salt.

In conclusion, by using directed evolution, we have selected variants with slightly enhanced activity in the temperature range of 20-30 °C. This provides a basis for further optimization of the ThermcoCas9 performance in future experiments. During completion of the here presented work, a similar approach has been published by Jennifer Doudna and co-workers, on the highly related GeoCas9 protein (172,173). Especially some mutations in the PAM interacting domain had a major effect on the R-loop formation kinetics, leading to a major improvement of the editing efficiency of human cells. This motivates us to continue our efforts to further improve

the performance of ThermoCas9, to make this an interesting tool for a wide range of genome editing applications.

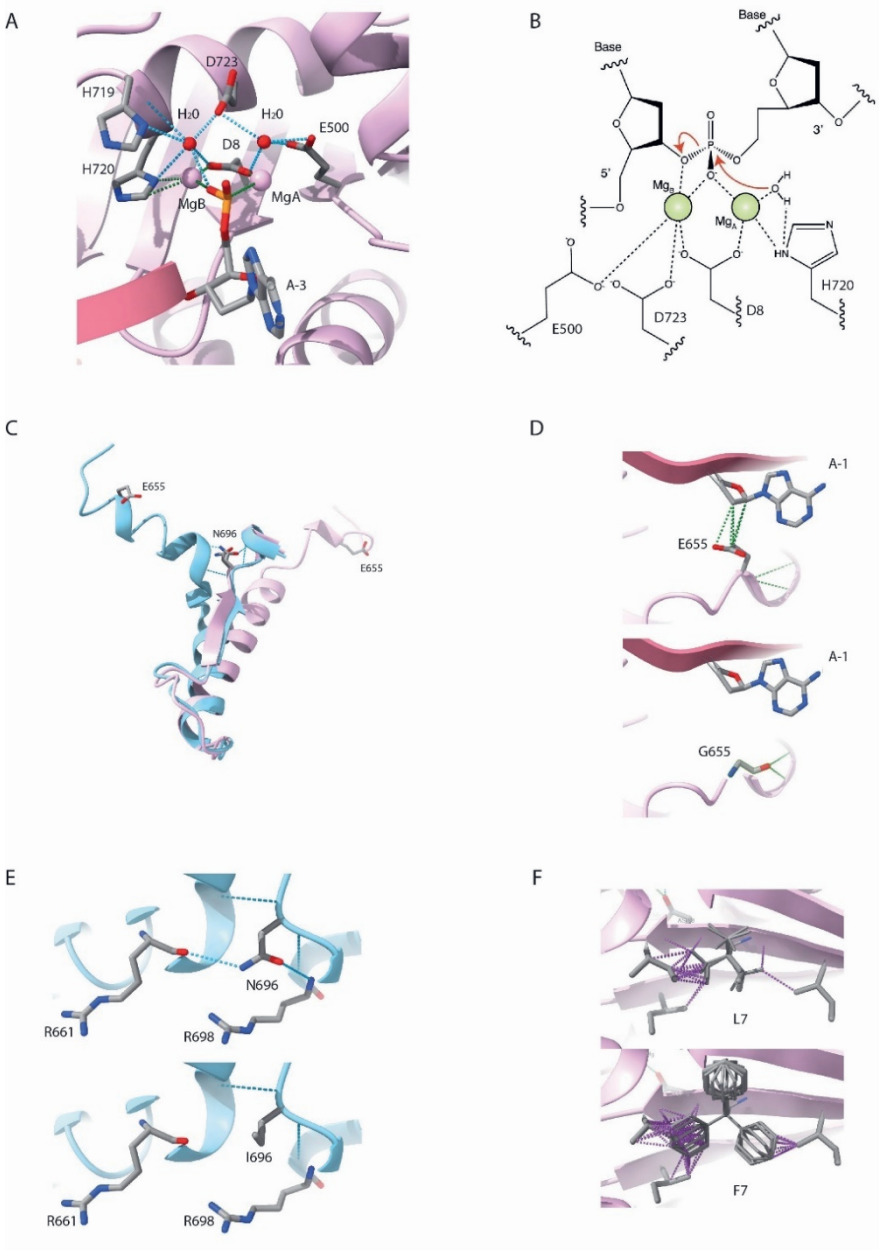
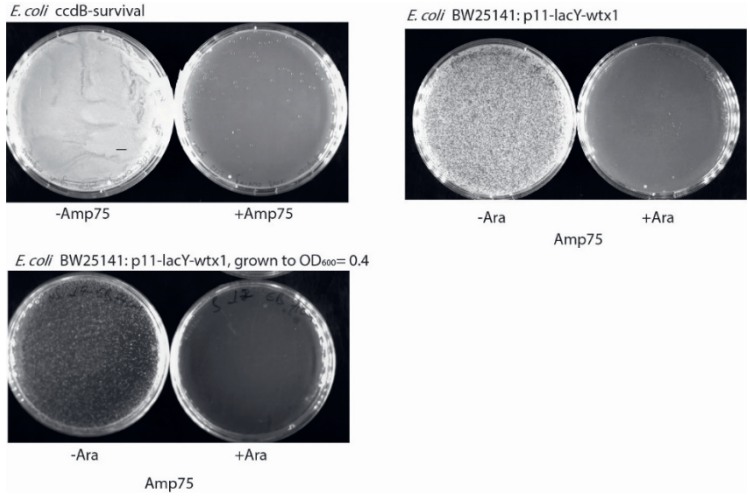


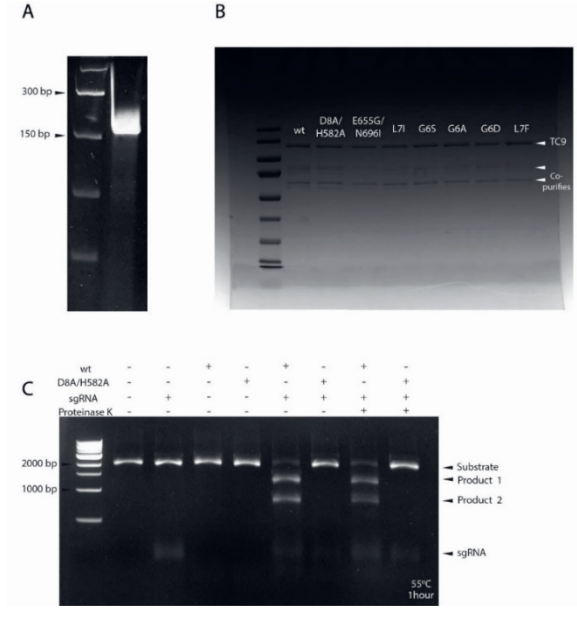
Figure 7. Structural Analysis of E655G/N696I and L7F ThermoCas9 mutants **(A)** Interactions of the catalytic triad residues of the RuvC domain with Mg^{+2}, H_2O and nucleotides of the displace DNA strand. blue dashed lines: h-bonds, green dashed lines: contacts. **(B)** RuvC DNA cleavage mechanism adapted from (171). **(C)** Structural alignment of the LII hinge domain in the open conformation (blue) and closed conformation (pink). **(D)** Comparison of E665 to G655 in closed conformation. Green dashed lines: contacts. **(E)** Comparison of N696 to I696 in open conformation. Blue dashed lines: h-bonds. **(F)** Comparison of rotamer library for L7 and F7 in closed conformation. Purple dashed lines: clashes. Prediction of h-bonds, contacts and clashes was performed in Chimera X software using the default settings.

Supplementary Material

Supplementary Figures



Supplementary Figure 1. Top left: Agar plates of *E. coli ccdB*-survival cured from p11-lacY-wtx1 for subsequent *ccdB* vector assembly and propagation. Top right: Agar plates of *E. coli* BW25141 transformed with p11-lacY-wtx1. Bottom right: Agar plates of *E. coli* BW25141: p11-lacY-wtx1 grown to OD_{600}=0.4.



Supplementary Figure 2. (A). 8% UREA-PAGE on which ThermoCas9, in vitro transcribed sgRNA was resolved. **(B)** SDS-PAGE gel on which ThermoCas9-6xhis mutants were resolved. **(C)**. Control in vitro cleavage assay.

Supplementary Table S1. Oligonucleotides used in this study

Named Description	**Bacgen**	**Oligos 5'-3'**
pTarget_*ccdB*:PS1_CCAA_23	BG29428	ctagAGAATTTATGCCCATTCACATCCCCATCCAAcatg
	BG29429	TTGGATGGGGATGTGAATGGGCATAAATTCT
pTarget_*ccdB*:PS1_CAAA_23	BG29432	ctagAGAATTTATGCCCATTCACATCCCCATCAAAcatg
	BG29433	TTTGATGGGGATGTGAATGGGCATAAATTCT
pTarget_*ccdB*:PS1_CACA_23	BG29434	ctagAGAATTTATGCCCATTCACATCCCCATCACAcatg
	BG29435	TGTGATGGGGATGTGAATGGGCATAAATTCT
pTarget_*ccdB*:PS1_CGCA_23	BG29430	ctagAGAATTTATGCCCATTCACATCCCCATCGCAcatg
	BG29431	TGCGATGGGGATGTGAATGGGCATAAATTCT
epPCR Insert Fw oligo	BG28246	CTTCATAAGCAGGCCATTTTGTCTG
epPCR insert Rv oligo	BG28247	ttaaattgctaacgcagtcaggcac
epPCR Backbone Fw oligo	BG28248	gtgcctgactgcgttagcaatttaactgtgataaacggaattgttatccg
epPCR Backbone Rv oligo	BG28249	CAGACAAAATGGCCTGCTTATGAAGCGGGCCATTTTTGTTTAATCC
Fw sgRNA ssDNA oligo	-	AGCTTGAAATAATACGACTCACTATAGGaAGATTATCAAAAAGGATCTTCACGTCATAGTTCCCCTGGAAACAGGGTTACTATGATAAGGGCTTTCTGCCTATAGGCAGACTGACCCGTGGCGTTGGGGATCGCCTATCGCCCGCTTTCTTCGGGCATTCCCCACTCTTAGGCGTTTT
Rv sgRNA ssDNA oligo	-	AAACGCCTAAGAGTGGGGAATGCCCGAAGAAAGCGGGCGATAGGCGATCCCCAACGCCACGGGTCAGTCTGCCTATAGGCAGAAAGCCCTTATCATAGTAACCCTGTTTCCAGGGGAACTATGACGTGAAGATCCTTTTTGATAATCTtCCTATAGTGAGTCGTATTATTTCAAGCTT
Fw sgRNA template PCR	-	AAGCTTGAAATAATACGACTCACTATAGG
Rv sgRNA template PCR	-	AAAACGCCTAAGAGTGGGG
Fw IVCA Substrate PCR oligo	BG32584	tagttaggccaccacttcaag
Rv IVCA Substrate PCR oligo	BG32585	tcatccatagttgcctgactcc
IVCA Linear Substrate	-	https://benchling.com/s/seq-I3XeUbtMfOzzbwt6bzvR?m=slm-EvDHY5t9O5IiTieyB5Tn

Chapter 5

Base editing in the target strand by the miniature Cas-effector Cas12f1

Thomas Swartjes*, Evgenios Bouzetos*, Belén Adiego Pérez, Raymond H.J. Staals, John van der Oost** and Wen Y. Wu**

* authors contributed equally
** corresponding authors

Manuscript in submission

Abstract

CRISPR-associated base editors have been established as genome editing tools that allow for base-conversions in targeted DNA sequences, without generating double strand breaks. Here we describe the development of new base editors based on CRISPR-Cas12f1, a miniature Cas protein of only 422 amino acids. Chimeric constructs have been generated by fusing a catalytically inactive dCas12f1 either to a cytosine deaminase or an adenine deaminase. Using these synthetic fusion proteins, systematic analyses have been performed on base editing of a target sequence on a plasmid in Escherichia coli. Interestingly, apart from the previously described base editing of the displaced non-target DNA strand, we also observed efficient editing of the target DNA strand. In addition to the small size of AsCas12f1 base editors, its unique editing profile makes it a valuable addition to the CRISPR-Cas toolbox.

Introduction

CRISPR-Cas is an adaptive immune system used by prokaryotic cells to defend themselves against invading genetic elements like plasmids and viruses (174,175). The immune system typically comprises an array of clustered regularly interspaced short palindromic repeats (CRISPR) and neighboring CRISPR-associated (Cas) genes (11,176). CRISPR-Cas systems are classified into two classes. Systems with a multi-protein Cas effector complex belong to Class 1 (types I, III, IV), whereas Class 2 systems employ a single, multi-domain effector (types II, V, VI) (9).

The best characterized DNA-targeting CRISPR-Cas systems are Cas9 (type II) and Cas12a (type V-A) (12,127,177,178). After recognition of an appropriate protospacer adjacent motif (PAM) (129), the Cas/crRNA (and tracrRNA in case of Cas9) ribonucleoprotein complex binds the dsDNA through base pairing of the crRNA guide and the complementary target DNA strand, while displacing the non-target strand (179–181). RNA-guided targeting of a complementary target DNA by Cas9 and Cas12a results in a double stranded break (DSB) in the DNA (127,177).

CRISPR-Cas systems can be heterologously expressed in a variety of hosts. The unique feature of programmable DNA target specificity by designing a matching crRNA has established CRISPR-Cas systems as highly efficient tools with applications ranging from strain engineering and gene therapy, to diagnostics (130,182). After introduction of a DSB in the target sequence, the actual outcome of the editing relies on the subsequent repair of the DSB. Induced DSBs can be repaired either by non-homologous end joining (NHEJ) or by homology directed repair (HDR) (183). While NHEJ can disrupt a gene's functionality (184), HDR allows for precise edits (183) when a DNA repair fragment is provided. However, HDR is generally less efficient compared to NHEJ, and is usually restricted to S and G2 phase of the cell cycle in eukaryotic cells (185). In addition, the dependency of both repair

systems on DSBs is considered a disadvantage, as this may result in undesired recombination events (186–189).

Base Editing is a genome editing approach that does not require the introduction of DSBs or co-delivery of DNA repair fragments (190). Reported base editors employ a Cas effector protein (e.g. Cas9, Cas12a) in which the nuclease activity has been disrupted. This Cas effector is fused to an adenine deaminase (e.g. TadA) or a cytosine deaminase (e.g. APOBEC1), resulting in an adenine base editor (ABE) or a cytosine base editor (CBE), respectively. The latter usually include a uracil DNA glycosidase inhibitor (UGI) domain to increase base editing efficiency by supressing the uracil DNA glycosylase enzyme that initiates the base excision repair pathway reversing a U:G pair back to C:G (191,192). After targeting and formation of an R-loop structure, base editing occurs through deamination of bases on the displaced, non-target strand, within a certain editing window. At what position base editing occurs likely depends on their nucleotide accessibility for the deaminase, and hence on the overall base editor architecture (187). Although base editing has also been demonstrated to occur with catalytically dead effectors (dCas), substantially increased efficiencies have been obtained with base editors that use a nickase variant (nCas9) (191,193). The nCas9 (RuvC-mutation) cleaves only the guide-complementary target strand. The nicked target strand is subsequently repaired by the mismatch repair pathway using the edited displaced strand as repair template (28,190–192). Currently existing base editors, such as Cas9 or Cas12a base editors are often large, exceeding the size capacity for viral vector delivery systems such as AAV, thereby obstructing their use for *in vivo* therapeutic editing (194). This limitation can be overcome by using smaller effector proteins for base editing.

One recently characterized small Cas protein is Cas12f1 (158,195) that functions as an asymmetrical homodimer (31, 32) (**Figure 1**). The Cas12f1 dimer is guided by a single crRNA-tracrRNA pair, recognizes a 5' T-rich PAM and binds and cleaves a dsDNA target. Cleavage of dsDNA occurs through

a single RuvC domain of the Cas12f1.1 monomer (196,197). Cas12f1 homologues range from 400-700 residues, and they have been used successfully for genome editing, including base editing, in mammalian and plant cells, especially after optimization by protein and sgRNA engineering (198,199).

Here we describe base editors (CBE and ABE) using the dCas12f protein of *Sulfoacidibacillus thermotolerans* (previously called *Acidibacillus sulfuroxidans*). At present, Cas12f1 base editors are the most compact base editors within the CRISPR-Cas toolbox. As they easily fit into an AAV capsid, they allow for *in vivo* gene therapy (146). The here developed Cas12f1 BEs were characterized in *E. coli,* resulting in high base editing efficiencies with a broad base editing window of the non-target strand. Interestingly, we describe the unprecedented, efficient editing of bases on the target strand as well.

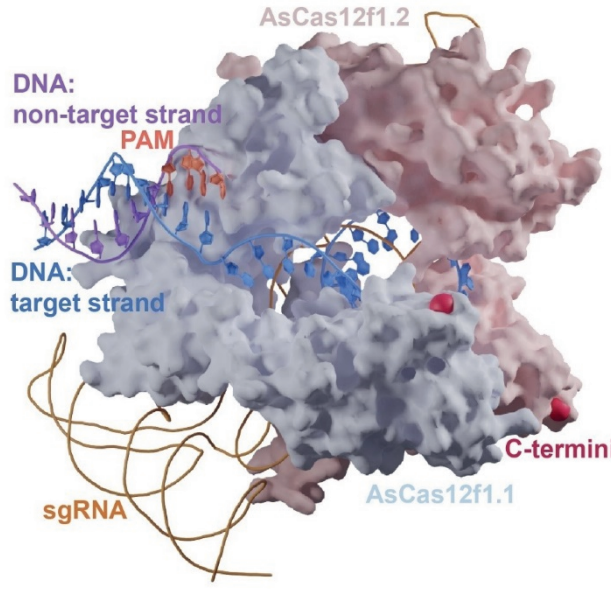


Figure 1. AsCas12f1 protein structure (PDB 8J12). C-termini are indicated with red dots to indicate where the deaminase domains were fused.

Material and Methods

Bacterial strains and growth conditions

Bacterial strains used for the cloning and propagation of plasmids in the current study are *E. coli* DH5α and DH10β. To assess base editing we used *E. coli* strain BW25113 lacking the *lacI, lacZ* genes and the type I-E CRISPR-Cas system (200). The *E. coli* strains were routinely cultured at 37°C and 220rpm, unless otherwise specified, in either lysogeny broth (LB) [10g/L peptone (Oxoid), 5g/L yeast extract (BD), 10g/L NaCl (Acros)] or M9TG minimal medium [1xM9 salts (Sigma), 10g/L tryptone (Oxoid), 5g/L glycerol (Acros)]. Plasmids were maintained with ampicillin (100mg/mL), chloramphenicol (35mg/mL), and/or kanamycin (50mg/mL) as needed.

Plasmid construction

We used a three-plasmid system (pCas, pGuide and pTarget) similar to previously published studies (200). The *asdcas12f1* gene with the D225A mutation (144) was amplified from a synthetically synthesized AsCas12f1 CRISPR locus (IDT gblock) and inserted into the pBAD33 (201) vector backbone under the control of the constitutive J23108 promoter to construct pCas-noBE. The pCas base editor plasmids (pCas- ABE/CBE) were constructed from this negative control plasmid by fusing TadA8e (202) (ABE) or APOBEC1 (CBEs) to the C-terminal end of AsdCas12f1 (Figure 1A). For CBEs, uracil glycosylase inhibitor (UGI) was fused to the C-terminal end of APOBEC1. When cloning pCas plasmids, especially pCas-CBE, we observed many mutations arising in the coding sequences of the BEs. This suggested that the cells experienced toxicity by the high expression levels of these proteins. To decrease toxicity and the resulting evolutionary pressure, we fused the last three amino acids (LVA) of the LVA ssrA degradation tag (203) to the C-terminus of all AsdCas12f1 (fusion) proteins in the pCas plasmids. A pGuide-RFP-entry plasmid was generated by inserting a sgRNA (144) downstream from the constitutive J23119 promoter

in a pBAD18 backbone (201). Spacer sequences were assembled into linear dsDNA by annealing ssDNA oligonucleotides and were inserted to the pGuide-RFP-entry plasmid by BbsI resctriction ligation. The pTarget Entry contains a *gfp* and a *mRFP* gene under the P_{lacIQ} and PTaq promoters, respectively, in a pAU66 plasmid backbone. Target sequences were similarly assembled as linear dsDNA by annealing ssDNA oligonucleotides and inserted in place of the mRFP gene by AatII/KpnI restriction ligation.

Base edition assay

E. coli cells (adapted BW25113, as described earlier) harboring pGuide plasmids and their corresponding pTarget plasmids were made chemically competent and transformed with the different pCas plasmids. To this end, 45μL of competent cells where mixed with 2μL 50ng/μL of pCas plasmid DNA, kept on ice for 30min, heat-shocked at 42^{o}C for 30s, kept on ice for 2min. Then, 150μL LB without antibiotics was added and the cell suspension was transferred to a 96-wells 2mL deep-well plate (Greiner) and incubated for 1h at 37^{o}C with short-stroke shaking at 900rpm for recovery. Different biological replicates originate from distinct batches of competent cells and were transformed with pCas separately. After recovery, the transformation mix was diluted: 2μL + 198μL in M9TG medium in 96-wells 2mL deep-well plate. The deep-well plate was then sealed using a gas-permeable membrane (Diversified Biotech, Breathe-EASIER) and grown incubated at 37°C with short-stroke shaking at 900rpm for 24h. For subsequent timepoints (48h and 72h), the cells were serially diluted twice, mixing 2μL with 198μL fresh M9TG medium in a deep-well plate. The bacteria were then incubated to grow 24h at 37°C. For the 72h timepoint, the previous step was done twice. For each timepoint, the remainder of the undiluted bacterial cultures was stored at -20^{o}C.

NGS Sample preparation

Diluted by mixing 10µL of the stored samples with 90µL mQ water and again stored at -20ºC for one week. At this point, due to a building-wide power outage, the stored diluted samples thawed and remained at ~22ºC for ~1.5 weeks, before analysis. We verified through Sanger sequencing comparable base editing before and after the power outage incident. A 150bp region surrounding the protospacer was PCR amplified by using 2µL diluted sample in 25µL PCR reactions using Q5® High-Fidelity 2X Master Mix (NEB). Forward primers used in the PCR reactions were internally barcoded using six additional nucleotides to distinguish between samples. PCR products with verified DNA concentrations were quantified by capillary gel electrophoresis using the Qiagen QIAxcel DNA high-sensitivity kit. PCR products were pooled in groups of 12, adjusting volumes to obtain equal concentrations between pooled conditions. Pooled DNA was purified using the Zymo kit DNA Clean & Concentrator-5 (D4004) according to manufacturer's protocol, eluting in 15µL mQ water. DNA concentrations were measured using Qubit™ dsDNA BR (Invitrogen), and DNA was further pooled in groups of 5 DNA pools, adjusting volumes to obtain equal concentrations between pooled conditions. This resulted in one pool for each replicate for which we verified roughly equal concentrations and purity by agarose gel electrophoresis. We then sent these three pools to Eurofins Genomics as three individual samples for adapter ligation, and subsequent NovaSeq 6000 paired-end Illumina sequencing.

NGS data analysis

To analyze the deep-sequencing data, the paired-end reads were programmatically merged using seqprep(204). The reads from different conditions were organized in separate files based on their internal barcodes. We filtered out reads that do not match the expected pattern of starting with a forward barcode followed by a forward primer annealing part and ending with the reverse primer annealing part. We also removed reads with low confidence base-calls (error probability of >0.01). Reads that did not

match the corresponding pTarget were also discarded. Finally, we analyzed the resulting reads for base editing at each position. The used scripts were created in-house.

Results and discussion

To establish base editing with the AsCas12f1 protein, we used a set of three plasmids: pCas, pTarget and pGuide. The pCas plasmids encode a catalytically inactivated variant (D225A) of AsCas12f1 (202), fused to either adenine deaminase TadA8e or cytosine deaminase APOBEC1 and UGI for adenine and cytosine base editing, respectively **(Figure 2A).** The pTarget plasmids each contained a 10bp-PAM-protospacer-10bp region with a target (A or C) nucleotide in every third position, except in the PAM **(Figure 2B, 2C, Supp. Table 1).**

For each protospacer, a pGuide plasmid was used expressing the corresponding single guide RNA (sgRNA). *Escherichia coli* cells harboring pTarget and pGuide were transformed with the pCas plasmid, after which they were incubated to allow base editing. Cells were sampled after 24, 48, and 72 hours, followed by target sequence amplification, and subsequent analysis by deep sequencing. We scored base editing at all positions, inferring either non-target (NTS) or target strand (TS) editing depending on the observed base conversion **(Supplementary Figure S1)**.

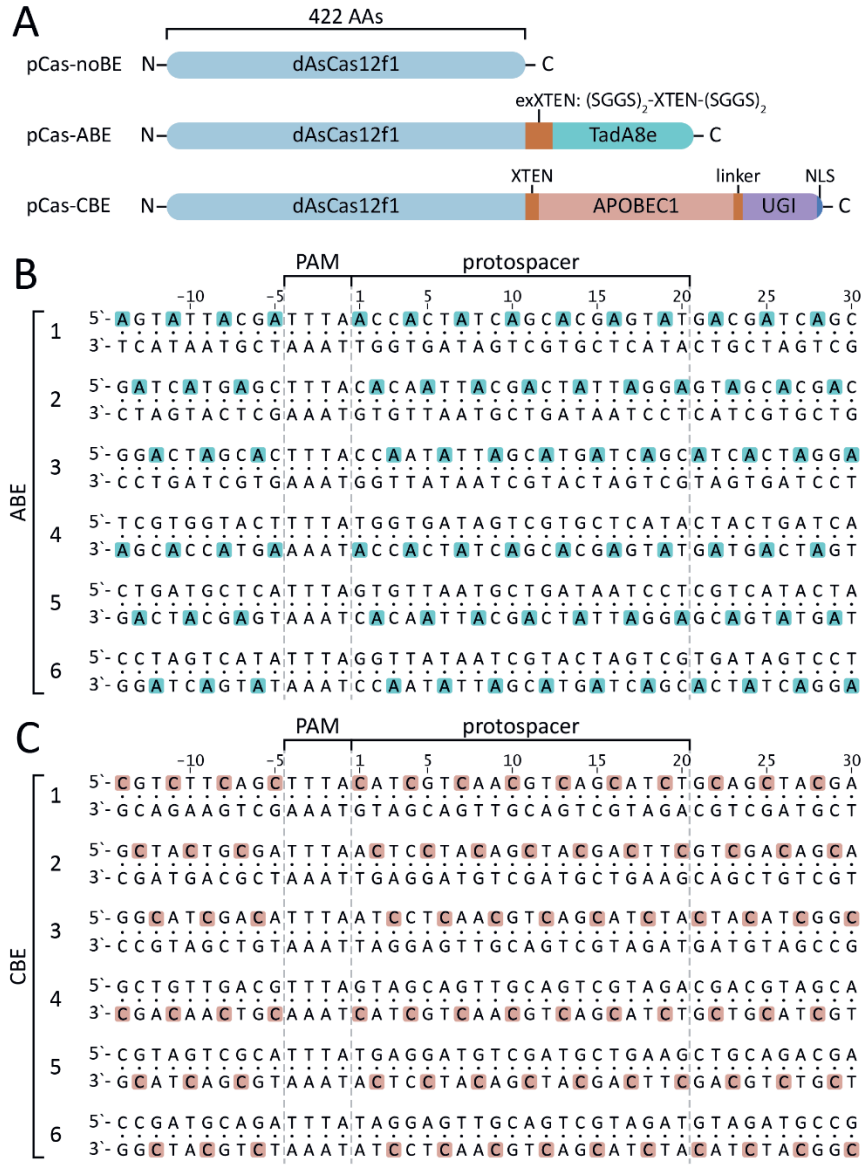


Figure 2. Protein fusion layout and nucleotide tiling strategy.
(A) Schematic overview of the fusion protein encoded on each pCas plasmid. The width of each element is proportional to their amino acid sequence. C- and N-termini are indicated for each fusion protein. Linkers are indicated in orange. AAs, amino acids; NLS, nuclear localization signal. **(B-C)** Target sequences of ABE (B) and CBE. **(C)** pTarget plasmids. The tiled bases are shown in blue (ABE) or pink (CBE).

We found highly efficient base conversion at many positions throughout the target sequences, for both adenine and cytosine base editing **(Supp. Figure 1)**. We detected neglectable base conversion numbers in conditions without a deaminase or with a non-targeting sgRNA **(Supp. Figure 2A)**. When considering the combined data from the different pTarget plasmids, the CBE specifically caused efficient base conversion at positions throughout the NTS of the protospacer **(Figure 3A)**. Base conversion by the ABE was mostly restricted to the PAM-proximal half for the NTS. Interestingly, we also observed efficient base conversion in the PAM-distal half of the protospacer on the TS for both base editors **(Figure 3A)**. These findings deviate substantially from those reported for ABEs that use Cas12f1 from *Candidatus Woesearchaeota* or an uncultured archaeon, where optimal editing is only reported for NTS positions in the PAM proximal part of the protospacer (27,200). This discrepancy may be due to structural differences between the Cas12f1 proteins or the various TadA mutants used in different studies.

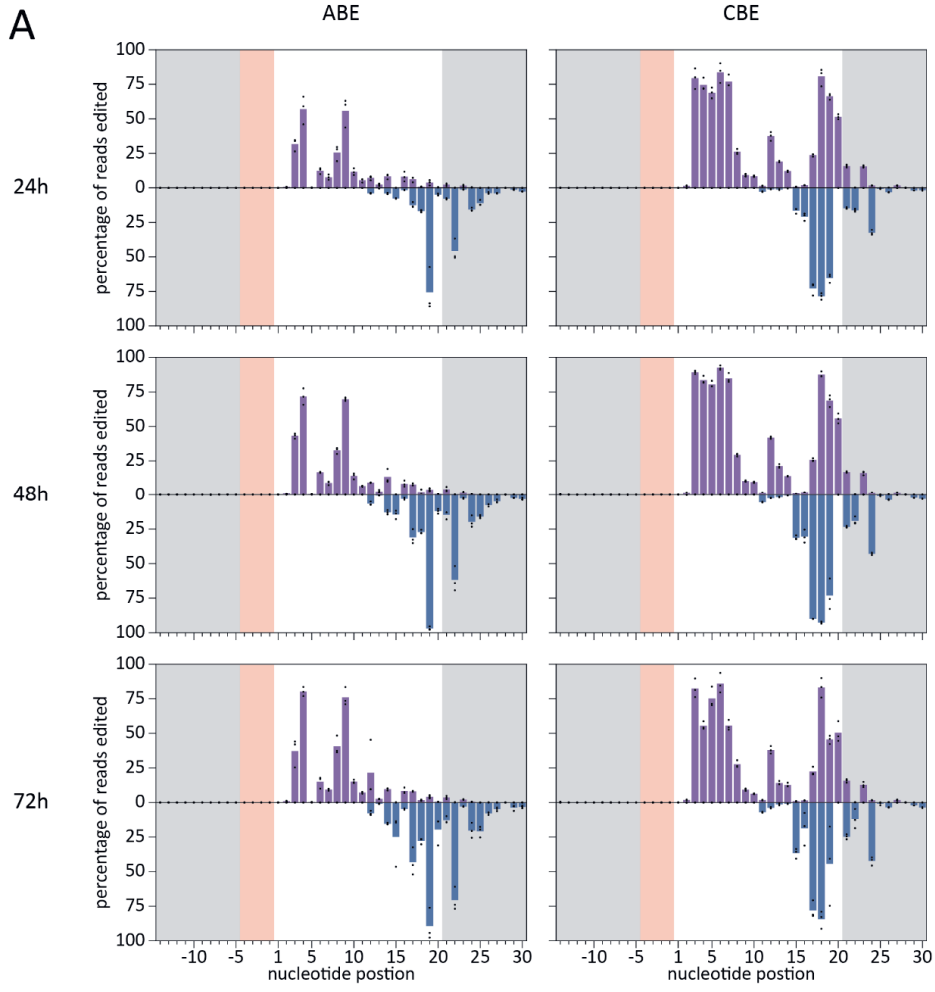

A
ABE
CBE
24h
48h
72h
percentage of reads edited
nucleotide postion
B
normalized total editing
adj. p-value: 0.0020
adj. p-value: 0.0095
adj. p-value: 0.0082
adj. p-value: 0.0046
replicates
1
2
3
timepoint

Figure 3. ABE and CBE editing at different positions and durations.
(A) Bar-charts showing base editing activity at tiled positions (Figure 1), combining data from all protospacers for ABE and CBE at three timepoints (plots per pTarget in Supp. Figure 1). The data from each biological replicate is shown as black dots, while the bars indicate the average of the three replicates. The pink-shaded area indicates the PAM, the grey-shaded areas are upstream from the PAM and downstream from the 20nt we refer to as protospacer. Purple bars: NTS editing. Blue bars: TS editing. **(B)** Bar-charts of the total editing (over all positions) by timepoint. The dots show total editing for single replicates, with the bars indicating the average across the three biological replicates. Adjusted p-values below 0.05 from Tukey's *post hoc* statistical test are indicated.

Base editing studies have commonly shown non-target strand base editing and therefore it was thought that only NTS nucleotides are available as deamination substrate due to TS-sgRNA heteroduplex formation (190,191,195). In most base editing applications, a nickase version of the Cas9 nuclease is used to cleave only the TS (e.g. SpCas9 D10A). In that way, retention of the edited NTS sequence is favored as template strand during DNA repair, probably concealing base editing of the TS (191). Base editing of the TS has only been reported using SpCas9 (D10A) cytosine base editing and was most prevalent (up to 6.3% editing) outside of the protospacer region or at off-target sequences (205). Reported TS cytosine base editing at on-target sites was less than 1% at individual nucleotides (205). Apart from efficient base editing of the TS, we also detected substantial editing outside of the protospacer in the PAM-distal end for both ABE and CBE **(Figure 3A)**. Similarly to TS base editing, such conversions downstream from the target sequence have rarely been reported (205). This may suggest that the melting of the dsDNA may expand beyond the sgRNA/target-strand heteroduplex, likely mediated by the AsCas12f1 protein, as that would expose downstream nucleotides to the deaminase.

Our sequence analyses indicate that base editing is highly sensitive to the neighboring nucleotide sequence. For instance, C-to-T editing by CBE on pTarget-5 resulted in very high editing at position 10 of the NTS, while editing at the same position in pTarget-1 is far lower **(Supp. Figure 1)**. The results presented here therefore show where editing can happen, but

do not exclude that editing might also occur elsewhere in a different sequence context.

When comparing total base editing levels over time, we found that ABE catalyzed A-to-G conversion increased through time **(Figure 3B)**. However, in case of CBE, C-to-T conversion increases from 24 to 48h, but then drops from 48 to 72h **(Figure 3B)**. We attribute this to the inherit cell toxicity of the APOBEC-1 cytidine deaminase (136,206). Cells that disrupt APOBEC-1 expression through random mutations are thought to have a substantial fitness advantage. After 72h, these cells (for which we expect lower cytosine base editing levels) have therefore been enriched, skewing the editing results.

The observed editing on the TS allows base editing at genomic positions previously not reported **(Figure 4)**. However, TS editing also increases the risk of undesired off-target edits. Investigation of off-target edits was outside the scope of this project and should be addressed in future studies, specifically those applying Cas12f1 base editing in therapeutics. In addition, as only 20nt long spacers were tested here, we do not know the effect that different spacer lengths could have on base editing. Increased gRNA-protospacer complementarity could further stabilize the RNA-DNA heteroduplex and possibly reduce base editing of the TS.

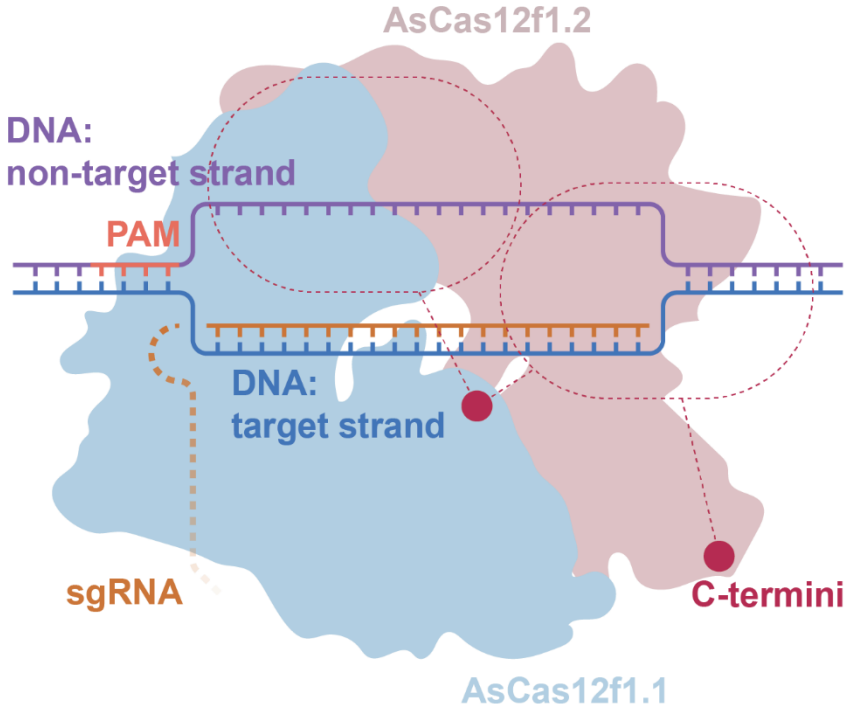


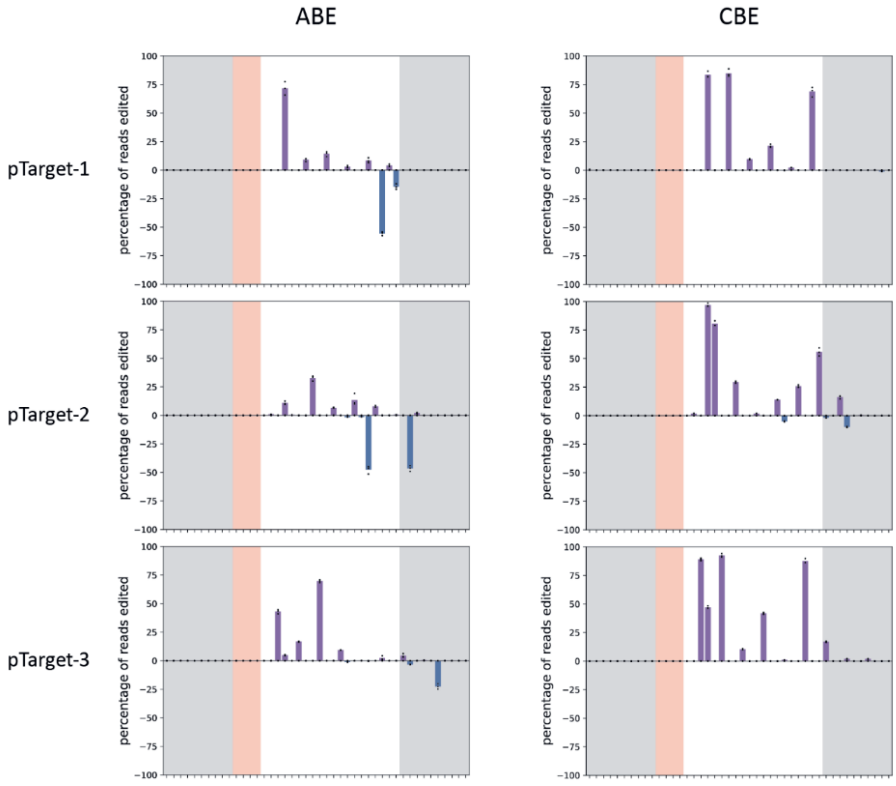
Figure 4. Working model of AsCas12f1 base editors.
The DNA non-target and target strand are shown by purple and blue color respectively. sgRNA is shown by orange color. The C-terminal of the Cas12f where the genetic fusion with the respected deaminases was made is shown with red color filled circles. The editing window of AsCas12f base editors are shown to surround the DNA strands with dashed red circles.

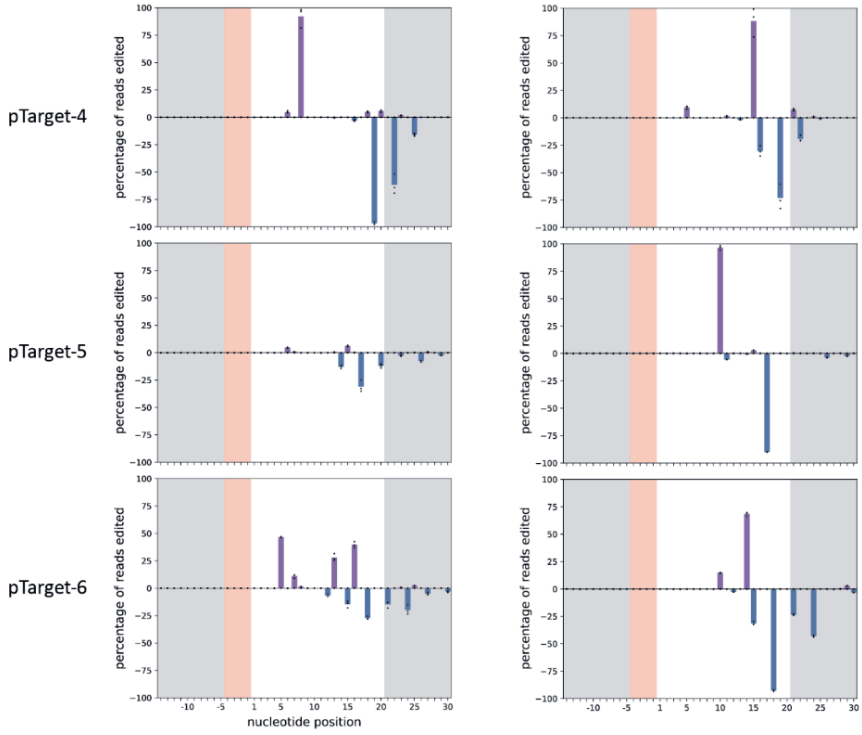
In conclusion, we have constructed adenine and cytosine base editors using a catalytically inactive AsCas12f1. These base editors are half the size of those relying on Cas9, potentially facilitating delivery by AAV. We observed high levels of base editing in *E. coli*, both on the DNA NTS as well as on the TS **(Figure 4)**. In addition, we found that editing also occurred downstream of the protospacer on both DNA strands. Such broad base editing patterns have not been previously reported to the same extent for other base editors. The here described AsCas12f1 ABE and CBE reveal increased base editing ranges and therefore are a valuable addition to the

current existing genome editing toolbox. Overall, our results show that alternative Cas effectors and derived synthetic variants may have unexpected and unique features that are not only interesting for fundamental reasons, but that might also be useful for certain applications.

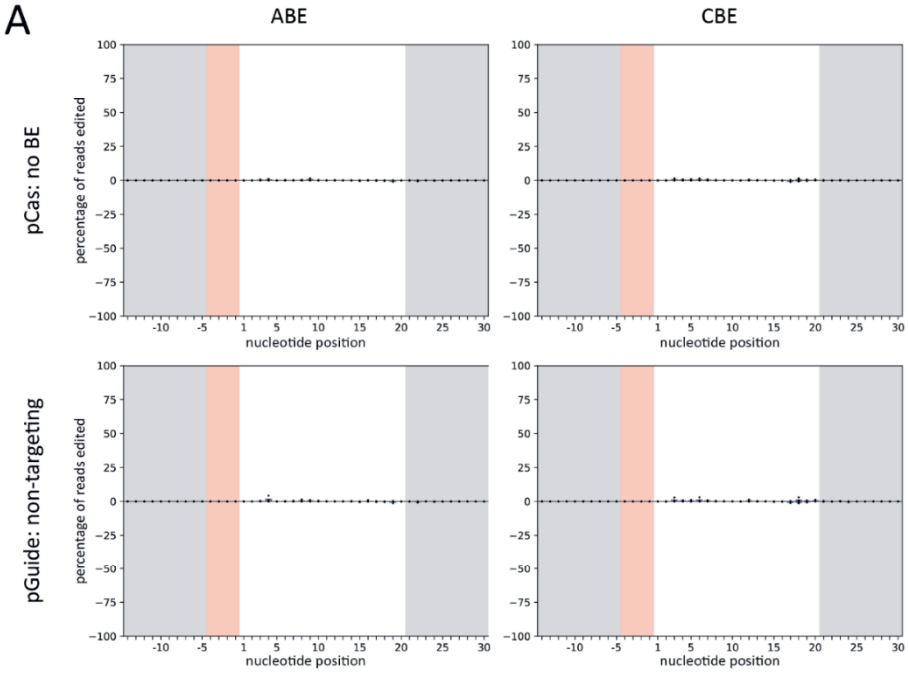
Supplementary material

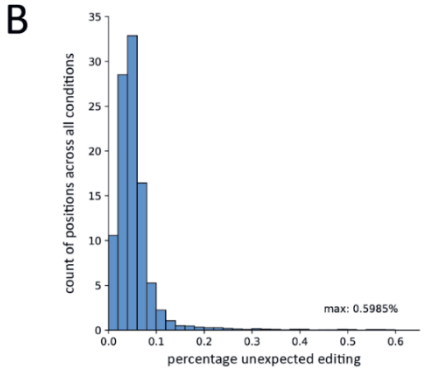
Supplementary Figures

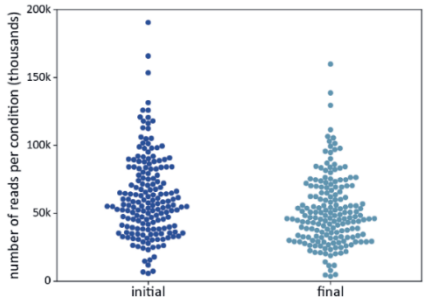




Supplementary figure 1. Base editing at individual pTarget plasmids. Bar-charts showing base editing activity at positions in single pTarget plasmids for ABE and CBE at after 48h. The data from each biological replicate is shown as black dots, while the bars indicate the average of the three replicates. The pink-shaded area indicates the PAM, the grey-shaded areas are upstream from the PAM and downstream from the 20nt we refer to as protospacer. Purple bars: non-target strand editing. Blue bars: target strand editing. In contrast to Figure


A
ABE
CBE
pCas: no BE
pGuide: non-targeting
percentage of reads edited
nucleotide position


B
count of positions across all conditions
max: 0.5985%
percentage unexpected editing


C
number of reads per condition (thousands)
initial
final

Supplementary Figure 2. Negative controls, unexpected editing, and read retention.
(A) Bar-charts showing base editing activity of negative controls at tiled positions (Figure 1), combining data from all protospacers for ABE and CBE after 72h. The data from each biological replicate is shown as black dots, while bars indicate the average of the three replicates. The pink-shaded area indicates the PAM, the grey-shaded areas are upstream from the PAM and downstream from the 20nt we refer to as protospacer. Purple bars: non-target strand editing. Blue bars: target strand editing. pCas: noBE indicates that a catalytically dead AsCas12f1 was used without any deaminase domain. pGuide: non-targeting indicates that a non-targeting guide was used. **(B)** Histogram of edits other than the expected base conversions by ABE or CBE for any position across all tested conditions. The maximum 'unexpected' editing at a single position in a single condition is indicated on the bottom right of the plot area. **(C)** Distribution of the number of sequencing reads (in thousands), with each dot representing one conditions across all three biological replicates. The initial distribution shows the numbers of reads directly after they were separated based on barcoding. The final distribution shows the numbers of reads left after cleaning and filtering.

Chapter 6

Summary and

General Discussion

Thesis Summary

CRISPR-Cas is a unique prokaryotic defence system that confers adaptive immunity. It encompasses a large variety of systems that are classified in two classes and multiple types and subtypes. It has been harnessed as a versatile molecular tool that revolutionized the field of genome engineering. Especially, Cas-effectors of Class 2 that consist of multi-domain proteins have been extensively studied and tested for genome engineering purposes. Although they are versatile, Cas nucleases fail under certain conditions to interfere with target substrates. Protein engineering through rational design and directed evolution has provided us with enhanced Cas variants and has enriched the CRISPR-Cas toolbox. In this thesis, focus is given on the development of novel directed evolution platforms, on engineering Cas9 variants with enhanced catalytic activity and on expanding the current CRISPR-Cas toolbox.

In chapter 2, we portray the performance of directed evolution using artificial microcompartments. First, we show how the principles of directed evolution are applied outside of the cell context. We continue by describing the types of artificial compartments that can substitute living cells and how they can be produced. Next, we describe how cell-free expression is performed inside artificial microcompartments, while retaining genotype to phenotype linkage. To close the directed evolution cycle, we demonstrate high-throughput screening schemes that are applicable for large random-mutagenesis gene libraries. Finally, we conclude with state-of-the-art developments in the field, and we speculate on the bright future of microfluidics assisted *in vitro* compartmentalization.

In chapter 3, we set off to establish a cell-free screening platform for the high throughput screening of CRISPR-Cas cleavage activity. Initially, using

a novel microfluidic chip design, we generated double bagged artificial compartments (water in oil in water droplets) at a frequency of ~500 Hz. We tested the biocompatibility of the compartments by internally expressing a green fluorescent protein using a cell-free *E. coli* extract. Next, we tested already described and prototyped genetic circuits that connect CRISPR-Cas activity to alterations in protein fluorescence. Finally, we generated two populations of compartments of known consistency: i) targeting gRNA, ii) non-targeting gRNA . We mixed the population in equal amounts and proceeded to sort for high CRISPR-Cas activity using a fluorescence-activated cell sorter. Consequent sequencing of the sorted droplets revealed high enrichment of the expected genotypes when compared with the pre-sorted droplets. Finally, we describe the necessary steps needed for the future adaption of our system into a genuine directed evolution platform tailored towards CRISPR-Cas activity optimization.

In chapter 4, we set to engineer a thermophilic Cas9 orthologue for increased catalytic activity at lower temperatures. First, we identified the selection conditions that will allow enrichment of Cas9 variants with enhanced cleavage properties. We opted to use a positive selection scheme based on the guided cleavage of a plasmid vector expressing the bacterial ccdB-toxin. This toxin induces a growth defect when expressed in *E. coli*. Next, we constructed libraries consisting of thermophilic Cas9 variants generated by random mutagenesis. These libraries were subsequently used to select for variants capable of rescuing the ccdB-induced growth phenotype of *E. coli* cells at 37°C. Surviving colonies were pooled together and Cas9 vectors were high-throughput sequenced. Enriched Cas9 variants were cloned into plasmid vectors upstream of His purification-tag. After purification of protein, the enriched Cas9 variants were characterized *in vitro* by probing their catalytic activity in a range of temperatures and compared it to the activity of the *wild-type* protein. We ended up with two

variants (E655G/N696I, L7F) that demonstrated increased catalytic activity at lower temperatures in comparison with the wild type.

In chapter 5, we developed miniature base editors based on the small Cas12f effector nuclease from *Acidibacillus sulphuroxidans* (AsCas12f). We tested for adenosine and cytosine base editing by our engineered AsCas12f base editor in *E. coli* using a plasmid vector encoding the target sequence. We used a systematic protospacer design in order to elucidate the editing window of the tested base editors. We observed efficient base editing on the displaced strand using the AsCas12f fused to either a cytosine or adenosine deaminase enzyme. Surprisingly, we also observed significant base conversions on the target strand, despite its theoretical unavailability due to its participation in forming the DNA/RNA heteroduplex. Additionally, substantial base conversion was detected outside of the protospacer region. Taken together, these results open new opportunities to broaden the applicability of Cas12-mediated base editing. We end the chapter with a discussion on possible factors that allow base conversion on the target strand and outside of the protospacer.

General Discussion

The optimization of CRISPR-Cas systems has not only resulted in a variety of improved systems tailored for specific applications but also provided fundamental insights into their natural properties. In this thesis, apart from developing new genome editing tools based on a miniature member of the CRISPR-Cas family, we evolved Cas9 variants with enhanced catalytic activity at lower temperatures and set off to develop a novel directed evolution platform. Here we discuss final experimental results of the thesis and underscore next steps that would promote and facilitate relevant scientific research. We also discuss discrepancies between *in vitro* and *in vivo* cleavage assays observed between Cas9 mutants characterized for increased catalytic activity in lower temperatures and speculate on underlying causes. We continue by illustrating the state-of-the-art genome editing technology and referring to CRISPR-diagnostics too. Finally, we put forward ethical considerations that arise from the rapid advance of genome editing technology.

Connecting CRISPR-activity to increased fluorescence by artificial genetic circuits

In **Chapter 3** we pursued the development of an *in vitro*, high throughput screening method for CRISPR-Cas activity. The screening system was based on fluorescent alterations of a reporter protein, indicative for RNA-guided DNA-interference. We first tested previously described genetic circuits in our laboratory setup. These circuits connect guided DNA interference with lower levels of GFP expression, thereby reducing protein fluorescence levels. (156).

We next sought to adapt for genetic circuits that connect guided DNA-interference with increased levels of GFP expression. We included the cI protein of phage lambda that binds to the operator sites (O_L) of the P70a promoter inhibiting gene transcription (207). We aimed to avoid excessive levels of cI protein giving time for SpCas9 and gRNA which are constitutively expressed to assemble into the RNP complex. Therefore, we expressed the *cI* gene from the P70a promoter, anticipating that cI expression would be self-regulated through autorepression. We observed sufficient repression of GFP expression and significant separation of fluorescent levels between targeting and non- targeting spacers **(Figure 6.1B)**. To improve separation between targeting and non-targeting gRNAs, we also expressed the *cI-ssrA* gene from the strong J23119 promoter that lacks any cI operator sites and therefore is not inhibited by the cI protein. We observed better separation of fluorescent levels using the J23119_cI-ssrA expression construct due to increased repression of GFP expression **(Figure 6.1 C)**. Therefore, we can conclude that autorepression of cI is not beneficial for the anticipated gene circuit and that straightforward expression of the cI protein allows for better separation between targeting and non-targeting spacers.

The expression constructs used in **Chapter 3** were assembled in plasmid vectors. Constructs that expressed genes from the P70a promoter had to be transformed to a different host strain (*E. coli* KL740 cI857+), since the strong and constitutive transcription driven by the P70a promoter triggers plasmid instability (208). Alternatively, expression constructs could also be directly assembled as synthetic gene fragments and amplified by routine PCR reactions. myTXTL is an *E. coli* cell-free extract and linear DNA fragments will suffer from poor stability due to the presence of RecBCD complex (208). The issue is easily circumvented by the addition of GamS protein or short Chi-site-containing oligonucleotides that will saturate the RecBCD complex protecting the expression constructs from degradation (209,210). The combination of synthetic gene fragments with DNA

degradation counter measures would surely facilitate and accelerate the prototyping of gene circuits in *E. coli* cell-free extracts.

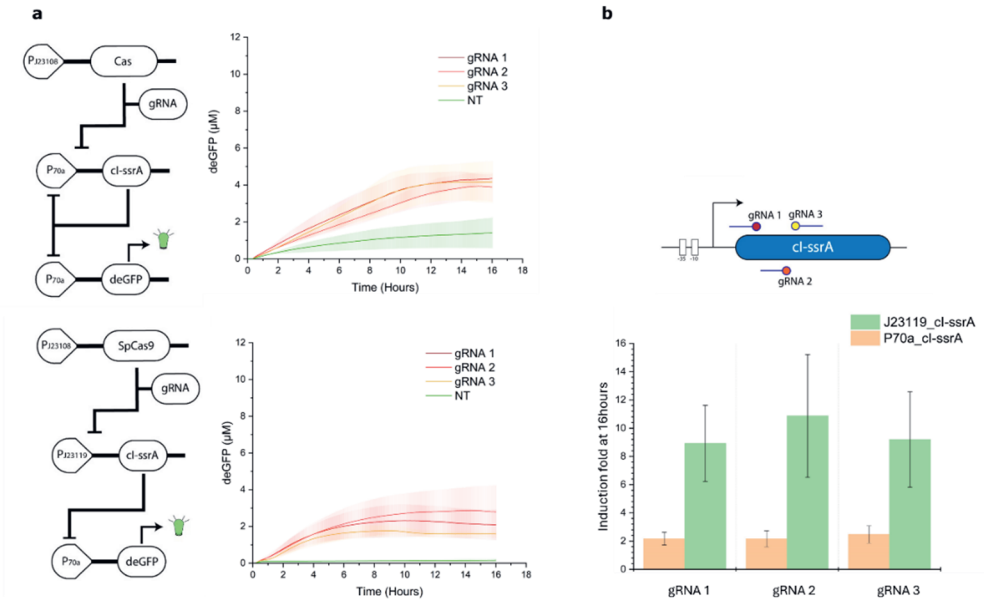


Figure 6.1. Fluorescent kinetic measurements of GFP concentration. **(A)** Schematics of genes circuits linking CRISPR-Cas activity to increased protein fluorescence (left) and corresponding protein fluorescence measurements (right). All kinetic measurements were made by supplementing myTXTL cell-free extract with 3nM: J23108_SpCas9, 1nM: P70a/J23119_cI-ssrA, 1nM: P70a_deGFP and 1nM J23119_gRNA. NT: non-targeting gRNA. Error bars represent the deviation from two independent repeats. **(B)** Targeting sites on the *cI-ssrA* expression template per gRNA and induction fold of protein fluorescence at 16hours. Error bars represent the deviation from two independent repeats

Cas9 mutants demonstrate different catalytic activity between *in vivo* and *in vitro* assays

In Chapter 4 we performed *in vitro* and *in vivo* cleavage assays with ThermoCas9 variants previously enriched for increased catalytic activity. We observed a discrepancy between *in vitro* and *in vivo* cleavage assays. Mutant E665G/N696I failed to rescue the toxin-induced growth phenotype when challenged with protospacer flanked by a 5'-N_4CAAA PAM on the toxin-expressing plasmid, while the L7F variant sufficiently rescued cell growth **(Figure 6.2 A)**. When the variants were tested *in vitro* for cleavage of a linear DNA substrate with a protospacer flanked by a 5'-N_4CAAA PAM, we observed higher catalytic activity from the mutant E665G/N696I **(Figure 6.2 B)**. Possible causes of the described discrepancy could be related to host-specific factors such as protein expression and the topology of the target DNA (i.e. negatively supercoiled plasmid versus a linear DNA substrate). Negative supercoiled DNA has been reported to facilitate DNA unwinding and hence R-loop formation (211). Moreover, it has been reported that the first 25 bases following the start codon play incremental role in the prediction of translation efficiency (212). The frequency per thousand of the phenylalanine used codon (TTT= 22.1) is significantly higher than leucine (CTT= 11.9). That factor would not play a role in the *in vitro* cleavage assay as ThermoCas9 has been expressed and provided as purified protein. In the *in vivo* assay, ThermoCas9 is antagonizing CcdB to rescue the growth phenotype. Faster expression and association with the 5'-N_4CAAA target substrate by the L7F mutant could be giving an edge to the L7F mutant over the E655G/N696I against 5'-N_4CAAA target substrate. These factors don't take place in the *in vitro* assay where the E655G/N696I mutant demonstrates higher catalytic activity than the L7F mutant.

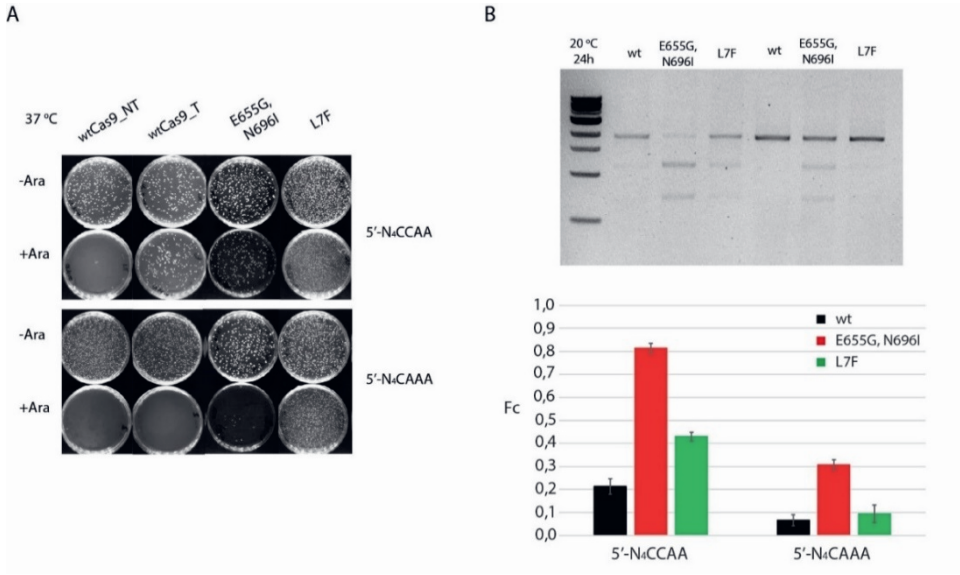


Figure 6.2. **Characterization of enriched ThermoCas9 variants by (A)** *in vivo* cleavage assay of ccdB-encoding plasmid vector and **(B)** *in vitro* cleavage assays of linear DNA template.

Towards efficient base conversion by miniature base editors in mammal cell lines

In **Chapter 5** we developed and characterized novel, miniature base editors based on the small AsCas12f effector nuclease in *E. coli*. Next to the anticipated editing on the non-target strand, we strikingly also observed \significant base editing on the target DNA strand. Cas12f effectors are significantly smaller proteins ranging from 400-700 aa (213). We speculated that the small size and melting of the dsDNA beyond the sgRNA/DNA heteroduplex could leave the target strand exposed to deamination. Additionally, Cas12f effectors form asymmetrical homodimers (136,196). Consequently, Cas12f-base editors will recruit one additional

copy of the deaminating enzyme in the proximity of the DNA-substrate increasing the chances of base edition even in unexpected places.

We further tested AsCas12f1 for its genome editing abilities in the mammalian HEK293T cell line. We only detected indel-formation but no base editing (Figure 6.3). A possible explanation is that the base editing efficiency was below the limit of detection for the read-out we used. Indeed, low base editing rates, which would be undetectable by Sanger's Sequencing, have been reported with Cas12f-base editors before (137,214,215). Meanwhile, protein and sgRNA engineering efforts have created Cas12f-systems with enhanced activity in mammalian cell lines, offering opportunities to increase base editing efficiencies with our AsCas12f1 base editor (198). Indeed, a Un1Cas12f orthologue engineered for enhanced substrate affinity catalysed highly efficient base deamination (216). It won't be long before for a similar approach might improve the editing capabilities of AsCas12f too.

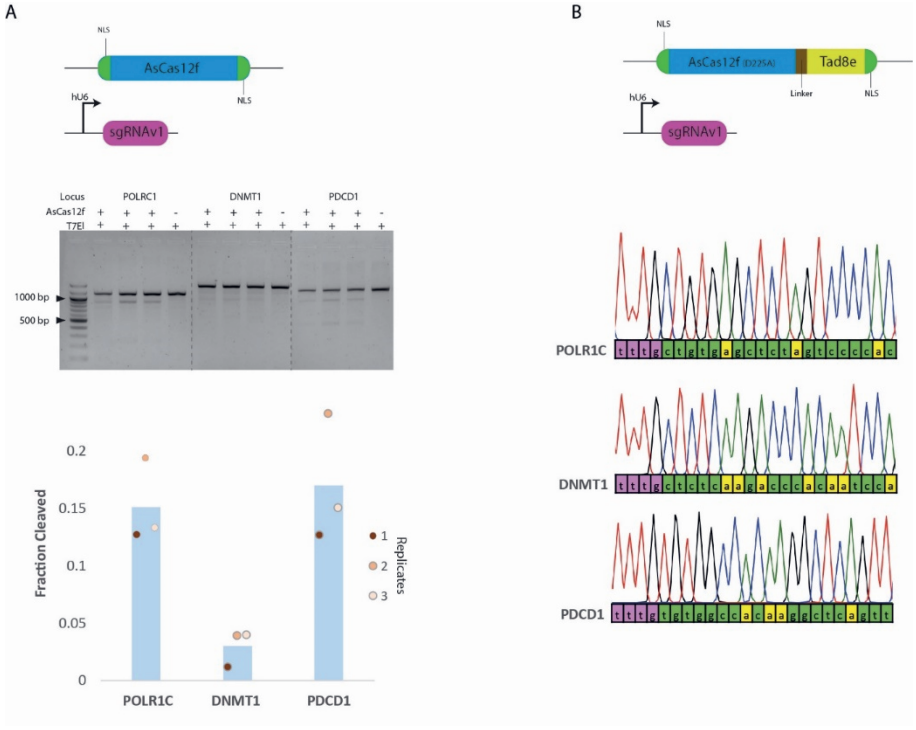


Figure 6.3. **Genome editing by AsCas12f in mammalian genome.**
A) Indel formation by AsCas12f1. Top: Constructs schematic encoding AsCas12f and sgRNAv1 (144). Middle: agarose gel of genomic target locus amplified by PCR and subjected to T7EI endonuclease assay. Bottom: Bar graph of T7EI cleaved fractions for each target locus, **B)** G->A Base Editing. Top: Constructs schematic encoding AsCas12f-Tad8e and sgRNAv1. Bottom: Sanger's Sequencing chromatographs of target genome locus amplicons. Purple boxes emphasize the PAM sequence and yellow boxes show candidate bases for A->G base conversion

State of the art gene editing tools and future perspectives

Precision, easiness of use and versatility are key characteristics that made CRISPR-Cas genome editing a successful and widely applied technique. Induction of DDSBs catalysed by Cas nucleases guided by an RNA molecule has been the basis of CRISPR-Cas genome engineering. Spontaneous repair of the DDSBs by the cell's DNA repair machinery, through the NHEJ and/or the HDR repair pathway, allows for targeted gene knock out and/or for insertion of desired changes, respectively. State of the art genome editing tools such as base and prime editors allow for the conversion of single bases and the introduction of precise changes without the need for DDSBs. These innovations promote the safety of CRISPR-Cas technology by reducing adverse effects, such as chromosomal translocations (186). Despite these improvements, Cas-independent off-target effects should still be under consideration (217,218). Nevertheless, the prospect of safe and precise gene therapy remains highly promising (219,220). Finally, guided insertion of large DNA fragments without DDSBs was made possible with the discovery of CRISPR-associated transposases (CASTs). CASTs systems natively include a set of cargo genes flanked on the one end by a locus of transposase 7-like proteins and by a CRISPR-array and a Cas-effector on the other end (221). The Cas-effector is catalytically inactive but still capable of target binding and shaping R-loop structures. First experiments were performed in *E. coli* and guided transposition of 2.2 kbp fragment was demonstrated at high frequencies (>50%) (221). However, the Type V-K

CASTs demonstrated low to no transposition activity when tested in mammalian cell lines (222). Promising results though, were shown with a Type I-F CAST system. When combined with the bacterial protease ClpX it demonstrated transposition frequence approximately at 1% (222). Nevertheless, the frequency of guided insertion by CAST systems remains comparable to HDR and significantly lags behind the editing efficiency achieved with base and prime editors.

One important challenge that state-of-the-art genome editing tools are still facing is in vivo gene delivery. Adeno-associated viral vectors (AAVs) have been granted safety clearance to be clinically applied as gene delivery vehicles (223). However, they come with a DNA capacity limit (~4.7 kbp) which is exceeded by Cas9 base and prime editors. Only the coding sequence of the SpCas9-BE3, without considering regulatory modules such as promoters and inverted terminal repeats, lengths at ~5.0 kbp. One solution for this issue is splitting the Cas9-editor into two parts. However, the approach is not ideal because the target cell must now be infected by two different viral capsids to reconstitute the gene editor (224). Miniature Cas-effectors such as those of Type V-F and Type V-M can overcome the limited capacity and get packaged into a single viral capsid. The analogous length of the coding sequence of a Cas12f1-BE3 or a Cas12m-BE3 would be approximately 2.3 kbp and 2.8 kbp, respectively (34) Members of the subtype V-F and subtype V-M have already been adapted to base editing tools and the first demonstration of gene delivery of a type V-F base editor by a single viral capsid was recently made (146,216).

CRISPR in diagnostics

CRISPR-Cas has been the basis of innovation in diagnostics too. First, subtype V-A effectors (Cas12a) were found to catalyse indiscriminate trans-cleavage of short ssDNA molecules (225). Trans-nuclease activity does not require the presence of a PAM motif or spacer complementarity, but it follows successful recognition and cis-cleavage of a target DNA template (activator) (226). Similar principles have been discovered for Type VI and

Type III Cas-effectors which catalyse cis-cleavage of RNA targets and trans-cleavage of short ssRNA oligonucleotides. (227,228).

Indiscriminate trans-cleavage has been the foundation for developing nucleic acid detections platforms based on CRISPR-Cas systems. Trans-nuclease activity was combined with short oligonucleotides, probed with a fluorophore/quencher pair. Guided interference with the target nucleic acid (e.g. virus DNA or RNA) acts as activator of the indiscriminate nuclease activity which separates the fluorophore from the quencher, generating a fluorescent signal upon detection of the target nucleic acid. Following this approach nucleic acid detection in the attomolar range has been reported (226,227,229).

Recently, trans-nuclease activity for Type II systems was also reported. In difference with Type V systems, Cas9 displayed strong preference for poly(A) and poly(C) ssDNA oligos but that does not restrict its application for nucleic acid detection purposes (230). However, the reasons behind this preference remain unsolved.

Bioethical Considerations

CRISPR-Cas systems have undoubtedly revolutionized the field of genome engineering with implications in fundamental genetic research, in microbial, plant and animal biotechnology and in human biomedicine (130). Bottlenecks such low specificity and/or catalytic activity that narrow the scope of CRISPR-Cas are well addressed through protein engineering methods such as directed evolution (231), as also showcased in this thesis.

CRISPR-Cas technology has already entered the sphere of therapeutical biotechnology as regulatory approvals have been granted for CRISPR-Cas9 gene therapy of sickle cell disease and β-thalassemia (232). The current cost of the treatment under the name CasGevy comes to a cost of $2 million per patient (233). Even if costs should be expected to drop, there is always the risk of exacerbating social inequalities as access to the state-of-the-art

CRISPR-Cas technology might be limited only to wealthier nations and individuals. Widespread use of the CRISPR-Cas technology, due to easiness of application, offers also space for non-therapeutic enhancements such as increased physical abilities. While scientists should be have the freedom to explore the potential of CRISPR-Cas in every direction that does not cause harm, the use of the technology beyond the academic sphere should be meticulously regulated, and the introduction of gene editing technologies into non-academic contexts should follow public debate.

Unfortunately, a bad precedence has already been set by the birth of two individuals that were genome edited for HIV resistance. Even though the goal of HIV resistance is benign, the execution of the editing and the lack of transparency surrounding the experiment were heavily criticized by experts of the field. Moreover, it is speculated that effects other than HIV resistance might also be present (234,235). The venture was regarded as premature and lacking transparency and clearly makes the showcase for the need of clearcut regulations and public debate before conducting such trials (236).

Conclusions

CRISPR-Cas systems undoubtedly earn a place in the great human milestones as acknowledged by the reward of the Nobel prize (https://www.nobelprize.org/prizes/chemistry/2020/press-release/). To further explore their potential, significant efforts in protein engineering of Cas-effectors have been made.

This thesis aimed to establish a new platform for directed evolution of CRISPR-Cas systems using cell-free methods. Despite encountering technical challenges, we successfully developed the foundational components of this envisioned platform. However, the functional adaptation of these components remains to be seen, particularly in discovering

CRISPR-Cas variants with desired properties. To broaden the scope of CRISPR-Cas systems, we conducted directed evolution using established methodologies, resulting in the enrichment of variants exhibiting enhanced catalytic properties *in vivo.* Finally, we contributed in the development of miniature gene editors whose small size facilitates their delivery.

Having add a brick to the house of the genome editing technology, it is now our responsibility to express our hopes that society and human legislation will utilize this exceedingly promising technology to the equitable improvement of human living conditions.

References

1. Aristotle., Cresswell R, Schneider JG. Aristotle's History of animals. In ten books. London: G. Bell; 1878.

2. Darwin C. The Origin of Species. Cambridge University Press; 2009.

3. Avery OT, MacLeod CM, McCarty M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. Journal of Experimental Medicine. 1944 Feb 1;79(2):137–58.

4. Hershey AD, Chase M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. Journal of General Physiology. 1952 Sep 20;36(1):39–56.

5. WATSON JD, CRICK FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 1953 Apr 25;171(4356):737–8.

6. CRICK F. Central Dogma of Molecular Biology. Nature. 1970 Aug;227(5258):561–3.

7. Andersson AF, Banfield JF. Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. Science (1979). 2008 May 23;320(5879):1047–50.

8. Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D. Germ Warfare in a Microbial Mat Community: CRISPRs Provide Insights into the Co-Evolution of Host and Viral Genomes. PLoS One. 2009 Jan 9;4(1):e4169.

9. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020 Feb 19;18(2):67–83.

10. Altae-Tran H, Kannan S, Suberski AJ, Mears KS, Demircioglu FE, Moeller L, et al. Uncovering the functional diversity of rare CRISPR-Cas systems with deep terascale clustering. Science (1979). 2023 Nov 24;382(6673).

11. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin E V., van der Oost J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science (1979). 2016 Aug 5;353(6299).

12. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (1979). 2012 Aug 17;337(6096):816–21.

13. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011 Mar 30;471(7340):602–7.

14. Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. Annu Rev Biophys. 2017 May 22;46(1):505–29.

15. O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature. 2014 Dec 28;516(7530):263–6.

16. Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature. 2016 Apr 20;532(7600):517–21.

17. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nat Biotechnol. 2017 Jan 5;35(1):31–4.

18. Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, et al. The crystal structure of Cpf1 in complex with CRISPR RNA. Nature. 2016 Apr 20;532(7600):522–6.

19. Swarts DC, van der Oost J, Jinek M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. Mol Cell. 2017 Apr;66(2):221-233.e4.

20. Swarts DC, Jinek M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. Mol Cell. 2019 Feb;73(3):589-600.e4.

21. Pardo B, Gómez-González B, Aguilera A. DNA Repair in Mammalian Cells. Cellular and Molecular Life Sciences. 2009 Mar 21;66(6):1039–56.

22. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol. 2017 Aug 17;18(8):495–506.

23. Jasin M, Rothstein R. Repair of Strand Breaks by Homologous Recombination. Cold Spring Harb Perspect Biol. 2013 Nov 1;5(11):a012740–a012740.

24. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol. 2016 Mar 20;34(3):339–44.

25. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell. 2013 Feb;152(5):1173–83.

26. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat Methods. 2013 Oct 25;10(10):973–6.

27. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May 19;533(7603):420–4.

28. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. 2017 Nov 23;551(7681):464–71.

29. Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019 Dec 5;576(7785):149–57.

30. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Mol Ther Nucleic Acids. 2015;4:e264.

31. Mir A, Edraki A, Lee J, Sontheimer EJ. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. ACS Chem Biol. 2018 Feb 16;13(2):357–65.

32. Creutzburg SCA, Wu WY, Mohanraju P, Swartjes T, Alkan F, Gorodkin J, et al. Good guide, bad guide: spacer sequence-dependent cleavage efficiency of Cas12a. Nucleic Acids Res. 2020 Apr 6;48(6):3228–43.

33. Karvelis T, Bigelyte G, Young JK, Hou Z, Zedaveinyte R, Budre K, et al. PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Res. 2020 May 21;48(9):5016–23.

34. Omura SN, Nakagawa R, Südfeld C, Villegas Warren R, Wu WY, Hirano H, et al. Mechanistic and evolutionary insights into a type V-M CRISPR–Cas effector enzyme. Nat Struct Mol Biol. 2023 Aug 17;30(8):1172–82.

35. Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar V V, Zheng Z, et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 2015 Dec 2;33(12):1293–8.

36. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y hoon, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nat Commun. 2018 Aug 6;9(1):3048.

37. Ma E, Chen K, Shi H, Stahl EC, Adler B, Trinidad M, et al. Improved genome editing by an engineered CRISPR-Cas12a. Nucleic Acids Res. 2022 Dec 9;50(22):12689–701.

38. Xu X, Chemparathy A, Zeng L, Kempton HR, Shang S, Nakamura M, et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol Cell. 2021 Oct;81(20):4333-4345.e4.

39. Chen Z. A highly sensitive selection method for directed evolution of homing endonucleases. Nucleic Acids Res. 2005 Oct 12;33(18):e154–e154.

40. Bouzetos E, Ganar KA, Mastrobattista E, Deshpande S, van der Oost J. (R)evolution-on-a-chip. Trends Biotechnol. 2022 Jan;40(1):60–76.

41. Odegrip R, Coomber D, Eldridge B, Hederer R, Kuhlman PA, Ullman C, et al. CIS display: *In vitro* selection of peptides from libraries of protein–DNA complexes. Proceedings of the National Academy of Sciences. 2004 Mar 2;101(9):2806–10.

42. Seelig B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. Nat Protoc. 2011 Apr 31;6(4):540–52.

43. Zahnd C, Amstutz P, Plückthun A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. Nat Methods. 2007 Mar 27;4(3):269–79.

44. Crook N, Abatemarco J, Sun J, Wagner JM, Schmitz A, Alper HS. In vivo continuous evolution of genes and pathways in yeast. Nat Commun. 2016 Oct 17;7(1):13051.

45. Gaudelli NM, Lam DK, Rees HA, Solá-Esteves NM, Barrera LA, Born DA, et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nat Biotechnol. 2020 Jul 13;38(7):892–900.

46. Smith GP. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. Science (1979). 1985 Jun 14;228(4705):1315–7.

47. McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains. Nature. 1990 Dec;348(6301):552–4.

48. Schmidt-Dannert C, Umeno D, Arnold FH. Molecular breeding of carotenoid biosynthetic pathways. Nat Biotechnol. 2000 Jul;18(7):750–3.

49. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nat Biotechnol. 2018 Mar 29;36(3):265–71.

50. Wu CH, Liu IJ, Lu RM, Wu HC. Advancement and applications of peptide phage display technology in biomedical science. J Biomed Sci. 2016 Dec 19;23(1):8.

51. Packer MS, Liu DR. Methods for the directed evolution of proteins. Nat Rev Genet. 2015 Jul 9;16(7):379–94.

52. Colin PY, Zinchenko A, Hollfelder F. Enzyme engineering in biomimetic compartments. Curr Opin Struct Biol. 2015 Aug;33:42–51.

53. Markel U, Essani KD, Besirlioglu V, Schiffels J, Streit WR, Schwaneberg U. Advances in ultrahigh-throughput screening for directed enzyme evolution. Chem Soc Rev. 2020;49(1):233–62.

54. Mayr LM, Fuerst P. The Future of High-Throughput Screening. SLAS Discovery. 2008 Jul;13(6):443–8.

55. Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. Nat Biotechnol. 1998 Jul;16(7):652–6.

56. Silverman AD, Karim AS, Jewett MC. Cell-free gene expression: an expanded repertoire of applications. Nat Rev Genet. 2020 Mar 28;21(3):151–70.

57. Griffiths AD, Tawfik DS. Miniaturising the laboratory in emulsion droplets. Trends Biotechnol. 2006 Sep;24(9):395–402.

58. Miller OJ, Bernath K, Agresti JJ, Amitai G, Kelly BT, Mastrobattista E, et al. Directed evolution by in vitro compartmentalization. Nat Methods. 2006 Jul;3(7):561–70.

59. Weng L, Spoonamore JE. Droplet Microfluidics-Enabled High-Throughput Screening for Protein Engineering. Micromachines (Basel). 2019 Oct 29;10(11):734.

60. Griffiths AD. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. EMBO J. 2003 Jan 2;22(1):24–35.

61. Autour A, Ryckelynck M. Ultrahigh-Throughput Improvement and Discovery of Enzymes Using Droplet-Based Microfluidic Screening. Micromachines (Basel). 2017 Apr 18;8(4):128.

62. Convery N, Gadegaard N. 30 years of microfluidics. Micro and Nano Engineering. 2019 Mar;2:76–91.

63. Shang L, Cheng Y, Zhao Y. Emerging Droplet Microfluidics. Chem Rev. 2017 Jun 28;117(12):7964–8040.

64. Wang BL, Ghaderi A, Zhou H, Agresti J, Weitz DA, Fink GR, et al. Microfluidic high-throughput culturing of single cells for selection based on extracellular metabolite production or consumption. Nat Biotechnol. 2014 May 6;32(5):473–8.

65. Mazutis L, Araghi AF, Miller OJ, Baret JC, Frenz L, Janoshazi A, et al. Droplet-Based Microfluidic Systems for High-Throughput Single DNA Molecule Isothermal Amplification and Analysis. Anal Chem. 2009 Jun 15;81(12):4813–21.

66. van Vliet LD, Colin PY, Hollfelder F. Bioinspired genotype–phenotype linkages: mimicking cellular compartmentalization for the engineering of functional proteins. Interface Focus. 2015 Aug 6;5(4):20150035.

67. Shi Z, Lai X, Sun C, Zhang X, Zhang L, Pu Z, et al. Step emulsification in microfluidic droplet generation: mechanisms and structures. Chemical Communications. 2020;56(64):9056–66.

68. Tay Y, Ho C, Drőge P, Ghadessy FJ. Selection of bacteriophage λ integrases with altered recombination specificity by in vitro compartmentalization. Nucleic Acids Res. 2010 Mar;38(4):e25–e25.

69. Takeuchi R, Choi M, Stoddard BL. Redesign of extensive protein–DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. Proceedings of the National Academy of Sciences. 2014 Mar 18;111(11):4061–6.

70. Ellefson JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat Biotechnol. 2014 Jan 3;32(1):97–101.

71. Czapinska H, Siwek W, Szczepanowski RH, Bujnicki JM, Bochtler M, Skowronek KJ. Crystal Structure and Directed Evolution of Specificity of NlaIV Restriction Endonuclease. J Mol Biol. 2019 May;431(11):2082–94.

72. Chung CHY, Cui B, Song R, Liu X, Xu X, Yao S. Scalable Production of Monodisperse Functional Microspheres by Multilayer Parallelization of High Aspect Ratio Microfluidic Channels. Micromachines (Basel). 2019 Sep 10;10(9):592.

73. Shim J uk, Ranasinghe RT, Smith CA, Ibrahim SM, Hollfelder F, Huck WTS, et al. Ultrarapid Generation of Femtoliter Microfluidic Droplets for Single-Molecule-Counting Immunoassays. ACS Nano. 2013 Jul 23;7(7):5955–64.

74. Abate AR, Weitz DA. Faster multiple emulsification with drop splitting. Lab Chip. 2011;11(11):1911.

75. Chong D, Liu X, Ma H, Huang G, Han YL, Cui X, et al. Advances in fabricating double-emulsion droplets and their biomedical applications. Microfluid Nanofluidics. 2015 Nov 9;19(5):1071–90.

76. Zinchenko A, Devenish SRA, Kintses B, Colin PY, Fischlechner M, Hollfelder F. One in a Million: Flow Cytometric Sorting of Single Cell-Lysate Assays in Monodisperse Picolitre Double Emulsion Droplets for Directed Evolution. Anal Chem. 2014 Mar 4;86(5):2526–33.

77. Larsen AC, Dunn MR, Hatch A, Sau SP, Youngbull C, Chaput JC. A general strategy for expanding polymerase function by droplet microfluidics. Nat Commun. 2016 Apr 5;7(1):11235.

78. Rampioni G, D'Angelo F, Leoni L, Stano P. Gene-Expressing Liposomes as Synthetic Cells for Molecular Communication Studies. Front Bioeng Biotechnol. 2019 Jan 17;7.

79. Spoelstra WK, Deshpande S, Dekker C. Tailoring the appearance: what will synthetic cells look like? Curr Opin Biotechnol. 2018 Jun;51:47–56.

80. Fujii S, Matsuura T, Sunami T, Kazuta Y, Yomo T. In vitro evolution of α-hemolysin using a liposome display. Proceedings of the National Academy of Sciences. 2013 Oct 15;110(42):16796–801.

81. Walde P, Cosentino K, Engel H, Stano P. Giant Vesicles: Preparations and Applications. ChemBioChem. 2010 May 3;11(7):848–65.

82. Sunami T, Sato K, Matsuura T, Tsukada K, Urabe I, Yomo T. Femtoliter compartment in liposomes for in vitro selection of proteins. Anal Biochem. 2006 Oct;357(1):128–36.

83. Deshpande S, Dekker C. On-chip microfluidic production of cell-sized liposomes. Nat Protoc. 2018 May 29;13(5):856–74.

84. Deshpande S, Caspi Y, Meijering AEC, Dekker C. Octanol-assisted liposome assembly on chip. Nat Commun. 2016 Jan 22;7(1):10447.

85. Deng NN, Yelleswarapu M, Huck WTS. Monodisperse Uni- and Multicompartment Liposomes. J Am Chem Soc. 2016 Jun 22;138(24):7584–91.

86. Weiss M, Frohnmayer JP, Benk LT, Haller B, Janiesch JW, Heitkamp T, et al. Sequential bottom-up assembly of mechanically stabilized synthetic cells by microfluidics. Nat Mater. 2018 Jan 1;17(1):89–96.

87. Staufer O, Antona S, Zhang D, Csatári J, Schröter M, Janiesch JW, et al. Microfluidic production and characterization of biofunctionalized giant unilamellar vesicles for targeted intracellular cargo delivery. Biomaterials. 2021 Jan;264:120203.

88. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. Proceedings of the National Academy of Sciences. 2010 Mar 2;107(9):4004–9.

89. van Loo B, Heberlein M, Mair P, Zinchenko A, Schüürmann J, Eenink BDG, et al. High-Throughput, Lysis-Free Screening for Sulfatase Activity Using *Escherichia coli* Autodisplay in Microdroplets. ACS Synth Biol. 2019 Dec 20;8(12):2690–700.

90. Madzak C. Yarrowia lipolytica: recent achievements in heterologous protein expression and pathway engineering. Appl Microbiol Biotechnol. 2015 Jun 7;99(11):4559–77.

91. Beneyton T, Thomas S, Griffiths AD, Nicaud JM, Drevelle A, Rossignol T. Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast Yarrowia lipolytica. Microb Cell Fact. 2017 Dec 31;16(1):18.

92. Beneyton T, Coldren F, Baret JC, Griffiths AD, Taly V. CotA laccase: high-throughput manipulation and analysis of recombinant enzyme libraries expressed in *E. coli* using droplet-based microfluidics. Analyst. 2014;139(13):3314–23.

93. Vallejo D, Nikoomanzar A, Paegel BM, Chaput JC. Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. ACS Synth Biol. 2019 Jun 21;8(6):1430–40.

94. Ma F, Chung MT, Yao Y, Nidetz R, Lee LM, Liu AP, et al. Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. Nat Commun. 2018 Mar 12;9(1):1030.

95. Debon A, Pott M, Obexer R, Green AP, Friedrich L, Griffiths AD, et al. Ultrahigh-throughput screening enables efficient single-round oxidase remodelling. Nat Catal. 2019 Sep 13;2(9):740–7.

96. Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD, et al. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. Nat Chem. 2017 Jan 29;9(1):50–6.

97. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, et al. Cell-free translation reconstituted with purified components. Nat Biotechnol. 2001 Aug 1;19(8):751–5.

98. Dopp BJL, Tamiev DD, Reuel NF. Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support in vitro protein synthesis in E. coli extract. Biotechnol Adv. 2019 Jan;37(1):246–58.

99. Courtois F, Olguin LF, Whyte G, Bratton D, Huck WTS, Abell C, et al. An Integrated Device for Monitoring Time-Dependent in vitro Expression From Single Genes in Picolitre Droplets. ChemBioChem. 2008 Feb 15;9(3):439–46.

100. Woronoff G, Ryckelynck M, Wessel J, Schicke O, Griffiths AD, Soumillion P. Activity-Fed Translation (AFT) Assay: A New High-Throughput Screening Strategy for Enzymes in Droplets. ChemBioChem. 2015 Jun 15;16(9):1343–9.

101. Stapleton JA, Swartz JR. Development of an In Vitro Compartmentalization Screen for High-Throughput Directed Evolution of [FeFe] Hydrogenases. PLoS One. 2010 Dec 6;5(12):e15275.

102. Fallah-Araghi A, Baret JC, Ryckelynck M, Griffiths AD. A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. Lab Chip. 2012;12(5):882.

103. Holstein JM, Gylstorff C, Hollfelder F. Cell-free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. ACS Synth Biol. 2021 Feb 19;10(2):252–7.

104. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990 Aug;346(6287):818–22.

105. Ryckelynck M, Baudrey S, Rick C, Marin A, Coldren F, Westhof E, et al. Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. RNA. 2015 Mar;21(3):458–69.

106. Paige JS, Wu KY, Jaffrey SR. RNA Mimics of Green Fluorescent Protein. Science (1979). 2011 Jul 29;333(6042):642–6.

107. Autour A, Westhof E, Ryckelynck M. iSpinach: a fluorogenic RNA aptamer optimized for *in vitro* applications. Nucleic Acids Res. 2016 Apr 7;44(6):2491–500.

108. Jia Y, Ren Y, Liu W, Hou L, Tao Y, Hu Q, et al. Electrocoalescence of paired droplets encapsulated in double-emulsion drops. Lab Chip. 2016;16(22):4313–8.

109. Abate AR, Hung T, Mary P, Agresti JJ, Weitz DA. High-throughput injection with microfluidics using picoinjectors. Proceedings of the National Academy of Sciences. 2010 Nov 9;107(45):19163–6.

110. O'Donovan B, Eastburn DJ, Abate AR. Electrode-free picoinjection of microfluidic drops. Lab Chip. 2012;12(20):4029.

111. Ghadessy FJ, Ong JL, Holliger P. Directed evolution of polymerase function by compartmentalized self-replication. Proceedings of the National Academy of Sciences. 2001 Apr 10;98(8):4552–7.

112. Zhu B, Mizoguchi T, Kojima T, Nakano H. Ultra-High-Throughput Screening of an In Vitro-Synthesized Horseradish Peroxidase Displayed on Microbeads Using Cell Sorter. PLoS One. 2015 May 20;10(5):e0127479.

113. Tran DT, Cavett VJ, Dang VQ, Torres HL, Paegel BM. Evolution of a mass spectrometry-grade protease with PTM-directed specificity. Proceedings of the National Academy of Sciences. 2016 Dec 20;113(51):14686–91.

114. Lindenburg L, Hollfelder F. "NAD-display": Ultrahigh-Throughput in Vitro Screening of NAD(H) Dehydrogenases Using Bead Display and Flow Cytometry. Angewandte Chemie International Edition. 2021 Apr 12;60(16):9015–21.

115. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. High-Throughput Screening of Enzyme Libraries: In Vitro Evolution of a β-Galactosidase by Fluorescence-Activated Sorting of Double Emulsions. Chem Biol. 2005 Dec;12(12):1291–300.

116. Uyeda A, Watanabe T, Kato Y, Watanabe H, Yomo T, Hohsaka T, et al. Liposome-Based in Vitro Evolution of Aminoacyl-tRNA Synthetase for Enhanced Pyrrolysine Derivative Incorporation. ChemBioChem. 2015 Aug 17;16(12):1797–802.

117. Fischlechner M, Schaerli Y, Mohamed MF, Patil S, Abell C, Hollfelder F. Evolution of enzyme catalysts caged in biomimetic gel-shell beads. Nat Chem. 2014 Sep 20;6(9):791–6.

118. Baret JC, Miller OJ, Taly V, Ryckelynck M, El-Harrak A, Frenz L, et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. Lab Chip. 2009;9(13):1850.

119. Sung YJ, Kim JYH, Choi H Il, Kwak HS, Sim SJ. Magnetophoretic sorting of microdroplets with different microalgal cell densities for rapid isolation of fast growing strains. Sci Rep. 2017 Sep 4;7(1):10390.

120. Goto H, Kanai Y, Yotsui A, Shimokihara S, Shitara S, Oyobiki R, et al. Microfluidic screening system based on boron-doped diamond electrodes and dielectrophoretic sorting for directed evolution of NAD(P)-dependent oxidoreductases. Lab Chip. 2020;20(4):852–61.

121. Sun S, Kennedy RT. Droplet Electrospray Ionization Mass Spectrometry for High Throughput Screening for Enzyme Inhibitors. Anal Chem. 2014 Sep 16;86(18):9309–14.

122. Kempa EE, Smith CA, Li X, Bellina B, Richardson K, Pringle S, et al. Coupling Droplet Microfluidics with Mass Spectrometry for Ultrahigh-Throughput Analysis of Complex Mixtures up to and above 30 Hz. Anal Chem. 2020 Sep 15;92(18):12605–12.

123. Holland-Moritz DA, Wismer MK, Mann BF, Farasat I, Devine P, Guetschow ED, et al. Mass Activated Droplet Sorting (MADS) Enables High-Throughput Screening of Enzymatic Reactions at Nanoliter Scale. Angewandte Chemie International Edition. 2020 Mar 9;59(11):4470–7.

124. Gielen F, Hours R, Emond S, Fischlechner M, Schell U, Hollfelder F. Ultrahigh-throughput–directed enzyme evolution by absorbance-activated droplet sorting (AADS). Proceedings of the National Academy of Sciences. 2016 Nov 22;113(47).

125. Sciambi A, Abate AR. Accurate microfluidic sorting of droplets at 30 kHz. Lab Chip. 2015;15(1):47–51.

126. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011 Mar 30;471(7340):602–7.

127. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences. 2012 Sep 25;109(39).

128. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. J Bacteriol. 2008 Feb 15;190(4):1390–400.

129. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (N Y). 2009 Mar 1;155(3):733–40.

130. van der Oost J, Patinios C. The genome editing revolution. Trends Biotechnol. 2023 Mar;41(3):396–409.

131. Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. Science (1979). 2018 Aug 31;361(6405):866–9.

132. Kleinstiver BP, Prew MS, Tsai SQ, Topkar V V., Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 22;523(7561):481–5.

133. Lee JK, Jeong E, Lee J, Jung M, Shin E, Kim Y hoon, et al. Directed evolution of CRISPR-Cas9 to increase its specificity. Nat Commun. 2018 Aug 6;9(1):3048.

134. Zhang L, Zuris JA, Viswanathan R, Edelstein JN, Turk R, Thommandru B, et al. AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. Nat Commun. 2021 Jun 23;12(1):3908.

135. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016 Jan 28;529(7587):490–5.

136. Hino T, Omura SN, Nakagawa R, Togashi T, Takeda SN, Hiramoto T, et al. An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. Cell. 2023 Oct;186(22):4920-4935.e23.

137. Xu X, Chemparathy A, Zeng L, Kempton HR, Shang S, Nakamura M, et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol Cell. 2021 Oct;81(20):4333-4345.e4.

138. Chen Z. A highly sensitive selection method for directed evolution of homing endonucleases. Nucleic Acids Res. 2005 Oct 12;33(18):e154–e154.

139. Bouzetos E, Ganar KA, Mastrobattista E, Deshpande S, van der Oost J. (R)evolution-on-a-chip. Trends Biotechnol. 2022 Jan;40(1):60–76.

140. Nuti N, Rottmann P, Stucki A, Koch P, Panke S, Dittrich PS. A Multiplexed Cell-Free Assay to Screen for Antimicrobial Peptides in Double Emulsion Droplets. Angewandte Chemie International Edition. 2022 Mar 21;61(13).

141. Lee YF. Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. Nucleic Acids Res. 2002 Nov 15;30(22):4937–44.

142. Doi N. In vitro selection of restriction endonucleases by in vitro compartmentalization. Nucleic Acids Res. 2004 Jul 7;32(12):e95–e95.

143. Zheng Y, Roberts RJ. Selection of restriction endonucleases using artificial cells. Nucleic Acids Res. 2007 Jun;35(11):e83.

144. Wu Z, Zhang Y, Yu H, Pan D, Wang Y, Wang Y, et al. Programmed genome editing by a miniature CRISPR-Cas12f nuclease. Nat Chem Biol. 2021 Nov 2;17(11):1132–8.

145. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat Methods. 2013 Oct 25;10(10):973–6.

146. Wu WY, Mohanraju P, Liao C, Adiego-Pérez B, Creutzburg SCA, Makarova KS, et al. The miniature CRISPR-Cas12m effector binds DNA to block transcription. Mol Cell. 2022 Dec;82(23):4487-4502.e7.

147. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. 2013 Nov 17;8(11):2180–96.

148. Romanowsky MB, Abate AR, Rotem A, Holtze C, Weitz DA. High throughput production of single core double emulsions in a parallelized microfluidic device. Lab Chip. 2012;12(4):802.

149. Stucki A, Vallapurackal J, Ward TR, Dittrich PS. Droplet Microfluidics and Directed Evolution of Enzymes: An Intertwined Journey. Angewandte Chemie International Edition. 2021 Nov 8;60(46):24368–87.

150. Vallapurackal J, Stucki A, Liang AD, Klehr J, Dittrich PS, Ward TR. Ultrahigh-Throughput Screening of an Artificial Metalloenzyme using Double Emulsions**. Angewandte Chemie International Edition. 2022 Nov 25;61(48).

151. Chen Y, Liu X, Zhang C, Zhao Y. Enhancing and suppressing effects of an inner droplet on deformation of a double emulsion droplet under shear. Lab Chip. 2015;15(5):1255–61.

152. Neumann SM, Scherbej I, van der Schaaf US, Karbstein HP. Investigations on the influence of osmotic active substances on the structure of water in oil emulsions for the application as inner phase in double emulsions. Colloids Surf A Physicochem Eng Asp. 2018 Feb;538:56–62.

153. Brower KK, Carswell-Crumpton C, Klemm S, Cruz B, Kim G, Calhoun SGK, et al. Double emulsion flow cytometry with high-throughput

single droplet isolation and nucleic acid recovery. Lab Chip. 2020;20(12):2062–74.

154. Chen C, Ganar KA, Deshpande S. On-Chip Octanol-Assisted Liposome Assembly for Bioengineering. Journal of Visualized Experiments. 2023 Mar 17;(193).

155. Shin J, Noireaux V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. J Biol Eng. 2010;4(1):8.

156. Marshall R, Maxwell CS, Collins SP, Jacobsen T, Luo ML, Begemann MB, et al. Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. Mol Cell. 2018 Jan;69(1):146-157.e3.

157. Mougiakos I, Mohanraju P, Bosma EF, Vrouwe V, Finger Bou M, Naduthodi MIS, et al. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nat Commun. 2017 Nov 21;8(1):1647.

158. Karvelis T, Bigelyte G, Young JK, Hou Z, Zedaveinyte R, Budre K, et al. PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Res. 2020 May 21;48(9):5016–23.

159. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science (1979). 2020 Apr 17;368(6488):290–6.

160. Ma S, Huck WTS, Balabani S. Deformation of double emulsions under conditions of flow cytometry hydrodynamic focusing. Lab Chip. 2015;15(22):4291–301.

161. Iizuka R, Tahara K, Matsueda A, Tsuda S, Yoon DH, Sekiguchi T, et al. Selection of green fluorescent proteins by in vitro compartmentalization using microbead-display libraries. Biochem Eng J. 2022 Nov;187:108627.

162. Wu H, Sun Y, Wang Y, Luo L, Song Y. Advances in miniature CRISPR-Cas proteins and their applications in gene editing. Arch Microbiol. 2024 May 23;206(5):231.

163. Trasanidou D, Potocnik A, Barendse P, Mohanraju P, Bouzetos E, Karpouzis E, et al. Characterization of the AcrIIC1 anti–CRISPR protein for Cas9–based genome engineering in E. coli. Commun Biol. 2023 Oct 13;6(1):1042.

164. Trasanidou D, Barendse P, Bouzetos E, de Haan L, Bouwmeester H, Staals RHJ, et al. Efficient Genome and Base Editing in Human Cells Using ThermoCas9. CRISPR J. 2023 Jun 1;6(3):278–88.

165. Bahassi EM, O'Dea MH, Allali N, Messens J, Gellert M, Couturier M. Interactions of CcdB with DNA Gyrase. Journal of Biological Chemistry. 1999 Apr;274(16):10936–44.

166. Fraikin N, Van Melderen L. Single-cell evidence for plasmid addiction mediated by toxin–antitoxin systems. Nucleic Acids Res. 2024 Jan 15;

167. Lee J, Jung M hee, Jeong E, Lee JK. Using Sniper-Cas9 to Minimize Off-target Effects of CRISPR-Cas9 Without the Loss of On-target Activity Via Directed Evolution. Journal of Visualized Experiments. 2019 Feb 26;(144).

168. Anders C, Jinek M. In Vitro Enzymology of Cas9. In 2014. p. 1–20.

169. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. PLoS One. 2015 Dec 30;10(12):e0146021.

170. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A Sequence Logo Generator: Figure 1. Genome Res. 2004 Jun;14(6):1188–90.

171. Bravo JPK, Liu MS, Hibshman GN, Dangerfield TL, Jung K, McCool RS, et al. Structural basis for mismatch surveillance by CRISPR–Cas9. Nature. 2022 Mar 10;603(7900):343–7.

172. Eggers AR, Chen K, Soczek KM, Tuck OT, Doherty EE, Xu B, et al. Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. Cell. 2024 May;

173. Chen K, Han H, Zhao S, Xu B, Yin B, Trinidad M, et al. Lung and liver editing by lipid nanoparticle delivery of a stable CRISPR-Cas9 RNP. bioRxiv. 2023 Nov 15;

174. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science (1979). 2007 Mar 23;315(5819):1709–12.

175. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science (1979). 2008 Aug 15;321(5891):960–4.

176. Wu WY, Lebbink JHG, Kanaar R, Geijsen N, van der Oost J. Genome editing by natural and engineered CRISPR-associated nucleases. Nat Chem Biol. 2018 Jul 18;14(7):642–51.

177. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell. 2015 Oct;163(3):759–71.

178. Swartjes T, Staals RHJ, van der Oost J. Editor's cut: DNA cleavage by CRISPR RNA-guided nucleases Cas9 and Cas12a. Biochem Soc Trans. 2020 Feb 28;48(1):207–19.

179. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014 Mar 29;507(7490):62–7.

180. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014 Sep 27;513(7519):569–73.

181. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell. 2014 Feb;156(5):935–49.

182. Doudna JA. The promise and challenge of therapeutic genome editing. Nature. 2020 Feb 13;578(7794):229–36.

183. Scully R, Panday A, Elango R, Willis NA. DNA double-strand break repair-pathway choice in somatic mammalian cells. Nat Rev Mol Cell Biol. 2019 Nov 1;20(11):698–714.

184. Ceccaldi R, Rondinelli B, D'Andrea AD. Repair Pathway Choices and Consequences at the Double-Strand Break. Trends Cell Biol. 2016 Jan;26(1):52–64.

185. Karanam K, Kafri R, Loewer A, Lahav G. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. Mol Cell. 2012 Jul;47(2):320–9.

186. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018 Sep 16;36(8):765–71.

187. Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions

at 17 sites in the mouse genome. Nat Commun. 2017 May 31;8(1):15464.

188. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar V V, Thapar V, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol. 2015 Feb 16;33(2):187–97.

189. Zhang L, Jia R, Palange NJ, Satheka AC, Togo J, An Y, et al. Large Genomic Fragment Deletions and Insertions in Mouse Using CRISPR/Cas9. PLoS One. 2015 Mar 24;10(3):e0120396.

190. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet. 2018 Dec 15;19(12):770–88.

191. Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, 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. Sci Adv. 2017 Aug 4;3(8).

192. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science (1979). 2016 Sep 16;353(6305).

193. Banno S, Nishida K, Arazoe T, Mitsunobu H, Kondo A. Deaminase-mediated multiplex genome editing in Escherichia coli. Nat Microbiol. 2018 Feb 5;3(4):423–9.

194. Dong JY, Fan PD, Frizzell RA. Quantitative Analysis of the Packaging Capacity of Recombinant Adeno-Associated Virus. Hum Gene Ther. 1996 Nov 10;7(17):2101–12.

195. Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science (1979). 2018 Nov 16;362(6416):839–42.

196. Takeda SN, Nakagawa R, Okazaki S, Hirano H, Kobayashi K, Kusakizako T, et al. Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol Cell. 2021 Feb;81(3):558-570.e3.

197. Xiao R, Li Z, Wang S, Han R, Chang L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease. Nucleic Acids Res. 2021 Apr 19;49(7):4120–8.

198. Kong X, Zhang H, Li G, Wang Z, Kong X, Wang L, et al. Engineered CRISPR-OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. Nat Commun. 2023 Apr 11;14(1):2046.

199. Bigelyte G, Young JK, Karvelis T, Budre K, Zedaveinyte R, Djukanovic V, et al. Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. Nat Commun. 2021 Oct 26;12(1):6191.

200. Leenay RT, Maksimchuk KR, Slotkowski RA, Agrawal RN, Gomaa AA, Briner AE, et al. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol Cell. 2016 Apr;62(1):137–47.

201. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995 Jul;177(14):4121–30.

202. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat Biotechnol. 2020 Jul 1;38(7):883–91.

203. Andersen JB, Sternberg C, Poulsen LK, Bjørn SP, Givskov M, Molin S. New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. Appl Environ Microbiol. 1998 Jun;64(6):2240–6.

204. St. John,J. (2016) SeqPrep.

205. Li X, Wang Y, Liu Y, Yang B, Wang X, Wei J, et al. Base editing with a Cpf1–cytidine deaminase fusion. Nat Biotechnol. 2018 Apr 19;36(4):324–7.

206. Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K, et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat Biotechnol. 2019 Mar 11;37(3):276–82.

207. Stayrook S, Jaru-Ampornpan P, Ni J, Hochschild A, Lewis M. Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. Nature. 2008 Apr;452(7190):1022–5.

208. Valdez-Cruz NA, Caspeta L, Pérez NO, Ramírez OT, Trujillo-Roldán MA. Production of recombinant proteins in E. coli by the heat inducible

expression system based on the phage lambda pL and/or pR promoters. Microb Cell Fact. 2010 Dec 19;9(1):18.

209. Sun ZZ, Yeung E, Hayes CA, Noireaux V, Murray RM. Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an *Escherichia coli* Based TX-TL Cell-Free System. ACS Synth Biol. 2014 Jun 20;3(6):387–97.

210. Marshall R, Maxwell CS, Collins SP, Beisel CL, Noireaux V. Short DNA containing χ sites enhances DNA stability and gene expression in *E. coli* cell-free transcription–translation systems. Biotechnol Bioeng. 2017 Sep 23;114(9):2137–41.

211. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, et al. CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. Mol Cell. 2012 Jun;46(5):595–605.

212. Nieuwkoop T, Terlouw BR, Stevens KG, Scheltema RA, de Ridder D, van der Oost J, et al. Revealing determinants of translation efficiency via whole-gene codon randomization and machine learning. Nucleic Acids Res. 2023 Mar 21;51(5):2363–76.

213. Xin C, Yin J, Yuan S, Ou L, Liu M, Zhang W, et al. Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. Nat Commun. 2022 Sep 24;13(1):5623.

214. Kluesner MG, Nedveck DA, Lahr WS, Garbe JR, Abrahante JE, Webber BR, et al. EditR: A Method to Quantify Base Editing from Sanger Sequencing. CRISPR J. 2018 Jun;1(3):239–50.

215. Zhang S, Song L, Yuan B, Zhang C, Cao J, Chen J, et al. TadA reprogramming to generate potent miniature base editors with high precision. Nat Commun. 2023 Jan 26;14(1):413.

216. Hu Y, Han L, Mo Q, Du Z, Jiang W, Wu X, et al. Engineering miniature CRISPR-Cas Un1Cas12f1 for efficient base editing. Mol Ther Nucleic Acids. 2024 Jun;35(2):102201.

217. Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science (1979). 2019 Apr 19;364(6437):289–92.

218. Doman JL, Raguram A, Newby GA, Liu DR. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat Biotechnol. 2020 May 10;38(5):620–8.

219. Porto EM, Komor AC, Slaymaker IM, Yeo GW. Base editing: advances and therapeutic opportunities. Nat Rev Drug Discov. 2020 Dec 19;19(12):839–59.

220. Koblan LW, Erdos MR, Wilson C, Cabral WA, Levy JM, Xiong ZM, et al. In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. Nature. 2021 Jan 28;589(7843):608–14.

221. Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin E V., et al. RNA-guided DNA insertion with CRISPR-associated transposases. Science (1979). 2019 Jul 5;365(6448):48–53.

222. Lampe GD, King RT, Halpin-Healy TS, Klompe SE, Hogan MI, Vo PLH, et al. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. Nat Biotechnol. 2024 Jan 29;42(1):87–98.

223. Wang JH, Gessler DJ, Zhan W, Gallagher TL, Gao G. Adeno-associated virus as a delivery vector for gene therapy of human diseases. Signal Transduct Target Ther. 2024 Apr 3;9(1):78.

224. Levy JM, Yeh WH, Pendse N, Davis JR, Hennessey E, Butcher R, et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. Nat Biomed Eng. 2020 Jan 14;4(1):97–110.

225. Li SY, Cheng QX, Liu JK, Nie XQ, Zhao GP, Wang J. CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. Cell Res. 2018 Apr 12;28(4):491–3.

226. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science (1979). 2018 Apr 27;360(6387):436–9.

227. Steens JA, Zhu Y, Taylor DW, Bravo JPK, Prinsen SHP, Schoen CD, et al. SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. Nat Commun. 2021 Aug 19;12(1):5033.

228. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, et al. C2c2 is a single-component programmable RNA-guided

RNA-targeting CRISPR effector. Science (1979). 2016 Aug 5;353(6299).

229. Kellner MJ, Koob JG, Gootenberg JS, Abudayyeh OO, Zhang F. SHERLOCK: nucleic acid detection with CRISPR nucleases. Nat Protoc. 2019 Oct 23;14(10):2986–3012.

230. Chen J, Chen Y, Huang L, Lin X, Chen H, Xiang W, et al. Trans-nuclease activity of Cas9 activated by DNA or RNA target binding. Nat Biotechnol. 2024 May 29;

231. Liu R, Liang L, Freed EF, Gill RT. Directed Evolution of CRISPR/Cas Systems for Precise Gene Editing. Trends Biotechnol. 2021 Mar;39(3):262–73.

232. Parums D V. Editorial: First Regulatory Approvals for CRISPR-Cas9 Therapeutic Gene Editing for Sickle Cell Disease and Transfusion-Dependent β-Thalassemia. Medical Science Monitor. 2024 Feb 21;30.

233. Sheridan C. The world's first CRISPR therapy is approved: who will receive it? Nat Biotechnol. 2024 Jan 21;42(1):3–4.

234. Cyranoski D. CRISPR-baby scientist fails to satisfy critics. Nature. 2018 Dec 28;564(7734):13–4.

235. Cohen J. Did CRISPR help—or harm—the first-ever gene-edited babies? Science (1979). 2019 Aug 1;

236. Ledford H. Beyond CRISPR babies: how human genome editing is moving on after scandal. Nature. 2023 Mar 2;

Acknowledgements

Even if embarking to the Phd thesis journey feels often as a solitary case, after its completion it gets clear that it would not be the same or as complete without the involvement of more than one individual. To that end, it is only appropriate to acknowledge the help and support of the following colleagues.

It is tradition to thank first the Phd supervisors, but I would like to start with **Christian** who even if came to be involved in the latter part of my Phd he had a significantly positive impact on Chapter 4 but most importantly, I would dare to say, to me as a scientist and professional. His excellent organizational skills, professional manners and analytical plus insightful way of thinking, make a great combination of a colleague that sets the standards high, makes always the good example

John & Raymond, you have been my supervisors through the Phd journey and that by itself deserves its regards.

John, you have been the reference point for so many years of the Bacterial Genetics group and that's not by chance. You are always ready to come up with project ideas and venture to new journeys which more than often conclude in exciting scientific stories. As your supervisee I have to say you were always ready to entertain my scientific quirks and offer the freedom to try and explore for myself. Finally, when I found myself in trouble you were there to support and provide with solutions and for that you have my big thanks!

Raymond, the person with all the funny and sometimes extravagant stories. I have to recall bumping to each other during our smoking breaks which I admit have been a good way of decompression from the lab experiments, work-meetings etc. You are a private person but I remember once that you let me look behind the curtain. Scientific wise, we had a few meetings through my Phd but what has sticked with me is your straightforward and accurate feedback. Always good recommendations and things to consider. You are a great scientist and a valuable mentor. PS, we should have a played a bit more chess, it would be fun.

Ricardo, the person with the cool ideas and bolt projects. You introduced me to protein structure analysis and gifted one very cool 3D printed protein structure. I admit I still pick it up sometimes and have a look at it. Also, thanks for the ''cheap'' philosophical debates, in the end they came to inspire Proposition 3. Finally, I think I will ever recall your 'Alexander Fleming' moment with the nickases, pretty cool story.

Ketan & Sid, we had a long run together trying to make the cell-free screening scheme. Even if, we did not manage to engineer a protein in the end, we surely made some beautiful droplets and built the base for the engineering CRISPR-proteins in vitro. I am sure it will lead to a nice publication!

Thomas, we worked together on the miniature base editors. You did some beautiful data analysis, someone can just tell by looking at the figures. Really happy that we worked together.

Wen, also very glad about our interaction. It is always nice talking projects and science with you. We did not get lucky to see some miniature base editing in the mammals' cells together but I have every faith that you are going to make it. Very much looking forward to it. Good luck to that and every other scientific journey you are going to venture into!

Belen & Rob, our technician duo that keeps this lab of rogue Phd students up and running. More than that, you were always there to help with questions, introduction to equipment and techniques. Thanks a lot!!

Despoina & Yannis, it would only be unfair to not mention you both. My journey to bacterial genetics lab started with you too and our fruitful work together concluded with me landing the Phd Candidate position. Thank you!!

Caglar & Frederica, good chess matches, nothing more is needed to be said.

Awesome office, Panos, Carina, Miguel, Suzan, Anneleen, Thomas, an Awesome office consists from awesome people. Thanks for being around, being patient with all the noise while listening music or my frustrated keyboarding. We have had goods chats, shared struggles and progress through our Phd journeys. It has been valuable!

The students, **Alexandra Maibaum**, **Kobus Youri**, **Ties Baljet**, **Roel Boonen**, **Giwrgos Athanasiou** and **Tim Verhoeven**, thank you all for putting me the trust to perform your thesis with me.

I want also to thank all the colleagues past and present that we shared the struggle of completing a Phd Thesis. This related to me best with **Lorenzo** so followed by **Kostas**, and **Panos**.

Finally big thanks to all past and present members of Bacterial Genetic group, **Maartje**, **Han**, **Diana**, **Carl**, **Afonso**, **Prarthana**, **Mehran**, **Sarah**, **Sjoerd**, **Laureen**, **Isabelle**, **Vittorio**, **Tim**, **Agatha**, **Charlotte**.

Finally, I would like to offer my regards to my teachers **Hraklis Ganotis** (High school Biology), **Labros Sergis** (High school Chemistry), **Maria Parisi** (High school Literature) and my parents **Aspasia Oikonomaki** and **Konstantinos Bouzetos** for their devotion in education.

Τέλος, θέλω να ευχαριστήσω τους δασκάλους μου, **Ηρακλή Γανωτή** (Βιολογία Γυμνασίου), **Λάμπρο Σέργη** (Χημεία Λυκείου), **Μαρία Παρίσι** (Έκθεση Λυκείου) και τους γονείς μου **Ασπασία Οικονομάκη** και **Κωνσταντίνο Μπουζέτο** για την αφοσίωση τους στην εκπαίδευση.

Overview of completed training activities

Category A: Discipline specific activities

Visiting Researcher Enrico Mastrobattista (2019)

wAGOningen 2019 Symposium

CRISPR Meeting 2021

CRISPR Meeting 2022

CRISPR Meeting 2023

Genome Engineering: CRISPR Frontiers 2024

Category B: General Courses

VLAG Phd Week (2019)

Adobe Illustrator-Scientific Artwork & Infographics (2019)

Career Orientation (2024)

Applied Statistics (2023)

Introduction to R (2023)

Category C: Assisting in Teaching and supervision activities

MSc Thesis Supervision (2021-2022)

Practical supervision Research Methods in Microbiology (MIB30303) (2020-2021)

Category D: Other activities

Phd Research Proposal preparation (2019)

Bacterial Genetics Group Weekly meeting (2019-2023)

MIB Seminar Monthly Meeting (2019-2020)

Phd Excursion to US (2022)

Scientific article revision (2022)