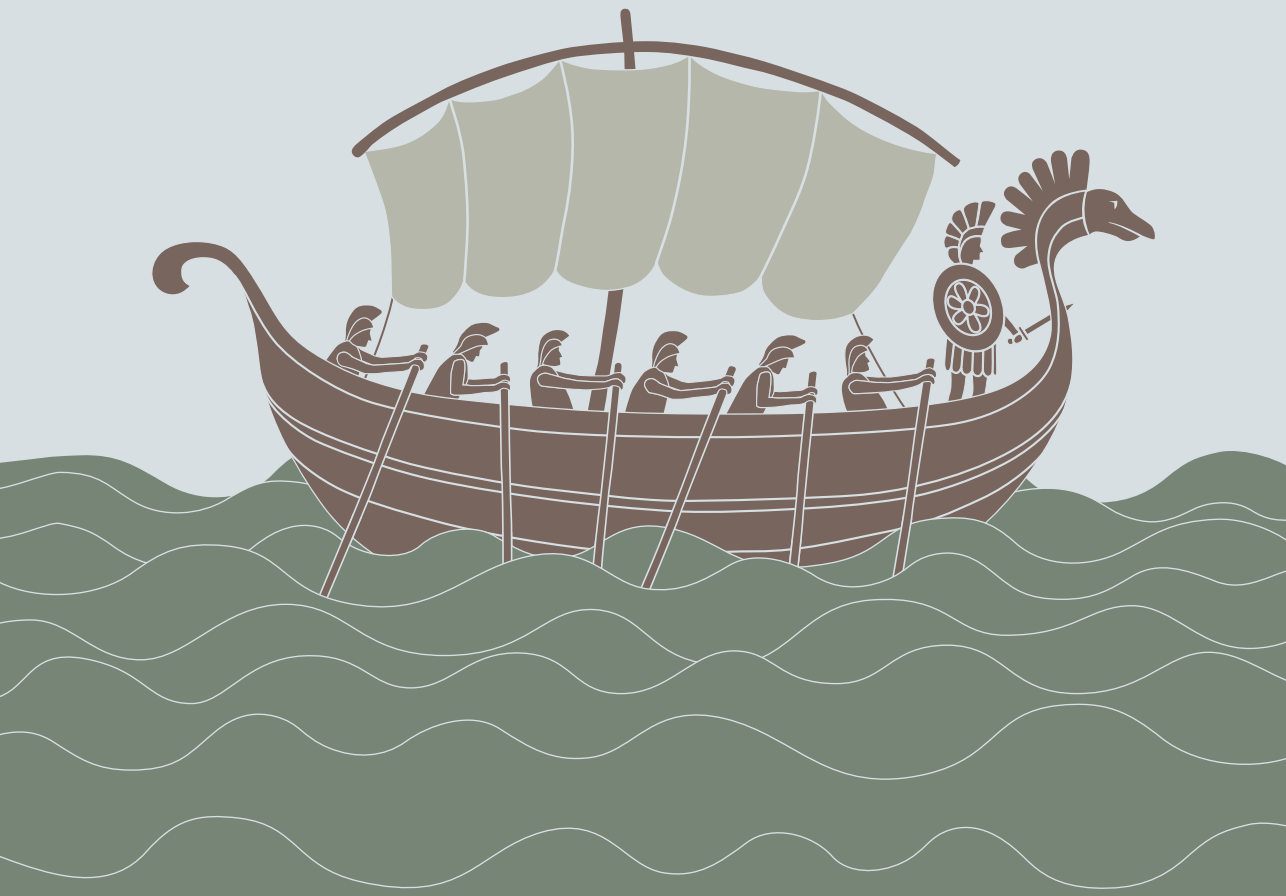
# The Journey of Prokaryotic Argonaute Proteins

## From fundamentals to application



Jorrit W. Hegge

# The Journey of Prokaryotic Argonaute Proteins

*From fundamentals to application*

**Jorrit W. Hegge**

**Thesis committee**

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

**Co-promotor**
Dr Raymond Staals
Assistant professor, Laboratory of Microbiology
Wageningen University & Research

**Other members**
Prof. Dr Tjakko Abee, Wageningen University & Research
Prof. Dr Thijn J. Brummelkamp, the Netherlands Cancer Institute, Amsterdam
Prof. Dr Dolf Weijers, Wageningen University & Research
Dr Stan J.J. Brouns, Delft University of Technology

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Argobiotechnology, Nutrition and Health Sciences).

# The Journey of Prokaryotic Argonaute Proteins

## *From fundamentals to application*

**Jorrit W. Hegge**

**Thesis**

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday the 10th of May 2019
at 4 p.m. in the Aula.

Jorrit W. Hegge

The Journey of Prokaryotic Argonaute Proteins - From fundamentals to application, 209 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summary in English

ISBN: 978-94-6343-414-0
DOI: https://doi.org/10.18174/468852

# Table of contents

# Chapter 1

## Preface and Outline

## It started in petunias flowers

Plant scientists in the Netherlands and the United States unintendedly stumbled on the importance of Argonaute proteins when working with petunia (*Petunia hybrida*) plants. (Krol et al., 1990; Napoli et al., 1990). Researchers from both countries tried to create petunias with more intense purple flowers by increasing the expression of a Chalcone synthase protein (CHS) that is essential in flower pigmentation pathways of petunias. They rationalized that introduction of extra copies of the gene encoding the CHS protein, would result in more RNA transcripts, more CHS protein and thus more purple pigment (Krol et al., 1990; Napoli et al., 1990). However, unexpectedly, they found that the petunias with the extra copies of the CHS gene developed white flowers instead (**Figure 1, left panel**). Unable to explain the puzzling result, they concluded that the natural and the synthetic copy of the CHS gene, somehow interfered with each other and hence called the observed phenomenon 'co-suppression' (Napoli et al., 1990). In the years after, co-suppression was shown not to be limited to petunias, as also in fungi the introduction of homologous transgenes caused 'quelling' of corresponding endogenous genes (Romano and Macino, 1992). In the following years, similar phenomena were observed in other plant species and in *Drosophila* (de Borne Dorlhac et al., 1994; Fray and Grierson, 1993; Pal-Bhadra et al., 1997). Besides affecting endogenous genes, it was also found that co-suppression could provide host resistance against viruses from which the introduced transgenes originated (Lindbo et al., 1993).

## RNA interference

The underlying mechanism of the co-suppression phenomena remained elusive until in 1998, in a Nobel prize winning experiment, Fire and Mello discovered that the trigger causing the silencing effect was double stranded RNA (dsRNA) (Fire et al., 1998). Working with the nematode *Caenorhabditis elegans* they found that it was not the introduction of single stranded sense RNA nor single stranded antisense RNA, but the introduction of a dsRNA hybrid of both that led to the effective silencing of genes that had sequence similarity to the introduced dsRNA. With their discovery, they ascribed the previously observed silencing phenomena to RNA interference (Fire et al., 1998), and triggered a series of discoveries that each identified parts of the RNA interference puzzle (**Figure 1, right panel)**.

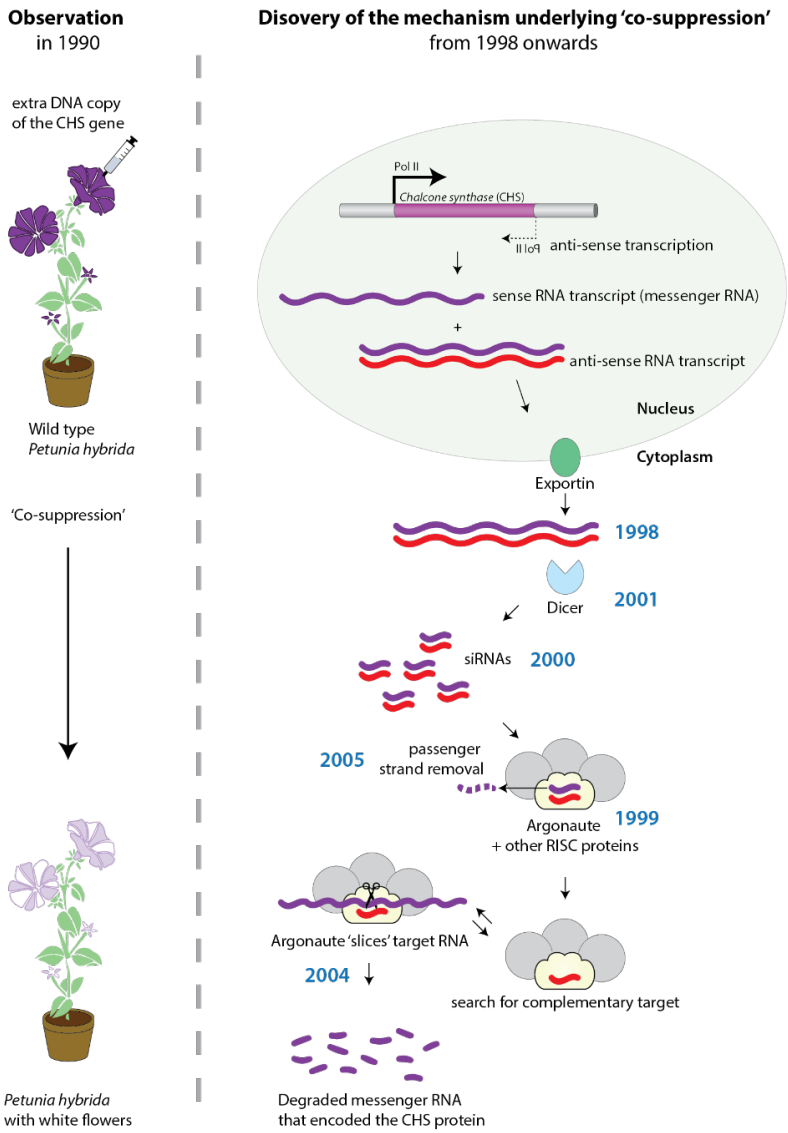



**Figure 1 | RNA interference and the effects of Argonaute were witnessed for the first time in petunias. Left panel** | In *Petunia hybrida* co-suppression of the CHS gene resulted in white flowers. **Right panel** | After the discovery of RNAi in 1998, a series of discoveries (years of the discovery in blue) provided insights into the RNAi mechanism. Sense and anti-sense transcripts form dsRNA, which triggers RNAi. Dicer processes the long dsRNA into small interfering RNAs (siRNA). These siRNA fragments are loaded into Argonaute and associated RISC proteins, which removes the passenger strand as a complex. The strand that remains bound (guide strand) is utilized to find complementary target RNA sequences. Bound targets, that are fully complementary to the guide, are cleaved by Argonaute.

Although the trigger of RNAi was now identified, the underlying mechanism through which dsRNA molecules silenced homologous mRNAs sequences remained unknown. Initially, helicase activity was believed to unwind the dsRNA after which the anti-sense strand could base pair with the messenger RNA transcript to prevent translation of the transcript into proteins (Sen and Blau, 2006). However, long anti-sense ssRNA intermediates were never found. Instead, small RNAs 25 nucleotides (nt) in length, were discovered in tomato plants and implicated to be involved in RNAi (Hamilton and Baulcombe, 1999). One year later, two other studies revealed that long synthetic dsRNA molecules were processed into small RNA fragments of 21-23 base pairs (bp) in size when incubated in a *Drosophila* embryo lysate (Hammond et al., 2000; Zamore et al., 2000). Combined this suggested that also the 742bp long sense-antisense RNA hybrids, that triggered RNAi in *C. elegans,* must have been processed into smaller intermediates in a similar way. The small dsRNA intermediates were named small interfering RNA (siRNA) (Elbashir et al., 2001). After the discovery that long dsRNAs were processed into siRNA, numerous other types of small RNAs (sRNA) emerged, such as microRNA (miRNA) and Piwi (P-element induced wimpy testis)-interacting RNA (piRNA) (Aravin et al., 2006; Girard et al., 2006; Lagos-Quintana et al., 2001; Lee and Ambros, 2001). Those sRNAs were categorized in different classes based on their biogenesis, protein binding partners and role (Carthew and Sontheimer, 2009). siRNAs originate from endogenous or exogenous RNA sources and interfere with the translation of complementary mRNA sequences. Although their silencing mechanism was not yet fully understood, siRNAs were produced synthetically as a tool in biological research to specifically silence target genes of interest (Elbashir et al., 2002; McCaffrey et al., 2003). miRNAs are encoded by the host genome and essential for the regulation of host gene expression (Bartel, 2004, 2009). piRNAs are an animal-specific class of small RNAs that are encoded on the genome and used to silence retrotransposons and other mobile genetic elements (Ozata et al., 2018).

## Argonaute

As only a few molecules of long dsRNA were necessary to trigger an effective RNAi response in *C. elegans*, Fire and Mello postulated the involvement of a protein component with catalytic activity (Fire et al., 1998). In the search for this catalytic component, Martinez and colleagues purified several binding partners of the short duplex siRNAs from human HeLa cells through co-immunoprecipitation using biotinylated siRNAs (Martinez et al.,

2002). Among others, two proteins of roughly 100 kDa were found and identified as Argonaute protein 1 and 2 (Ago1 and Ago 2). Two years later, Liu and coworkers confirmed that it was Argonaute that formed the catalytic core of RNAi (Liu et al., 2004). They demonstrated that all four Argonaute (Ago1-Ago4) proteins that are found on the human genome could bind siRNAs, but that only Ago2 was able to form a cleavage competent RNA Induced Silencing Complex (RISC) with siRNAs (Liu et al., 2004). Also in other organisms, Argonaute became recognized as the RNA-guided 'slicer' protein that was solely responsible for cleavage of target RNAs complementary to the small guide RNA (Rand et al., 2004; Song et al., 2004).

Strikingly, the Argonaute protein was 'discovered' already in 1998, in a mutagenesis study of *Arabidopsis thaliana* (Bohmert et al., 1998). One of the mutated *Arabidopsis thaliana* plants developed severe leaf abnormalities which reminded the authors of the tentacles of the pelagic octopus *Argonauta argo*. Therefore, the protein encoded by the mutated gene (AGO1), was named Argonaute (**Figure 2A, B**). The name of the octopus was, in turn, derived from the Greek myth 'Jason and the Argonautes' as the shell of the female octopus resembled the boat 'Argo' sailed by Jason and the Argonautes (**Figure 2B, C**).

Ago proteins are encoded on most sequenced eukaryotic genomes and are highly conserved between different species (Shabalina and Koonin, 2008). Based on their phylogeny, function and mechanism, eukaryotic Agos (eAgos) can be divided into several clades, among which the AGO and PIWI-clade proteins (Meister, 2013). Ago proteins from the AGO-clade are ubiquitously expressed and interact with siRNA and miRNA to regulate host gene expression and mediate defense against foreign genetic elements (Carthew and Sontheimer, 2009). Proteins of the PIWI-clade are animal-specific, and exclusively expressed in germ cells where they interact with piRNAs to regulate the activity of retrotransposons and viruses (Ozata et al., 2018).

Eukaryotic genomes commonly encode several copies of Argonaute genes (Höck and Meister, 2008). For example, mammalian genomes (i.e. humans) encode four AGO-clade and four PIWI-clade proteins, whereas the plant *A. thaliana* encodes ten AGO-clade proteins. The yeast *Saccharomyces cerevisiae* is one of the few organisms that apparently has lost its Ago gene during evolution, together with other genes encoding key proteins involved in the RNAi machinery (Ender and Meister, 2010).

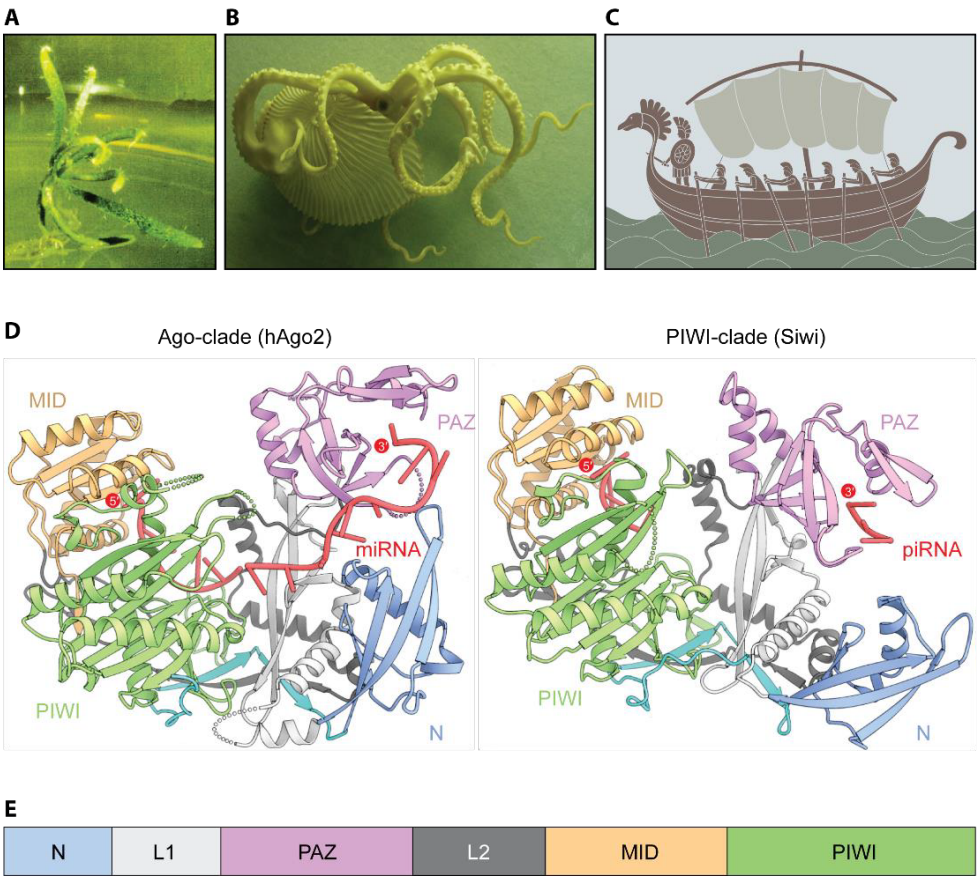



**Figure 2 | Argonaute proteins were first mentioned in an *Arabidopsis thaliana* mutagenesis study**. **A** | an AGO1 mutant *Arabidopsis thaliana* seedling of 35 days old (Bohmert et al., 1998). **B** | Model of the octopus *Argonauta argo* from the American Museum of Natural History. **C** | Drawing of the Argo ship with Jason and the Argonautes (adapted image of the thesis cover by Philip Patenall). **D** | Crystal structure of AGO-clade (PDB:4W5N) and PIWI-clade (PDB:5GUH) Argonaute proteins. (adapted from Matsumoto et al. 2016). **E** | Typical domain architecture of AGO and PIWI-clade Argonaute proteins.

In 2012, the first full length crystal structures of eukaryotic Ago proteins from the AGO-clade were solved (Elkayam et al., 2012; MacRae et al., 2012; Nakanishi et al., 2012). Four years later, the first structure of a PIWI-clade protein followed (Matsumoto et al., 2016). Despite having different functionalities, the overall structures of AGO and PIWI-clade Agos are very similar (**Figure 2D**). Both proteins adopt a bi-lobed architecture comprising a N-terminal lobe (N-PAZ domain) and a C-terminal lobe (MID-PIWI domain) connected by two

linkers domains (L1-2) (MacRae et al., 2012; Matsumoto et al., 2016) **(Figure 2D, E**). The PAZ and MID domain contain a binding pocket that anchors 3' or 5' end of the guide RNA, respectively (Faehnle and Joshua-Tor, 2010; Song et al., 2003). The PIWI domain contains the catalytic center that is formed by a tetrad of four amino acids (DED-D/H/N), which coordinate two magnesium ions essential for RNA cleavage (Kaya et al., 2016; Song et al., 2004). Not all Ago proteins contain a complete catalytic tetrad. For example, three human Ago proteins (Ago 1,3 and 4), are inactive and unable to cleave RNA targets (Liu et al., 2004). The N-domain plays an important role in the release of the cleaved target RNA strand for AGO-clade proteins (Kwak and Tomari, 2012). In contrast, some PIWI-clade proteins are unable to autonomously release the cleaved RNA target strand with the N-terminal domain and require assistance of a specific helicase (Matsumoto et al., 2016).

## MicroRNAs (miRNAs)

The first described miRNA, lin-4 from *C. elegans*, was already reported before the discovery of RNAi (Lee et al., 1993; Wightman et al., 1993). Researchers revealed that lin-4 did not encode a protein but rather a transcript that could repress the expression of another gene (lin-14). Initially, the lin-4 transcript was called a small temporal RNA (stRNA). However after the discovery of RNAi, stRNAs were redefined as miRNAs and found on genomes of other animals, plant species and certain viruses (Lagos-Quintana et al., 2001; Lee and Ambros, 2001; Pfeffer et al., 2004; Reinhart et al., 2002). In humans, more than 60% of all genes are regulated by miRNAs (Friedman et al., 2009). Like other endogenous sRNAs, the biogenesis miRNAs starts with the transcription by RNA polymerase II **(Figure 3**) (Lee et al., 2004). The miRNA transcript, (named pri-miRNA) forms a typical secondary structure with one or more hairpins and an imperfectly base paired stem loop (Lee et al., 2002). In the nucleus this pri-miRNA structure is recognized by a microprocessor complex, including Drosha and DGCR8, which processes the pri-miRNA to form the pre-miRNA (Lee et al., 2003).

Exportin-5 recognizes the pre-miRNA and exports it from the nucleus into the cytoplasm (Lund et al., 2004; Yi et al., 2003). Next, Dicer, a RNaseIII-like endonuclease, cleaves the hairpin of the pre-miRNA and generates short miRNA duplexes of ~22nt in length, containing a 5'-end phosphate and 2nt overhangs at the 3' end (Bernstein et al., 2001; Knight and Bass, 2001). Those short miRNA duplexes are ultimately loaded into the Argonaute protein to form the RISC (RNA-induced Silencing Complex), where one of the two strands,

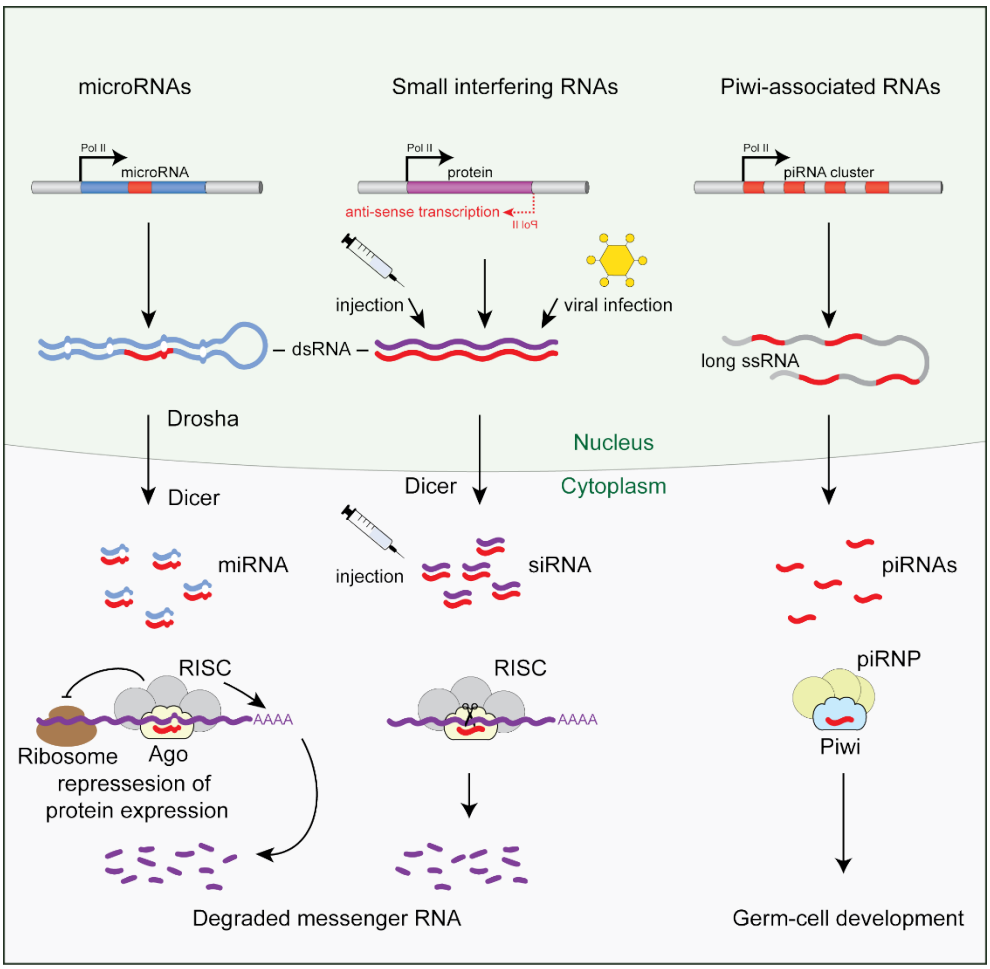



**Figure 3 | Overview of the origin of three of well-studied forms of small RNA . MicroRNA (miRNAs)** are encoded on the genome and form dsRNA hairpins with imperfect stem-loops upon transcription. These structures are processed by both Drosha and Dicer to form miRNAs. **Small interfering RNAs (siRNAs)** are generated by Dicer from a long dsRNA precursor that is formed by perfect base pairing between sense and natural anti-sense transcripts (NATs). These siRNA precursors can also originate from dsRNA viruses or injections. Both siRNAs and miRNAs are loaded onto the Argonaute proteins of the Ago clade. **Piwi-associated RNAs (piRNAs)** are processed in a Drosha and Dicer independent pathway from long ssRNA precursors. piRNAs are typically located in clusters on the genome and serve as inheritable memory to repress invading transposable elements to maintain genome integrity. In contrast to the other two small RNAs, piRNAs are loaded into the Argonaute proteins of the PIWI-clade. Figure is adapted from (Großhans and Filipowicz, 2008).

"the passenger strand" is unwound and removed from the complex (Matranga et al., 2005). The strand that remains bound is named "the guide strand", which directs the Argonaute and

associated RISC proteins towards complementary RNA targets that are often found in the 3' UTR region of mRNAs. The level of translation repression depends on the degree of complementarity between the guide and target. When the mRNA target is bound to the guide with perfect complementary, as is the case for most plant miRNA and some animal miRNAs, the mRNA is cleaved by the Argonaute protein and can no longer be used for protein translation (Bartel, 2009). However, most animal miRNAs bind their mRNA targets imprecisely with several mismatches. These mismatches prevent Argonaute from cleaving the target RNA (Bartel, 2009). Instead Argonaute forms a binding scaffold for the recruitment of additional silencing factors, which repress the translation of mRNA (Jonas and Izaurralde, 2015). Due to the imprecise targeting, miRNAs can have multiple target locations, sometimes on different mRNAs and thus can regulate the expression of multiple genes (Bartel, 2009).

### Small interfering RNA (siRNA)

The origin of siRNA can be either endogenous or exogenous. Endogenous siRNAs (endo-siRNAs) are transcribed from the genome either as hairpins with perfect paired stem-loops, or as hybrids of sense (mRNA) and natural antisense transcripts (NATs) (**Figure 3**) (Carthew and Sontheimer, 2009; Faghihi and Wahlestedt, 2009; Kim et al., 2009). Although NATs are usually several orders of magnitude less abundant than sense (mRNAs) transcripts (Ozsolak et al., 2010), sense-NAT hybrids were shown to contribute to low levels of endo-siRNAs in fully differentiated somatic cells (Werner et al., 2014). In fact, the formation of sense-NAT hybrids is believed to be the underlying trigger for the silencing effect that was observed in petunias in the early 90's (**Figure 1**) (Steensels and Verstrepen, 2016). In the cytoplasm, endo-siRNA follow the same processing pathway as miRNAs. Dicer processes the dsRNA precursors to generate small RNA duplexes with a 5'-end phosphate and 2nt 3'-end overhangs (Carthew and Sontheimer, 2009). In mouse oocytes, loss of Dicer decreased the siRNA levels and increased the levels of retrotransposons transcripts (Tam et al., 2008; Watanabe et al., 2008). In some organisms additional endo-siRNAs are generated via the secondary siRNA biogenesis pathway. In this process, an Argonaute-siRNA complex binds to a mRNA transcript (sense-strand) and recruits a RNA-dependent RNA polymerase (RDRP) for the transcription of the anti-sense strand (Sijen et al., 2007). The dsRNA that is formed is processed by Dicer similar to other endo-siRNAs. The pool of secondary siRNAs greatly amplifies the silencing response (Sijen et al., 2007).

Most exogenous siRNAs (exo-siRNAs) are either synthetic or originate from viral infections (Golden et al., 2008). There are several ways to introduce synthetic exo-siRNA into a cell (Kim and Rossi, 2007). Long dsRNAs can be introduced into the cytoplasm, where it is processed into short siRNAs by the RNAi machinery. Similarly, short dsRNA (21-23nt in length) can be introduced into the cytoplasm where they are, without a processing step, directly bound by Argonaute and associated RISC proteins (Salomon et al., 2010). Alternatively, siRNAs can be introduced indirectly on a DNA plasmid, as was unwittingly the case during the petunia experiment in 1990. Transcription of the DNA in both directions (sense and anti-sense) results in dsRNAs, which are exported to the cytoplasm and processed by Dicer into short siRNAs.

Regardless of the origin, small siRNA duplexes of 21-22nt in length are ultimately loaded into Argonaute to form the RISC complex. The stability of the base pairs at the 5'-end of siRNA defines the faith of each strand and determines which strand will reside in the complex as a guide strand, and which will be cleaved by Argonaute to be eliminated from the complex (Khvorova et al., 2003; Schwarz et al., 2003). The siRNA-directed RISC is able to catalyze target RNA cleavage in a multiple turnover manner, which means that after slicing the cleaved RNA products will be released, enabling the RISC to bind and cleave new RNA targets using the same guide strand (Haley and Zamore, 2004; Hutvágner and Zamore, 2002). This way, a single siRNA-directed RISC can cleave up to 50 different complementary target strands (Haley and Zamore, 2004).

## PIWI-associated RNA (piRNA)

The piRNAs are an animal-specific class of small RNAs that are mainly found in the germline. Whereas siRNAs and miRNAs are generated from long dsRNA precursors, piRNAs are derived from long strands of non-coding RNA (lncRNA) (**Figure 3**). piRNAs are typically encoded on specific loci of the genome, the so called piRNA clusters. In flies, these piRNA clusters resemble graveyards of transposon remnants that are no longer able to transpose (Czech and Hannon, 2016). The underlying mechanism behind the acquisition of new transposons into these piRNA clusters remains elusive. However, a model exists which proposed that active transposons are trapped into piRNA clusters after which they decay (Bergman et al., 2006). In the cytoplasm the long transcripts of these piRNA clusters are fragmented by Zucchini nucleases to form primary piRNAs. The cellular pool of piRNAs is

enriched for piRNAs with a 5‘-end uracil and mainly bind Ago proteins of the PIWI-clade. Once bound to a PIWI-clade Ago, the 3‘-end of the piRNA is trimmed to yield piRNAs of 21-35nt in size. After methylation of the 3‘-end of the piRNA guide by the Hen1 enzyme the maturation of the piRNA-Piwi complex is completed (Kawaoka et al., 2011). In fly germ cells, mature piRNA-Piwi complexes mediate cleavage of complementary RNA transcripts of retrotransposons or other mobile genetic elements to preserve germline integrity. Cleavage of RNA targets results in products that can be loaded into Ago3 and serve as new piRNAs (secondary piRNAs). Since the sequence of the secondary piRNAs are complementary to the original transcript of piRNA cluster, from which the primary piRNAs were derived, these secondary piRNA-Ago3 complexes can cleave the original long piRNA cluster transcript to generate additional piRNAs. This mechanism is termed the ‘ping pong cycle’ and enables an amplification of the piRNA mediated silencing response. Ultimately, the fate of mature piRNA-Piwi complexes depend on their binding partners. For example, the Piwi-clade Ago MIWI2 of murine enters the nucleus, whereas the murine MILI and MIWI Ago proteins act in the cytoplasm (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2001). Overall, piRNA-guided PIWI-complexes silence transposable elements, regulate gene expression and impede viral infections.

Interestingly, the piRNA-based immunity in the germline of animals shows many similarities with the recently discovered prokaryotic immune system, called CRISPR (Clustered Regularly Interspaced Palindromic Repeats), which bacteria and archaea use to fend of viral invaders (Brouns et al., 2008)(Barrangou et al., 2007; Marraffini and Sontheimer, 2010a; Rotem Sorek, 2008). CRISPR provides adaptive immunity and utilizes, analogous to Argonaute, RNA guides to find, bind and degrade complementary invading DNA (Brouns et al., 2008). After an infection, fragments of the invader are incorporated in the genome, in so called CRISPR-arrays (analog of the piRNA clusters). These arrays can be transferred to the progeny and serve as a genetic memory of past invasions.

## Argonaute in prokaryotes

Two years after the initial discovery in eukaryotes, Argonaute proteins were also found to exist in bacteria and archaea (Cerutti et al., 2000) (**Figure 4**). Based on a bioinformatics analysis prokaryotic Argonautes (pAgos) were predicted to be involved in host defense as they were commonly found in genomic neighborhoods that also encoded proteins involved

in the defense against foreign genetic elements (Makarova et al., 2009). Furthermore, protein sequence alignments revealed that short pAgos exist that lost their N-terminal lobe, formed by the N and PAZ domains (Makarova et al., 2009). Long pAgos encompass the same four domain architecture (N-PAZ-MID-PIWI) as eAgos.

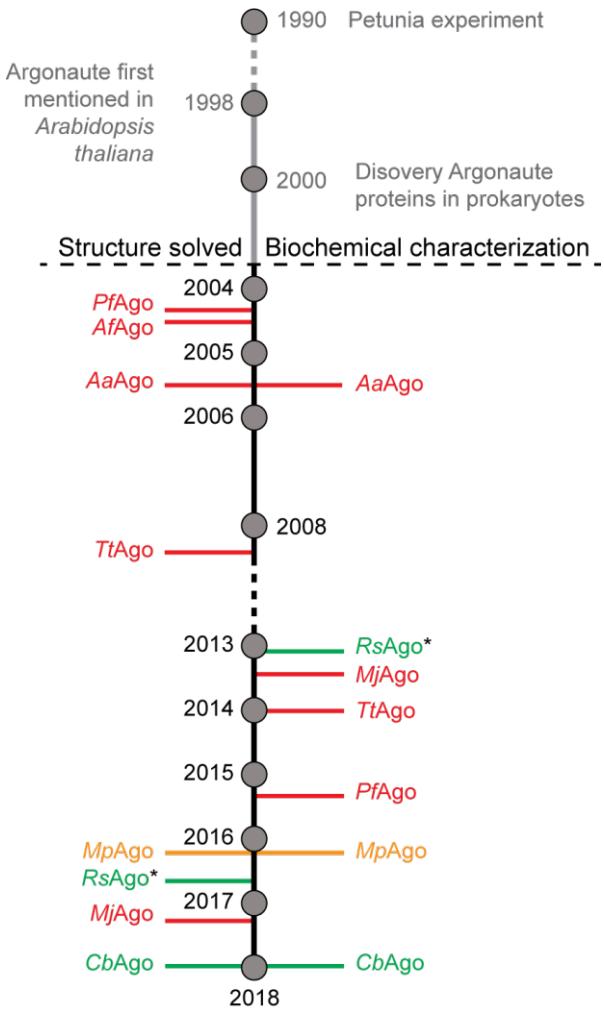



**Figure 4 | Milestones in the field of prokaryotic Argonaute proteins.** The colors indicate the optimum temperatures of the characterized Argonaute protein; red = 65°C >, orange = 45-60°C and green = 30-40°C. **Rs*Ago is catalytically inactive due to mutations in the conserved DEDX active site. *Aa*Ago, *Aquifex aeolicus* Argonaute; *Af*Ago, *Archaeoglobus fulgidus* Argonaute; *Mj*Ago, *Methanocaldococcus jannaschii* Argonaute; *Mp*Ago, *Marinitoga piezophila* Argonaute; *Pf*Ago, *Pyrococcus furiosus* Argonaute; *Rs*Ago, *Rhodobacter sphaeroides* Argonaute; *Tt*Ago, *Thermus thermophilus* Argonaute. *Cb*Ago is the Argonaute of *Clostridium butyricum* that has been described in this thesis. It is the first example of a pAgos that efficiently catalyzes DNA-guided DNA cleavage at 37°C.

Argonaute of the Archaeon *Pyrococcus furiosus* (*Pf*Ago) was the first full-length Argonaute structure that was solved (Song et al., 2004) eight years before elucidation of the first eAgo structure. From this time onwards, seven other pAgo structures have been solved (**Figure 4**). The structures of pAgos showed similarities to the structure of a PIWI-clade Ago (Matsumoto et al., 2016). In proteins from both clades, recognition of the 5'-phosphate of the guide strand is coordinated by a metal ion in the binding pocket of the MID domain (Matsumoto et al., 2016; Sheng et al., 2014). Agos from the Ago-clade however,

instead utilize a conserved lysine to neutralize the negative charge of the 5'-phosphate of the guide RNA (Elkayam et al., 2012; MacRae et al., 2012).

In 2005, Yuan and coworkers reported the first biochemical characterization of the prokaryotic Argonaute derived from *Aquifex aeolicus* (*Aa*Ago) (Yuan et al., 2005). *Aa*Ago was shown to mediate RNA cleavage, but in contrast to eAgos, it utilized DNA guides rather than RNA guides. More evidence for differences in the nucleic acid binding partners between eukaryotic and prokaryotic Agos followed in 2013, when it was found that *in vivo* the catalytically inactive Argonaute of *Rhodobacter sphaeroides* (*Rs*Ago) associated with short ssRNA transcript fragments and short complementary DNA molecules (Olovnikov et al., 2013). At the time it was not clear yet whether *Rs*Ago used these associated DNAs as guides to target RNA transcripts or whether the co-purified RNA molecules were used as guide to target DNA. One year later, a big breakthrough followed when it was demonstrated that the archaeal and bacterial Argonaute of *Methanocaldococcus jannaschii* (*Mj*Ago) and *Thermus thermophilus* (*Tt*Ago), respectively, exhibit DNA-guided DNA cleavage activity (Swarts et al., 2014b; Zander et al., 2014). Thus far, eAgos had exclusively been shown to mediate RNA-guided RNA interference. In line with the prediction, DNA-targeting pAgos were shown to be involved in host defense to fend of DNA invaders such as plasmids (Olovnikov et al., 2013; Swarts et al., 2015a, 2014b) In 2016, when the structure of *Rs*Ago was solved, biochemical assays also confirmed that *Rs*Ago targets DNA, but in a RNA-guide dependent manner (Miyoshi et al., 2016). How the previously *in vivo* associated small DNA targets were cleaved still remains enigmatic, as *Rs*Ago is catalytically inactive. Currently several mechanistic insights exist for *Tt*Ago and *Rs*Ago which are explained **in chapter 2**. Furthermore, the DNA-targeting behavior of several prokaryotic Argonautes raised the idea to repurpose these pAgos for DNA-editing applications (Also reviewed in **chapter 2**) (Hegge et al., 2018a).

Catalytically active DNA-targeting pAgos have been exclusively characterized from (hyper)-thermophilic bacteria and archaea (**Figure 4**; highlighted in red and orange). Although promising applications exist for these kind of DNA-targeting pAgos (two of which are described in this thesis), their thermophilic nature limits their use for *in vivo* applications at moderate temperatures (37°C). Therefore, we explored to pool of mesophilic Agos, which resulted in an extensive characterization of the Argonaute from *Clostridium butyricum* (*Cb*Ago) which mediates DNA-guided DNA cleavage at 37°C.

The field of the prokaryotic Argonaute proteins is still in its infancy compared to the field of eAgos. Given the variation of pAgos with respect to their protein and nucleic acid binding partners (i.e. DNA and RNA), and domain architecture (i.e. short vs long pAgos), it is not unlikely that similar to eAgos, also pAgos have a plethora of different functions and mechanisms that yet need to be discovered.

# Thesis outline



**Chapter 1** provides a brief history of Argonaute proteins (Agos). Agos were originally discovered in plants, but are also found in the other domains of life. They utilize guide nucleic acids to bind complementary target nucleic acids. Some Agos cleave their target nucleic acids by their endonucleolytic activity while other variants only bind their target and form a binding scaffold for other effector proteins. Overall, Agos are crucial for the regulation of both endogenous and exogenous nucleic acids inside the cell.

**Chapter 2** reviews the prokaryotic Agos (pAgos) that have biochemically been characterized to date. It furthermore describes the diversity that is present within the pAgo family, of which many variants yet need to be explored. Based on protein sequence alignments, the pAgo family can be classified into three major groups: Long pAgos which contain the same (N-PAZ-MID-PIWI) domain architecture found in eukaryotic Agos (eAgos), short pAgos and PIWI-RE proteins. The N-terminal and PAZ domains, that are found in long pAgos, are either lacking or substituted by alternative yet to be characterized domains in short pAgos and PIWI-RE proteins. Of these three pAgo groups, research has mainly focused on long-pAgos variants derived from thermophiles. Given the vast number of unexplored pAgos in addition to the functional and mechanistic variation that the characterized pAgos display, it is not unlikely that the pAgo family might form a source of a new generation of genome editing tools.

In **chapter 3** a biochemical and structural characterization of the pAgo from the mesophilic bacterium *Clostridium butyricum* (*Cb*Ago) is described. Like other pAgos, *Cb*Ago appears to be implicated in host defense as it mainly acquired plasmid-derived DNA guides upon expression in *Escherichia coli*. Interestingly, these co-purified DNA have a strong bias towards a deoxyadenosine at the 5'-end, and thymidines in the sub-seed segment (guide nucleotide 2 to 4). *In vitro Cb*Ago can be reprogrammed with short synthetic DNA guides to cleave both ssDNA and AT-rich (AT ≥65%) dsDNA at moderate temperatures (37°C). Although attempts to exploit *Cb*Ago for genome editing purposes in mammalian cells were not yet successful, the characterization described in this chapter represents a valuable step towards the use of pAgos for a range of genetic applications.

**Chapter 4** describes a biophysical characterization of *Cb*Ago. Like other DNA binding proteins, *Cb*Ago requires a fast and efficient search mechanism to find potential target sites in an enormous pool of cellular DNA. Hence *Cb*Ago combines a gliding mechanism with intersegmental jumps while scanning the target DNA. During gliding *Cb*Ago diffuses in lateral direction through a cloud of counter ions while maintaining loose contact with the target DNA. In order to bypass obstacles, such as secondary structures and protein barriers, *Cb*Ago is capable of making intersegmental jumps after which it continues gliding until it finds a complementary DNA target.

C**hapter 5** describes a novel pAgo-based diagnostic application termed, NAVIGATER (enriched **N**ucleic **A**cids **Vi**a DNA-**G**uided **A**rgonaute of ***T**hermus th**er**mophilus*). Using the pAgo from *Thermus thermophilus* (*Tt*Ago), the detection of the rare KRAS G12D mutant allele (in blood samples of several pancreatic cancer patients) could significantly be improved by means of NAVIGATER. In those patient samples, mutant alleles are vastly outnumbered by wild type alleles. To enrich the mutant allele, which only differed by a single nucleotide mutation from the wild type alleles, we took advantage of the ability of *Tt*Ago to cleave the WT allele with nucleotide specificity, while leaving the mutant allele untouched. By the collective use of NAVIGATER with rare mutant allele detection methods, such as digital droplet PCR (ddPCR) and Peptide Nucleic Acid-Mediated PCR (PNA-PCR), the sensitivity of those well-established methods could be improved by nearly a 60 fold.

**Chapter 6** describes a novel one-tube cloning method, termed "Jason cloning". This method relies on a two-step thermal-cycle (95-45°C) reaction, in which the pAgo from *Pyrococcus furious* (*Pf*Ago) works concertedly with a thermostable ligase (Ampligase™). We successfully ligated a pre-assembled insert of 100 base pairs into pUC19. At 45°C, Ampligase ligates an insert into a backbone vector, which was linearized by *Pf*Ago at 95°C thru compatible overhangs. Although Jason cloning was shown successful, currently optimizations are on-going trying to further enhance the flexibility and efficiency of the cloning method.

**Chapter 7** discusses and concludes the work in this thesis from fundamental research (chapter 3 & 4) to applications (chapter 5 & 6). It furthermore elaborates on potential cellular functions and future directions of pAgos.

# Chapter 2

# Prokaryotic Argonaute proteins: novel genome-editing tools?

Jorrit W. Hegge*, Daan C. Swarts* and John van der Oost

*contributed equally

Adapted from - Prokaryotic Argonaute proteins: novel genome-editing tools?
*Nature reviews microbiology (volume 16 | January 2018)*

## Abstract

Argonaute proteins constitute a highly diverse family of nucleic acid-guided proteins. They were first discovered in eukaryotes as key proteins in RNA interference systems, but homologous prokaryotic Argonaute proteins (pAgos) have also been found in archaea and bacteria. In this Progress article, we focus on long pAgo variants, a class of pAgos that are involved in nucleic acid-guided host defense against invading nucleic acids, and discuss the potential of pAgos in genome editing.

# Introduction

Argonaute proteins were initially discovered in eukaryotes, in which they are key players in RNA interference (RNAi) pathways (Bohmert et al., 1998; Ketting, 2011). As the functional core of the RNA-induced silencing complex (RISC), eukaryotic Argonaute proteins (eAgos) use small single-stranded RNA (ssRNA) molecules as a guide to target complementary RNA sequences (Hammond et al., 2000; Ketting, 2011; Swarts et al., 2014a). The translation of the targeted RNA is silenced either directly, by binding and cleavage of the target RNA, or indirectly, by binding to the target RNA followed by the recruitment of additional silencing proteins (Hannon, 2002; Hutvagner and Simard, 2008). As such, eAgos can post-transcriptionally regulate gene expression, defend their host against invading RNA viruses and preserve genome integrity by reducing the mobility of transposons (Ketting, 2011).



Following the discovery of homologous prokaryotic Argonaute proteins (pAgos), the first full-length Argonaute structures in archaea and bacteria were resolved (Song et al., 2004; Yuan et al., 2005). Moreover, it was reported that horizontal gene transfer resulted in a rather uneven distribution of genes that encode pAgos, which can be found in ~32% and ~9% of the sequenced archaeal and bacterial genomes, respectively (Swarts et al., 2014a). As archaea and bacteria lack RNAi pathways (Shabalina and Koonin, 2008), the physiological functions of pAgos have long remained elusive. Recent studies of selected pAgos have revealed unique mechanistic features and have uncovered insights into their physiological roles (Kaya et al., 2016; Olovnikov et al., 2013; Swarts et al., 2015a, 2014b, 2015b; Wang et al., 2009; Willkomm et al., 2017a; Zander et al., 2014). Interestingly, it has been shown that at least some pAgos mediate nucleic acid-guided cleavage of DNA targets (Kaya et al., 2016; Swarts et al., 2015a, 2014b; Zander et al., 2014). Given that the nucleic acid-guided DNase activity of the CRISPR-associated protein Cas9 has revolutionized the genome-editing field, the intriguing possibility that pAgo variants might be suitable candidates for biotechnological applications was postulated (Swarts et al., 2014a). In this Progress article, we describe the most recent pAgo-related discoveries and discuss the potential of pAgos for the purposes of genome editing.

## Domain architecture of long pAgos

On the basis of their domain architecture, the family of pAgos can be divided into three classes: long pAgos, short pAgos and PIWI-RE proteins (P-element induced wimpy testis

with conserved R (Arg) and E (Glu) residues) (**Figure 1a**). Of these classes, only the long pAgo class has been studied experimentally (Kaya et al., 2016; Miyoshi et al., 2016; Olovnikov et al., 2013; Swarts et al., 2015a, 2014b, 2015b, 2017a, Wang et al., 2008a, 2008b; Willkomm et al., 2017a; Zander et al., 2014, 2017). Despite poor sequence conservation and functional segregation, the general domain architecture of long pAgos is very similar to that of eAgos. Both eAgos and long pAgos form a bi-lobed scaffold, in which one lobe consists of the amino-terminal (N) and PIWI–Argonaute–Zwille (PAZ) domains, whereas the other lobe consists of the middle (MID) and PIWI domains (Song et al., 2004; Wang et al., 2008b; Yuan et al., 2005) (**Figure 1b**). The MID and PAZ domains usually form binding pockets that facilitate the anchoring of the 5′ and 3′ ends of an oligonucleotide guide, respectively (Ma et al., 2005; Song et al., 2003; Wang et al., 2008b). On target binding, the PIWI domain of catalytically active Agos (that contain the DEDX motif, in which X denotes D, H or N (Kaya et al., 2016)) mediates target-strand cleavage. This motif is often incomplete in eAgos and pAgos that for their function seem to rely on target binding rather than on target cleavage (Liu et al., 2004; Olovnikov et al., 2013; Swarts et al., 2014b). Interestingly, genes that encode pAgo variants in which this motif is incomplete are often clustered with genes that encode putative nucleases (Makarova et al., 2009), which suggests that these pAgos might be indirectly involved in target cleavage. Similar to the N-terminal domain of eAgos, the N-terminal domain of at least some pAgos is predicted to function as a 'wedge' between the guide and the target strands to facilitate the dissociation of cleaved target strands (Kwak and Tomari, 2012).

## Biological functions of pAgos

In 2009, a bioinformatics analysis revealed that genes that encode pAgos are found in the same operon as genes that encode proteins that are involved in host defense (for example, endonucleases that are associated with restriction–modification systems), which suggests that pAgos might have a role in protecting their hosts from invading nucleic acids, such as plasmids (Makarova et al., 2009). Indeed, this hypothesis has been confirmed for at least some long pAgos (Olovnikov et al., 2013; Swarts et al., 2015a, 2014b, 2015b). Remarkably, whereas eAgos exclusively mediate RNA-guided RNA interference, some pAgos bind to short DNA guides and, in some cases, can directly target DNA (Swarts et al., 2015a, 2014b; Zander et al., 2017) (**Table 1**). This indicates that a larger functional and mechanistic

diversity exists in pAgos than in eAgos. Currently, the long pAgos of *Thermus thermophilus* (*Tt*Ago) and *Rhodobacter sphaeroides* (*Rs*Ago) are the best characterized. Although both are involved in host defense, they exhibit distinct mechanisms, as discussed below.



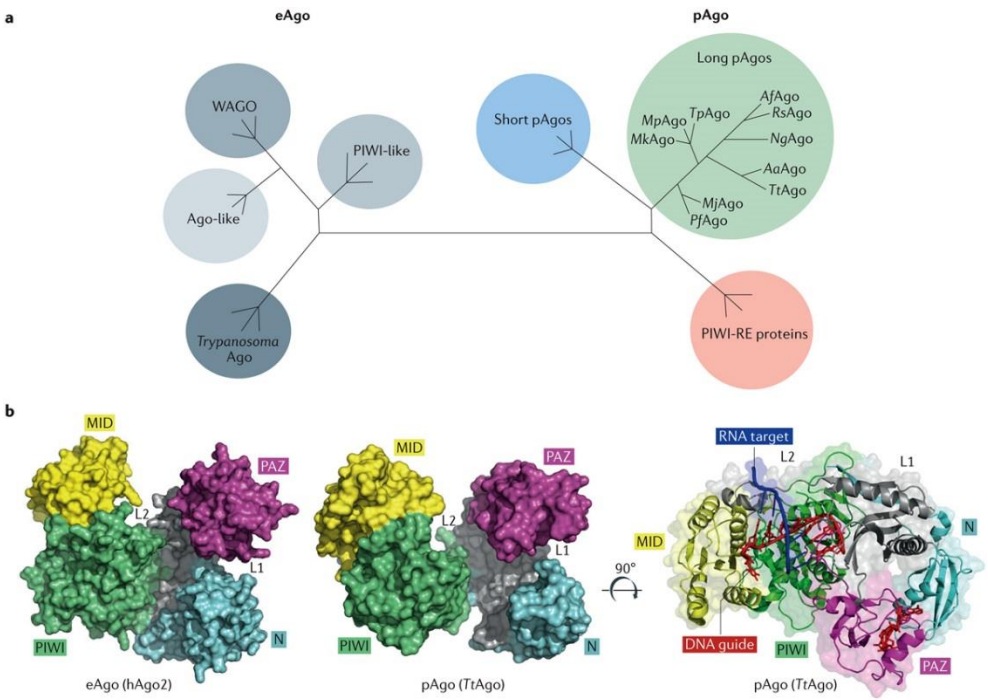



**Figure 1 | Argonaute evolution and structure. a** | Schematic phylogenetic tree based on the amino acid alignments of conserved middle (MID) and P-element induced wimpy testis (PIWI) domains from eukaryotic Argonaute proteins (eAgos) and prokaryotic Argonaute proteins (pAgos). **b** | The crystal structure of human Argonaute 2 (hAgo2; RCSB Protein Data Bank (PDB) entry 4Z4C), indicated as eAgo (left). The crystal structure of *Thermus thermophilus* Argonaute (*Tt*Ago; PDB entry 3F73), indicated as pAgo (middle). Both structures have the conserved bi-lobed scaffold between eAgos and pAgos. Ternary complex of *Tt*Ago (rotated 90° compared with the structure shown in the middle) with a small interfering DNA (siDNA) guide (red) bound to an RNA target (blue) (PDB entry 3F73) (right)(Wang et al., 2008a). All of the distinct protein domains are shown in different colors (MID in yellow, PAZ (PIWI– Argonaute– Zwille) in purple, amino-terminal domain (N) in cyan, PIWI in green, and linker domains L1 and L2 in grey). *Aa*Ago, *Aquifex aeolicus* Argonaute; *Af*Ago, *Archaeoglobus fulgidus* Argonaute; *Mj*Ago, *Methanocaldococcus jannaschii* Argonaute; *Mk*Ago, *Methanopyrus kandleri* Argonaute; *Mp*Ago, *Marinitoga piezophila* Argonaute; *Ng*Ago, *Natronobacterium gregoryi* Argonaute; *Pf*Ago, *Pyrococcus furiosus* Argonaute; *Rs*Ago, *Rhodobacter sphaeroides* Argonaute; *Tp*Ago, *Thermotoga profunda* Argonaute; WAGO, worm-specific Argonaute; PIWI-RE, PIWI domain-containing protein with conserved R and E residues. Part **a** is adapted with permission from (Swarts et al., 2014a)

**Table 1 | Guide and target preferences of characterized eAgo and long pAgo proteins.**

| eAgo and long pAgo proteins | | Activity | | | |
|---|---|---|---|---|---|
| **Host** | **Argonaute name** | **RNA-guided RNA interference** | **RNA-guided DNA interference** | **DNA-guided RNA interference** | **DNA-guided DNA interference** |
| Eukaryotes | eAgo | + | – | – | – |
| *Marinitoga piezophila** | *Mp*Ago | + | + | – | – |
| *Thermotoga profunda** | *Tp*Ago | – | + | – | – |
| *Rhodobacter sphaeroides* | *Rs*Ago | – | + | – | – |
| *Aquifex aeolicus* | *Aa*Ago | – | – | + | – |
| *Natronobacterium gregoryi*‡ | *Ng*Ago | – | – | unknown | unknown |
| *Thermus thermophilus* | *Tt*Ago | – | – | + | + |
| *Methanocaldococcus jannaschii* | *Mj*Ago | – | – | – | + |
| *Pyrococcus furiosus* | *Pf*Ago | – | – | – | + |

Plus signs indicate activity that has been demonstrated. Minus signs indicate either inactivity under tested conditions or untested. eAgo, eukaryotic Argonaute protein; pAgo, prokaryotic Argonaute protein. *Determined using 5′-hydroxylated guides. ‡Activity under investigation.

## DNA-guided DNA interference

The first evidence that *Tt*Ago is involved in host defense was obtained after the discovery that a *T. thermophilus* strain, which was selected for its high natural uptake of foreign DNA, had a *Tt*Ago gene that was disrupted by the insertion of a transposon (Swarts et al., 2014b). Further investigation showed that *Tt*Ago does not regulate the expression of other genes that are involved in host defense at the mRNA level. Instead, it directly interferes with invading nucleic acids at the DNA level by mediating DNA-guided DNA interference (Swarts et al., 2015b) (**Figure 2a**). *In vivo*, *Tt*Ago associates with small interfering DNA (siDNA) guides (13–25 nucleotides in length) that are used by *Tt*Ago to bind to and cleave cognate DNA targets. The siDNAs that associate with *Tt*Ago *in vivo* are preferentially derived from plasmid DNA, which indicates that *Tt*Ago can distinguish between invading DNA and genomic DNA (Swarts et al., 2014b). It was recently shown that *Tt*Ago uses its catalytic site to generate its own guides using a mechanism known as ‘chopping’ (Swarts et al., 2017a). *In vitro*, guide-free *Tt*Ago chops DNA with a certain degree of instability (for example, AT-rich double-stranded DNA (dsDNA) or DNA with mismatched DNA stretches) and generates small dsDNA fragments that are subsequently loaded onto *Tt*Ago. It is hypothesized that, after the binding of the small dsDNA molecule to *Tt*Ago, the strand that is complementary to the guide strand (the passenger strand) is released, whereas *Tt*Ago remains associated with the guide strand (siDNA) (Swarts et al., 2017a). Once the *Tt*Ago–siDNA complex is formed, it can bind to and cleave its cognate single-stranded target DNA. Furthermore, two *Tt*Ago–siDNA complexes with complementary siDNA guides can introduce dsDNA breaks (Swarts et al.,

2014b) (**Figure 2a**). Through this mechanism, *Tt*Ago interferes with invading DNA that is taken up by natural competence systems and it also decreases intracellular plasmid levels. In addition, it was shown that *Tt*Ago can cleave RNA *in vitro* (Swarts et al., 2014b) and that it decreases the levels of intracellular RNA that are transcribed from invading DNA *in vivo* (Swarts et al., 2015b). However, it remains to be shown whether this observed decrease in RNA levels is directly caused by the targeting of the RNA or whether it is instead an indirect consequence of DNA interference.



Similar to *Tt*Ago, long pAgos from the archaea *Pyrococcus furiosus* (*Pf*Ago) and *Methanocaldococcus jannaschii* (*Mj*Ago) mediate DNA-guided DNA cleavage *in vitro* (Swarts et al., 2015a; Willkomm et al., 2017a; Zander et al., 2014). In addition, *Pf*Ago has been shown to decrease the efficiency of plasmid transformation in *P. furiosus* (Swarts et al., 2015a). Although the physiological role of *Mj*Ago has not been studied, it has been shown that, similar to *Tt*Ago, *Mj*Ago generates siDNA by chopping longer DNA substrates (Zander et al., 2017). Interestingly, chromatinized DNA derived from *M. jannaschii* is not chopped by *Mj*Ago (Zander et al., 2017), which demonstrates how this pAgo can distinguish between invader DNA and its own genomic DNA. Although chromatization of the genomic DNA has only been shown to protect against degradation for *Mj*Ago, it is tempting to speculate that other genome-packing proteins can have the same function in other archaea and bacteria.

It is noteworthy that for both *Tt*Ago and *Mj*Ago, canonical target cleavage that is mediated by siDNA is substantially more efficient than guide-independent DNA chopping (Zander et al., 2017). Thus, although DNA chopping might provide some level of innate immunity against invading nucleic acids, siDNA-guided DNA interference enables the efficient removal of invader DNA. In addition, pAgo–siDNA complexes might target multiple copies of the same invader DNA and so may potentially interfere with previously encountered invader DNA for multiple generations.

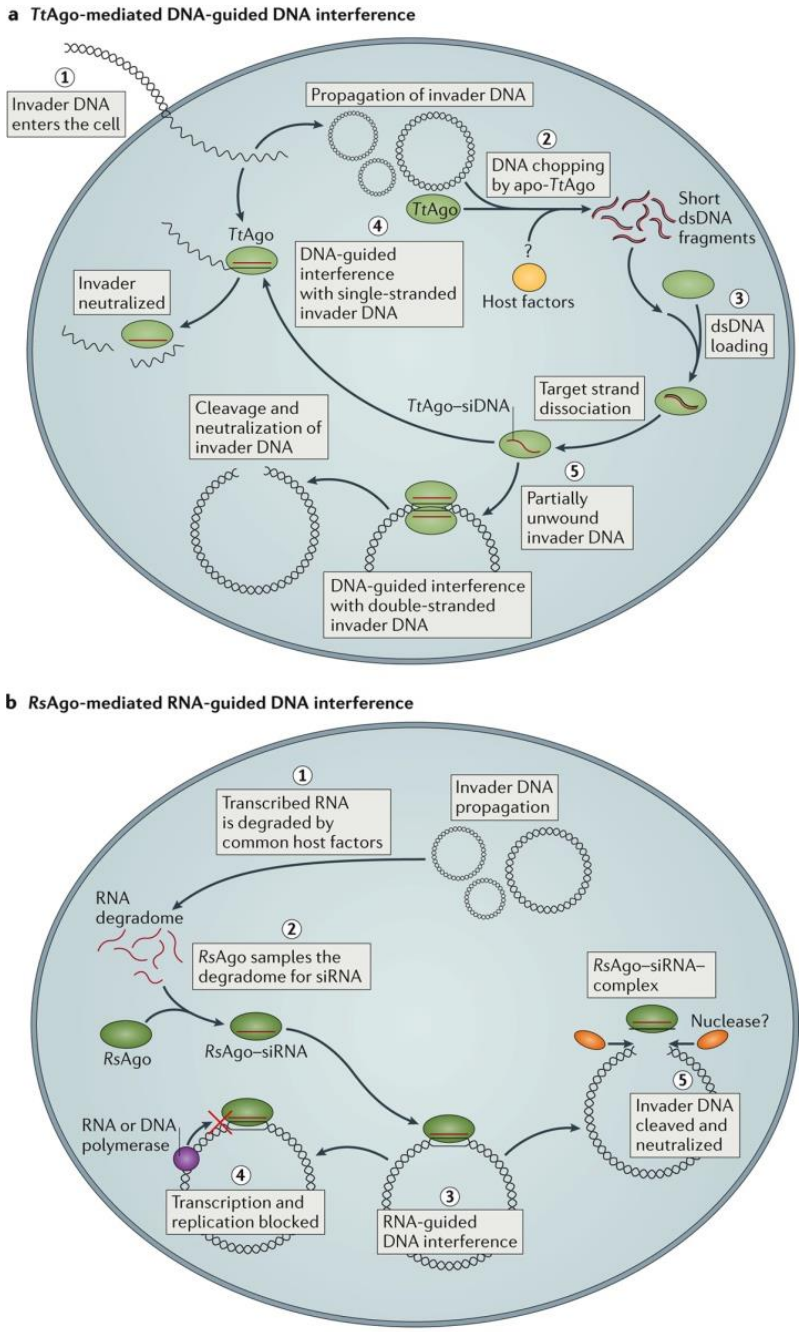

a TtAgo-mediated DNA-guided DNA interference
1
Invader DNA enters the cell
Propagation of invader DNA
2
DNA chopping by apo-TtAgo
TtAgo
Short dsDNA fragments
TtAgo
4
DNA-guided interference with single-stranded invader DNA
?
Host factors
Invader neutralized
3
dsDNA loading
Target strand dissociation
Cleavage and neutralization of invader DNA
TtAgo–siDNA
5
Partially unwound invader DNA
DNA-guided interference with double-stranded invader DNA
b RsAgo-mediated RNA-guided DNA interference
1
Transcribed RNA is degraded by common host factors
Invader DNA propagation
RNA degradome
2
RsAgo samples the degradome for siRNA
RsAgo–siRNA–complex
Nuclease?
RsAgo
RsAgo–siRNA
5
Invader DNA cleaved and neutralized
RNA or DNA polymerase
4
Transcription and replication blocked
3
RNA-guided DNA interference

**Figure 2 | Prokaryotic Argonaute protein-mediated DNA interference**. **a |** *Thermus thermophilus* Argonaute (*Tt*Ago)-mediated DNA-guided DNA interference. Invading DNA enters the cell through phage infection, conjugation or natural transformation (step 1) (Blesa et al., 2015). Natural transformation is shown, during which DNA enters the cell as single-stranded DNA (ssDNA) (Averhoff, 2009). Following the propagation of the invading DNA, guide-free *Tt*Ago (apo-*Tt*Ago) chops the DNA, possibly after being stimulated by other host factors (indicated by a question mark; step 2) (Swarts et al., 2017a). This generates short double-stranded DNA (dsDNA) fragments that are loaded onto *Tt*Ago, which selects for a guanine in the passenger (non-guide) strand opposite the first nucleotide of the guide strand, a cytosine (step 3, specific nucleotides not shown) (Swarts et al., 2017a). After cleavage and release of the passenger strand, *Tt*Ago is loaded with small interfering DNAs (siDNAs; red) (Swarts et al., 2017a). The *Tt*Ago–siDNA complex binds to and cleaves nucleic acids that are complementary to the siDNA. The complex can cleave ssDNA (step 4) (for example, during natural transformation), mRNA and dsDNA targets (step 5). The complex can cleave dsDNA targets by individually targeting two partially unwound strands of the DNA using two *Tt*Agos (Swarts et al., 2014b). **b** | *Rhodobacter sphaeroides* Argonaute (*Rs*Ago)-mediated RNA-guided DNA interference. Invader DNA is transcribed into mRNA that is eventually degraded into small RNA fragments, probably by common host factors (step 1) (Olovnikov et al., 2013). The RNA degradome is sampled by *Rs*Ago, which results in the loading of a small interfering RNA (siRNA; red) guide (step 2) (Miyoshi et al., 2016; Olovnikov et al., 2013). The *Rs*Ago–siRNA interference complex binds to partially unwound dsDNA complementary to the siRNA (step 3) (Miyoshi et al., 2016; Olovnikov et al., 2013). This can result in the blocking of transcription and replication (step 4), or invader DNA degradation by unknown host factors (indicated by a question mark; step 5) (Olovnikov et al., 2013). It should be noted that *Rs*Ago is encoded by an operon that additionally encodes a predicted nuclease, which possibly has a role in the cleavage of the invader DNA (Olovnikov et al., 2013). Note that steps 2, 4 and 5 are based on a proposed mechanism for which no direct experimental evidence currently exists.



## RNA-guided DNA interference

*Rs*Ago contains an incomplete DEDX motif and is thus predicted to be catalytically inactive. The gene that encodes *Rs*Ago is clustered with a gene that encodes a putative DNA nuclease that is predicted to contribute to the degradation of nucleic acids targeted by *Rs*Ago (Makarova et al., 2009). Nevertheless, despite the absence of a catalytic motif and the predicted nuclease partner, heterologous expression of *Rs*Ago in *Escherichia coli* resulted in lower concentrations of intracellular plasmids and a lower expression of plasmid-encoded genes than in *E. coli* cells in which *Rs*Ago was not expressed (Olovnikov et al., 2013). Similar to *Tt*Ago, *Rs*Ago does not affect the expression of endogenous genes that are involved in host defense, which suggests that it instead directly mediates host defense (Olovnikov et al., 2013) (**Figure 2b**).

*In vivo*, *Rs*Ago associates with small interfering RNA (siRNA) guides (14–19 nucleotides in length) that are sense-oriented, which suggests that they are derived from degraded mRNA (Olovnikov et al., 2013). In addition, as *Rs*Ago binds to siRNA guides when heterologously

expressed in *E. coli* (that is, in the absence of the genetically colocalized nuclease), guide generation does not exclusively rely on the genetically colocalized nuclease or on other *R. sphaeroides*-specific host factors (Olovnikov et al., 2013). The siRNAs are enriched in ribosomal RNA sequences and, interestingly, are also enriched in sequences that are derived from transposon-encoded or plasmid-encoded mRNA (Olovnikov et al., 2013). It has been suggested that *Rs*Ago samples the RNA degradome to generate its guide siRNAs, but how this enrichment is achieved is currently unknown.

Moreover, *Rs*Ago–siRNA complexes co-purify with DNA molecules that are 22–24 nucleotides in length and are enriched in exogenous sequences derived from plasmid DNA, genome-encoded transposon or phage genes (Olovnikov et al., 2013). Complementarity with the RNA guides indicates that these DNA molecules are co-purified target strands. This is in agreement with the observation that *Rs*Ago has a higher affinity for RNA guides than for DNA guides, and a higher affinity for DNA targets than for RNA targets (Miyoshi et al., 2016).

Elucidation of the structure of *Rs*Ago revealed the amino acids that specifically interact with the 2′-hydroxyl groups of RNA guides and showed that, in contrast to eAgos, *Pf*Ago and *Tt*Ago, the PAZ domain of *Rs*Ago seems to lack a binding pocket for the 3′ end of the guide (Miyoshi et al., 2016) (**Figure 3a, b**). These structural deviations might enhance the stable binding of the *Rs*Ago–siRNA complex to DNA targets, as these deviations potentially enable full base-pairing between the siRNA and the DNA target strand without inducing conformational changes; in other pAgos, such conformational changes are associated with target cleavage (Sheng et al., 2014). Although binding of the *Rs*Ago–siRNA complex might be sufficient to silence invading DNA by inhibiting transcription and/or replication, it has been proposed that DNA target binding can also result in the recruitment of an unknown nuclease that cuts the targeted DNA sequence at both ends, generating the observed DNA fragments that co-purify with *Rs*Ago–siRNA complexes (Olovnikov et al., 2013). Potentially, the nuclease that genetically colocalizes with *Rs*Ago fulfils this role; however, it should be noted that these small DNA molecules also co-purify with *Rs*Ago–siRNA complexes in the absence of this nuclease (Olovnikov et al., 2013). This suggests that, although this nuclease could contribute to *Rs*Ago-mediated DNA interference in *R. sphaeroides*, it is not essential.

## Other long pAgos

Other long pAgos that have been characterized include Argonautes from the bacteria *Archaeoglobus fulgidus* (*Af*Ago), *Aquifex aeolicus* (*Aa*Ago), *Marinitoga piezophila* (*Mp*Ago) and *Thermotoga profunda* (*Tp*Ago). At least *in vitro*, *Af*Ago has a preference for ssDNA over ssRNA as a guide and binds to DNA targets (Parker et al., 2005). However, it has not been reported whether this pAgo can cleave DNA targets. *Aa*Ago mediates DNA-guided RNA cleavage, although DNA targeting has not been tested (Yuan et al., 2005). Genes that encode *Mp*Ago and *Tp*Ago cluster with genes that encode CRISPR–Cas enzymes, which suggests a functional link (Kaya et al., 2016). In contrast to most other pAgos, which generally use guides with a 5′-phosphate, both *Mp*Ago and *Tp*Ago use small RNA guides with a 5′-hydroxyl group to cleave cognate ssDNA targets (Kaya et al., 2016). In addition, *Mp*Ago can use a 5′-hydroxylated RNA guide to cleave ssRNA targets (Kaya et al., 2016). Thus, although the role of *Mp*Ago and *Tp*Ago might be similar to that of other long pAgos, their mechanism of guide generation and binding is distinct (Kaya et al., 2016).



Besides these partially characterized pAgos, other long pAgos remain unexplored. Bioinformatics analyses have shown that some genes that encode long pAgos are fused to genes that encode Schlafen-like ATPases, whereas others co-occur in predicted operons with genes that encode Mrr, Sir2, Cas4-like or PLD nucleases (Makarova et al., 2009) (**Figure 3b**). It is anticipated that these pAgo-associated enzymes are involved in guide generation, double-stranded target DNA unwinding and/or target degradation.

## Short pAgos and PIWI-RE proteins

Short pAgos consist of only a MID domain and a PIWI domain in which the catalytic tetrad is incomplete (Makarova et al., 2009) (**Figure 3c**). Although they lack the N-terminal and PAZ domains, genes that encode short pAgos are typically associated with a gene that encodes the uncharacterized APAZ (analogue of PAZ) domain (Makarova et al., 2009). Genes that encode APAZ domains are always fused to a predicted nuclease domain from the Mrr, Sir2 or TIR families (Makarova et al., 2009). In other short pAgo variants, the predicted APAZ–nuclease domains are additionally fused to the short pAgo (Makarova et al., 2009). Alternatively, the Sir2–APAZ domains are fused either to a Schlafen-like ATPase domain, or to both the Mrr and the TIR domains (Swarts et al., 2014a). It is predicted that short pAgos

function as a platform for nucleic acid-guided target recognition, whereas the associated APAZ–nuclease enzymes have a role in guide generation and/or target degradation.

Similar to short pAgos, PIWI-RE proteins also lack the N-terminal and PAZ domains, although the MID and PIWI domains are conserved, which suggests that they provide a platform for nucleic acid-guided target recognition (Burroughs et al., 2013) (**Figure 3d**). In contrast to short pAgos, some PIWI-RE proteins have a complete catalytic tetrad, which suggests that they can cleave their targets (Burroughs et al., 2013). Furthermore, PIWI-RE proteins contain a uniquely conserved 'domain X' at the N-terminal end, which is unrelated to the PAZ and N-terminal domains in long pAgos (Burroughs et al., 2013). This domain might provide an interaction platform for the helicases and nucleases that co-occur with most PIWI-RE proteins (Burroughs et al., 2013). Most of the PIWI-RE proteins cluster on operons with both a DinG-type helicase and a predicted nuclease. DinG-type helicases act on RNA–DNA duplexes that are primarily found during replication and the transcription of phages and plasmids (Burroughs et al., 2013). These helicases are predicted to be involved in making double-stranded targets accessible for PIWI-RE–guide complexes, and/or in guide–target dissociation, whereas the co-occurring nuclease is anticipated to be involved in guide generation and/or target cleavage (Burroughs et al., 2013).

## Genome editing by pAgo

Owing to the programmable and DNA-targeting nature of some long pAgos, it is tempting to speculate about applications of pAgos for DNA-editing purposes. Below, we describe how pAgos have been used to cleave dsDNA targets *in vitro* and discuss the advantages and disadvantages of pAgos compared with established genome-editing tools.

Both *Tt*Ago and *Pf*Ago can be loaded with siDNA guides to specifically linearize dsDNA targets *in vitro* (Enghiad and Zhao, 2017; Swarts et al., 2015a, 2014b, 2017a; Zander et al., 2017). As the pAgo–guide complexes currently described only mediate the cleavage of single-stranded targets, two pAgos with complementary guides need to be incubated with the target DNA to generate a double-stranded break. The mechanism of dsDNA target cleavage has been studied in detail for *Pf*Ago, providing a method to generate dsDNA breaks with any desired overhang in a plasmid backbone or DNA insert (Enghiad and Zhao, 2017). As guide-free pAgos can also mediate DNA chopping, pAgos should be saturated with siDNAs to

prevent off-target DNA cleavage (Swarts et al., 2017a; Zander et al., 2017). Considering that the synthesis of siDNA guides is relatively cheap and can theoretically be designed to specifically target any desired location, this technique can potentially replace restriction enzymes.

CRISPR–effector complexes, consisting of Cas9 or Cas12a (also known as Cpf1) and a variable RNA guide, are well-established as programmable genome-editing tools (Hsu et al., 2013; Zetsche et al., 2015). These complexes can generate dsDNA breaks in genomes that can be repaired by non-homologous end-joining or homologous recombination processes (Sander and Joung, 2014). Besides complementarity between the RNA guide and the target DNA, CRISPR–effector complexes require a protospacer adjacent motif (PAM) next to the targeted DNA sequence; this requirement limits the sites that are targetable by CRISPR–effector complexes (Hsu et al., 2013; Zetsche et al., 2015).



Long pAgos that cleave ssDNA are possible alternatives to CRISPR–effector complexes as genome-editing tools. Unlike Cas9 and Cas12a, pAgos do not require a PAM for target cleavage, which enables them to target more sequences than CRISPR–effector complexes. Furthermore, pAgos typically use DNA guides that are smaller (15–24 nucleotides in length) than typical CRISPR–effectors RNA guides (Cas9 and Cas12a use RNA guides that are 100 and 43 nucleotides in length, respectively). Short DNA guides are cheaper to synthesize than long RNA guides and potentially enable the high-throughput generation of guides and, consequently, high-throughput genome-editing screenings. In addition, with a size of 75–85 kDa, long pAgos are about half of the size of Cas9 and Cas12a, which could make delivery into the desired host more efficient. However, in contrast to Cas9 and Cas12a, the generation of dsDNA breaks requires a pair of pAgo–guide complexes.

The possibility of generating dsDNA breaks *in vitro* by using a pair of long pAgos with complementary guides was initially only described for pAgos that were derived from thermophiles, which are mainly active at increased temperatures, probably because the dsDNA is partially unwound at those temperatures (Enghiad and Zhao, 2017; Swarts et al., 2015a, 2014b). Obviously, this makes these pAgos less suitable as a standalone genome-editing tool in mesophilic organisms. Lack of helicase activity probably explains why there are currently no characterized pAgos that efficiently cleave dsDNA at mesophilic temperatures *in vitro*. *In vivo*, these and other mesophilic pAgos potentially rely on natural

processes that initiate the local unwinding of dsDNA (such as RNA transcription or DNA replication) for efficient dsDNA targeting. It has already been shown that pAgos efficiently interfere with dsDNA plasmid propagation in bacterial cells (Swarts et al., 2014b, 2015b). However, targeting of dsDNA by long mesophilic pAgos requires further investigation. At least some genes that encode pAgos co-occur with, or are fused to, genes that encode predicted helicases that potentially enable pAgos to target dsDNA (Makarova et al., 2009; Swarts et al., 2014a).

The search for a mesophilic pAgo that targets DNA resulted in the characterization of *Natronobacterium gregoryi* Argonaute (*Ng*Ago) (Gao et al., 2016). In one study, it was reported that *Ng*Ago displays DNA-guided dsDNA cleavage at 37°C, and that it is possible to use *Ng*Ago for genome editing in human cell lines (Gao et al., 2016). However, this work is heavily disputed as many laboratories could not reproduce the described results (Cyranoski, 2016; Javidi-Parsijani et al., 2017; Lee et al., 2016). A recent study showed that *Ng*Ago can be programmed to downregulate specific genes in zebrafish (Qi et al., 2016). As gene knockdown was also observed in the absence of the catalytic DEDX tetrad of *Ng*Ago, it has been suggested that gene knockdown is a result of DNA binding by the *Ng*Ago–siDNA complex, which induces transcriptional silencing (Qi et al., 2016). However, preliminary data from another study suggest that, in contrast to earlier findings, *Ng*Ago mediates DNA-guided RNA cleavage and thus post-transcriptionally regulates gene expression, and so resembles eukaryotic RNAi pathways (Kim et al., 2017). However, these data require further validation. Whereas RNAi is an efficient system to knock down genes of interest in many organisms, some organisms lack a functional RNAi system that can be hijacked for gene knockdown purposes, or suffer from toxicity due to off-target effects of RNAi-mediated gene knockdown (Qi et al., 2013). *Ng*Ago–siDNA complexes can potentially function as an alternative to RNAi in these organisms.

The delivery of pAgo–siDNA complexes into cells poses another potential limitation for the use of pAgos as a tool for genome editing or gene knockdown purposes. CRISPR–effector complexes can be delivered into cells by protein transfection, but as the complex also efficiently assembles *in vivo*, Cas9 or Cas12a and/or the RNA guide can be expressed in the cell after viral-mediated transfection or plasmid transduction (Hsu et al., 2013). As at least some pAgos mediate DNA chopping in the absence of siDNA, these pAgos could potentially

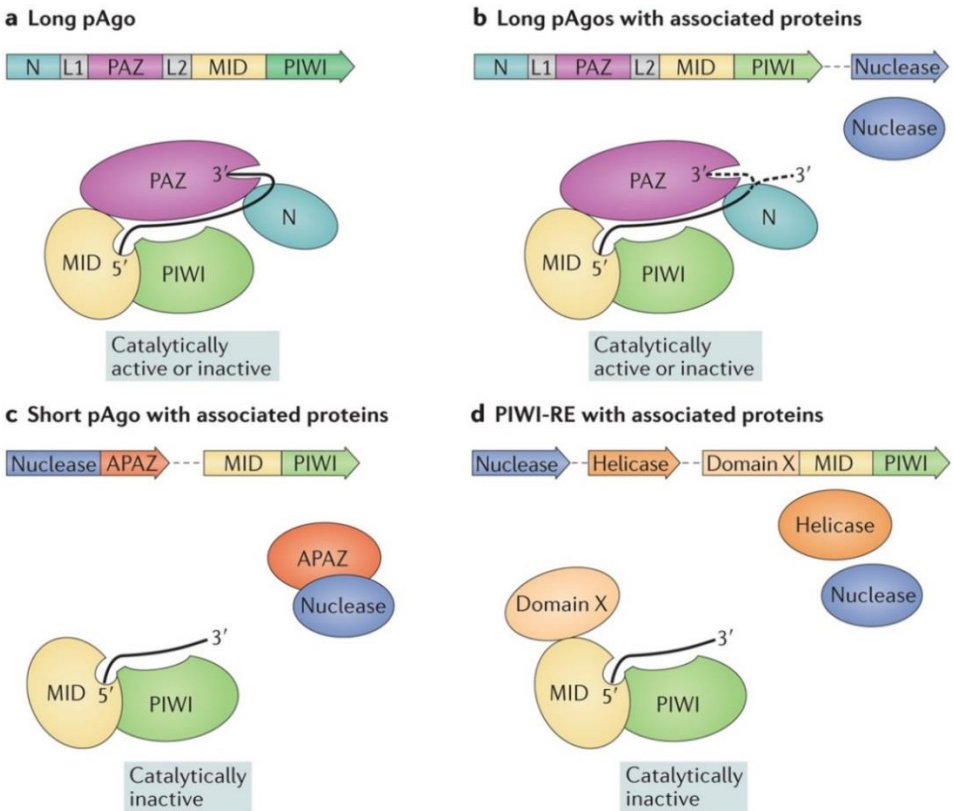



**Figure 3 | Domain architectures of prokaryotic Argonaute proteins.** For each prokaryotic Argonaute protein (pAgo) class, a schematic representation of a pAgo-encoding operon and a schematic representation of the predicted protein scaffold and functions are provided. In the schematic operon, individual genes are indicated by individual arrows. In the schematic predicted protein scaffold, protein domains are represented by distinct colors and the guide is indicated by a black line and shows the predicted anchoring of its 5′ and/or 3′ ends to the pAgo scaffold. The associated proteins most likely have a role in the unwinding and/or cleaving of the target nucleic acids. Catalytically active long pAgos that contain a complete catalytic DEDX tetrad in the P-element induced wimpy testis (PIWI) domain or catalytically inactive long pAgos are shown (part **a**). Catalytically active or inactive long pAgos (part **b**) that are associated with nucleases from the Mrr, Sir2, Cas4-like or PLD protein families are shown. Not all catalytically inactive long pAgos bind to the 3′ end of the guide in the PIWI–Argonaute–Zwille (PAZ) domain (indicated by the dashed line). Short pAgos (part **c**) that are associated with an APAZ (analogue of PAZ) domain and proteins from the Mrr, Sir2 or TIR family are shown. PIWI-RE proteins (part **d**) that are associated with a DinG-type helicase and predicted nuclease are shown. Domain X indicates a conserved globular region the biochemical function of which is still unknown (Burroughs et al., 2013). Of note, genes that encode PIWI-RE proteins with an intact catalytic site also exist (not shown), but they are often found as standalone genes and lack operon-associated proteins (Burroughs et al., 2013). MID, middle; N, amino-terminal domain; PIWI-RE, PIWI domain-containing protein with conserved R and E residues.

cause toxicity through nonspecific DNA degradation. Furthermore, as siDNAs cannot be expressed *in vivo*, the transduction of *in vitro*-assembled pAgo–siDNA complexes is theoretically the only suitable method for cellular delivery. Even if pAgos can be expressed *in vivo* without causing toxicity (for example, by mutating residues that are involved in DNA chopping), the transfection of synthetic siDNAs potentially interferes with RNAi pathways, as eAgos can also bind to ssDNA molecules (Smalheiser and Gomes, 2015). Future studies are required to explore feasible delivery methods. Nevertheless, at least for *Ng*Ago, the transfection of *in vitro*-assembled pAgo–siDNA complexes was shown to be suitable for efficient cellular delivery (Qi et al., 2016).

## Conclusions

During the past 3 years, considerable progress has been made in elucidating the role and mechanisms of pAgo proteins. Several studies have revealed that pAgos constitute a diverse class of nucleic acid-guided nucleases, of which at least a subset has a role in host defense by interfering with invading nucleic acids. The long pAgo class has been studied in the most detail, whereas other pAgo classes are as yet uncharacterized. Further studies are required to shed light on the functions and mechanisms of these uncharacterized pAgo classes. Interestingly, long pAgos display variations in their guide and target specificity. Most long pAgos seem to target DNA using either DNA or RNA guides. This DNA-targeting feature led to the suggestion of the potential use of pAgos as programmable genome-editing tools. Although *Tt*Ago and PfAgo can be used to selectively cleave dsDNA targets, they function optimally at increased temperatures, thus making them less suitable for the genome editing of cells from mesophilic organisms. Other pAgos, such as *Mp*Ago, cleave ssDNA targets but not dsDNA targets, probably because of the lack of helicase activity. The mesophilic *Rs*Ago has been demonstrated to mediate RNA-guided DNA interference *in vivo*, but does not seem to catalyze DNA cleavage.

Currently, not a single characterized pAgo displays robust genome-editing capabilities. However, it is notable that of the equally diverse pool of CRISPR–Cas enzymes only a few are suitable for genome-editing applications. Therefore, given the high functional and mechanistic variation that pAgos display, the vast pools of uncharacterized long pAgos, short pAgos and PIWI-RE proteins potentially contain hidden gems that can overcome the bottlenecks of the currently characterized pAgos.

# Chapter 3

## DNA-guided DNA cleavage by *Clostridium butyricum* Argonaute

Jorrit W. Hegge, Daan C. Swarts, Stanley D. Chandradoss, Tao Ju Cui, Jeroen Kneppers, Martin Jinek, Chirlmin Joo, John van der Oost

Adapted from –
'DNA-guided DNA cleavage at moderate temperatures by *Clostridium butyricum* Argonaute'
Nucleic acids Research, accepted for peer-review

# Abstract

Prokaryotic Argonaute proteins (pAgos) constitute a diverse group of endonucleases of which some mediate host defense by utilizing small interfering DNA guides (siDNA) to cleave complementary invading DNA. This activity can be repurposed for programmable DNA cleavage. However, currently characterized DNA-cleaving pAgos require elevated temperatures (≥65°C) for their activity, making them less suitable for applications that require moderate temperatures, such as genome editing. Here we report the functional and structural characterization of the siDNA-guided DNA-targeting pAgo from the mesophilic bacterium Clostridium butyricum (CbAgo). CbAgo displays a preference for siDNAs that have a deoxyadenosine at the 5'-end and thymidines in the sub-seed segment (siDNA nucleotides 2-4). Furthermore, CbAgo mediates DNA-guided DNA cleavage of AT-rich double stranded DNA at moderate temperatures (37°C). This study demonstrates that certain pAgos are capable of programmable DNA cleavage at moderate temperatures and thereby expands the scope of the potential pAgo–based applications.

# Introduction

Eukaryotic Argonaute proteins (eAgos) play a key role in RNA interference (RNAi) processes (Joshua-Tor and Hannon, 2011; Ketting, 2010; Meister, 2013). As the core of the multiprotein RNA-induced silencing complex (RISC), eAgos bind small non-coding RNA molecules as guides to direct the RISC complex towards complementary RNA targets (Bartel, 2009; Meister, 2013; Pratt and MacRae, 2009). Reflecting their physiological function, variation among eAgos is observed with respect to the presence or absence of a catalytic site, and to their potential to interact with other proteins (Kuhn and Joshua-Tor, 2013). Depending on the eAgo and on the sequence complementarity between guide and target RNA, eAgo-guide complexes either catalyze endonucleolytic cleavage of the target RNA (Hutvagner and Simard, 2008) or indirectly silence the target RNA by repressing its translation and promoting its degradation through recruitment of additional silencing factors (Ketting, 2011). Independent of the mechanism, eAgo-mediated RNA binding generally results in sequence-specific silencing of gene expression. As such, eAgos can coordinate various cellular processes by regulating intracellular RNA levels.



Prokaryotes also encode Argonaute proteins (pAgos) (Makarova et al., 2009; Swarts et al., 2014a). Various pAgos share a high degree of structural homology with eAgos as both pAgos and eAgos adopt the same four domain (N-PAZ-MID-PIWI) architecture (Makarova et al., 2009; Song et al., 2004; Swarts et al., 2014a; Wang et al., 2008a). Despite their structural homology, several recently characterized pAgos have distinct functional roles and different guide and/or target preferences compared to eAgos. For example, several pAgos have been implicated in host defense by directly targeting DNA instead of RNA (Olovnikov et al., 2013; Swarts et al., 2015a, 2014b; Zander et al., 2017). One of the best characterized mechanisms that pAgos utilize is DNA-guided DNA interference, which is demonstrated for pAgos from *Thermus thermophilus* (*Tt*Ago), *Pyrococcus furiosus* (*Pf*Ago), and *Methanocaldococcus jannaschii* (*Mj*Ago) (Swarts et al., 2015a, 2014b, 2015b, 2017a; Willkomm et al., 2017a; Zander et al., 2014, 2017). These pAgos use 5'-end phosphorylated small interfering DNAs (siDNAs) for recognition and successive cleavage of complementary DNA targets. This mechanism enables both *Tt*Ago and *Pf*Ago to mediate host defense against invading nucleic acids. Prokaryotes lack homologs of eukaryotic enzymes that are involved in guide biogenesis (Shabalina and Koonin, 2008). Instead, both *Tt*Ago and *Mj*Ago - besides the canonical siDNA-dependent target cleavage termed 'slicing' - exhibit an alternative nuclease

activity termed 'chopping' (Swarts et al., 2017a; Zander et al., 2017). Chopping facilitates autonomous generation of small DNA fragments from dsDNA substrates. Subsequently, these DNA fragments generated during chopping can serve as siDNAs for canonical slicing (Swarts et al., 2017, Zander et al., 2017).

*Tt*Ago and *Pf*Ago can be programmed with short synthetic siDNA which allows them to target and cleave dsDNA sequences of choice *in vitro* (Swarts et al., 2015a, 2014b). This activity has enabled the repurposing of *Pf*Ago as an universal restriction endonuclease for *in vitro* molecular cloning (Enghiad and Zhao, 2017). In addition, a diagnostic TtAgo-based application termed NAVIGATER (Nucleic Acid enrichment Via DNA Guided Argonaute from *Thermus thermophilus*) was developed which enables enhanced detection of rare nucleic acids with single nucleotide precision (Song et al., 2019). In analogy with the now commonly used CRISPR-Cas9 and CRISPR-Cas12a enzymes (Fellmann et al., 2017; Knott and Doudna, 2018; Savić and Schwank, 2016), it has also been suggested that pAgos could be repurposed as next-generation genome editing tools (Hegge et al., 2018a). However, due to the thermophilic nature (optimum activity temperature ≥65°C) and low levels of endonuclease activity at the relevant temperatures (20-37°C), it is unlikely that the well-studied *Tt*Ago, *Pf*Ago and *Mj*Ago are suitable for genome editing. The quest for a pAgo that can cleave dsDNA at moderate temperatures has resulted in the characterization of the Argonaute protein from *Natronobacterium gregory* (*Ng*Ago), which was claimed to be the first pAgo suitable for genome editing purposes (Gao et al., 2016). However, the study reporting this application has been retracted after a series of reproducibility issues (Cyranoski, 2016; Gao et al., 2016; Lee et al., 2016). Instead, it has been suggested that *Ng*Ago targets RNA rather than DNA (Kim et al., 2017).

Although considerable efforts have been made to elucidate the mechanisms and biological roles of pAgos, efforts have mainly focused on pAgo variants from (hyper)thermophiles. This has left a large group of mesophilic pAgos unexplored. We here report the characterization of the Argonaute protein from the mesophilic bacterium *Clostridium butyricum* (*Cb*Ago). We demonstrate that *Cb*Ago can utilize siDNA guides to cleave both ssDNA and dsDNA targets at moderate temperatures (37°C). In addition, we have elucidated the macromolecular structure of *Cb*Ago in complex with a siDNA guide and complementary ssDNA target in a catalytically competent state. *Cb*Ago displays an unusual preference for siDNAs with a deoxyadenosine at the 5'-end and thymidines in the sub-seed segment (siDNA nt 2-4). The

programmable DNA endonuclease activity of *Cb*Ago provides a foundation for the development of pAgo-based applications at moderate temperatures.

# Results

## *Cb*Ago mediates siDNA-guided ssDNA cleavage

*Cb*Ago was successfully expressed in *Escherichia coli* from a codon-optimized gene using a T7-based pET expression system and purified (**Figure S1A**). To determine the guide and target binding characteristics of *Cb*Ago, we performed single-molecule experiments using Förster resonance energy transfer (FRET). We immobilized either Cy5-labeled single stranded RNA or DNA targets (FRET acceptor) on a polymer-coated quartz surface (**Figure 1A**). Next we introduced *Cb*Ago in complex with either a Cy3-labeled siRNA or siDNA guide (FRET donor) and recorded the interactions. Strikingly, *Cb*Ago could utilize both siRNAs and siDNAs to bind DNA or RNA targets (**Figure 1B**). To test which guide is preferentially bound by *Cb*Ago we performed a competition assay in which *Cb*Ago was immobilized into the microfluidic chamber, and an equimolar mixture of siDNA and siRNAs was introduced. While only short-lived interactions (average dwell time: 0.48 seconds) were observed for siRNA, siDNA was strongly bound (average dwell time: 44 seconds) by *Cb*Ago (**Figure 1C**). This results suggests that *Cb*Ago utilizes siDNA rather than siRNA as a guide.



*Cb*Ago is phylogenetically closest related to the clade of halobacterial pAgos, among which also pAgo from *Natronobacterium gregoryi* (*Ng*Ago) can be found (**Figures 1D**, **S2**). A multiple sequence alignment of *Cb*Ago with other pAgos (**Figure S1B**) suggests that *Cb*Ago contains the conserved DEDX catalytic residues (where X can be a D, H or N) which are essential for nuclease activity in 'slicing' Agos (Nakanishi et al., 2012). In the case of *Cb*Ago, this concerns residues D541, E577, D611 and D727.

To confirm whether *Cb*Ago indeed is an active nuclease, we performed *in vitro* activity assays in which *Cb*Ago was loaded with either synthetic siDNAs or siRNAs (21 nucleotides in length). Next the complexes were incubated at 37°C with 45-nucleotide complementary single stranded RNA or DNA target oligonucleotides. While no activity was found in any of the combinations in which siRNAs or target RNAs were used, *Cb*Ago was able to cleave target DNAs in a siDNA-dependent manner (**Figure 1E**).

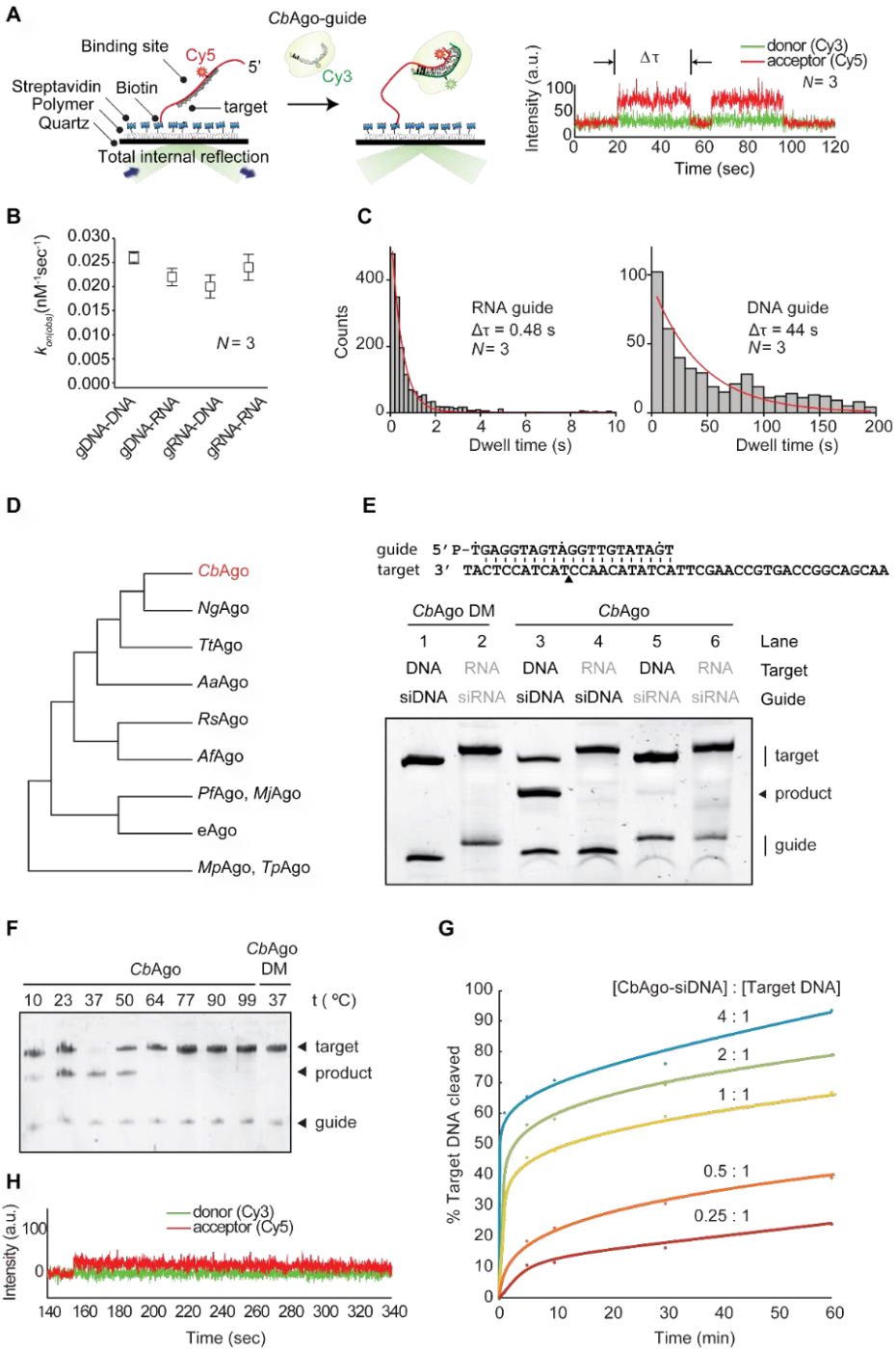

A
CbAgo-guide
Binding site
Cy5
5'
Streptavidin
Polymer
Quartz
Biotin
target
Cy3
Total internal reflection
donor (Cy3)
acceptor (Cy5)
N= 3
Intensity (a.u.)
Δτ
Time (sec)
B
N= 3
gDNA-DNA
gDNA-RNA
gRNA-DNA
gRNA-RNA
C
RNA guide
Δτ = 0.48 s
N= 3
DNA guide
Δτ = 44 s
N= 3
Counts
Dwell time (s)
D
CbAgo
NgAgo
TtAgo
AaAgo
RsAgo
AfAgo
PfAgo, MjAgo
eAgo
MpAgo, TpAgo
E
guide 5′ P-ṪGAGGTAGTAĠGTTGTATAĠT
target 3′ TACTCCATCATCCAACATATCATTCGAACCGTGACCGGCAGCAA
CbAgo DM
CbAgo
Lane
Target
Guide
DNA
RNA
siDNA
siRNA
target
product
guide
F
CbAgo
CbAgo DM
t (°C)
target
product
guide
G
[CbAgo-siDNA] : [Target DNA]
4 : 1
2 : 1
1 : 1
0.5 : 1
0.25 : 1
% Target DNA cleaved
Time (min)
H
donor (Cy3)
acceptor (Cy5)
Intensity (a.u.)
Time (sec)

**Figure 1 | *Cb*Ago exhibits DNA-guided DNA endonuclease activity at 37°C. A** | Left: Overview of the single molecule assay to determine the binding characteristics of *Cb*Ago. Right: FRET diagram of a *Cb*Ago-siDNA complex that has 3 complementary base pairs (2-4nt) to the DNA target. Indicated is the dwell time ($\Delta\tau$). **B** | Comparison of the binding rates ($k_{on}$) of *Cb*Ago in complex with siDNA or siRNA to bind DNA or RNA targets. The rates are similar for each nucleic acid type guide and target. *N* is the number of base paired nucleotides. **C** | Dwell time histograms showing *Cb*Ago preferentially binds siDNAs in siDNA-siRNA competition experiments. **D** | Schematic phylogenetic tree of characterized pAgos. **E** | *Cb*Ago exhibits DNA-guided DNA endonuclease activity. Upper panel: Sequence of the synthetic let7 miRNA-based siDNA guide and target DNA sequences that were used for the *in vitro* assays. Lower panel: *Cb*Ago, guides and targets were mixed in a 1:1:1 molar ratio and incubated for 1 h at 37°C. Catalytic mutant *Cb*AgoDM was used as a control. Cleavage products were analyzed by denaturing polyacrylamide electrophoresis. **F** | *Cb*Ago displays highest activity at 37°C. *Cb*Ago and siDNA were mixed and pre-incubated at various temperatures for 10 minutes. Next, target DNA was added and the sample was incubated for 1 h at the same temperature. *Cb*AgoDM was used as a control. Cleavage products were analyzed by denaturing polyacrylamide electrophoresis. **G** | Quantified data of a *Cb*Ago-mediated siDNA-guided ssDNA cleavage turnover experiment using 5 pmol target DNA and increasing concentrations of *Cb*Ago-siDNA (1.25-20 pmol). h, FRET diagram showing that a cleavage compatible *Cb*Ago-siDNA remains bound to a fully complementary target DNA (*N*=21) during the entire the measurement (340 seconds).



In agreement with the predicted DEDD catalytic site (**Figure S1B**), alanine substitutions of two of aspartic acids (D541A, D611A) in the expected catalytic tetrad abolished the nuclease activity, demonstrating that the observed siDNA-guided ssDNA endonucleolytic activity was indeed catalyzed by the DEDD catalytic site. To further investigate the full temperature range at which *Cb*Ago is active, we performed additional cleavage assays at temperatures ranging from 10-95°C. While *Cb*Ago displayed the highest activity at its physiologically relevant temperature (37°C), *Cb*Ago also catalyzed siDNA-guided target DNA cleavage at temperatures as low as 10°C and as high as 50°C (**Figure 1F**). When *Cb*Ago-siDNA complexes and target ssDNA substrates (45nt) were mixed in equimolar amounts, cleavage of the target DNA was not complete after one hour incubation (**Figure 1E**). Therefore, we investigated the substrate turnover kinetics of *Cb*Ago by monitoring the cleavage assays in a time course using variable *Cb*Ago:siDNA:target DNA ratios (**Figure 1G**). A rapid burst of activity was observed during the first minute, likely indicating the first target binding and cleavage event. This stage was followed by a slow steady state, suggesting that under these conditions the *Cb*Ago-siDNA complex slowly dissociates from the cleaved target DNA product before being able to bind and cleave a new target DNA strand. The cleavage kinetics were confirmed using single-molecule assays which demonstrated that the *Cb*Ago-siDNA complex remains bound to the DNA target (*N*=21) for several minutes (**Figure 1H**), which prevents *Cb*Ago-siDNA complexes from binding and cleaving new DNA targets. Thus, while

*Cb*Ago functions as a multi-turnover nuclease enzyme, its steady-state rate is limited by product release.

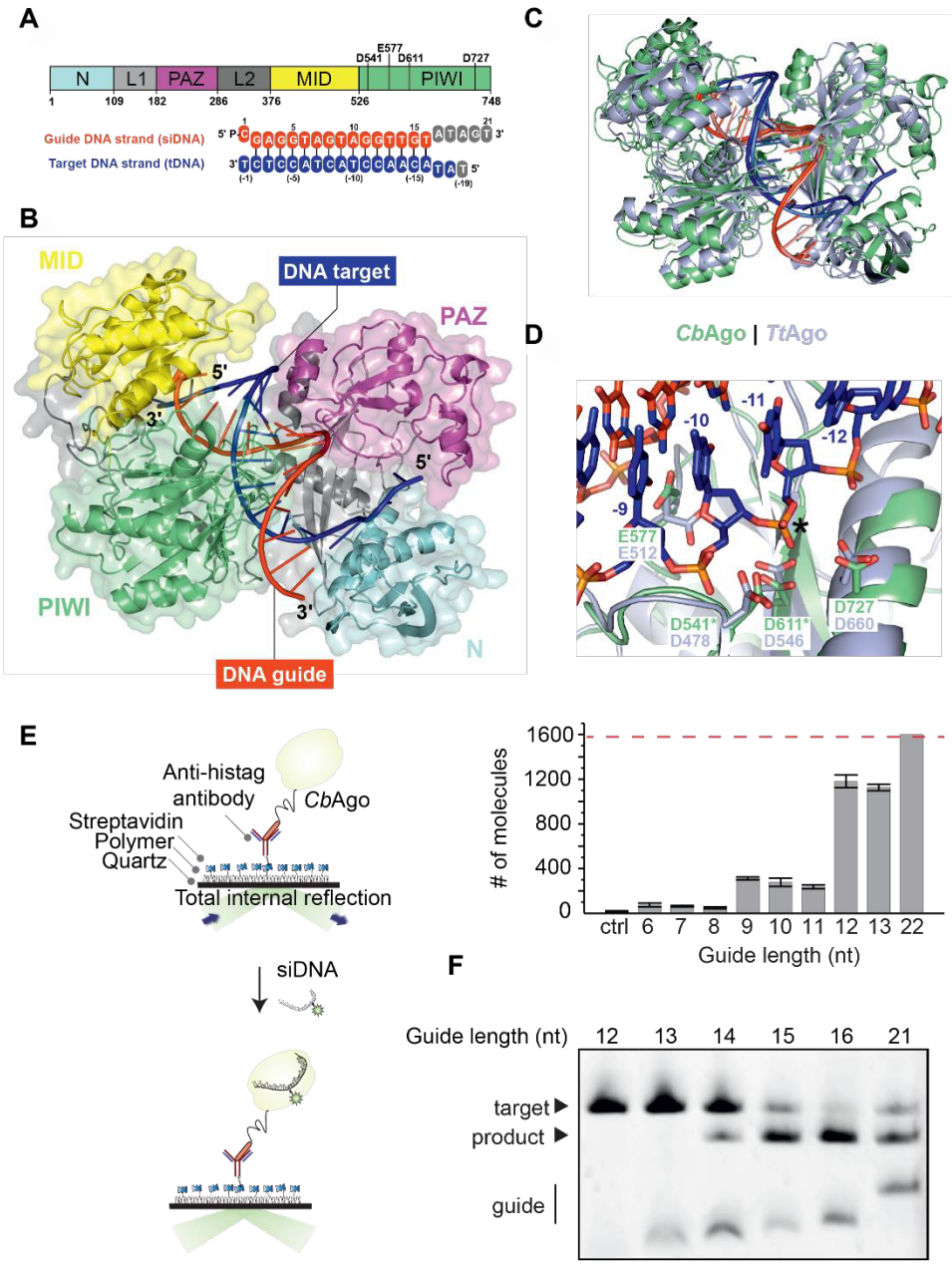


**Figure 2 | Structure of *Cb*Ago in complex with a siDNA and a DNA target. A** | Upper panel: Schematic diagram of the domain organization of *Cb*Ago. L1 and L2 are linker domains. Lower panel: Sequences of the siDNA (red) and target DNA (blue). Nucleotides that are unordered in the structure are colored grey. See also **Table S1. B** | Overall structure of the *Cb*Ago-siDNA-target DNA complex. Domains are colored according to the color scheme in panel (**A**). **C** | Structural alignment of *Cb*Ago (green) and *Tt*Ago (light purple; PDB: 4NCB). Core Root Mean Square Deviation of 3.0 Å over 563 residues. **D** | Close-up view of the aligned DDED catalytic sites of *Cb*Ago (green) and *Tt*Ago (light purple; PDB: 4NCB). Modelled side chains of D541 and D611 in *Cb*Ago are indicated with green asterisks. The glutamate finger of both pAgos (E512 in *Tt*Ago or E577 in *Cb*Ago) are inserted into the catalytic site. The scissile phosphate between nucleotide -10 and -11 of the target DNA strand (blue) is indicated with a black asterisk. **E** | Total internal reflection microscopy (TIRM) was used to determine the minimal length for siDNA to be bound by *Cb*Ago. Left panel: Graphical overview of the TIRM method. Right panel: Histogram with TIRM results demonstrated that synthetic siDNAs of at least 12nt in length are efficiently bound by *Cb*Ago. The red line indicates the total number of countable molecules within the microscope image. In **Figure S5**, the raw microscope images are given**. F** | *Cb*Ago mediates target DNA cleavage with siDNAs as short as 14 nucleotides. *Cb*Ago was incubated with siDNA and target DNA in a 1:1:1 ratio. Cleavage products were analyzed by denaturing polyacrylamide electrophoresis.



## Structure of *Cb*Ago in the cleavage-competent conformation

To investigate the molecular architecture of *Cb*Ago in light of its biochemical activity, we crystalized *Cb*AgoDM in complex with both a 21-nt siDNA and a 19-nt DNA target, and solved the structure of the complex at 3.54 Å resolution (**Figure 2**, **Table S1**). Like other Agos, *Cb*Ago adopts a bilobed conformation in which one of the lobes comprises the N-terminal, linker L1, and PAZ domains, which are linked by linker L2 to the other lobe comprising the MID and PIWI domains. Nucleotides 2-16 of the siDNA constitute a 15 base-pairs A-form-like duplex with the target DNA (**Figure 2A**). The 5'-terminal nucleotide of the siDNA is anchored in the MID domain pocket, where the 5'-phosphate group of the siDNA makes numerous interactions with MID domain residues and the C-terminal carboxyl group of *Cb*Ago (**Figure S3**). To test whether the interactions with the 5'-phosphate group of the siDNA are important for *Cb*Ago activity, we performed target DNA cleavage assays in which we used siDNAs with a 5' phosphate or a 5' hydroxyl group (**Figure S4**). As observed for other pAgos (Kaya et al., 2016; Willkomm et al., 2018), *Cb*Ago is able to utilize both siDNAs for target DNA cleavage, but it cleaves target DNA much more efficiently when the siDNA contains a 5' phosphate group. This is in agreement with the siDNA-protein interactions observed in the crystal structure. Furthermore, the backbone phosphates of the siDNA seed segment form hydrogen-bonding and ionic interactions with specific residues in the MID, PIWI and L1 domains (**Figure S3**). At the distal end of the siDNA-target DNA duplex, the N-domain residue His35 caps the duplex by stacking onto the last base pair. After

this point, the remaining 3'-terminal nucleotides of the siDNA are unordered, while the target DNA bends away from the duplex and enters the cleft between the N-terminal and PAZ domains. In agreement with other ternary pAgo complexes (Sheng et al., 2014; Wang et al., 2009; Zander et al., 2014), the PAZ domain pocket, which normally binds the 3' end of the guide in a binary Ago-guide complex, is empty.

*Cb*Ago is phylogenetically closely related to *Tt*Ago (**Figure 1D**). However, *Cb*Ago is 63 amino acids (9.2%) longer than *Tt*Ago (748 amino acids vs. 685 amino acids) and *Cb*Ago and *Tt*Ago share only 23% sequence identity. Superposition of the *Cb*Ago complex structure with the structure of *Tt*Ago bound to a siDNA and DNA target (PDB: 4NCB) (**Figure 2C**) reveals that the macromolecular architecture and conformation of these *Tt*Ago and *Cb*Ago structures are highly similar (Core root mean square deviation of 3.0 Å over 563 residues), with differences found mostly in the loop regions. This agrees with the fact that loops of thermostable proteins are generally more compact and shorter (Russell et al., 1997; Thompson and Eisenberg, 1999). In the *Tt*Ago structure, which is thought to represent a catalytically competent state, a 'glutamate finger' side chain ($Glu512^{TtAgo}$) is inserted into the catalytic site completing the catalytic DDED tetrad (Sheng et al., 2014). Similarly, the corresponding residue in *Cb*Ago (Glu577) is located within a flexible loop and is positioned near the other catalytic residues (**Figure 2D**; Asp541, Asp611, and Asp727). All pAgos and eAgos characterized to date cleave the target strand in between nucleotide 10 and 11 of the target strand. In line with the consensus, the catalytic residues of *Cb*Ago perfectly align with the scissile phosphate linking these nucleotides in our structure (**Figure 2D**). This observation implies that this structure represents the cleavage competent conformation of *Cb*Ago.

Only 15 siDNA-target DNA base pairs are formed in the complex, which suggests that additional siDNA-target DNA binding is not essential for target DNA cleavage. To determine the minimum siDNA length that CbAgo requires for target binding, we performed single-molecule fluorescence assays. First, CbAgo was immobilized on a surface and next it was incubated with 5'-phosphorylated Cy3-labelled siDNAs (**Figure 2E**). These assays demonstrate that *Cb*Ago can bind siDNAs with a minimal length of 12 nucleotides. Next, we determined the minimum siDNA length for *Cb*Ago-siDNA mediated target DNA cleavage (**Figure 2F**). In line with the observation that the *Cb*Ago adopts a cleavage-competent confirmation when only 14 base pairs are formed, *Cb*Ago can cleave target DNAs when

programmed with siDNAs as short as 14 nt (forming 13 siDNA-target DNA base pairs) under the tested conditions. This resembles the activity of *Pf*Ago, *Mj*Ago, and *Mp*Ago, which require siDNAs with a minimal length of 15 nt to catalyze target DNA cleavage (Kaya et al., 2016; Swarts et al., 2015a; Zander et al., 2017). Only *Tt*Ago has been reported to mediate target DNA cleavage with siDNAs as short as 9 nt (Wang et al., 2008a).

**_Cb_Ago associates with plasmid-derived siDNAs _in vivo_**

It has previously been demonstrated that certain pAgos co-purify with their guides and/or targets during heterologous expression in *E. coli* (Olovnikov et al., 2013; Swarts et al., 2014b). To determine whether *Cb*Ago also acquires siDNAs during expression, we isolated and analyzed the nucleic acid fraction that co-purified with *Cb*Ago. Denaturing polyacrylamide gel electrophoresis revealed that *Cb*Ago co-purified with small nucleotides with a length of ~12-19 nucleotides (**Figure 3A**). These nucleic acids were susceptible to DNase I but not to RNase A treatment, indicating that *Cb*Ago acquires 12-19 nucleotide long siDNAs *in vivo*, which fits with its observed binding and cleavage activities *in vitro* (**Figure 1, 2**).



We cloned and sequenced the siDNAs that co-purified with *Cb*Ago to determine their exact length and sequence. The majority of the siDNAs had a length of 16 nucleotides and are complementary to the plasmid used for expression of *Cb*Ago (**Figure 3B, C**). Likewise the siRNAs and siDNAs that co-purify with respectively *Rhodobacter sphaeroides* (*Rs*Ago) and *Tt*Ago are also mostly complementary to their expression plasmids (Olovnikov et al., 2013; Swarts et al., 2014b). As both *Tt*Ago and *Rs*Ago have been demonstrated to interfere with plasmid DNA, this suggests that also *Cb*Ago might play a role in protecting its host against invading DNA. However, no significant reduction of plasmid content could be detected during or upon expression of *Cb*Ago in *E. coli* (**Figure S6**). We also investigated whether *Cb*Ago co-purified with nucleic acids that were enriched for certain motifs. Sequence analysis revealed that most siDNAs co-purified with *Cb*Ago contain a deoxyadenosine at their 5' ends (**Figure 3D**). In addition, we observed an enrichment of thymidine nucleotides in the three positions directly downstream of the siDNA 5' end (nt 2-4) (**Figure 3D**).

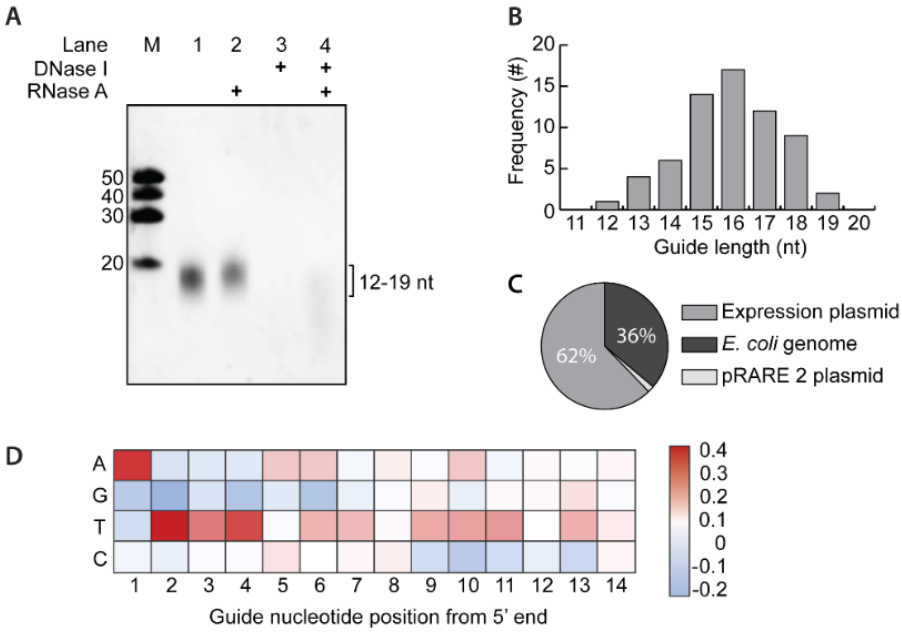



**Figure 3 | *Cb*Ago associates with small plasmid-derived siDNA *in vivo*. A** | Nucleic acids that co-purified with *Cb*Ago were treated with either RNase A, DNase I or both, and were analyzed by denaturing polyacrylamide gel electrophoresis. **B** | Histogram displaying the length of DNA co-purified with *Cb*Ago as determined by sequencing. **C** | Sequenced nucleic acids that co-purified with *Cb*Ago are mostly complementary to the *Cb*Ago expression plasmid. **D** | Heat map showing the base preference of the co-purified nucleic acids at each position. The red squares indicate bases that were more often found compared to a random distribution (25%); blue squares indicate bases that were less frequently found.

## The sequence of the siDNA affects *Cb*Ago activity

To investigate if the 5'-terminal nucleotide of the siDNA affects the activity of *Cb*Ago, we performed cleavage assays. *Cb*Ago was loaded with siDNA guides with varied nucleotides at position 1 (g1N) and incubated with complementary target DNAs (**Figure 4A**). Surprisingly, the highest cleavage rates were observed when *Cb*Ago was loaded with siDNAs containing a 5'-T, followed by siDNAs containing 5'-A. *Cb*Ago bound 5'-G or 5'-C siDNAs displayed slightly lower initial cleavage rates. Also for other pAgos the g1N preference observed *in vivo* is not reflected in the *in vitro* activities; *Tt*Ago (which preferentially co-purifies with g1C siDNAs) as well as *Pf*Ago and *Mp*Ago (of which the *in vivo* g1N preferences are unknown) demonstrate no clear preference for a specific g1N during *in vitro* cleavage reactions (Kaya et al., 2016; Swarts et al., 2014b, 2017a). Instead, the preference of

*Tt*Ago for 5'-C siDNAs is determined by specific recognition of a guanosine nucleotide in the corresponding position (t1) in the target DNA (Swarts et al., 2017a). Indeed, *Tt*Ago structures and models have revealed base-specific interactions with target strand guanine, while base-specific interactions with the 5'-terminal cytidine in the siDNA are less obvious (Swarts et al., 2017a). Similarly, we observe no obvious base-specific interactions with the 5'-terminal cytidine in the structure of the *Cb*Ago complex (**Figure S7**). When we investigated potential base-specific interactions with the base at the opposing target strand t1 position, we observed that the t1 thymine base is not placed in the t1 binding pocket as has been observed in *Tt*Ago, *Rs*Ago and hAGO2 (Liu et al., 2018; Schirle et al., 2015; Swarts et al., 2017a). Instead, the thymine bases is flipped and stacks on Phe557 that also caps the siDNA-target DNA duplex (**Figure S7**). At present, we are unable to rationalize the preferentially co-purification of 5'-adenosine siDNAs with *Cb*Ago.



In order to characterize the seed segment of *Cb*Ago, and to test whether the seed length changes depending on the nature of the guide and the target (*i.e.* DNA vs. RNA), we performed additional single-molecule binding assays. The length of seed was determined based on the minimal number of complementary nucleotide pairs between guide and target that were required to achieve a stable binding event. We first tested the sub-seed (nt 2-4), a 3-nt motif involved in initial target recognition in hAgo2 (Chandradoss et al., 2015; Schirle et al., 2014). When only the sub-seed segment of the siDNA is complementary to the DNA and RNA targets, *Cb*Ago-siDNA complexes bound to the DNA target with an average dwell time 58-fold longer compared to RNA target binding (**Figure 4B**). When nt 2-7 of the guide were complementary to the target, the *Cb*Ago-siDNA complex stably bound to both to target DNA and RNA beyond our observation time of 300 s. This suggests *Cb*Ago prefers DNA targets above RNA targets and that the seed segment of the siDNAs bound by *Cb*Ago comprises nucleotides 2-7.

Next, we set out to investigate whether *Cb*Ago displays a preference for siDNAs with a TTT sub-seed (nt 2-4) *in vitro*, similar to the observed sequence preference for siDNAs that co-purified with *Cb*Ago *in vivo*. *Cb*Ago was incubated with siDNAs in which the sub-seed was varied and complementary target DNAs were added. In contrast to the 5'-base preference, the TTT sub-seed preference that we observed *in vivo* is also reflected *in vitro*: *Cb*Ago displays the highest target cleavage rates when programmed with TTT sub-seed siDNAs (**Figure 4C**).

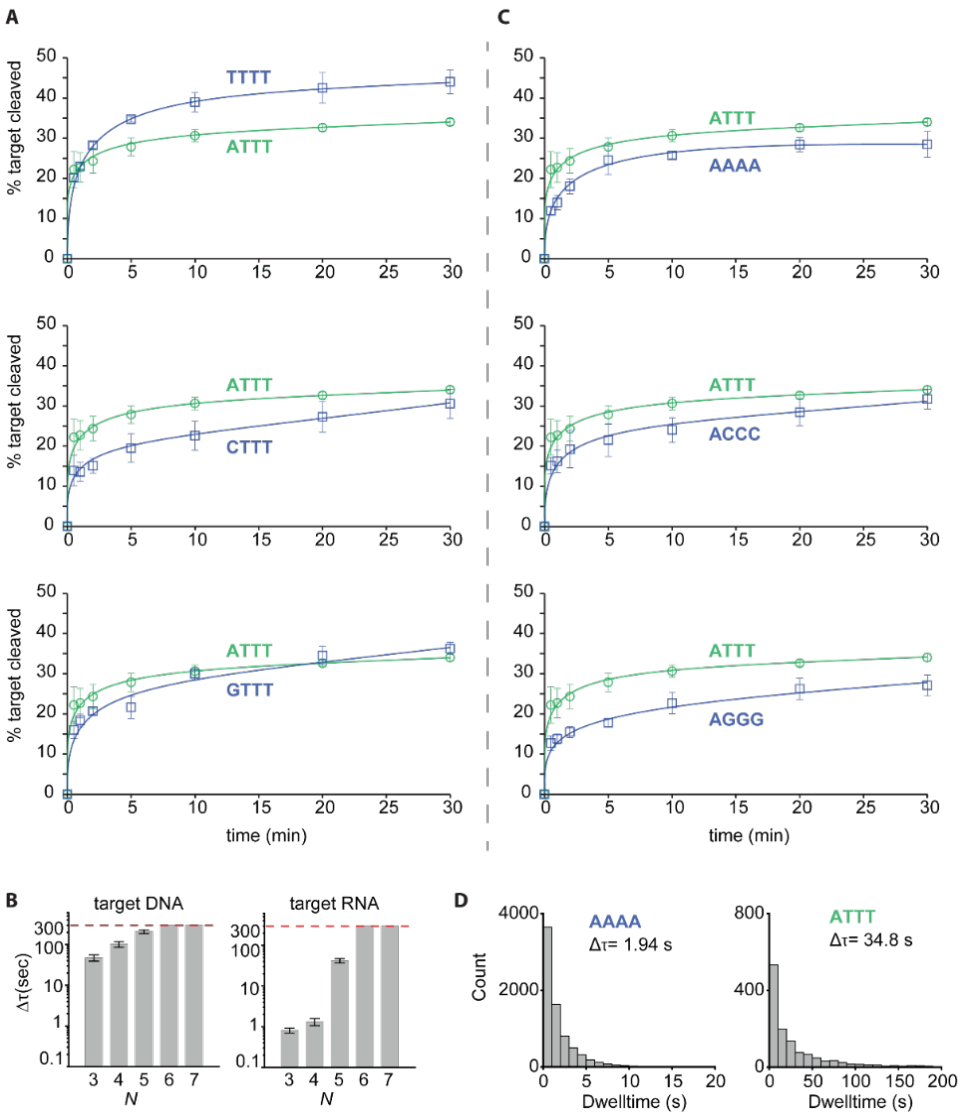



**Figure 4 | The siDNA sequence affects *Cb*Ago activity. A** | *Cb*Ago has no strong 5'-end nucleotide preference. *Cb*Ago was incubated with siDNA with varied 5'-end and incubated with complementary DNA targets. Cleavage products were analyzed by denaturing polyacrylamide electrophoresis and quantified. Graphs display the amount of target DNA cleaved. Error bars indicate the standard variation of three independent experiments. **B** | Histograms displaying dwell time of *Cb*AgoDM-siDNA complexes binding either DNA or RNA targets with a varied sequence complementarity (N = number of complementary nucleotides between the siDNA and the target, starting at nt2, thus *N* 3= nt 2-4). The photobleaching limit is reached where the signal is deactivated (300s). **C** | *Cb*Ago preferentially utilizes siDNAs with a TTT sub-seed segment. *Cb*Ago was incubated with siDNA with varied sub-seed segments

(nt 2-4) and incubated with complementary DNA targets. Cleavage products were analyzed by denaturing polyacrylamide electrophoresis and quantified. Graphs display the amount of target DNA cleaved. Error bars indicate the standard variation of three independent experiments. **D** | Histograms displaying dwell time of *Cb*AgoDM in complex with a 5'-ATTT siDNA or 5'-AAAA siDNA binding to a target DNA. Interactions that are on average ~18-fold longer than *Cb*Ago in complex with siDNAs containing a 5'-AAAA motif.

To confirm these findings, we performed single-molecule assays in which we compared the target binding properties of *Cb*Ago-siDNA complexes containing siDNAs with either a TTT or an AAA sub-seed segment. These assays demonstrate that the dwell time of *Cb*Ago loaded with a TTT sub-seed siDNA on a target was 18-fold longer compared to *Cb*Ago loaded with siDNA containing an AAA sub-seed (**Figure 4D**). Combined, these data indicate that *Cb*Ago displays a preference for siDNAs containing a TTT sub-seed segment.



### A pair of *Cb*Ago-siDNA complexes can cleave double stranded DNA

Thermophilic pAgos have successfully been used to generate double stranded DNA breaks in plasmid DNA (Swarts et al., 2015a, 2014b). As each pAgo-siDNA complex targets and cleaves a single strand of DNA only, two individual pAgo-siDNA complexes are required for dsDNA cleavage, each targeting another strand of the target dsDNA. Although all pAgos characterized so far appear to lack the ability to actively unwind or displace a dsDNA duplex substrate, it has been proposed that, at least *in vitro*, thermophilic pAgos rely on elevated temperatures (>65°C) to facilitate local melting of the dsDNA targets to target each strand of the DNA individually. However, *Cb*Ago is derived from a mesophilic organism and we therefore hypothesize that it is able to mediate protection against invading DNA at moderate temperatures (37°C). To test if *Cb*Ago can indeed cleave dsDNA targets at 37°C, we incubated apo-*Cb*Ago and pre-assembled *Cb*Ago-siDNA complexes with a target plasmid. Previous studies showed that the 'chopping' activity of siDNA-free apo-*Tt*Ago and apo-*Mj*Ago can result in plasmid linearization or degradation, respectively (Swarts et al., 2017a; Zander et al., 2017). We observed that apo-*Cb*Ago converted the plasmid substrate from a supercoiled to open-circular state, possibly by nicking one of the strands, but did not observe significant linearization or degradation of the plasmid DNA (**Figure 5A**). When the plasmid was targeted by *Cb*Ago loaded with a single siDNA, we also observed loss of supercoiling (**Figure 5A**). As this activity was not observed with nuclease-deficient *Cb*AgoDM, we conclude that apo-*Cb*Ago and *Cb*Ago-siDNA complexes generate nicks in dsDNA plasmid targets with their DEDD catalytic site. When using two *Cb*Ago-siDNA complexes, each targeting one strand of the plasmid, we observed that a fraction of the target plasmid DNA

becomes linearized (**Figure 5A**). This implies that *Cb*Ago-siDNA complex-mediated nicking of each of the target plasmid DNA strands resulted in the generation of a double stranded DNA break. Next, we investigated if the spacing between the two siDNAs affects the ability of *Cb*Ago to cleave the plasmid. The most efficient plasmid linearization was achieved when the siDNAs were orientated exactly or almost opposite to each other (**Figure 5A**).

Finally, we investigated whether the GC-content of the target DNA plays a role during DNA targeting by *Cb*Ago. For *Tt*Ago, it has been observed that AT-rich DNA is cleaved more efficiently that GC-rich DNA (Swarts et al., 2017a). To test if such preference also exists for *Cb*Ago, we designed a target plasmid containing 16 gene fragments of 100 base pairs complementary to sequences from the human genome, with an increasing GC content (**Figure 5B**). *Cb*Ago-siDNA complexes were only able to generate dsDNA in gene fragments with a GC-content of 31% or lower (**Figure 5C**). This indicates that, at least *in vitro*, the GC-content is an important factor that determines target DNA cleavage by *Cb*Ago.

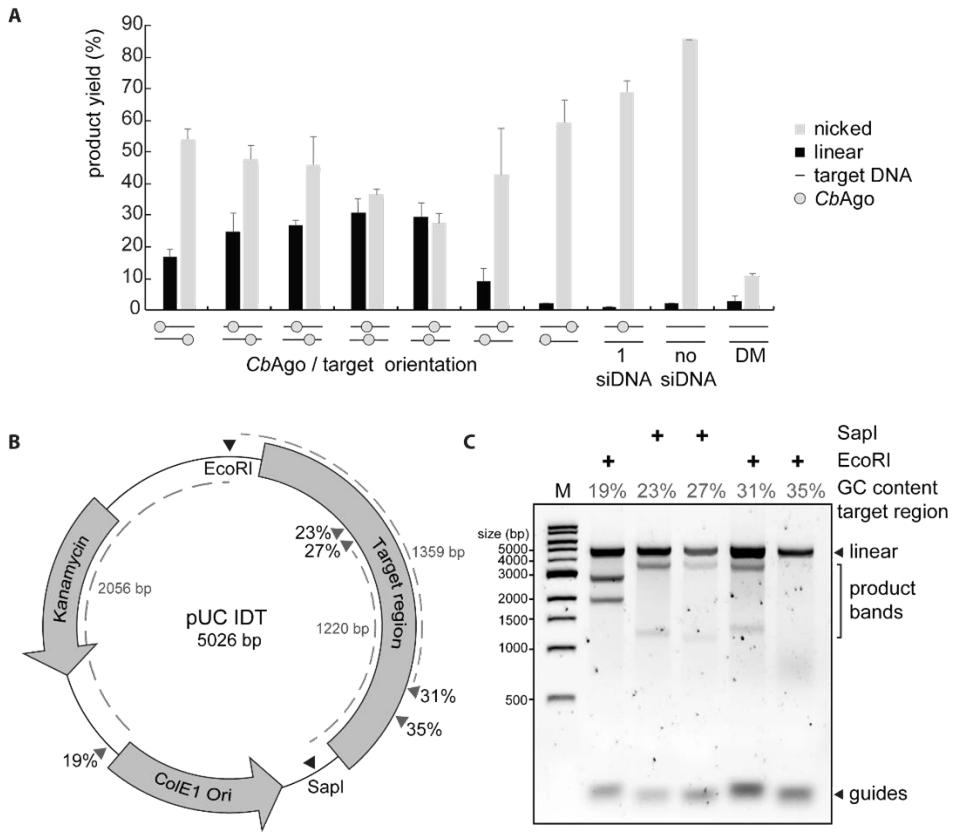



**Figure 5 | Double stranded plasmid DNA cleavage by *Cb*Ago. A |** Two *Cb*Ago-siDNA complexes can generate double stranded DNA breaks in plasmid DNA. *Cb*Ago-siDNA complexes were pre-assembled and incubated with target plasmid DNA. Cleavage products were analyzed by agarose gel electrophoresis and quantified. The spacing between both *Cb*Ago-siDNA target sites affects the linearization efficiency (nucleotide spacing between the predicted cleavage sites: +15 nt, +10 nt, +5 nt, 0 nt, -5 nt, -10 nt, -15 nt, a single siDNA, no siDNA). With 0 nt spacing, both *Cb*Ago-siDNA complexes are exactly on top of each other. **B** | Schematic overview of the pUC IDT target plasmid. Arrows indicate target sites while percentages indicate the GC-content of the 100 bp segments in which these target sites are located. **C** | Pre-assembled *Cb*Ago-siDNA complexes targeting various pUC IDT segments were incubated with pUC IDT. Cleavage products were incubated with EcoRI or SapI and were further analyzed by agarose gel electrophoresis. The GC-content of the segments in which the target sites were located are indicated by the percentage.

# Discussion

Several prokaryotic Argonaute proteins have been demonstrated to protect their host against invading nucleic acids, such as plasmid DNA (Olovnikov et al., 2013; Swarts et al., 2015a, 2014b). Similar to *Tt*Ago and *Rs*Ago, *Cb*Ago co-purifies with guides which are preferentially acquired from the plasmid used for its heterologous expression in *E. coli*. In addition, *Cb*Ago mediates programmable DNA-guided DNA cleavage *in vitro*. This suggests that, similar to the phylogenetically related *Tt*Ago, also *Cb*Ago can interfere with plasmid DNA via DNA-guided DNA interference.

Sequencing of the nucleic acids that co-purified with *Cb*Ago revealed that *Cb*Ago preferentially associates with siDNAs with a 5'-ATTT-3' sequence at their 5' end. It was previously shown that the guide RNA utilized by eAgos can be divided into functional segments. These segments are (from 5' to 3') the anchor nucleotide (nt 1), the seed (nt 2-8) and sub-seed segments (nt 2-4), and the central (nt 9-12), 3' supplementary (nt 13-16) and tail (nt 17-21) segments (Schirle et al., 2014; Wee et al., 2012). Extending this knowledge to the siDNAs that co-purified with *Cb*Ago, *Cb*Ago preferentially associates with siDNAs that have a 5'-terminal adenosine anchor (nt 1) and a T-rich sub-seed. In RNAi pathways, the preference for a specific 5'-terminal nucleotide is important for guide RNA loading into a subset of eAgos (Aravin et al., 2006; Frank et al., 2010, 2012) Similarly, several pAgos including RsAgo, *Tt*Ago, and now *Cb*Ago also preferentially associate with specific 5'-terminal nucleotides *in vivo* (Olovnikov et al., 2013; Swarts et al., 2014b). However, for both *Cb*Ago and *Tt*Ago, there is no clear preference for siDNAs with that specific 5'-base during cleavage assays *in vitro*. Rather than having a functional importance, the preference of pAgos for a specific nucleotide at the siDNA 5' end might be a consequence of siDNA generation and/or loading, as has been demonstrated for *Tt*Ago (Swarts et al., 2017a). Several studies on human Ago2 have described the importance of the sub-seed segment (nt 2-4) in its RNA guides (Chandradoss et al., 2015; Salomon et al., 2015; Schirle et al., 2014). For hAgo2, a complete match between the guide RNA sub-seed segment and the target RNA triggers a conformational change that first exposes the remainder of the seed (nt 5-8), and eventually the rest of the guide. This facilitates progressive base paring between the guide RNA and the target (Klein et al., 2017). However, a specific nucleotide preference in the sub-seed segment, as we have observed for *Cb*Ago, has not been described for any other Argonaute protein. The preference for the T-rich sub-seed is not only observed in the *in vivo* acquired siDNAs, but

also plays a clear role during target binding and cleavage assays *in vitro*. This may reflect a structural preference for these thymidines in the cleft of the PIWI domain. We have not been able to obtain diffracting crystals of *Cb*Ago in complex with siDNAs that have a 5'-ATTT-3' sequence at the 5'-end. Future research will thus be necessary to determine the structural basis the apparent preference for these nucleotides at these positions. We hypothesize that this bias might reflect the mesophilic nature of *Cb*Ago, which might have better access to AT-rich dsDNA fragments, both for siDNA acquisition and for target cleavage.

Several DNA-targeting pAgos have been repurposed for a range of molecular applications among which a cloning, recombineering and nucleic acid-detection method (Enghiad and Zhao, 2017; Lapinaite et al., 2018; Song et al., 2019; Xie et al., 2019). Additionally, the potential repurposing of pAgos for genome editing applications has previously been discussed (Hegge et al., 2018a). However, all characterized DNA-cleaving pAgos to date originate from thermophilic prokaryotes and are solely active at elevated temperatures, which limits the potential repurposing of pAgos for applications that require moderate temperatures, such as genome editing. The biochemical characterization of *Cb*Ago reported herein is the first example of a pAgo that catalyzes siDNA-guided dsDNA cleavage at 37°C, indicating that the pool of mesophilic pAgos contains candidates that – in theory – can be utilized for potential applications that require moderate temperatures such as genome editing. If *Cb*Ago or other mesophilic pAgos could be harnessed for genome editing, they will have certain advantages over the currently well-established genome editing tools CRISPR-Cas9 and CRISPR-Cas12a. While CRISPR-based genome editing tools can be programmed with a guide RNA to target DNA sequences of choice, target DNA cleavage additionally requires the presence of a protospacer adjacent motif (PAM) next to the targeted sequence (5'-NGG-3' for Cas9 and 5'-TTTV-3' for Cas12a) (Swarts and Jinek, 2018). This limits the possible target sites of Cas9 and Cas12a. In contrast, pAgos do not require a PAM for DNA targeting, which would make them much more versatile tools compared to CRISPR-associated nucleases. However, PAM binding by Cas9 and Cas12a also promotes unwinding of dsDNA targets (Anders et al., 2014; Swarts et al., 2017b; Yamano et al., 2017) which subsequently facilitates strand displacement by the RNA guide, and eventually R-Loop formation. The absence of such mechanism in pAgos might explain their limited nuclease activity on dsDNA targets.

Here, we have demonstrated that *Cb*Ago does not strictly rely on other proteins when targeting AT-rich dsDNA sequences in vitro. As such, this study provides a foundation for future efforts to improve double stranded DNA target accessibility of pAgos and to facilitate the further development of pAgo-based applications at moderate temperatures.

# Methods

## Plasmid construction

The *Cb*Ago gene was codon harmonized for *E. coli* Bl21 (DE3) and inserted into a pET-His6 MBP TEV cloning vector (obtained from the UC Berkeley MacroLab, Addgene #29656) using ligation-independent cloning (LIC) using oligonucleotides oDS067 and oDS068 (**Table S4**) to generate a protein expression construct that encodes the *Cb*Ago polypeptide sequence fused to an N-terminal tag comprising a hexahistidine sequence, a maltose binding protein (MBP) and a Tobacco Etch Virus (TEV) protease cleavage site.

## Generation of the Double mutant

*Cb*Ago double mutant (D541A, D611A) was generated using an adapted Quick Directed Mutagenesis Kit instruction manual (Stratagene). The primers were designed using the web-based program primerX (http://bioinformatics.org/primerx).



## *Cb*Ago expression and purification

The *Cb*Ago WT and DM proteins were expressed in *E. coli* Bl21 (DE3) Rosetta™ 2 (Novagen). Cultures were grown at 37°C in LB medium containing 50 µg ml-1 kanamycin and 34 µg ml-1 chloramphenicol until an OD600nm of 0.7 was reached. *Cb*Ago expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. During the expression cells were incubated at 18°C for 16 hours with continues shaking. Cells were harvested by centrifugation and lysed by sonication (Bandelin, Sonopuls. 30% power, 1s on/2s off for 5min) in lysis buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM imidazole, supplemented with a EDTA free protease inhibitor cocktail tablet (Roche). The soluble fraction of the lysate was loaded on a nickel column (HisTrap Hp, GE healthcare). The column was extensively washed with wash buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 30 mM imidazole. Bound protein was eluted by increasing the concentration of imidazole in the wash buffer to 250 mM. The eluted protein was dialyzed at 4°C overnight against 20 mM HEPES pH 7.5, 250 mM KCl, and 1mM dithiothreitol (DTT) in the presence of 1mg TEV protease (expressed and purified according to Tropea et al. 2009 (Tropea et al., 2009)) to cleave of the His6-MBP tag. Next the cleaved protein was diluted in 20mM HEPES pH 7.5 to lower the final salt concentration to 125 mM KCl. The diluted protein was applied to a heparin column (HiTrap Heparin HP,

GE Healthcare), washed with 20 mM HEPES pH 7.5, 125 mM KCl and eluted with a linear gradient of 0.125-2 M KCl. Next, the eluted protein was loaded onto a size exclusion column (Superdex 200 16/600 column, GE Healthcare) and eluted with 20 mM HEPES pH 7.5, 500mM KCl and 1 mM DTT. Purified *Cb*Ago protein was diluted in size exclusion buffer to a final concentration of 5 µM. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

## Co-purification nucleic acids

To 500 pmole of purified *Cb*Ago in SEC buffer CaCl2 and proteinase K (Ambion) were added to final concentrations of 5 mM CaCl2 and 250 µg/mL proteinase K. The sample was incubated for 4 hours at 65°C. The nucleic acids were separated from the organic fraction by adding Roti phenol/chloroform/isoamyl alcohol pH 7.5-8.0 in a 1:1 ratio. The top layer was isolated and nucleic acids were precipitated using ethanol precipitation by adding 99% ethanol in a 1:2 ratio supplied with 0.5% Linear polymerized acrylamide as a carrier. This mixture was incubated overnight at -20°C and centrifuged in a table centrifuge at 16,000 g for 30 min. Next, the nucleic acids pellet was washed with 70% ethanol and solved in 50 µL MilliQ water. The purified nucleic acids were treated with either 100 µg/mL RNase A (Thermo), 2 units DNase I (NEB) or both for 1 hour at 37°C and resolved on a denaturing urea polyacrylamide gel (15%) and stained with SYBR gold.

## Single stranded Activity assays

Unless stated otherwise 5 pmoles of each *Cb*Ago, siDNA and target were mixed in a ratio of 1:1:1, in 2x reaction buffer containing 20 mM Tris-HCl (pH 7.5) supplemented with 500 µM $MnCl^{2+}$. The target was added after the *Cb*Ago and siDNA had been incubation for 15 min at 37°C. Then the complete reaction mixture was incubated for 1 hour at 37°C. The reaction was terminated by adding 2x RNA loading dye (95% formamide, 0.025% bromophenol blue, 5 mM ETDA) and heating it for 5 minutes at 95°C. After this the samples were resolved on a 20% denaturing (7 M Urea) polyacrylamide gel. The gel was stained with SYBR gold nucleic acid stain (Invitrogen) and imaged using a G:BOX Chemi imager (Syngene).

## Double stranded Activity assay

In two half reactions 12.5 pmoles of *Cb*Ago was loaded with either 12.5 pmoles of forward or reverse siDNA in reaction buffer containing 10 mM Tris-HCl, 10 µg/ml BSA and 250 µM $MnCl_2$. The half reactions were incubated for 15 min at 37°C. Next, both half reactions were mixed together and 120 ng target plasmid was added after which the mixture was incubated for 1 hour of 37°C. After the incubation the target plasmid was purified from the mixture using a DNA clean and concentrate kit (DNA Clean & Concentrator™-5, Zymogen) via the supplied protocol. The purified plasmid was subsequently cut using either EcoRI-HF (NEB) or SapI-HF (NEB) in Cutsmart buffer (NEB) for 30 min at 37°C. A 6x DNA loading dye (NEB) was added to the plasmid sample prior to resolving it on a 0.7% agarose gel stained with SYBR gold (Invitrogen).



## Crystallization

To reconstitute the *Cb*Ago DM-siDNA-target DNA complex, siDNA and target DNA were pre-mixed at a 1:1 ratio, heated to 95°C, and slowly cooled to room temperature. The formed dsDNA duplex (0.5 M) was mixed with *Cb*Ago DM in SEC buffer at a 1:1:4 ratio (*Cb*Ago DM:duplex DNA), and $MgCl_2$ was added to a final concentration of 5 mM. The sample was incubated for 15 minutes at 20°C to allow complex formation. The complex was crystallized at 20°C using the hanging drop vapor diffusion method by mixing equal volumes of complex and reservoir solution. Initial crystals were obtained at a *Cb*Ago DM concentration of 5 mg/ml with a reservoir solution consisting of 4 M sodium formate. Data was collected from crystals grown obtained using a complex concentration of 4.3 mg/ml and reservoir solution containing 3.8 M sodium formate and 5 mM $NiCl_2$ at 20°C. For cryoprotection, crystals were transferred to a drop of reservoir solution and flash-cooled in liquid nitrogen.

X-ray diffraction data were measured at beamline X06DA (PXIII) of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). Data were indexed, integrated, and scaled using AutoPROC (Vonrhein et al (2011)). Crystals of the *Cb*Ago-siDNA-target DNA complex diffracted to a resolution of 3.55 Å and belonged to space group $P6_3\,2\,2$, with one copy of the complex in the asymmetric unit. The structure was solved by molecular replacement using Phaser-MR (McCoy et al., 2007). As search model, the structure of *Tt*Ago in complex with guide and target DNA strands (PDB: 5GQ9) was used after removing loops and truncating amino acid side chains. Phases obtained using the initial molecular

replacement solution were improved by density modification using phenix.resolve (Terwilliger, 2004) and phenix.morph_model (Terwilliger et al., 2013). The atomic model was built manually in Coot (Emsley et al., 2010) and refined using phenix.refine (Afonine et al., 2012). The final binary complex model contains *Cb*Ago residues 1-463 and 466-748, guide DNA residues 1–16, and target DNA residues (-18)–(-1).

## Structure analysis

Core Root Means Square Deviations (rmsd) of structure alignments were calculated using Coot SSM superpose (Krissinel et al 2004). Intramolecular interactions were analyzed using PDBePISA (Krissinel and Henrick, 2007). Figures were generated using PyMOL (Schrödinger).

## Single-Molecule Experimental Set-Up

Single-molecule fluorescence FRET measurements were performed with a prism-type total internal reflection fluorescence microscope. Cy3 and Cy5 molecules were excited with 532 nm and 637 nm wavelength, respectively. Resulting Cy3 and Cy5 fluorescence signal was collected through a 60X water immersion objective (UplanSApo, Olympus) with an inverted microscope (IX73, Olympus) and split by a dichroic mirror (635dcxr, Chroma). Scattered laser light was blocked out by a triple notch filter (NF01-488/532/635, Semrock). The Cy3 and Cy5 signals were recorded using an EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology) with exposure time 0.1 s. All single-molecule experiments were done at room temperature (22 ± 2C).

## Fluorescent DNA and RNA preparation

The RNAs with amine-modification (amino-modifier C6-U phosphoramidite, 10-3039, Glen Research) were purchased from STPharm (South Korea) and DNAs with amine-modification (internal amino modifier iAmMC6T) Ella biotech (Germany). The guide and target strands were labeled with donor (Cy3) and acceptor (Cy5), respectively, using the NHS-ester form of Cy dyes (GE Healthcare). 2012).1 µL of 1 mM of DNA/RNA dissolved in MilliQ H20 is added to 5 µL labeling buffer of (freshly prepared) sodiumtetraborate (380 mg/10mL, pH 8.5). 1 µL of 20 mM dye (1 mg in 56 µL DMSO) is added and incubated overnight at room

temperature in the dark, followed by washing and ethanol precipitation. The labeling efficiency was ~100%.

### Single-molecule sample preparation

A microfluidic chamber was incubated with 20 µL Streptavidin (0.1 mg/mL, Sigma) for 30 sec. Unbound Streptavidin was washed with 100 µL of buffer T50 (10 mM Tris-HCl [pH8.0], 50 mM NaCl buffer). The fifty microliters of 50 pM acceptor-labelled target construct were introduced into the chamber and incubated for 1 min. Unbound labeled constructs were washed with 100 µL of buffer T50. The *Cb*Ago binary complex was formed by incubating 10 nM purified *Cb*Ago with 1 nM of donor-labeled guide in a buffer containing 50 mM Tris-HCl [pH 8.0] (Ambion), 1 mM MnCl2, and 100 mM NaCl (Ambion) at 37°C for 20 min. For binding rate ($k_{on}$) measurements, the binary complex was introduced into the fluidics chamber using syringe during the measurement. The experiments were performed at the room temperature (23 ± 1°C).



For fluorescence Guide Loading Experiments before immobilizing *Cb*Ago on the single-molecule surface, 1 µL of 5 µM His-tagged apo-*Cb*Ago was incubated with 1 µL of 1 µg/ml biotinylated anti-6x His antibody (Abcam) for 10 min. Afterward, the mixture was diluted 500x in T50 and 50 µL were loaded in the microfluidic channel for 30 s incubation, followed by washing with 100 µL of T50 buffer. Cy3-labeled ssDNA (0.1 ) was applied to the microfluidic chamber in imaging buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM MnCl2, 1 mM Trolox ((±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid), supplemented with an oxygen-scavenging system (0.5 mg/mL glucose oxidase (Sigma), 85 mg/mL catalase (Merck), and 0.8% (v/v) glucose (Sigma)).

### Single-molecule data acquisition and analysis

CCD images of time resolution 0.1 or 0.3 sec were recorded, and time traces were extracted from the CCD image series using IDL (ITT Visual Information Solution). Co-localization between Cy3 and Cy5 signals was carried out with a custom-made mapping algorithm written in IDL. The extracted time traces were processed using Matlab (MathWorks) and Origin (Origin Lab).

The binding rate ($k_{on}$) was determined by first measuring the time between when *Cb*Ago binary complex was introduced to a microfluidic chamber and when the first *Cb*Ago- guide docked to a target; and then fitting the time distribution with a single-exponential growth curve, $A(1 - e^{-k_{on}t})$. The dissociation rate was estimated by measuring the dwell time of a binding event. A dwell time distribution was fitted by single-exponential decay curve ($Ae^{-t/\Delta\tau}$).

## Fluorescence competition experiments

MBP-tagged *Cb*Ago was immobilized on the quartz surface using an anti-MBP antibody. An equimolar mixture of let7 DNA guide (Cy3 labeled) and let7 RNA guide (Cy5 labeled) in imaging buffer was introduced to the microfluidic chamber. After 5 minutes, 10 snapshots of independent fields of view with simultaneous illumination were collected to estimate the amount of guide molecules bound to protein. Movies were taken for 200 s (2000 frames) at continuous illumination of Cy3 and Cy5 molecules to determine the dwell times of the binding events. Dwell times were binned in a histogram and fitted with a single exponential decay curve.

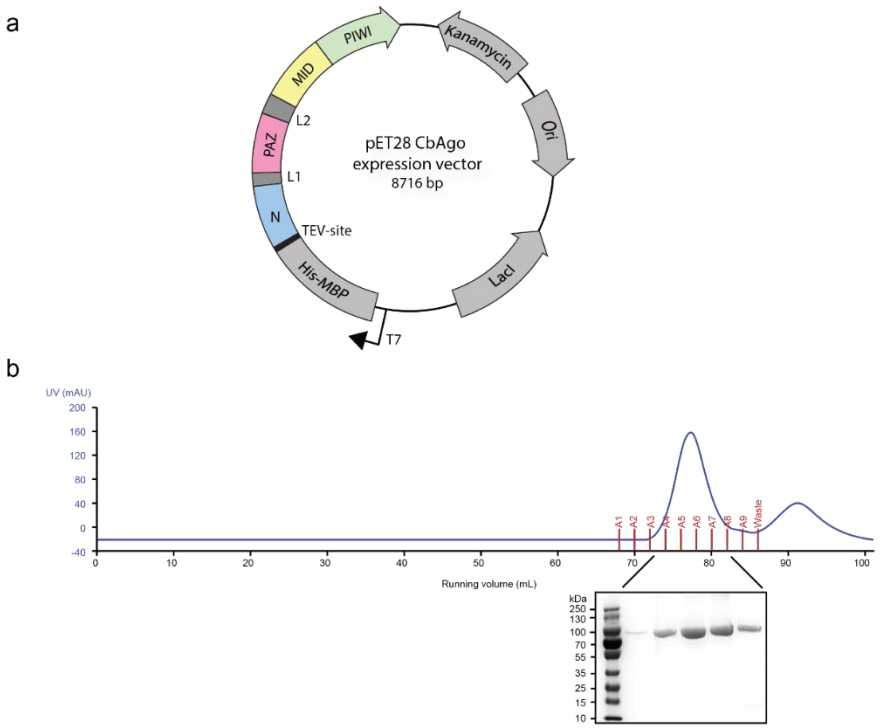
## FRET targeting experiments of ATTT and AAAA guide target combinations

100 pM of target construct annealed with biotin handle were flushed in the microfluidic chamber. After incubation of 1 min, the microfluidic chamber was rinsed with 100 µL T50 buffer. 10 nM of apo-*Cb*Ago was loaded with 1 nM of ATTT seed DNA guide or with AAAA seed DNA guide at 37°C for 30 minutes in imaging buffer after which the mixture is introduced inside the microfluidic chamber. Movies of 200 s were taken at continuous illumination of the Cy3 signal. Site specific protein target interactions were identified as FRET signals and were further analyzed.

## Acknowledgements

We are grateful to Meitian Wang, Vincent Olieric, and Takashi Tomizaki at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) for assistance with X-ray diffraction measurements. This work was supported by grants from the Netherlands Organization of Scientific Research (NWO; ECHO grant 711013002 and NWO-TOP grant 714.015.001) to J.v.d.O. A Swiss National Science Foundation (SNSF) Project Grant to M.J. (SNSF 31003A_149393) and by long-term postdoctoral fellowships from the European Molecular Biology Organization (EMBO) to D.C.S (ALTF 179-2015 and aALTF 509-2017). M.J. is International Research Scholar of the Howard Hughes Medical Institute and Vallee Scholar of the Bert L & N Kuggie Vallee Foundation. C.J. was supported by Vidi (864.14.002) of the Netherlands Organization for Scientific research.

# Supplementary figures




c

| | | D | E | D | X |
|---|---|---|---|---|---|
| RsAgo | - *Rhodobacter Sphaeroides* | VVGMGLAEL | ECEYEGYSD | VFHAHRPLK | IFYSERIAE |
| hAgo 2 | - *Homo sapiens* | FLGADVTHP | QHRQEIIQD | IFYRDGVSE | AYYAHLVAF |
| PfAgo | - *Pyrococcus furiosus* | IIGIDVAPM | EQRGESVDM | LLLRDGRIT | VHYAHKFAN |
| MpAgo | - *Marinitoga piezophila* | YIGIDLSHD | LELNEKMNL | FILRDGRFI | LHIANKVAL |
| MjAgo | - *Methanocaldococcus jannaschii* | IMGLDTGLG | OAOGERLHL | LFLRDGFIQ | IHYADKFVK |
| TtAgo | - *Thermus thermophilus* | AVGFDAGGR | AQAGERIPQ | LLLRDGRVP | LHLADRLVK |
| NgAgo | - *Natronobacterium gregoryi* | FIGIDVSRS | PQLGEKLQS | VIHRDGFMN | TAYADQAST |
| CbAgo | - *Clostridium butyricum* | FIGLDVGTR | PQSGEKIAE | VIHRDGFSR | TGYADKICK |
| CbAgo DM | - *Clostridium butyricum* double mutant | FIGLAVGTR | PQSGEKIAE | VIHRAGFSR | TGYADKICK |

**Figure S1 | *Cb*Ago harbors a catalytic DEDD tetrad. a |** Schematic representation of the pET28 based expression vector of *Cb*Ago. *Cb*Ago consists of the typical (N-PAZ-MID-PIWI) domains. A his (6x)-MBP tag was N-terminally fused to *Cb*Ago linked with a cleavable TEV site. *Cb*Ago was under the control of a T7 promotor. **b** | Size exclusion diagram showing the elution peak of *Cb*Ago. The purity of the purified *Cb*Ago was determined using SDS-PAGE. **c** | Multiple sequence alignment of a part of the PIWI domain from *Cb*Ago with several other biochemically or structurally characterized Ago proteins.

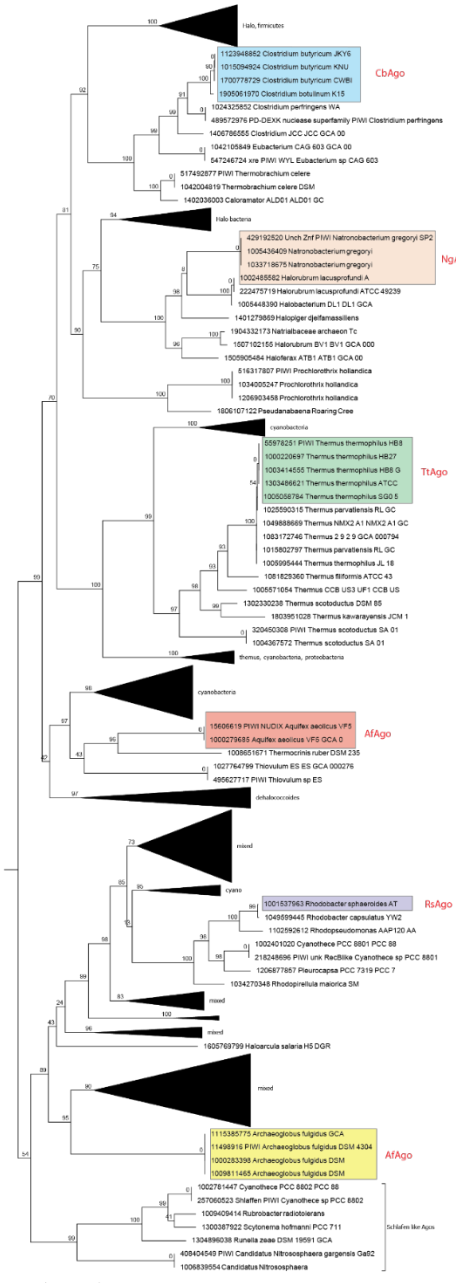



**Figure S2 | Phylogenetic tree of prokaryotic Argonautes.** Maxium-likehood phylogenetic trees were constructed by Kira Makarova (personal communication). The colored boxes indicate the Argonaute proteins that have been characterized biochemically.

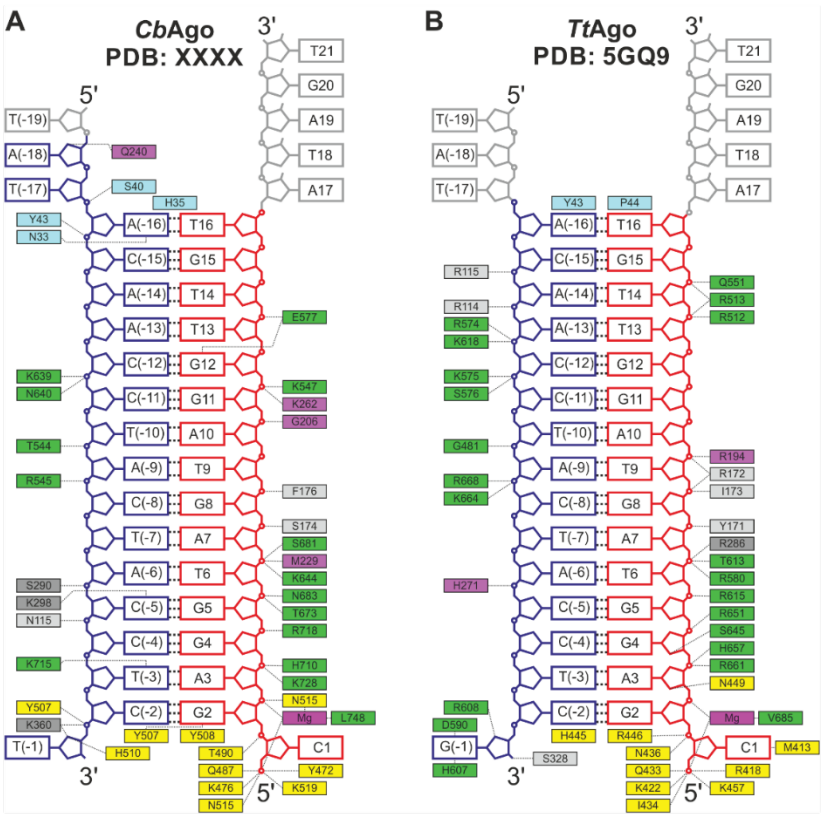



**Figure S3 | Comparison of siDNA-target DNA interactions with *Cb*Ago and *Tt*Ago.** Schematic representation of hydrogen bonding interactions between *Cb*Ago or *Tt*Ago and their guide-target nucleic acids using the ternary protein structures of both pAgos. The residues of *Tt*Ago and *Cb*Ago are colored according to their domains (see **Figure 2A**). Nucleotides colored grey are not ordered in the structure. Base pairs are indicated with thick dashed lines, while other hydrogen bonds are indicated with thin dashed lines.

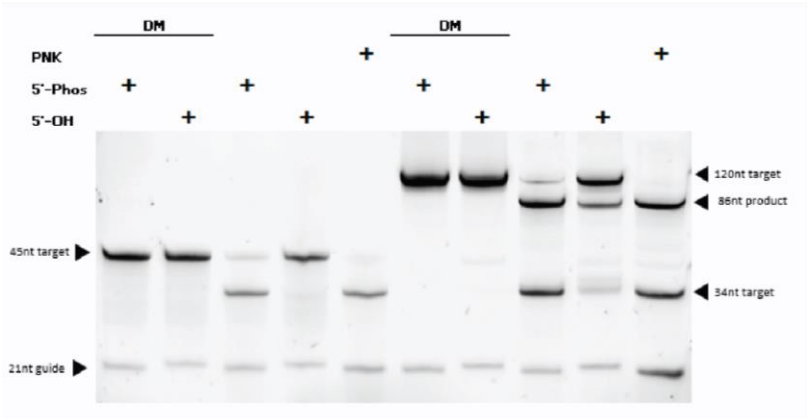



**Figure S4 | *Cb*Ago prefers 5' phosphorylated guides over 5' hydroxylated guides.** *Cb*Ago is unable to cleave a short 45nt target DNA using a 5'OH siDNA. The 45nt target DNA is efficiently cleaved into two products (34nt and 10nt; 10nt not visible on gel) using a 5'phosphorylated siDNA (either phosphorylated using PNK or ordered at IDT). *Cb*Ago cleaves a longer 120nt target DNA into two product (86 and 34nt) using a 5'OH siDNA, but less efficient than with the 5'phosphorylated siDNA.



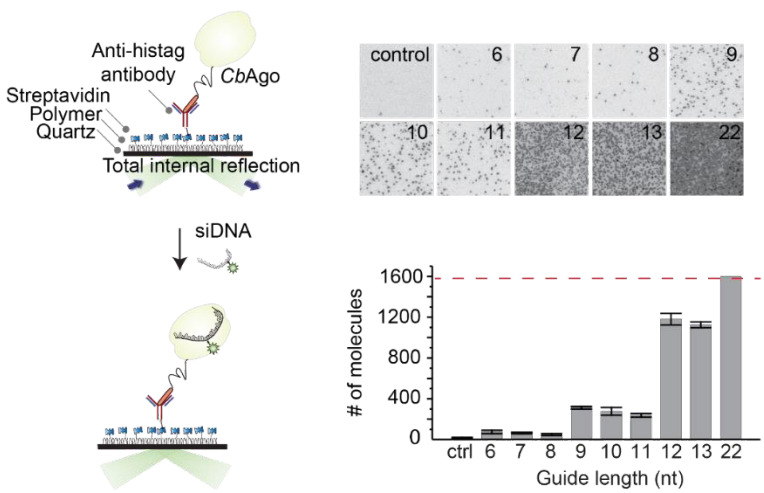



**Figure S5 | Total internal reflection microscopy (TIRM) was used to determine the minimal length for siDNA to be bound by *Cb*Ago. Left panel |** Graphical overview of the TIRM method. **Right panel |** Microscope snapshots of the polymer-coated quartz surface after siDNAs of different lengths were loaded. Each black dot represents a bound siDNA. The total number of bound siDNA (black dots) per microscope image are displayed in the histogram.

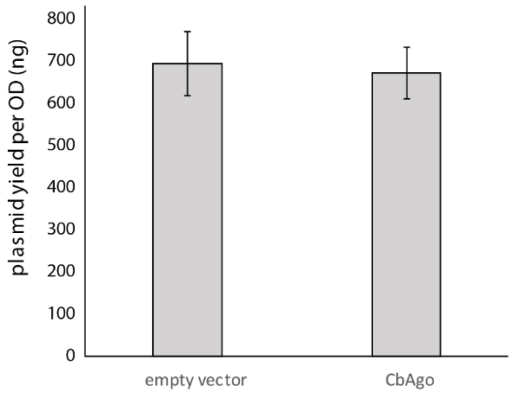



**Figure S6 | Expression of *Cb*Ago does not affect the plasmid content in *E. coli*.** In biological triplicates of both strains protein expression was induced by the addition of ITPG. Four hours after expression the plasmids were isolated and quantified. No significant differences in plasmid content could be observed.

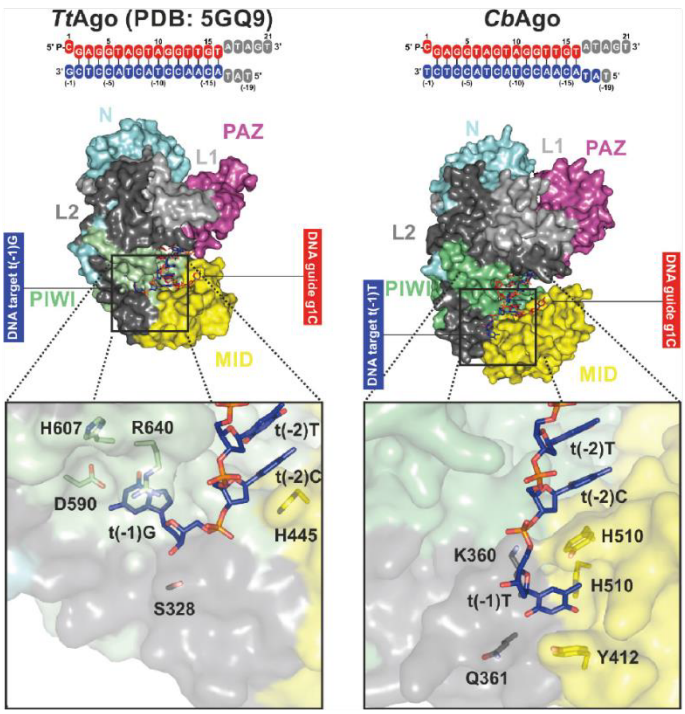



**Figure S7 | Simulations that compare the interaction between the t1N binding site of *Cb*Ago and *Tt*Ago with the t1N residues of the target strand based on molecular dynamics simulation models.** In *Cb*Ago the T1 thymine base is not placed in the t1 binding pocket as has been observed in *Tt*Ago (which binding pocket is formed by D590, R640). Instead, the t1 thymine base of *Cb*Ago is flipped.

# Supplementary tables

**Table S1 | Crystallographic Data Collection and Refinement Statistics**

| **Dataset** | |
|---|---|
| *Cb*Ago | D541A, D611A |
| siDNA | oDS313 (Let7a-based siDNA) |
| target DNA | oDS314 (Let7a-based target DNA) |
| X-ray source | SLS PXIII |
| Space group | $P6_3$ 2 2 |
| Cell dimensions | |
| a, b, c (Å) | 181.45, 181.45, 142.99 |
| α, β, γ, (°) | 90, 90, 120 |
| Wavelength (Å) | 1.00 |
| Resolution (Å)* | 68.86-3.548 (3.675-3.548) |
| $R_{merge}$* (%) | 28 (187.6) |
| CC1/2 (%)* | 99.9 (86.1) |
| I/σ*I** | 17.05 (2.47) |
| Completeness (%)* | 99.89 (99.94) |
| Redundancy* | 39 (38) |
| **Refinement** | |
| Resolution (Å) | 68.86-3.548 |
| No. reflections | 17344 |
| $R_{work}$ / $R_{free}$ | 0.247/0.276 |
| **No. Atoms** | |
| Macromolecules | 6,730 |
| Ligands | 16 |
| Solvent | 3 |
| **B Factors** | |
| Mean (Å$^2$) | 110.40 |
| Macromolecules (Å$^2$) | 110.41 |
| Ligands (Å$^2$) | 111.11 |
| Water (Å$^2$) | 83.70 |
| **RMSDs** | |
| Bond lengths (Å) | 0.0044 |
| Bond angles (°) | 0.720 |
| **Ramachandran plot** | |
| Favored (%) | 95.69% |
| Allowed (%) | 4.31% |
| Outliers (%) | 0% |
| **Molprobity** | |
| Clashscore | 16.8 |

**Table S2 | Target sequences used during the *in vitro* cleavage assays.**

| Oligo | Sequence (5'- 3') | description |
|---|---|---|
| **BG13030** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACAAAT<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | ATTT DNA target |
| **BG13032** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACAAAA<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | TTTT DNA target |
| **BG13034** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACAAAG<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | CTTT DNA target |
| **BG13036** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACAAAC<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | GTTT DNA target |
| **BG13038** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACTTTT<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | AAAA DNA target |
| **BG13040** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACGGGT<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | ACCC DNA target |
| **BG13042** | TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACCCCT<br>TATAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT | AGGG DNA target |
| **BG7023** | AAACGACGGCCAGTGCCAAGCTTACTATACAACCTACTACCTCAT | Let-7 based DNA target |
| **BG8385** | AAACGACGGCCAGUGCCAAGCUUACUAUACAACCUACUACCUCAU | Let-7 based RNA target |
| **oDS314** | TATACAACCTACTACCTCT | Let-7 based target crystallization |
| **SM021** | TTTTTTTTTTTTTTTTTTTTTTAACTATACAACCTACTACCTCTTTTTT<br>TTTTTTTTTTTTTTTTTTTTTTTTTGGCGACGGCAGCGAGGC | Single perfect target |
| **SM022** | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTACCTCTTTTTT<br>TTTTTTTTTTTTTTTTTTTTTTTGGCGACGGCAGCGAGGC | Single target 8nt |
| **SM023** | [BiotinTEG]GCCTCGCTGCCGTCGCCA | Biotin handle |
| **SM024** | UUUUUUUUUUUUUUUUUUUUUUUU*UUUUUUUACUACCUC | let-7 single target 8nt |
| **SM025** | [phos]UUUUUUUUUUUUUUUUUUUUUUUUUUUUUU[biotin] | U30 + biotin |
| **SM029** | CCCCCCCCCCTCCCCCCCCTACTTTTCCCCCCCCCCCCCCCCCCCCCC<br>CCCCCCCTGGCGACGGCAGCGAGGC | AAAA target seed |
| **SM030** | CCCCCCCCCCTCCCCCCCCTACAAATCCCCCCCCCCCCCCCCCCCCCC<br>CCCCCCCTGGCGACGGCAGCGAGGC | ATTT target seed |

**Table S3 | Guide sequences used during the *in vitro* cleavage assays.**

| oligo | sequence (5'- 3') | description |
|---|---|---|
| **BG13031** | ATTTGTAGTAGGTTGTATAGT | DNA guide ATTT |
| **BG13033** | TTTTGTAGTAGGTTGTATAGT | DNA guide TTTT |
| **BG13035** | CTTTGTAGTAGGTTGTATAGT | DNA guide CTTT |
| **BG13037** | GTTTGTAGTAGGTTGTATAGT | DNA guide GTTT |
| **BG13039** | AAAAGTAGTAGGTTGTATAGT | DNA guide AAAA |
| **BG13041** | ACCCGTAGTAGGTTGTATAGT | DNA guide ACCC |
| **BG13043** | AGGGGTAGTAGGTTGTATAGT | DNA guide AGGG |
| **BG11523** | ATACTGAATTTGTATGTTTTG | DNA (FW) guide pUCDRB 23% |
| **BG11524** | TATTAATATATGCTGGCAAAT | DNA (FW) guide pUCDRB 27% |
| **BG11525** | GTATGTATTAGTGTGCTAGAT | DNA (FW) guide pUCDRB 31% |
| **BG11526** | AAAGATTGTTTCCTGTGCTTT | DNA (FW) guide pUCDRB 35% |
| **BG11539** | CAAAACATACAAATTCAGTAT | DNA (RV) guide pUCIDT 23% |
| **BG11540** | ATTTGCCAGCATATATTAATA | DNA (RV) guide pUCIDT 27% |
| **BG11541** | ATCTAGCACACTAATACATAC | DNA (RV) guide pUCIDT 31% |
| **BG11542** | AAAGCACAGGAAACAATCTTT | DNA (RV) guide pUCIDT 35% |
| **BG10631** | TTACTCATATATACTTTAGAT | DNA (FW) guide pUC19 -15 |
| **BG10632** | CATATATACTTTAGATTGATT | DNA (FW) guide pUC19 -10 |
| **BG10633** | ATACTTTAGATTGATTTAAAA | DNA (FW) guide pUC19 -5 |
| **BG10634** | TTAGATTGATTTAAAACTTCA | DNA (FW) guide pUC19 0 |
| **BG10635** | TTGATTTAAAACTTCATTTTT | DNA (FW) guide pUC19 5 |
| **BG10636** | TTAAAACTTCATTTTTAATTT | DNA (FW) guide pUC19 10 |
| **BG10637** | ACTTCATTTTTAATTTAAAAG | DNA (FW) guide pUC19 15 |
| **BG10646** | ATCTAAAGTATATATGAGTAA | DNA (RV) guide pUC19 -15 |
| **BG10647** | AATCAATCTAAAGTATATATG | DNA (RV) guide pUC19 -10 |
| **BG10648** | TTTTAAATCAATCTAAAGTAT | DNA (RV) guide pUC19 -5 |
| **BG10649** | TGAAGTTTTAAATCAATCTAA | DNA (RV) guide pUC19 0 |
| **BG10650** | AAAAATGAAGTTTTAAATCAA | DNA (RV) guide pUC19 5 |
| **BG10651** | AAATTAAAAATGAAGTTTTAA | DNA (RV) guide pUC19 10 |
| **BG10652** | CTTTTAAATTAAAAATGAAGT | DNA (RV) guide pUC19 15 |
| **BG8050** | [phos]UGAGGUAGUAGGUUGUAUAGU | RNA guide based on let-7 miRNA |
| **BG6790** | [phos]TGAGGTAGTAGGTTGTATAGT | DNA guide based on let-7 miRNA |
| **BG5640** | [phos]TGAGGTAGTAGGTTGT | DNA guide based on BG6790 |
| **BG5713** | [phos]TGAGGTAGTAGGTTG | DNA guide based on BG6790 |
| **BG5714** | [phos]TGAGGTAGTAGGTT | DNA guide based on BG6790 |
| **BG5715** | [phos]TGAGGTAGTAGGT | DNA guide based on BG6790 |
| **BG5641** | [phos]TGAGGTAGTAGG | DNA guide based on BG6790 |
| **oDS313** | [phos]CGAGGTAGTAGGTTGTATAGT | Let-7 miRNA based DNA guide crystallization |

| | | |
|---|---|---|
| **SM001** | [phos]CGAGGTAGTAGGTTGTATAGTT | Perfect guide |
| **SM002** | [phos]CGAGTATTTTTTTTTTTTTTT | 3nt match (nt 2-4) (nt 6, T changed to A) |
| **SM003** | [phos]CGAGGATTTTTTTTTTTTTTT | 4nt match (nt 2-5) (nt 6, T changed to A) |
| **SM004** | [phos]CGAGGTTTTTTTTTTTTTTT | 5nt match (nt 2-6) |
| **SM005** | [phos]CGAGGTATTTTTTTTTTTTTT | 6nt match (nt 2-7) |
| **SM006** | [phos]CGAGGTAGATTTTTTTTTTTT | 7nt match (nt 2-8) (nt 9, T changed to A) |
| **SM007** | [phos]CCATCT | 6nt |
| **SM008** | [phos]CCATCTG | 7nt |
| **SM009** | [phos]CCATCTGT | 8nt |
| **SM010** | [phos]CCATCTGTC | 9nt |
| **SM011** | [phos]CCATCTGTCT | 10nt |
| **SM012** | [phos]CCATCTGTCTT | 11nt |
| **SM013** | [phos]CCATCTGTCTTA | 12nt |
| **SM014** | [phos]CCATCTGTCTTAG | 13nt |
| **SM015** | [phos]GAGGUAGU*AGGUUGUAUAGUU | Mature let-7a perfect guide |
| **SM016** | [phos]GAGUAUUU*UUUUUUUUUUUUU | Mature let-7a 3nt match (nt 2-4) (nt 6, U changed to A) |
| **SM017** | [phos]GAGGAUUU*UUUUUUUUUUUUU | Mature let-7a 4nt match (nt 6, U changed to A) |
| **SM018** | [phos]GAGGUUUU*UUUUUUUUUUUUU | Mature let-7a 5nt match |
| **SM019** | [phos]GAGGUAUU*UUUUUUUUUUUUU | Mature let-7a 6nt match |
| **SM020** | [phos]UGAGGUAGAU*UUUUUUUUUUUU | Mature let-7a 7nt match (nt 9, U changed to A; U* is at nt 10) |
| **SM027** | [phos]AAAAGTAGTTTTTTTTTTTTT | AAAA guide seed |
| **SM028** | [phos]ATTTGTAGTTTTTTTTTTTTT | ATTT guide seed |

U* in italic is the uridine that is conjugated with amino-modifier C6-U phosphoramidite.

**Table S4 | Primers used to construct the expression plasmids.**

| Oligo | Sequence (5'-3') | description |
|---|---|---|
| **BG6808** | GTTTCATTGGGCTAGCTGTGGGTACCCGTG | *Cb*Ago D541A fw |
| **BG6809** | CACGGGTACCCACAGCTAGCCCAATGAAAC | *Cb*Ago D541A rv |
| **BG6810** | TTGTGATTCATCGTGCTGGGTTTTCTCGTG | *Cb*Ago D611A fw |
| **BG6811** | TCACGAGAAAACCCAGCACGATGAATCACA | *Cb*Ago D611A rv |
| **oDS067** | TACTTCCAATCCAATGCAAATAATCTGACCTTTGAGGCTTTTG | LIC cloning fw |
| **oDS068** | TTATCCACTTCCAATGTTATTACAGAAAGAATAGACGATTATCAAC | LIC cloning rv |
| **SM026** | AAAAAAAAAAAAGAGGTAGTAAAA | DNA splint 8nt match |

**Table S5 | Plasmids.**

| Plasmid | description | vendor |
|---|---|---|
| **p*Cb*Ago1** | Codon harmonized *C. butyricum* Ago gene in pET28 expression plasmid | GenScript (USA) |
| **p*Cb*Ago2** | p*Cb*Ago1, with substituted active site *Cb*Ago DM (D541A, D661A) | |
| **pUC19** | Standard pUC19 vector | NEB |
| **pDRB** | pUC19 with 16x 100bp Gene fragments human genome | IDT |

# Chapter 4

## Argonaute glides past cellular obstacles during target search

Tao Ju Cui, Misha Klein, Jorrit W. Hegge, Stanley D. Chandradoss, John van der Oost, Martin Depken, Chirlmin Joo

Adapted from –
'Argonaute glides past cellular obstacles during target search'
Nature communications, accepted for peer-review

# Abstract

Argonaute (Ago) proteins are key players in gene regulation and host defense in eukaryotes and prokaryotes. For specific interference, Ago proteins rely on base pairing between small nucleic acid guides and complementary target sequences. To efficiently scan nucleic acid chains for potential targets, Ago must bypass both secondary structures and protein barriers. Using a DNA-targeting Argonaute from the mesophilic bacterium *Clostridium butyricum* (*Cb*Ago), we employed single-molecule imaging and kinetic modeling to reveal that lateral diffusion is mediated mainly through protein-nucleic acid interactions, rather than interactions between the guide and targeted strand. This interaction enables *Cb*Ago to scan for targets with high efficiency but without maintaining tight contact with the DNA backbone, which allows *Cb*Ago to "glide" short distances over secondary structures along the nucleic-acid chain. Moreover, it is demonstrated that the target search process is coupled to intersegmental jumps over longer distances, which reduces scanning redundancy and allows for bypassing more substantial barriers, such as other DNA-binding proteins.

# Introduction

Target recognition by oligonucleotide guides is essential in cellular development, differentiation and immunity (Eulalio et al., 2008; Swarts et al., 2014a). Argonaute (Ago) proteins are key mediators of target interference processes, by utilizing short oligo-nucleotides (~20-30 nt) as guides for finding complementary target sequences (He and Hannon, 2004; Rivas et al., 2005). The guide-target interaction initiates through Watson-Crick base pairing at the "seed" segment at the 5' part of the guide, after which target binding propagates downstream along the guide, resulting in target interference (Bartel, 2009).

Eukaryotic Argonautes use RNA guides to target RNA. Prokaryotic Agos (pAgo) have recently been demonstrated to vary with respect to the nature of their guide and target (Olovnikov et al., 2013; Swarts et al., 2014b, 2017a)**.** Depending on the pAgo type, it uses either DNA or RNA guides to target single-stranded DNA (ssDNA), RNA (ssRNA) or both (Swarts, Makarova, *et al.*, 2014, Hegge 2018). In **chapter 3**, we show that an Ago from the mesophilic bacterium *Clostridium butyricum* (*Cb*Ago) utilizes DNA guides to cleaves DNA at moderate temperatures (37°C), highlighting the general potential of pAgos for genome editing purposes (Hegge et al., 2018a). To make effective use of pAgo for such applications, it is crucial to understand the underlying protein-DNA interactions that pAgo utilizes for target interrogation.



In the context of a living cell Ago-guide complexes need to search long stretches of nucleic acids before finding a complementary target. In a previous biophysical study (Chandradoss et al., 2015) we suggested that a human Argonaute 2 (hAGO2) uses lateral diffusion along RNA for target search. Yet, the nature of such a mechanism remains unclear, as lateral diffusion alone would lead to excessive re-sampling of potential target sites and to problems at various roadblocks that are present on the target nucleic acids (Gerland et al., 2002; Kolomeisky and Veksler, 2012).

For other DNA-binding proteins, such as transcription factors (TFs), a multi-step process termed facilitated diffusion (**Table S1**) has been proposed. Use of such a mechanism would lead to reduced sampling redundancy, and the possibility to circumvent obstructions when TFs search for their targets. After recruiting the DNA non-specifically from solution, the protein diffusively scans only a limited section (Berg et al., 1981; Halford and Marko, 2004; Hammar et al., 2012; Von Hippel and Berg, 1989), until it dissociates and returns to solution,

after which the protein binds to a new section. During diffusive scanning, the TF interrogates DNA sites to find a specific target. In addition to complete dissociation into solution, intersegmental jumping, where a protein transfers between two spatially close-by segments, has been shown to occur for the DNA binding restriction enzyme EcoRV (van den Broek et al., 2008). The complexity of the target search further increases due to the fact that nucleic acids can be covered with other proteins (Wang et al., 2013) and structural elements such as hairpins and plectonemes (Travers and Muskhelishvili, 2005). It is presently unknown whether nucleic acid-guided proteins like Agos utilize facilitated diffusion.

DNA/RNA-guided target search depends on short-ranged (Jones et al., 2017; Li et al., 2009b; Sternberg et al., 2014) and short-lived interactions (Chandradoss et al., 2015; Globyte et al., 2018b, 2018a; Jones et al., 2017; Xue et al., 2017). A method of high spatiotemporal resolution is therefore needed to understand how Ago proteins localize their target within an extensive and crowded target pool. Here we make use of single molecule Förster Resonance Energy Transfer (FRET) to elucidate the mechanism of ssDNA target search by *Cb*Ago, a mesophilic bacterial Ago. We show that *Cb*Ago does not remain in tight contact with the DNA backbone, enabling it to bypass secondary structures along the nucleic-acid chain by "gliding" over them without apparent loss in its ability to recognize its target. After gliding locally, the protein is able to reach distant sites (>100 nt) along the DNA through intersegmental jumps, and then starts gliding again. These different modes of lateral diffusion allow Ago to rapidly search through facilitated diffusion, as well as to bypass substantial obstacles during DNA scanning.

# Results

## Single-molecule kinetics of *Cb*Ago binding

To elucidate the complex target search mechanism, we make use of the high spatial sensitivity of single-molecule FRET (Förster resonance energy transfer). We study a minimal Argonaute complex that consists of *Cb*Ago, loaded with a 22 nt DNA guide (small interfering DNA, siDNA) (Hegge, Swarts et al 2018). By using total internal reflection (fluorescence, TIRF) microscopy, we recorded the interactions of *Cb*Ago-siDNA with target DNA. Target DNA was immobilized on a PEG-coated quartz surface in a microfluidic chamber through biotin-streptavidin conjugation. Guide-loaded *Cb*Ago was introduced to the microfluidic chamber by flow. The target was composed of a 3 nt sub-seed motif embedded within a poly-thymine sequence and labelled with an acceptor dye (Cy5) (**Figure 1A**). The guide construct was labelled at nt 9 from the 5'-end with a donor dye (Cy3) (**Figure 1B**). A 532-nm laser excitation resulted in donor excitation when the protein loaded with the guide DNA interacted with the target DNA. Once the *Cb*Ago-siDNA complex became bound to the target, the proximity of the donor dye to the acceptor dye on the target resulted in high FRET efficiency. This was followed by a sudden disappearance of the signal, indicating that the complex dissociated from the target and diffused into the free solution. Freely diffusing molecules move too rapidly (~µs) in and out of the evanescent field for the current time resolution of the experimental setup (100 ms) and were therefore not recorded. Likewise, when a DNA target that lacks complementary to the seed motif of the guide was used, only transient interactions were detected (**Figure 1C**).



To observe target search that involves intrinsically transient interactions, we determined the optimal target DNA motif for recording binding events. The optimal motif should provide binding events longer than our detection limit of 100 ms, but still lead to dissociation events within the time of our measurement (200 s). To determine the optimal motif, the complementarity between guide and target was incrementally extended from nt 2 to 8 of the guide, showing a gradually increasing dwell time of the Ago-siDNA complex (**Figure S1C**). We found that increasing the number of complementary base pairs above 6 resulted in stable binding beyond the photobleaching time (**Figure S1C**). To maintain weak interactions, we continued our experiments using a siDNA with three-base complementarity with the target (nt 2-4) (**Figure 1D**). Our estimation of the photobleaching rate ($1.4 \times 10^{-3}$ $s^{-1}$) was an order

of magnitude lower than the dissociation rate (2.7 x $10^{-2}$ $s^{-1}$) (**Figure S1D**), suggesting that photobleaching does not affect our estimation of the dissociation rate.

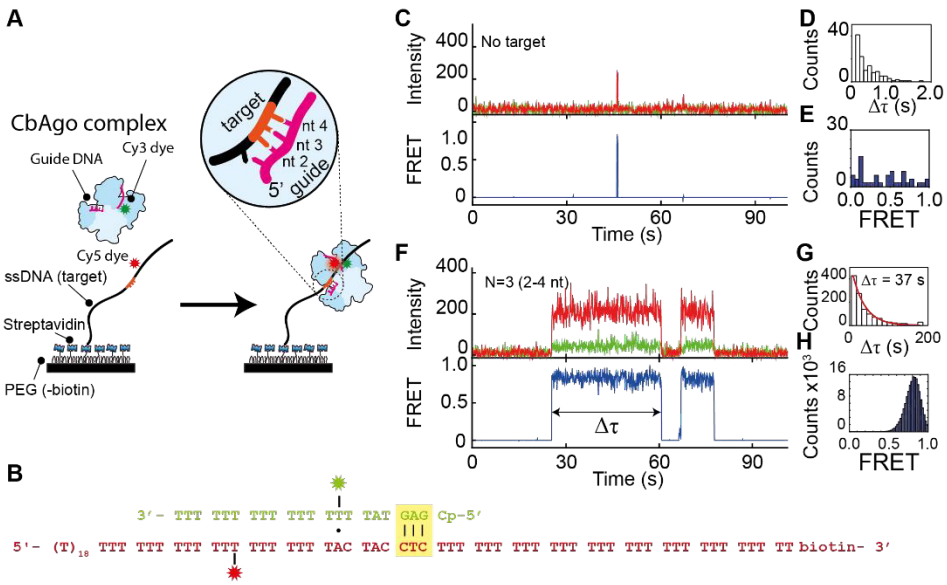



**Figure 1 | Single molecule imaging of target binding by siDNA:*Cb*Ago complex. A** | Immobilization scheme of the Argonaute-guide DNA complex. ssDNA is immobilized on a pegylated quartz slide surface. Presence of the Ago-siDNA complex is detected by specific binding to target site (light yellow) resulting in high FRET. **B** | Sequences of guide (green) and target DNA. Guide is labelled on the 9th nucleotide position. **C** | Representative FRET trace of a single molecule experiment at 100mM NaCl showing a transient interaction between *Cb*Ago and a poly-T strand. Time resolution is 100 ms. **D** | Dwell time distribution of the Argonaute in absence of target motif. **E** | FRET values of the transient interactions of (D). **F** | Representative FRET trace of a single molecule experiment showing the interaction between *Cb*Ago and a 2-4 nt motif. **G** | Dwell time distribution Dwell time distribution of N=3 binding events with the mean dwell time of 37 s. **H** | FRET histogram of binding events, showing a single FRET population for N=3 (2-4 nt) at E=0.78.

## Lateral diffusion of *Cb*Ago

It was previously shown that an Ago-guide complex does not directly bind a specific target site from solution, but rather binds non-specifically to random positions along a surfaced-immobilized nucleic acid construct (Chandradoss et al., 2015). Such non-specific interactions of *Cb*Ago-siDNA along target DNA are too short-lived to resolve in the absence of a canonical target motif (**Figure 1C**), and in the presence canonical target motif there was still no lateral diffusion visible (**Figure 1F**). As we were unable to resolve lateral diffusion by *Cb*Ago from non-specifically bound regions to the target, we questioned whether the

observed stable signal for three complementary base pairs is due to stable binding to the target or contains lateral excursions below our time resolution. In case of the latter, measured apparent dwell times (**Figure 1G**) would consist of the combined dwell times of many target escapes through lateral diffusion, each followed by rapid recapture below the detection limit, before *Cb*Ago eventually unbinds from the DNA (**Figure S1G**). It is shown (See **supplementary Kinetic Modelling**) that such a process of repeated recapture would result in an exponential distribution of apparent dwell times, in accordance with **Figure 1G**.

To overcome the temporal resolution limit, we adopted a tandem target assay (Chandradoss et al., 2015; Ragunathan et al., 2012). While lateral diffusive excursions from a trap are too short-lived to resolve in the presence of only a single target, a second target can trap an excursion for long enough to be observed. We placed two identical optimal targets ($N_1$ and $N_2$) separated by 22 nt (**Figure 2A**) along the DNA construct. Both targets base pair only with the first three nucleotides (nt 2-4) of the guide bound by *Cb*Ago. As the second target is located further away from the acceptor dye, binding the second target results in a lower FRET efficiency than binding the first target. The respective distance and FRET efficiency between the first binding site ($N_1$) and the acceptor dye (Cy5) remained the same as for the single target assay (E~0.78), while an additional peak occurs at a lower FRET efficiency for the second target (E~0.43, **Figure 2D**). This difference in FRET values allows us to determine which of the two sites *Cb*Ago-siDNA is bound to (**Figure 2B**). After binding to one of the target sites, a majority of the binding events (87.8%) resulted in *Cb*Ago-siDNA shuttling to the other target without loss of FRET signal. Under our standard experimental condition (100 mM NaCl), an average of 13.5 shuttling events occur per binding event (**Figure 2E**). When the experiment was repeated with guides and targets with increased complementary to each other, a 6-nt match (nt 2-7), only 15.1% of the traces showed the shuttling signature within our time window (**Figure S2F**), establishing that the shuttling signature is controlled by *Cb*Ago-ssDNA-motif interactions. With a 6-nt match, the target is so strongly bound and it is less likely that we observe a shuttling event within our observation window.

Interestingly, the average dwell time of the first target (**Figure 1G**) decreased from 37 s to 1.7-1.8 s after adding a second target in its vicinity (**Figure 2C**). This observation is in agreement with our lateral diffusion model, since with close by targets, each sub-resolution diffusive excursion is more likely to be caught in the opposing target.

To further test our claim that the transition between targets occur through lateral diffusion, we extract the time between each shuttling event from traces using single-molecule analysis software (Van De Meent et al., 2014). The average time ($\Delta\tau_{\text{shuttle}}$) between shuttling events is recorded, the shuttling rate is estimated ($k_{\text{shuttle}} = 1/\Delta\tau_{\text{shuttle}}$), and the results are compared to model predictions below.

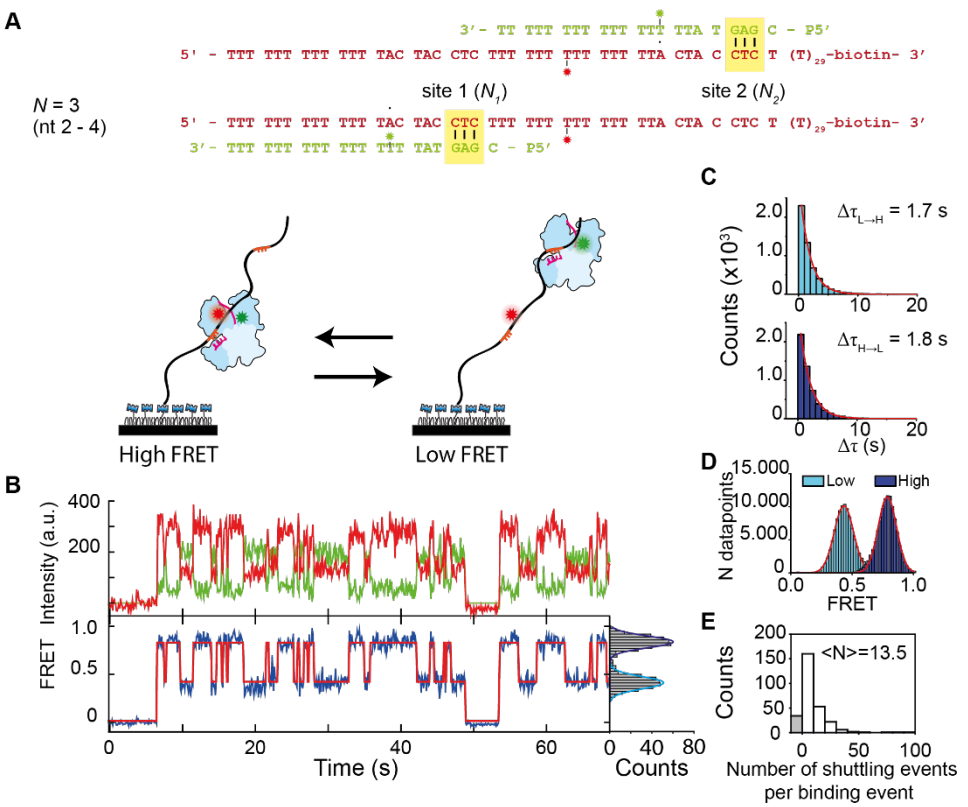



**Figure 2 | Shuttling signature of *Cb*Ago appears in presence of two targets.** **A** | In the top right corner the DNA sequence of guide and target for 22 nt separation between targets. Here the distance is defined as the distance from beginning of a target to the beginning of the next target. The placement of the second target (N2) results in the appearance of an additional FRET signal, with lower FRET efficiency. **B** | (Top) Representative shuttling trace of a 22 nt separation tandem target at 100 mM NaCl for N=3. (Bottom) The corresponding FRET states (blue) with the fitted HMM trace on top (red). (Right) FRET histogram of the respective time trace. Time resolution is 100 ms. **C** | Dwell time distributions of respectively the transitions from low FRET state to high FRET state (top) and vice versa (bottom). **D** | FRET histograms of respective states, with peaks at 0.43 and 0.78. **E** | Shuttling event distribution for the same conditions (N=309). Bin size = 10. On average 13.5 shuttling events take place before dissociation. The grey bar (N=33) marks binding events followed by dissociation (no shuttling).

## Kinetic modelling of lateral diffusion

To determine how lateral diffusion contributes to the shuttling rate, we kinetically model how the shuttling rate depends on the distance between traps. The DNA construct is modelled as a series of binding sites along which *Cb*Ago will perform an unbiased random walk between neighboring nucleotides. The rate of stepping away from the target is $k_{\text{escape}}$ in both directions, while at non-specific sites (poly-T) the stepping away is assumed to be near instantaneous—an approximation justified by the fact that lateral excursions are never resolved in the experiments. Hence, the apparent shuttling rate equals the rate at which the protein escapes the initial trap, $k_{\text{escape}}$, multiplied by the probability to get captured by, shuttle into the other trap ($P_S$). In the **supplementary Kinetic Modelling** we show this probability to scale proportionally with the distance $x_{\text{Target}}$ between the targets start nucleotides, resulting in a simple relationship between the average dwell time and the distance between targets.

$$\Delta\tau_{shuttle}(x_{\text{Target}}) = \frac{1}{P_S(x_{\text{Target}}) \times k_{escape}} = \frac{x_{\text{Target}}}{k_{escape}} \qquad (1)$$

In support of this model, we observed that the apparent shuttling rate $(1/\Delta\tau_{shuttle}(x_{\text{Target}}))$ decreases when the distance between the targets increases (11, 15, 18 and 22 nt) (**Figure 3**). A fit to **Equation 1** reveals that *Cb*Ago-siDNA complexes escape the target site at a rate of 12 times per second ($k_{escape}$ = 11.9 $s^{-1}$).

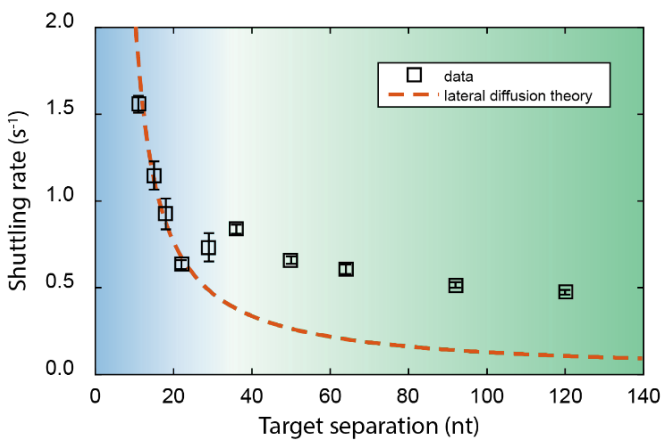



**Figure 3 | *Cb*Ago shuttling behavior differs across short and large distances.** Shuttling rate is plotted versus distance between targets. The orange dotted line indicates a minimum assumption model where unbinding is ignored due to the dwell time per state being much smaller than the unbinding time. Plotted error bars are the 95% percentile of $10^5$ bootstrapped dwell times. The blue region indicates where only lateral diffusion is present, where data points are fitted with $k_{escape}$= 11.9 $s^{-1}$. This theory breaks down for larger distances (green).

## Ago probes for targets during lateral diffusion

Next, we placed a third target on the tandem construct (**Figure 4A**), keeping the distance between each set of neighboring targets similar to the target separation in the previous assay (11 nt). We observed three different FRET levels, corresponding to *Cb*Ago getting trapped at the three different targets (**Figure 4B**).

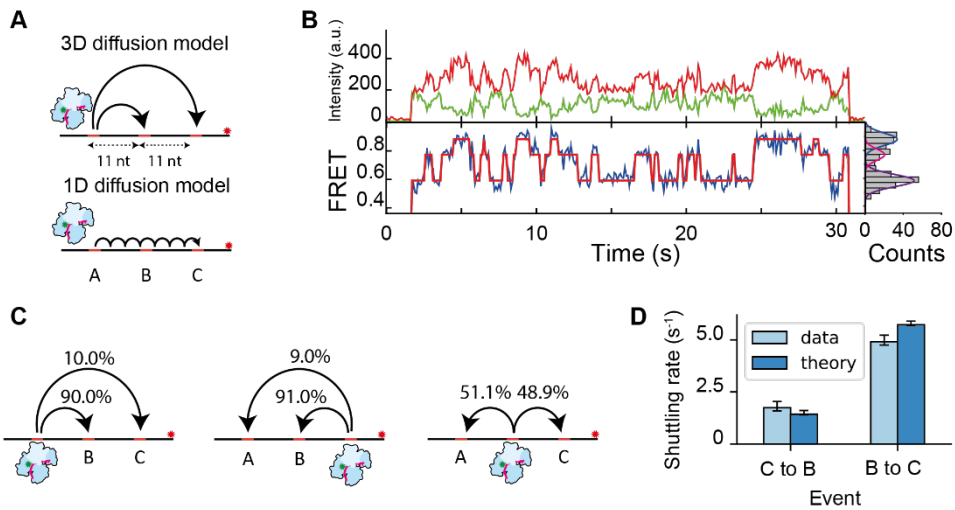



**Figure 4 | *Cb*Ago undergoes short range diffusion through correlated jumps. A** | Models for target translocation at short range. In the 3D diffusion model, target dissociation occurs from A followed by random 3D diffusion through solution. In effect, the neighboring two targets (B and C) will compete for binding. In the lateral diffusion model, the *Cb*Ago complex will have to bypass the adjacent target B before binding to target C. **B |** Representative FRET trace showing the shuttling behavior between three targets. Top: donor (green) and acceptor (red) intensities. Bottom: FRET trace (blue) and HMM assigned states (red). Right: The fitted states from this data trace. **C |** Transition probabilities from one state to the other two derived from the HMM software. **D |** Experimental values of the shuttling rate of the three target construct were compared against the parameter-free theoretical model that only uses the $k_{escape}$= 20 $s^{-1}$ from Figure 3. Error bars indicate the 95% confidence interval acquired from 105 bootstraps.

Using Hidden Markov Modelling (HMM), states can be assigned (**Figure 4B**) and transition probabilities can be extracted (**Figure 4C**). If *Cb*Ago returns to solution between binding targets, transitions between any pair of targets will be equally probable, resulting in equal effective rates between all targets. However, if lateral diffusion dominates, transitions between adjacent sites will be favored. The transition probabilities (**Figure 4C**) indicate that

most transitions between the two outer targets (from state A to C, or from C to A) proceed through the intermediate target site (state B). The rate to transfer from B to either A or C is greater than that of the opposite path (A or C to B). Using the fitted escape rate from above, $k_{escape} = 11.9\ s^{-1}$, we predict similar times based on our theoretical model for lateral diffusion (**Figure 4D, supplementary Theoretical Modelling**). With no more free-parameters remaining for this prediction, we take this experimental agreement with our prediction as further evidence of lateral diffusion.

It is noteworthy that there are about 10% direct transitions from A to C and C to A without any intervening dissociation. The exponential distribution of the dwell times (**Figure S4C**) suggests that at our current time resolution these 10% may be either due to missed events or due to the existence of an additional diffusive mode through which Ago is able to bypass the intermediate target. To conclude, Ago uses lateral diffusion to repeatedly scan DNA segments locally.

## Ago target search is unhindered by structural and protein barriers



To further probe the mechanism of lateral diffusion, a Y-fork structure (DNA junction) was introduced as a road block between two targets (**Figure 5A**), while keeping their separation the same as **Figure 3**. The construct was designed such that the labelled target was partially annealed at the stem with a biotinylated target, thus only annealed constructs were observable on the surface of the microfluidic device. When *Cb*Ago binds to either of the two targets, it can reach the other target only by crossing the junction. Our measurement showed that there was no significant difference in shuttling rate between the standard tandem-target construct and the Y-fork construct (**Figure 5B**), indicating that the Y-fork does not impede any of the lateral diffusion modes present. We observed that the *Cb*Ago-siDNA complex is not able to stably bind to dsDNA, suggesting that the protein cannot simply go around the junction **(Figures S1B, S5).** Thus, our result suggests that the Ago-siDNA complex does not maintain tight contact with the DNA during lateral diffusion. A weak interaction with the DNA molecule instead allows *Cb*Ago-siDNA to move past the junction.

Next, we questioned whether *Cb*Ago is also able to overcome larger barriers, such as proteins which are not traversable through lateral diffusion alone. Lin28, a sequence-specific inhibitor

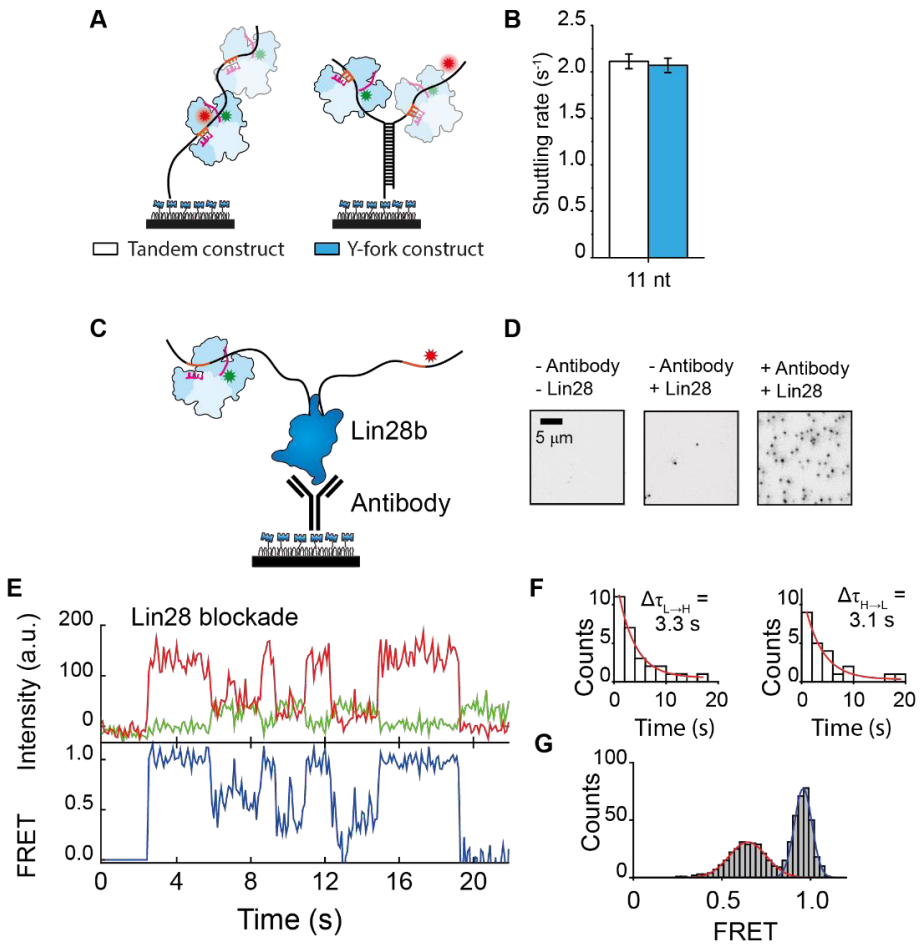



**Figure 5 | Argonaute can overcome structural and protein barriers. A |** Schematic drawing of the Y-fork assay (right). *Cb*Ago does not interact readily with the dsDNA junction in the middle so the presence of the junction may interfere with the diffusion. **B |** The shuttling rate of the Y-fork junction (blue bar) compared with the tandem assay (white bar). **C |** Schematic drawing of the his-Lin28b blockade assay. Immobilization happens through a biotin-anti-His antibody. **D |** An EMCCD image of the acceptor channel. (Left) In absence of Lin28 protein and antibody with Cy5 labeled DNA. (Middle) In absence of antibody, but in presence of Lin28 protein and Cy5 labeled DNA. (Right) In presence of antibody, Lin28 protein and Cy5 labeled DNA. **E |** Example of a shuttling trace with Lin28b located in between two targets with an exposure time of 100 ms. **F |** Individual dwell times from low FRET state to high FRET state (left) and vice versa (right). **G |** FRET histogram (N=46) fit with two Gaussian functions (E=0.64 for red fit and E=0.95 for dark blue fit).

of let-7 miRNA biogenesis, has been found to associate sequence specifically to RNA and DNA (Nam et al., 2011).His-tagged Lin28 was immobilized on the surface of the microfluidic chamber after which a fluorescent ssDNA fragment was added containing a central Lin28 binding motif and an Ago target motifs on either side (**Figure 5C**). Although the presence of the protein blockade (**Figure 5E**) lowered the shuttling from 0.60 $s^{-1}$ to 0.27 $s^{-1}$, it did not preclude Ago from reaching the distal site (**Figure 5F**). Since short-range lateral movement is now blocked by the protein barrier, Ago's ability to move between targets suggests that the target search process also allows for intersegmental jumps, in accordance with our observation that the middle target is sometimes skipped when transitioning between the outer targets in **Figure 4D**. Noticeably, the presence of the protein blockade gives rise to FRET fluctuations, broadening the FRET peak (**Figure 5G**).

## Intersegmental jumps allow Ago to rapidly access distant DNA segments

Lateral diffusion is not expected to dominate across large distances, as the steep $1/x_{Target}$ decay in the rate of reaching the neighboring (partial) targets (Equation 1) would render the search process prohibitively slow. However, when *Cb*Ago was studied with tandem targets that were separated 36 nt or more, we observed that the shuttling still persisted across larger distances (**Figure 3, Table S2**). Together with the evidence of intersegmental jumping above, and the fact that the ssDNA can easily be coiled back to bring the second target close to the Ago protein (Chen et al., 2012), we speculate that there is a second mechanism of lateral diffusion**:** after local scanning for the target through sliding, the *Cb*Ago complex transfers to a different part of the segment that has looped back into proximity of *Cb*Ago. This intersegmental jumping mechanism would enable *Cb*Ago to travel to new sites without dissociating, and rescanning of the same segment would be avoided (Halford and Marko, 2004; Kolomeisky and Veksler, 2012).

In order to test this hypothesis, we altered the ionic strength of the buffer solution. Based on the dependence of the single-target off-rate on the ionic strength (**Figure S1E**), we hypothesized that counterions are not completely expunged when binding to the target sequence. If so, the rate of the intersegmental jumps would also be dependent on salt concentration: increasing the ionic strength should make *Cb*Ago more likely to escape the target motif and more likely to jump to the distant segment. However, at the same time, increased electrostatic screening would also result in a lower binding rate to a nonspecific

strand. Depending on which effect is stronger, the shuttling rate may go either up or down when ionic strength is increased. However, if the shuttling took place without disturbing the ionic cloud, the effect from ions would be negligible and therefore the shuttling rate should remain the same.

We used dual-target constructs with 15-nt separation and 64-nt separation (**Figure 6**), taken from the two different regions in **Figure 3** (indicated by blue and green shading). At a separation of 64 nt, we observed a 13-fold increase of the observed shuttling rate relative to 10 mM NaCl concentration with increasing salt concentration up to 200 mM NaCl. This rate increase suggests that the protein indeed traverses the ion cloud upon long-range shuttling, and that the looseness of association allows for counterions to facilitate dissociation from the target more than it hinders association through screening of the DNA strand.

**Ago uses a gliding mode for local scanning**

In contrast, we observed that, for the dual-target construct with 15-nt separation, the shuttling rate (when corrected for the dissociation rate) changed roughly only two-fold when the ionic strength of the solution was altered within our range of 10 mM to 200 mM NaCl (**Figure 6)**, a modest change compared to 13-fold of the dual-target constructs with 64-nt separation. We hypothesize that during short-ranged lateral diffusion, Ago forces little rearrangements in the ion cloud, and is thus only marginally affected by ionic strength. When targets are close together, this is the main mode of shuttling. Comparatively, for the 64-nt construct, the complex is unlikely to reach the distal site through short-ranged lateral diffusion only, and travels outside the cloud, explaining the strong salt dependence.

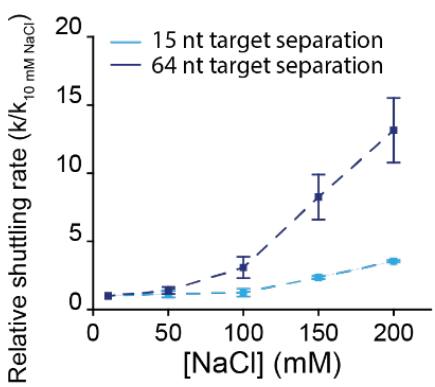



**Figure 6 | Argonaute target search is characterized by gliding and intersegmental jumps.** The relative change in shuttling rate of two constructs from (A), 15 nt separation (light blue) and 64 nt separation (dark blue), normalized against $k_{shuttling}$ at 10 mM NaCl. Errors of the ratio were determined through bootstrapping $10^5$ times the ratio of $k/k_{10\ mM\ NaCl}$.

In our model, for short distances *Cb*Ago travels predominantly within the ionic cloud, while our Y-fork experiments show that *Cb*Ago need not be in constant physical contact with its substrate during this process (**Figure 4B**). Since neither sliding—characterized by constant binding and complete displacements of counterions, nor micro-dissociations/associations—characterized by longer range jumps (Berg et al., 1981) completely capture the properties of this mode of diffusion, we use the term "gliding" to emphasize that it takes short steps, but it does so while being loosely associated with the DNA.

In conclusion, lateral diffusion during *Cb*Ago target search is governed by two distinct modes. For short distances, lateral diffusion takes place through a gliding process characterized by loose contact with the DNA strand while remaining within the ionic cloud. This allows the protein to overcome secondary structures. For larger distances, *Cb*Ago is able to utilize intersegmental jumps to nearby segments.

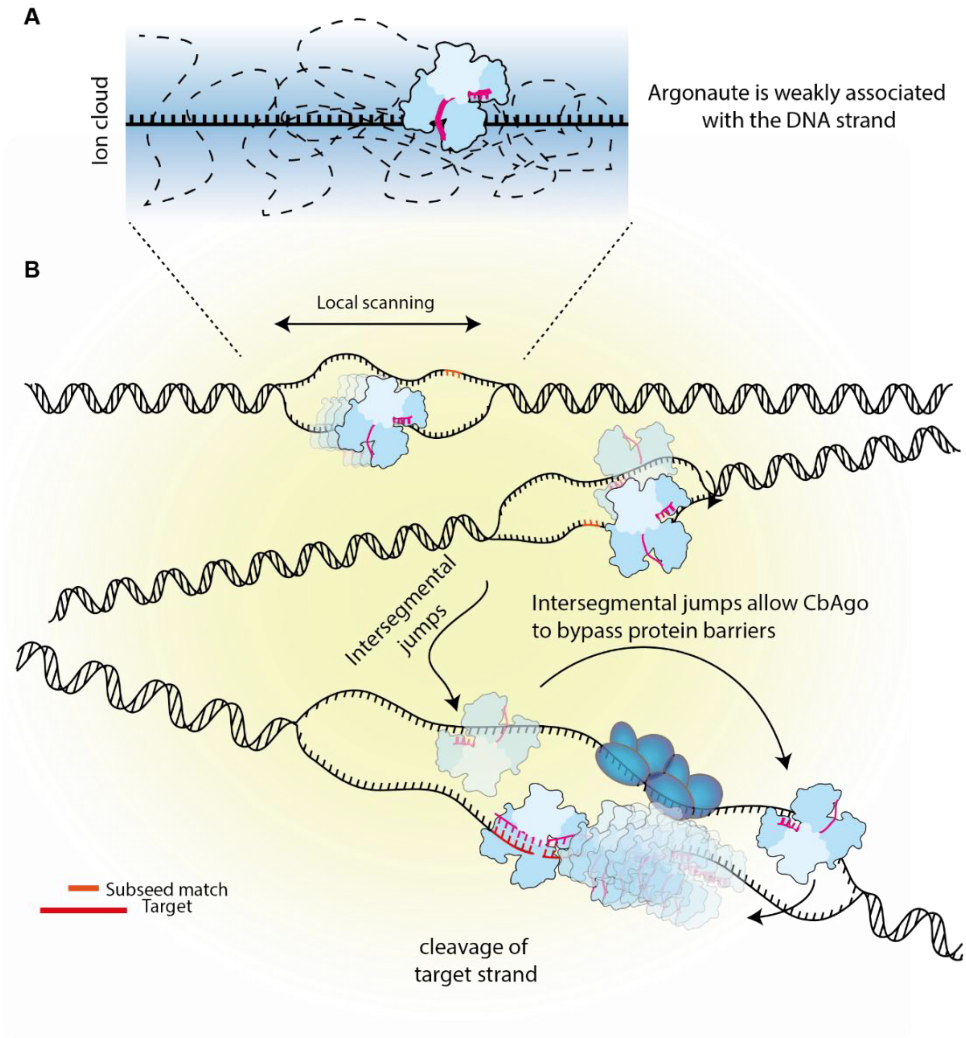



**Figure 7 | Biophysical model of target search. A |** The *Cb*Ago complex utilizes short transient interactions with nucleic acid strands to rapidly sample the adjacent sites for possible targets. **B** | Weak interaction with the nucleic acid strand allows Argonaute to travel to neighboring strands while protein obstacles can be overcome through intersegmental jumps. At the same time, travelling to other DNA segments through intersegmental jumps reduces redundancy in target scanning.

## Discussion

Within a vast number of potential targets, Ago-guide complexes have to minimize the time spent on non-targets as speed and timing of regulation is crucial for the development of the cell and the host defense (Banerjee and Slack, 2002). Our single-molecule study shows that Argonaute from *C. butyricum* (*Cb*Ago) uses a target search mechanism distinct from other known DNA binding proteins. Since the *Cb*Ago searches on ssDNA, not dsDNA, it cannot make use of the structural regularity of double stranded DNA. Bacterial Ago utilizes gliding to scan locally for complementary DNA targets, a mode where the protein is not tightly associated with DNA but stays within the ion cloud (**Figure 7A**). To the best of our knowledge, this mode of lateral diffusion has not been reported for any DNA interacting proteins. In addition, we show that *Cb*Ago is able to move to a new segment via intersegmental jumps, avoiding redundant scanning of the same segment (**Figure 7B**) and allowing *Cb*Ago to bypass roadblocks.

In literature, short-range lateral diffusion of target search typically consist either of tight-contact translocation (sliding), or of weakly-bound translocation (hopping) (Berg et al., 1981). Our experiments obtained with DNA junctions and protein barriers rule out tight association with DNA strand and thus would suggest that Ago uses hopping (a series of microscopic dissociations and associations) for a dominant mechanism. Unexpectedly, we observed that the weakly interacting short-range lateral diffusion depends only mildly on ionic strength. The absence of a strong ionic strength dependence indicates that, when Ago makes a short exclusion away from a target site, it remains within the ion cloud, only slightly rearranging the ions in its vicinity. Within our experimental conditions, the characteristic Debye distance is estimated to be a couple of nanometers, which is thick enough to accommodate the profile of Ago. Therefore, we propose that Ago is "gliding" through the Debye cloud, displacing some of the counter ions, but not all.

Formation of ssDNA rarely occurs *in vivo*. The fact that infection of prokaryotes by mobile genetic elements (plasmids and viruses) often proceeds via a ssDNA-state may reflect the physiological importance of ssDNA targeting by pAgo. Upon entry in the infected cell, ssDNA binding and recombination proteins may associate with the invading nucleic acid, and DNA polymerase will start to generate the second strand. In addition, it is anticipated that secondary structures will be formed. This will generate road blocks that may affect

scanning by defense systems such as restriction enzymes and Argonaute. In case of ssRNA, both in prokaryotes and in eukaryotes, it is well known that complex secondary structures can be formed by base pairing different anti-parallel RNA segments (Beaudoin et al., 2018; Bevilacqua et al., 2016; Ding et al., 2014; Vandivier et al., 2016). The presence of secondary structures suggest that gliding is necessary for Agos to search along ssRNA. Based on the functional and structural and functional similarities of prokaryotic Agos and eukaryotic Agos (Chandradoss et al., 2015; Swarts et al., 2014a), we expect eAgo to also glide past RNA secondary structures, minimizing time spent trapped at such structures.

The effect of lateral diffusion on the total target search time is dependent on the roughness of the energy landscape that the DNA binding protein encounters once it binds non-specifically. Theoretical predictions point out that if the roughness exceeds 2 $k_BT$, the energy landscape would prevent a protein to diffuse laterally further than a few nucleotides (Slutsky and Mirny, 2004). Here, we expect the variation in the landscape in our poly-T sequence to be minimal for the target search, so that lateral diffusion is able to occur over distances larger than few nucleotides. However, the effect of *in vivo* DNA sequences on the target search remains unexplored, and we expect nucleic acid-guided proteins to encounter a rugged energy landscape during sequence interrogation (Klein et al., 2017). We have inferred a 12 $s^{-1}$ escape rate from the 3-nt GAG guide sequence (**Figure 3**), indicating that if a strand were to consists only of GAG in repeating order, an effective diffusion coefficient is $D = \frac{dx^2}{2dt} = \frac{nt^2}{2(2*k_{escape})^{-1}} = \mathrm{nt}^2 k_{escape} = 12 \frac{\mathrm{nt}^2}{s}$. Changing the number of base-paring nucleotides as well as the identity of nucleotides in the guide/target could provide insights into how sequence variation would affect the rate of diffusion.

Compared to proteins that are known to interact tightly with dsDNA, such as restriction enzymes or DNA repair proteins are structurally adapted to associate with dsDNA (e.g. the clamp like structure of Msh2-Msh6 (Gorman et al., 2007)). However, the currently known structure of pAgos (Wang et al., 2008b; Willkomm et al., 2017b) indicate that these proteins lack the capability to firmly embrace and/or unwind dsDNA This by itself is not surprising, since dsDNA is substantially more conducive for tight interactions than flexible single stranded nucleic acids. Lacking a tight interaction, lateral diffusion for *Cb*Ago is expected to dominate only at short distances, and recurring binding and dissociation events are expected

to dominate at long distances. Additionally, since the guide strand only provides the specificity needed for accurate targeting, lateral diffusion could be reliant on the non-specific surface interactions with the protein. We envision that the positive surface charge distribution inside the Ago cleft could orientate the Ago with the guide towards the negatively charged nucleic acid strand, thereby promoting target interrogation while traveling along the target strand. It is unknown whether Ago is able to scan each base during this process or whether it skips over nucleotides. For our triple target construct, we have observed that 90% of the time the middle target traps Ago. It will be of interest to investigate whether this level of effective efficiency of target recognition is achieved by a low trapping efficiency offset by repeated passes over the target before Ago is eventually trapped.

For a longer range target search, we have observed that at 120 nt separation, the shuttling rate remains well above what would be expected for lateral diffusion (**Figure 3**). We hypothesize that coiling of the ssDNA (persistence length ~ 1 nm) may bring distant segments in close proximity, allowing intersegmental jumps over longer distances (beyond ~30 nt target separation). Presumably, Ago cannot use intersegmental jumping for covering shorter distances, as noted by the "dip" in **Figure 3,** but the Ago-guide complex might efficiently translocate across longer distances from one place to the other, to bridge sites that are separated >30 nt. Further theoretical modelling is required in order to establish to what extent partitioning different length scales will allow nucleic acid-guided proteins to traverse large distances to speed up the search process (Mirny et al., 2009; Slutsky and Mirny, 2004; Wunderlich and Mirny, 2008).



Inside the cell, DNA binding proteins may form roadblocks that hinder Ago from lateral scanning of nucleic acid sequences. The intersegmental jumping may be an effective strategy of bypassing such obstacles. Therefore, we propose a model in which Ago locally scans for a target whenever it encounters patches of ssDNA, before jumping to a neighboring DNA segment (**Figure 7B**) and repeating the process until it encounters its target. These patches of ssDNA may be the result of local melting, transcription bubbles, replication forks, or the aforementioned invasion of viruses and plasmids.

We hypothesize that similar target search strategies may be used by Agos from different families, which are structurally and functionally similar (Swarts et al., 2014a). For example, in RNA induced transcriptional silencing (RITS), guide-loaded AGO1 binds to a transcript

after which other proteins are recruited for heterochromatin assembly (Holoch and Moazed, 2015; Moazed, 2009). Similarly, in the piRNA pathway of *Drosophila* Aubergine and AGO3 from the PIWI subfamily associate with piRNA in germline cells to bind and cleave transposon transcripts, generating new guides for targeting transposons with Aubergine and Piwi, safeguarding the genomic integrity of the germline cell (Brennecke et al., 2007; Gunawardane et al., 2007; Li et al., 2009a). Likewise, the Piwi protein in mice is responsible for transcriptional silencing by binding to nascent RNA which induces heterochromatin formation (Aravin et al., 2008). In each of these functions, the reliance on guide-complementary sequential target search likely necessitates the usage of facilitated diffusion strategies to optimize the search time for proper regulation of cell development or gene stability.

## Acknowledgements

We thank Ian MacRae for critical reading and Malwina Szczepaniak, Margreet Docter, Dimitri de Roos, Anna Haagsma and Jan Wignand for technical support. C.J. was supported by Vidi (864.14.002) of the Netherlands Organization for Scientific research. M.K. and M.D. were supported by the Netherlands Organization for Scientific Research, as part of the Frontiers in Nanoscience program. M.D. acknowledges financial support from a TU Delft startup grant. J.v.d.O. was financially supported by two grants from the Netherlands Organization of Scientific Research (NWO; ECHO grant 711.013.002 and NWO-TOP grant 714.015.001).

# Methods

## Purification of *Cb*Ago

The *Cb*Ago gene was codon harmonized for *Escherichia coli* Bl21 (DE3) and inserted into a pET-His6 MBP TEV cloning vector (Addgene plasmid # 29656) using ligation-independent cloning. The *Cb*Ago protein was expressed in *E. coli* Bl21 (DE3) Rosetta™ 2 (Novagen). Cultures were grown at 37ºC in LB medium containing 50 µg ml-1 kanamycin and 34µg ml-1 chloramphenicol till an OD600 nm of 0.7 was reached. *Cb*Ago expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. During the expression cells were incubated at 18ºC for 16 hours with continues shaking. Cells were harvested by centrifugation and lysed, through sonication (Bandelin, Sonopuls. 30% power, 1s on/2s off for 5 min) in lysis buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM imidazole, supplemented with a EDTA free protease inhibitor cocktail tablet (Roche). The soluble fraction of the lysate was loaded on a nickel column (HisTrap Hp, GE healthcare). The column was extensively washed with wash buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 30 mM imidazole. Bound protein was eluted by increasing the concentration of imidazole in the wash buffer to 250mM. The eluted protein was dialyzed at 4ºC overnight against 20 mM HEPES pH 7.5, 250mM KCl, and 1mM dithiothreitol (DTT) in the presence of 1 mg TEV protease (expressed and purified according to Tropea et al. 2009 (Tropea et al., 2009)) to cleave of the His6-MBP tag. Next the cleaved protein was diluted in 20 mM HEPES pH 7.5 to lower the final salt concentration to 125 mM KCl. The diluted protein was applied to a heparin column (HiTrap Heparin HP, GE Healthcare), washed with 20 mM HEPES pH 7.5, 125 mM KCl and eluted with a linear gradient of 0.125-2 M KCl. Next, the eluted protein was loaded onto a size exclusion column (Superdex 200 16/600 column, GE Healthcare) and eluted with 20 mM HEPES pH 7.5, 500 mM KCl and 1 mM DTT. Purified *Cb*Ago protein was diluted in size exclusion buffer to a final concentration of 5 µM. Aliquots were flash frozen in liquid nitrogen and stored at -80ºC.

## Purification of His-tagged Lin28b

The protein was prepared following the protocol of Yeom et al.(Yeom et al., 2011).

**Single molecule experimental setup**

Single molecule FRET experiments were performed with an inverted microscope (IX73, Olympus) with prism-based total internal reflection. Excitation of the donor dye Cy3 is done by illuminating with a 532nm diode laser (Compass 215M/50mW, Coherent). A 60X water immersion objective (UPLSAPO60XW, Olympus) was used for collection of photons from the Cy3 and Cy5 dyes on the surface, after which a 532 nm long pass filter (LDP01-532RU-25, Semrock) blocks the excitation light. A dichroic mirror (635 dcxr, Chroma) separates the fluorescence signal which is then projected onto an EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology). All experiments were performed at an exposure time of 0.1 s at room temperature (22 ± 0.1°C)

**Fluorescent dye labeling of nucleic acid constructs**

All DNA constructs were ordered from ELLA Biotech. Nucleic acid constructs that have an internal amino modification were labeled with fluorescent dyes based on the CSHL protocol (Joo and Ha, 2012). 1 µL of 1 mM of DNA/RNA dissolved in MilliQ H20 is added to 5 µL labeling buffer of (freshly prepared) sodium bicarbonate (84 mg/10mL, pH 8.5). 1 µL of 20 mM dye (1 mg in 56 µL DMSO) is added and incubated overnight at 4°C in the dark, followed by washing and ethanol precipitation. Concentration of nucleic acid and labeling efficiency was determined with a Nanodrop spectrophotometer.

**Single molecule chamber preparation**

Quartz slides were coated with a polyethylene-glycol through the use of amino-silane chemistry. This is followed by assembly of microfluidic chambers with the use of double sided scotchtape. For a detailed protocol, we refer to (Chandradoss et al., 2014). Further improvement of surface quality occurs through 15 min incubation of T50 and 5% Tween20 (Cai and Wind, 2016) after which the channel is rinsed with 100 µL T50 buffer. Streptavidin (5 mg/mL) was diluted in T50 to 0.1 mg/mL. 50 µL of this solution is then flowed inside the chamber. This is followed by incubation for 1 min followed by rinsing with approximately 10-fold the volume of the chamber with T50 (10 mM Tris-HCl [pH 8.0], 50 mM NaCl). 100 pM of DNA/RNA target with biotin construct is then flushed in the chamber, followed by 1 min incubation. This is followed subsequently by rinsing with T50. The chamber is subsequently flushed with *Cb*Ago buffer, containing 50 mM Tris-HCl [pH 8.0], 1 mM

Trolox, 1 mM MnCl2 and 100 mM NaCl. Guide-loading of apo-*Cb*Ago occurs by incubation of the protein (10 nM) with 1 nM guide construct in a buffer containing 50 mM Tris-HCl [pH 8.0], 1 mM Trolox, 1 mM MnCl2, 100 mM NaCl and 0.8% glucose at 37°C for 30 min. Following incubation, glucose oxidase and catalase is added (0.1 mg/mL glucose oxidase) after which the sample is flushed in the microfluidic chamber containing the DNA targets.

### Lin28 assay

Immobilization of Lin28b occurred in the following way: 50 µl of streptavidin (0.1 mg/mL) in T50 is flowed inside the chamber and incubated for 1 minute. After this, the chamber is rinsed with approximately 100 µL of T50. 1 µl of Anti-6X His tag® antibody (Biotin) diluted 100-fold in T50 and subsequently flowed inside the chamber. After 5 minutes, the chamber is rinsed with 100 µL of T50. Stock of Lin28b (100 µM) is diluted to 100 nM and incubated with the target DNA (10 nM) and 10 mM $MgCl_2$ for 5 minutes, after which the solution is flushed inside the chamber, followed by incubation of 5 minutes. Lastly, the *Cb*Ago buffer is flushed inside the chamber. Guide-loading of apo-*Cb*Ago occurs in the same way as described above (**Single molecule chamber preparation**) after which the *Cb*Ago:siDNA complex is also flushed inside the chamber.



## Quantification and statistical analysis

### Data acquisition and analysis

Fluorescence signals are collected at 0.1-s exposure time unless otherwise specified. For 7-nt target separation, 30-ms exposure time is used. Time traces were subsequently extracted through IDL software using a custom script. Prior to data collection, the location of targets (Cy5 labeled) are found by illuminating the sample with the 637nm laser. Through a mapping file, it subsequently collects the individual intensity hotspots in both the donor and acceptor channel and pairs them up through the mapping file, after which the traces are extracted. During the acquisition of the movie, the green laser is used. Only at the end, the red laser is turned on once more to check for photobleaching of the red dye. Traces containing the fluorescence intensity from the donor and acceptor signal are manually pre-selected occurs through the use of MATLAB (Mathworks), disregarding artefacts caused by non-specific binding, additional binding to neighboring regions and photobleaching.

## Determination of dissociation rate

Binding of Argonaute complex to a single target results in a sudden increase of acceptor signal. The length of these interactions was quantified through a custom script in MATLAB 2015b based on a thresholding algorithm. Briefly, a histogram was made of every data trace, from which the lowest population was fitted with a Gaussian peak. The resulting mean value and standard deviation are then used to distinguish binding events. Intensities that exceeded five times the standard deviation of the baseline (noise) were recognized as a potential binding event. Events that were recognized as potential binding events were marked by the script with a marker for individual checking. Subsequently, the duration of these events were collected and plotted in Origin. Some interactions (at low ionic strength) (**Figure S1F**) were beyond the observation window of our setup. Hence only a lower limit of the dissociation time could be given. The collected dwell times were bootstrapped through custom code using standard bootstrap algorithms provided by MATLAB. From the resulting distribution, the 95% percentile confidence interval is taken as the error.

## HMM analysis

For assigning states to the FRET traces, a HMM software package is used from Van der Meent et al (Van De Meent et al., 2014), which can be found on their github repository (https://ebfret.github.io/**)**. Their software package is optimized for immobilized donor dye molecules on the surface. Here, we immobilize the acceptor dye molecule and hence when no molecule is present, the zero intensity signal in both channels results in large variations in FRET signal, which will result in false positives for the ebFRET software.

Increasing the donor signal and hence artificially creating an extra stable "zero FRET state" is adequate for our purposes, as the distinction between bound and unbound molecules is still made. For the analysis of shuttling traces from constructs where the sub-seed targets are located far away, the low FRET bound state becomes almost indistinguishable from donor only. Here, this method proves adequate in separating the two states (**Figure S3**).

After assigning states to the collected data, the dwell times for low FRET → high FRET and vice versa are extracted. The experimental data shows that there is only one rate-limiting step, in accordance with our theoretical analysis shown below. Using maximum likelihood estimation, the lifetime $\Delta\tau_{shuttle}$ of the single-exponential distribution $P(t) =$

$\tau_{shuttle}^{-1} e^{-t/\Delta\tau_{shuttle}}$ was extracted (the empirical average dwell time equals the ML estimator of $\Delta\tau_{shuttle}$). The 95% confidence interval was extracted using empirical bootstrapping.

# Kinetic modeling

## Binding times single-target including recapture events follow single-exponential distribution

We here build a kinetic model for the lateral diffusion by *Cb*Ago. Since Argonaute can in principle bind to any sequence along the DNA, we assume the binding sites to be located single nucleotide apart. Further, we shall here only explicitly take gliding/sliding into account, which is represented as an unbiased random walk with unit step length. If the protein is bound at the designed 3-nt sub-seed 'target' it can move to either of its neighbors at a rate of $k_{escape}$ or unbind from the ssDNA at a rate of $k_{ub}$. When bound elsewhere movement and dissociation are assumed to happen instantaneously. To derive in what manner these undetectable movements contribute to the observed dwell time distribution $(p_{bound}(\Delta t))$ we count all possible paths that the protein can take to dissociate following initial association to the sub-seed. Using its Laplace Transform, $P_{ub}(s) = \mathcal{L}\{p_{bound}(\Delta t)\}$ multi-step paths are represented by simple products of individual paths (rather than their convolutions).



$$P_{ub} = \sum_{m=0}^{\infty} \left( \frac{2k_{escape}}{s + 2k_{escape} + k_{ub}} P_{RC} \right)^m \frac{k_{ub} + 2k_{escape}(1 - P_{RC})}{s + 2k_{escape} + k_{ub}} = \frac{k_{ub} + 2k_{escape}(1 - P_{RC})}{s + k_{ub} + 2k_{escape}(1 - P_{RC})} \qquad (2)$$

Here, $P_{RC}$ denotes the probability to get recaptured at the target after having entered either flank. In other words, it is the probability to not dissociate from the flank. The sum on the left hand side of **Equation 2** therefore accounts for the protein escaping from and getting recaptured at the target an arbitrary amount of times. The two terms outside the sum represent the probability distributions to unbind from either the target directly or after having escaped one final time, respectively. Taking the inverse Laplace transform, we derive the observed dwell time distribution.

$$P_{bound}(\Delta t) = \mathcal{L}^{-1}\left\{\frac{k_{ub} + 2k_{escape}(1-P_{RC})}{s + k_{ub} + 2k_{escape}(1-P_{RC})}\right\} \equiv \hat{k}_{ub} e^{-\hat{k}_{ub}\Delta t} \quad (3)$$

Hence, despite the multitude of possible bound states along the DNA the protein can reside in, the observed distribution remains single-exponential. The apparent dissociation rate follows

$$\hat{k}_{ub} = k_{ub} + 2k_{escape}(1-P_{RC}) \quad (4)$$

Given the assay selects for events that get (re-)captured, the observed rate is greater than its intrinsic value.

**Shuttling rate due to gliding alone**

We seek to explain to what extend gliding contributes to the observed shuttling rate from the tandem-target assay. Given under the current experimental conditions about 13 shuttle events occur prior to unbinding, we shall ignore unbinding in the following analysis.

To get the distribution of shuttle times $p(\Delta t_{shuttle})$ we count all possible paths that lead the protein from one sub-seed to the other. If the two 3-nt nucleotide long sub-seeds are separated by $x_{poly-T}$ thymine nucleotides, the shuttle times are distributed as (setting $x_{\text{Target}} = x_{poly-T} + 3 \geq 3$):

$$P_{shuttle}(s, x_{\text{Target}}) = \sum_{m=0}^{\infty}\left(\frac{k_{escape}}{s + 2k_{escape}}\left(1 + P_R\left(x_{\text{Target}}\right)\right)\right)^m \frac{k_{escape}}{s + 2k_{escape}} P_S\left(x_{\text{Target}}\right) \quad (5)$$

The two terms within the sum represent recapture events at the initial trap via either the flanking sequence or the poly-T stretch in between the traps. Finally, the term outside the sum accounts for successful shuttling events.

Once the protein has left the initial trap $P_R$ and $P_S$ denote the distributions for either returning back to the initial trap or shuttling/making it across to the other.

$$P_R\left(x_{\text{Target}}\right)=\sum_{m=0}^{\infty}\left(\frac{1}{2}P_R\left(x_{\text{Target}}-1\right)\right)^m\frac{1}{2} \tag{6}$$

$$P_S\left(x_{\text{Target}}\right)=\sum_{m=0}^{\infty}\left(\frac{1}{2}P_R\left(x_{\text{Target}}-1\right)\right)^m\frac{1}{2}P_S\left(x_{\text{Target}}-1\right) \tag{7}$$

Inverting the Laplace transformation of **Equation 5** we obtain:

$$p(\Delta t_{shuttle})=\mathcal{L}^{-1}\left\{\frac{k_{escape}P_S(x_{\text{Target}})}{s+k_{escape}P_S(x_{\text{Target}})}\right\}=k_{escape}P_S(x_{\text{Target}})e^{-k_{escape}P_S(x_{\text{Target}})\Delta t_{shuttle}} \tag{8}$$

Hence, the observed dwell time distributions are indeed single exponential. In terms of the microscopic model the average time is set by the escape rate from the trap modified by the probability to make it across once outside of it $P_S(x_{\text{Target}})$. By solving the recursion relations shown in **Equations 6 and 7** this probability can be shown to be inversely proportional to the distance between the two traps:

$$P_S(x_{\text{Target}})=\frac{1}{x_{\text{Target}}} \tag{9}$$

Taken together, the observed shuttling rate and time equal:

$$\Delta\tau_{shuttle}=\frac{1}{k_{escape}P_S(x_{\text{Target}})}=\frac{x_{\text{Target}}}{k_{escape}} \tag{10}$$

$$k_{shuttle}(x_{\text{Target}})=\frac{k_{escape}}{x_{\text{Target}}} \tag{11}$$

Note that $x_{\text{Target}}\geq 3$, as the two sub-seeds cannot overlap. A fit of **Equation 11** to the experimental data for $x_{\text{Target}}$ of 11nt, 15nt, 18nt and 22nt in **figure 3A** (also see S.I. **Table 2**) were used to estimate the value of $k_{escape}$ for *Cb*Ago.

## Shuttling rate triple-target assay

For the assay using three sub-seed targets, we seek both the time needed to glide from any of the outer targets to the inner one ( $C \rightarrow B$ ) and the average time needed to glide along the opposite path $(B \rightarrow C)$. The former is equal to the time measured on the tandem target construct, denoted above as $\Delta\tau_{shuttle}$ (**Equation 10**, $\Delta\tau_{CB} = \Delta\tau_{shuttle}$ ). To calculate the effective rate of transitioning from target B to target C ( $k_{BC} = \Delta\tau_{BC}^{-1}$ ) we use a similar approach as above. In Laplace space, the dwell time distribution is given by:

$$P_{BC}(s, x_{\text{Target}}, k_{escape}) = \sum_{m=0}^{\infty} \left( \frac{2k_{escape}}{s + 2k_{escape}} P_R(x_{\text{Target}}) \right)^m \frac{k_{escape}}{s + 2k_{escape}} P_S(x_{\text{Target}}) \qquad (12)$$

The sum accounts for all paths that return to target B. Given the equal distances between all targets on the construct the probability to not make it across to either A or C are equal, which gives rise to the factor of two. The factor outside the sum accounts for the fact that the protein must eventually leave B and make it across to C. Using the same technique as shown above, the average time to make it from B to C solely through gliding equals:

$$\Delta\tau_{BC}(x_{\text{Target}}) = \frac{x_{\text{Target}}}{4k_{escape}} \rightarrow k_{BC} = \frac{4k_{escape}}{x_{\text{Target}}} = 4 \times k_{CB} \qquad (13)$$

## Predicting shuttle rate triple-target assay using bootstrapping

After using the data from the tandem target assay to estimate $k_{escape}$ using a fit to **Equation 11** there are no more free parameters remaining when predicting the data for the triple-target assay based on **Equation 13** (**Figure 3E**). We bootstrapped the dwell time distributions acquired using the original tandem target assay (distances of 11nt, 15nt, 18nt and 22nt). For each of the $10^5$ bootstrap samples we calculated new values for $\Delta\tau_{shuttle}$ and repeated the fit to **Equation 11** to obtain an error estimate in the fitted value of the escape rate. An error estimate for the experimental values of $\Delta\tau_{BC}$ and $\Delta\tau_{CB}$ were obtained using $10^5$ bootstrap samples of the dwell time distributions measured using the triple-target assay. Finally, **Figure 3E** shows the 95% confidence intervals. All analysis was performed with a custom code written in Python.

# Supplemental information

**Table S1 | Characteristics of target search mechanisms.** In the facilitated diffusion model, the protein binds DNA from solution. This is followed by lateral diffusion after which it reaches the target or unbinds into solution again and the cycle begins anew. The modes described below are those that complement 3D diffusion and facilitate the target search.

Lateral /1D diffusion

| Name | Characteristics |
|---|---|
| Sliding | Tight association with the nucleic acid at all times. It has been observed with dsDNA by restriction enzymes, repair enzymes and others (Blainey *et al.*, 2006, 2009; Gorman *et al.*, 2007). This movement could be simply one-dimensional, or coupled with the grooves of dsDNA, making a rotational movement around the helical axis. |
| Hopping | This mode is characterized by microdissociations (Berg, Winter and von Hippel, 1981) which are correlated along the contour of the nucleic acid strand. The idea is that counterions condense during unbinding of the enzyme. Experimental evidence of hopping resides mostly in a decreased time spent on non-specific nucleic acids, hence a higher diffusion coefficient is observed. |
| Gliding | For CbAgo, we have observed a mode that is neither fully characterized by sliding or by hopping as the counterions are not completely condensed during unbinding and no tight association is present throughout the search. |
| (Intersegmental) Jumping | Similar to hopping, this movement results in uncorrelated rebinding to a strand. Allowing the protein to access segments that are sequentially distant speeds up enhances the target search rate (van den Broek *et al.*, 2008; Lomholt *et al.*, 2009). |
| Intersegmental transfer | Proteins that have multiple binding sites are able to bind momentarily to two strands before letting one go. In such a hand-over process, proteins are able to cover larger distances. An example would be transcription factors (Ruusala and Crothers, 1992; Doucleff and Clore, 2008) |



**Table S2 | Dwell times of different two target DNA constructs for several distances.** The upper bound and lower bound are estimated through 20000 bootstraps of the acquired dwell times. Related to **Figure 3**, **S3**.

| Target distance | Lifetime | Lowerbound lifetime | Upperbound lifetime | Shuttling rate | Lower bound shuttling rate | Upper bound shuttling rate |
|---|---|---|---|---|---|---|
| 11 | 0.47 | 0.46 | 0.49 | 2.11 | 2.04 | 2.19 |
| 15 | 0.83 | 0.81 | 0.87 | 1.19 | 1.15 | 1.24 |
| 18 | 1.17 | 1.11 | 1.24 | 0.85 | 0.81 | 0.90 |
| 22 | 1.79 | 1.74 | 1.86 | 0.56 | 0.54 | 0.57 |
| 29 | 1.36 | 1.30 | 1.42 | 0.73 | 0.70 | 0.77 |
| 36 | 1.19 | 1.16 | 1.23 | 0.84 | 0.81 | 0.86 |
| 50 | 1.52 | 1.46 | 1.57 | 0.66 | 0.64 | 0.68 |
| 64 | 1.65 | 1.59 | 1.71 | 0.61 | 0.59 | 0.63 |
| 92 | 1.94 | 1.85 | 2.02 | 0.52 | 0.49 | 0.54 |
| 120 | 2.11 | 2.03 | 2.19 | 0.47 | 0.46 | 0.49 |

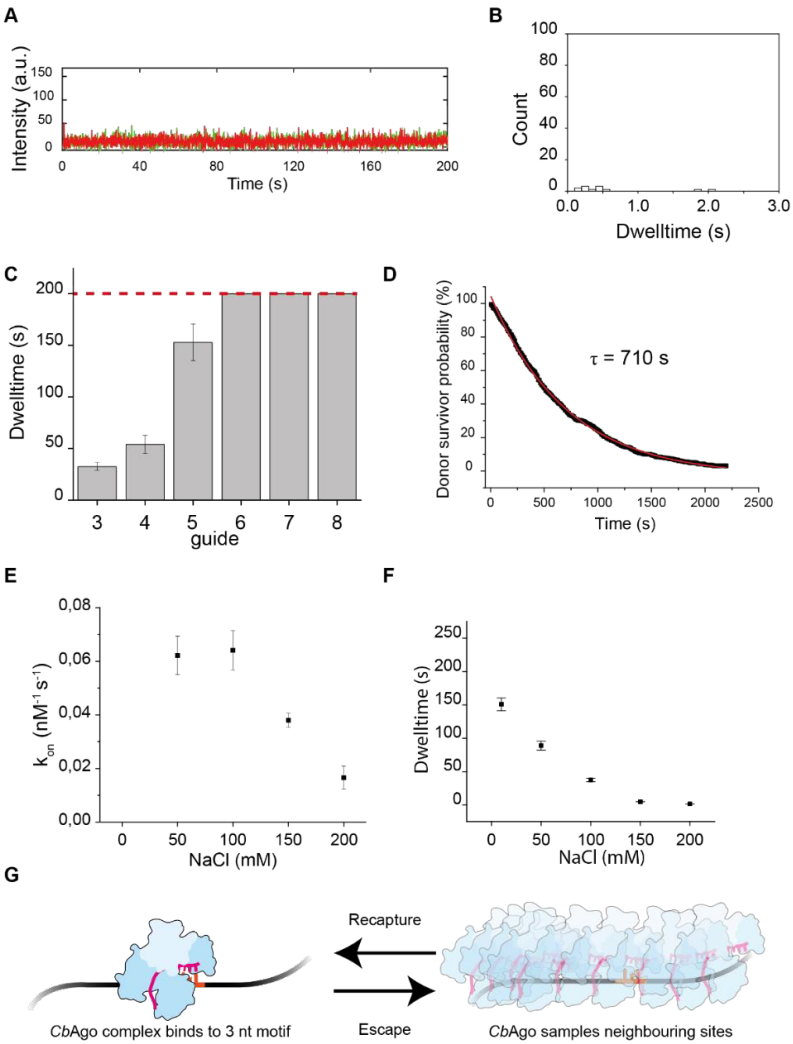



**Figure S1 | Single molecule interactions of *Cb*Ago:siDNA (2-4 nt) at different conditions. Related to Figure 1.** **A |** Representative trace single-molecule interaction of *Cb*Ago with dsDNA DNA target immobilized on the surface (~300 per FoV). **B |** Dwell time distribution of *Cb*Ago-guide 3-dsDNA target interactions. Number of molecules: 540. **C |** Average dwell time of protein bound to target versus guide length. **D |** Survival plot of donor only (Cy3) constructs in standard experimental conditions (100 mM NaCl). Mean donor bleaching time was obtained by single exponential fitting to survival probability plot. **E |** Binding rate for different salt concentrations, acquired by taking the time in between binding events. **F |** Salt titration of *Cb*Ago and a single-stranded single target DNA construct (N=3) at 10, 50, 100, 150 and 200 mM NaCl concentration. Total measurement time = 250 s. Error bars are indicating the 95% percentile of 20.000 empirical bootstraps of the mean dwell time. **G |** Schematic image indicating the dynamic escape and recapture events of *Cb*Ago

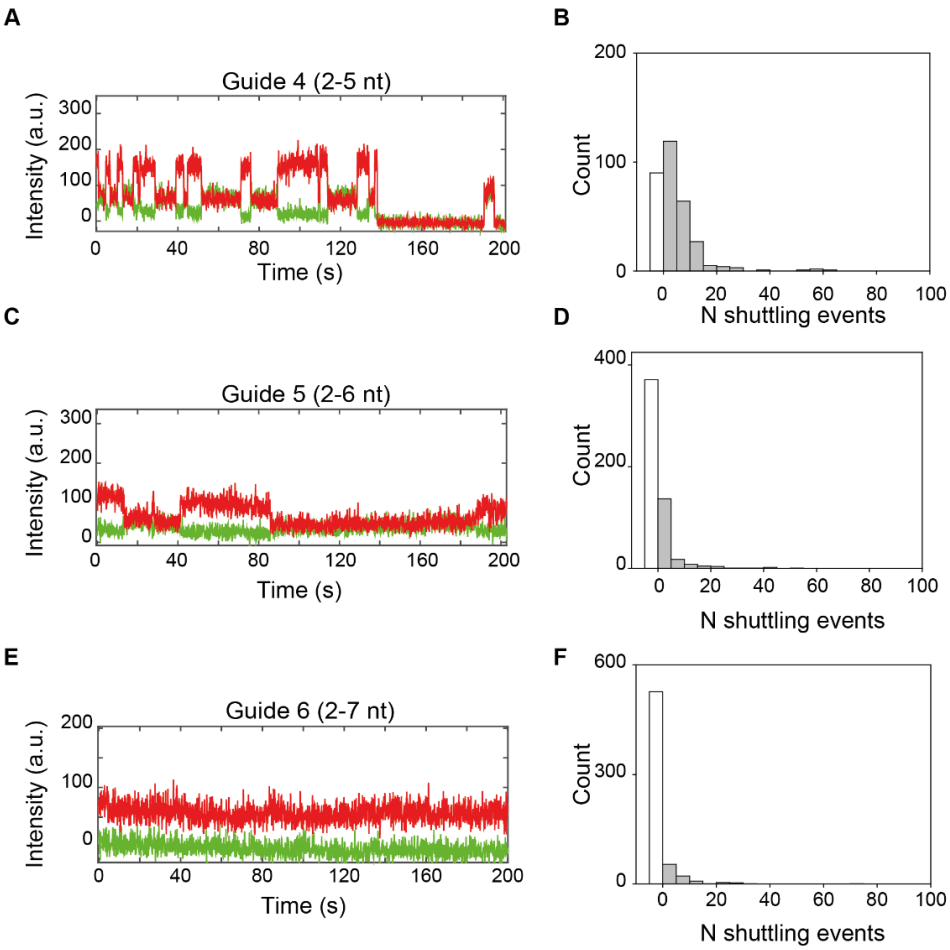



**Figure S2 | Single-molecule interactions of *Cb*Ago with guide 4, 5, 6 and tandem target (22 nt separation). Related to Figure 1. A |** Representative trace of binding events by *Cb*Ago with guide 4 (nt 2-5). Duration of observation 200 s. **B |** Shuttling event distribution for guide 4 (nt 2-5). Bin size = 10. The white bar represents binding (no shuttling) events followed by dissociation. N=317. **C |** Representative trace of binding events by *Cb*Ago with guide 5 (2-6). **D |** Shuttling event distribution for guide 5 (2-6 nt). Bin size = 10. The white bar represents events that consists of single molecule binding followed by dissociation. N = 550. **E |** Representative trace of guide 6 (2-7 nt) interaction. **F |** Shuttling event distribution for guide 6. The white bar represents events that consists of single molecule binding followed by dissociation. N = 621.

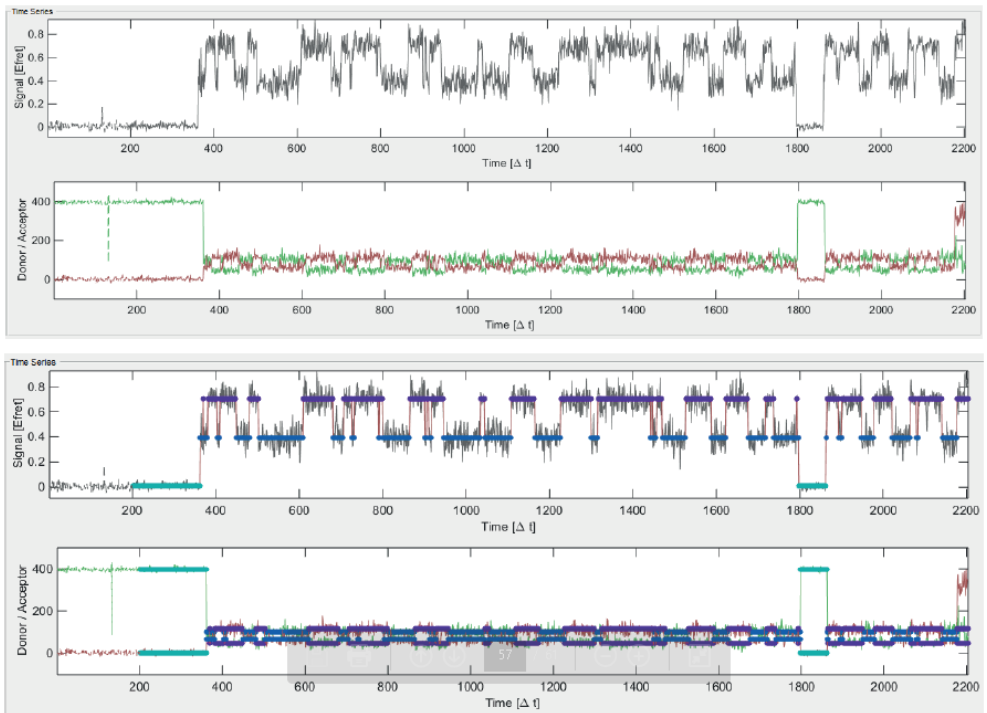



**Figure S3 | Example of HMM software applied to data trace. Related to Figure 2 and Figure 3 and Star Methods. Top |** An example shuttling trace of *Cb*Ago in the user interface of ebFRET. The donor and acceptor intensities plotted versus time. The donor intensity is enhanced artificially, resulting in an extra zero FRET state (upper subfigure). **Bottom |** The donor, acceptor and FRET intensities overlaid with states resulting from the Hidden Markov Modeling. The HMM analysis program recognizes the unbound state as an extra state (light blue), while low FRET and high FRET are respectively assigned dark blue and purple.

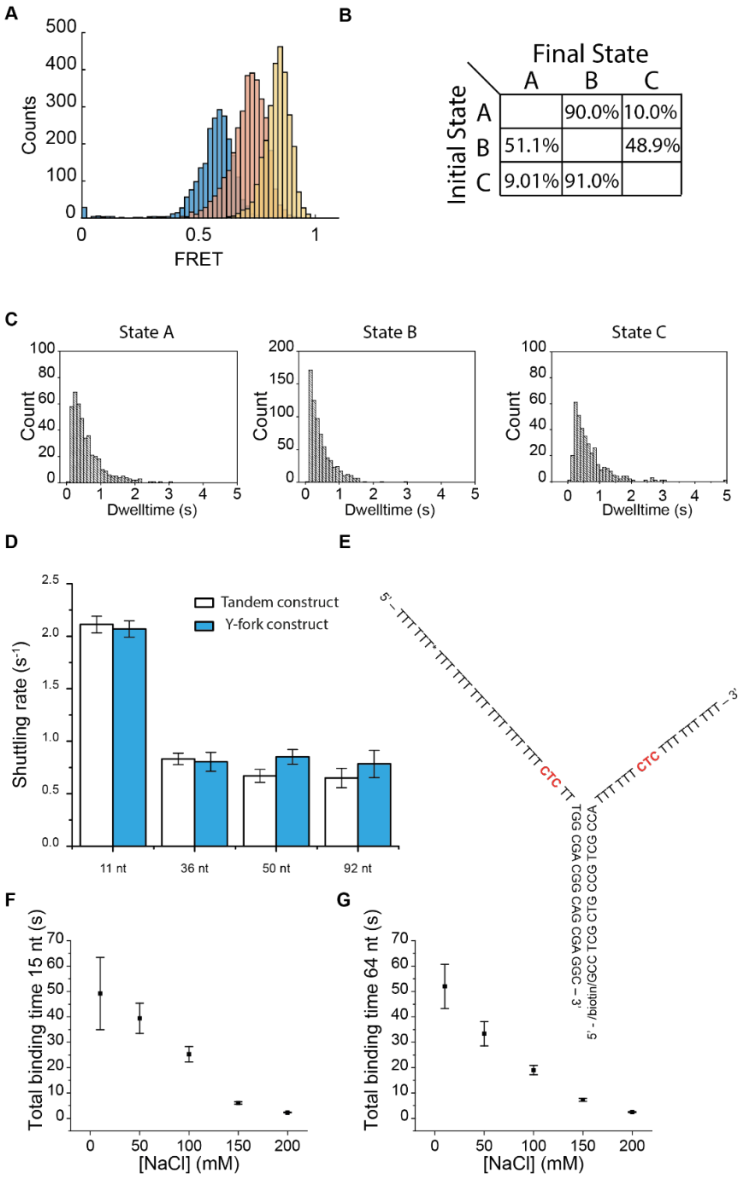



**Figure S4 | Three target assay, Y-fork assay and total binding time of salt titration of 15 nt and 64 nt target separation. Related to Figure 3 & Figure 4. A |** FRET histogram of three-target assay. N = 168 molecules**. B |** Dwell time histograms for respectively the low FRET, mid FRET and high FRET state of the three target assay. **C |** Shuttling rate of Y-fork constructs (blue) compared to tandem target assay (white). The error bars are given by the 95% percentile of 20.000 bootstrapped mean dwell times. **D |** Schematic of the Y-fork assay for 11 nt. The sub-seed targets are marked in red. **E |** Total binding time of target 15 vs salt concentration for 10, 50, 100, 150 and 200 mM [NaCl]. **F |** Total binding time of target 64 vs salt concentration for 10, 50, 100, 150 and 200 mM [NaCl]

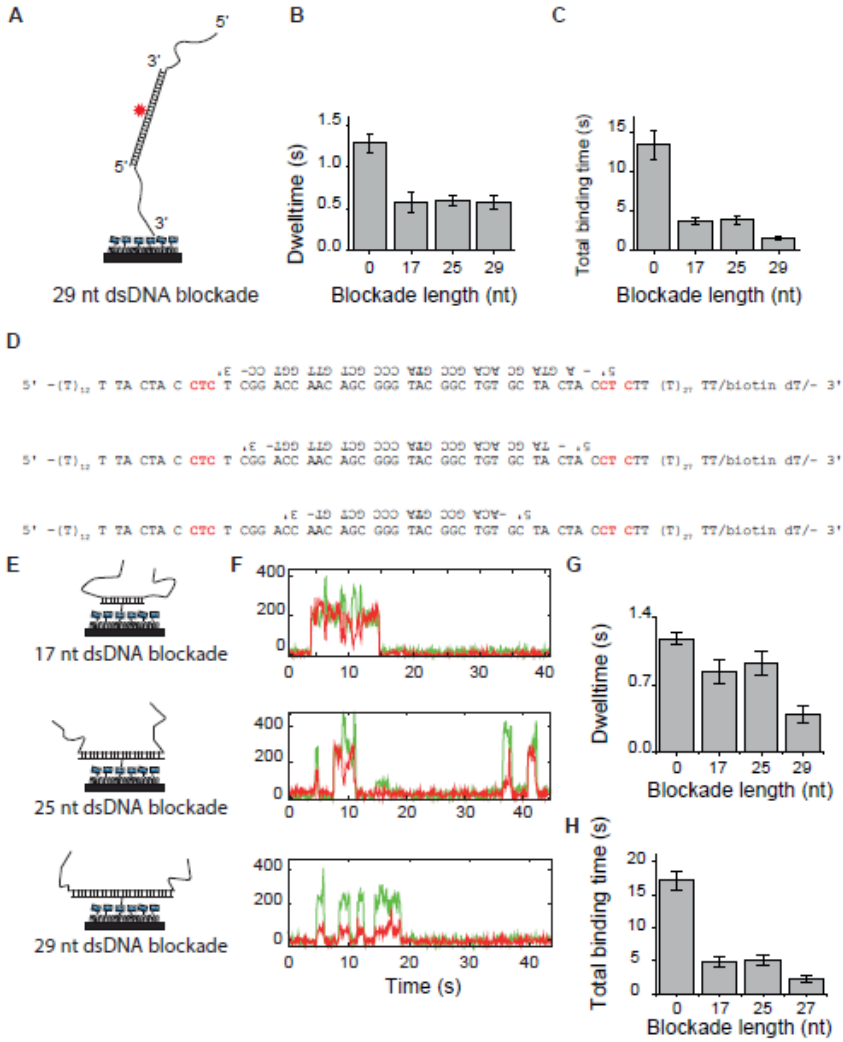



**Figure S5 | Interactions of *Cb*Ago with dsDNA between targets (related to Figure 4). A |** Schematic image of a dsDNA blockade in between targets. Analogous to the tandem construct, the dsDNA region is put in between two targets. The Cy5 dye is placed such that binding to one site will result in a higher FRET efficiency compared to binding to the other site. **B |** Dwell time per state for construct (A) with no dsDNA, 17 nt segment, 25 nt segment and 29 nt segment. Error bars are given by the 95% confidence interval of $10^5$ bootstrapped samples. **C |** Total binding time of above constructs. Error bars are given by the 95% confidence interval of $10^5$ bootstrapped samples. **D |** Sequence of tandem dsDNA constructs. The bold thymine is contains a fluorescent Cy5 dye. **E |** Schematic of the immobilization scheme of the control construct with dsDNA. **F |** Representative traces of *Cb*Ago interacting with constructs from (E). **G |** Individual dwell time plotted versus the length of the dsDNA blockade. Error bars are given by the 95% confidence interval of $10^5$ bootstrapped samples. **H |** Total binding time plotted versus the dsDNA blockade length. Error bars are given by the 95% confidence interval of $10^5$ bootstrapped samples.

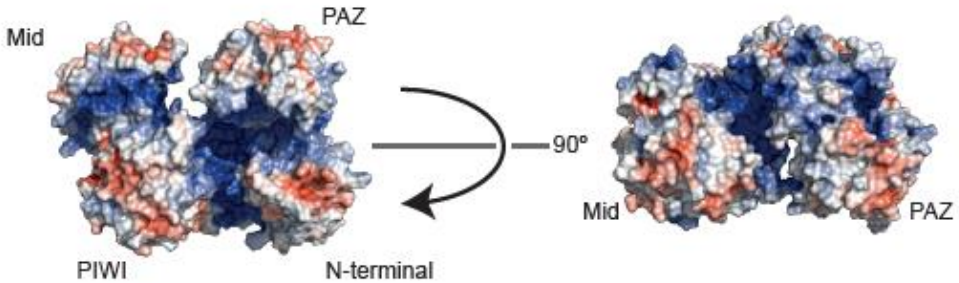



**Figure S6 | Coulombic surface coloring of *Clostridium butyricum* Argonaute (*Cb*Ago).** The crystal structure of *Cb*Ago (3.23 Å resolution) reveals the charge distribution. The cleft that contains the guide DNA and the target DNA is highly positively charged (blue). Crystal structure is taken from (Hegge et al., 2018b)

# Chapter 5

## *Thermus thermophilus* Argonaute repurposed for diagnostics

Jinzhao Song*, Jorrit W. Hegge*, Michael G. Mauk, Junman Chen, Neha Bhagwat, Jacob E. Till, Lotte T. Azink, Jing Peng, Moen Sen, Jazmine Mays, Erica Carpenter, John van der Oost and Haim H. Bau

*contributed equally

Adapted from –
'Highly specific enrichment of rare nucleic acids using *Thermus thermophilus* Argonaute'
Submitted and on Biorxiv

# Abstract

Rare nucleic acids in human plasma often contain essential diagnostic information but are difficult to detect due to sequence homology with the more abundant wild type nucleic acids. Here we describe a new enrichment method that utilizes the DNA-guided Argonaute nuclease from *Thermus thermophilus* (*Tt*Ago), to specifically cleave fully complementary wild type nucleic acids while sparing rare nucleic acids. Our method, named NAVIGATER (Nucleic Acid enrichment Via DNA-Guided Argonaute from *Thermus thermophilus*), enables us to greatly increase the fractions of rare nucleic acids with single nucleotide precision thereby enhancing the sensitivity of downstream detection methods such as, ddPCR, sequencing, and clamped enzymatic amplification. We demonstrate a 60-fold enrichment of the rare KRAS G12D allele in blood samples from pancreatic cancer patients and could detect rare KRAS, EGFR, and BRAF mutant alleles at allelic fractions as low as 0.01%. Ultimately, we combined NAVIGATER with LAMP (Loop mediated isothermal amplification) in a smartphone-connected cup system to enable it for field applications.

# Introduction

Detection of disease-related nucleic acid biomarkers in blood is increasingly used as a powerful and minimally-invasive instrument (liquid biopsy) for diagnostic, prognostic and treatment purposes (Crowley et al., 2013). However, detection of alleles of clinical interest is often challenged by their low concentration and sequence homology with the more abundant 'healthy' wild type (WT) nucleic acids. Recently, various tools have been developed for the specific detection of nucleic acids, among which methods that involve CRISPR-associated (CRISPR-Cas) nucleases (Aalipour et al., 2018; Chen et al., 2018; East-Seletsky et al., 2016; Gootenberg et al., 2017, 2018; Gu et al., 2016; Lee et al., 2017). Several CRISPR-Cas nucleases, which mainly attracted attention as genome editing tools (Barrangou and Doudna, 2016; Komor et al., 2017; Wu et al., 2018), were shown to exhibit a secondary nonspecific ribonuclease activity upon target recognition. This activity was repurposed to rapidly amplify a reporter signal, upon the specific recognition of target nucleic acids (Chen et al., 2018; East-Seletsky et al., 2016; Gootenberg et al., 2017, 2018; Gu et al., 2016). Besides, CRISPR-Cas nucleases were used to enrich oncogenic sequences by eliminating WT sequences (Aalipour et al., 2018; Lee et al., 2017). Despite spectacular progress, a limitation of using CRISPR-Cas for diagnostics relates to the requirement of a protospacer-adjacent motif (PAM), which is absent in many sequences of clinical interest.

Analogous to CRISPR-Cas, Argonaute (Ago) proteins are nucleic acid-guided endonucleases (Hegge et al., 2018a). However, Ago nucleases lack the PAM requirement and are therefore more versatile than CRISPR-Cas nucleases. Ago from the thermophilic bacterium *Thermus thermophilus* (*Tt*Ago) utilizes short 5′-phosphorylated single-stranded DNA guides to selectively cleave complementary target nucleic acids (RNA or DNA) (Swarts et al., 2014b, 2014a, Wang et al., 2008a, 2009). By optimizing reaction conditions to maximize the cleavage efficiency of *Tt*Ago on wild type (WT) alleles, while sparing rare alleles that differ from the WT alleles by a single nucleotide, we created a PAM-independent *Tt*Ago based enrichment method (**Figure 1**), termed NAVIGATER (Nucleic Acid enrichment Via DNA-Guided Argonaute from *Thermus thermophilus*). This method enables us to increase the rare allelic fractions of several oncogenic mutants, such as KRAS, EGFR, and BRAF. We demonstrate that NAVIGATER significantly improves the sensitivity of several downstream detection methods such as droplet digital PCR (ddPCR) (Taly et al., 2013), Peptide Nucleic Acid-Mediated PCR (PNA-PCR) (Choi et al., 2010), PNA-Loop Mediated Isothermal

Amplification (LAMP) (Choi et al., 2010), Xeno nucleic Acid clamp PCR (XNA-PCR) (Powell and Zhang, 2016), and Sanger sequencing.

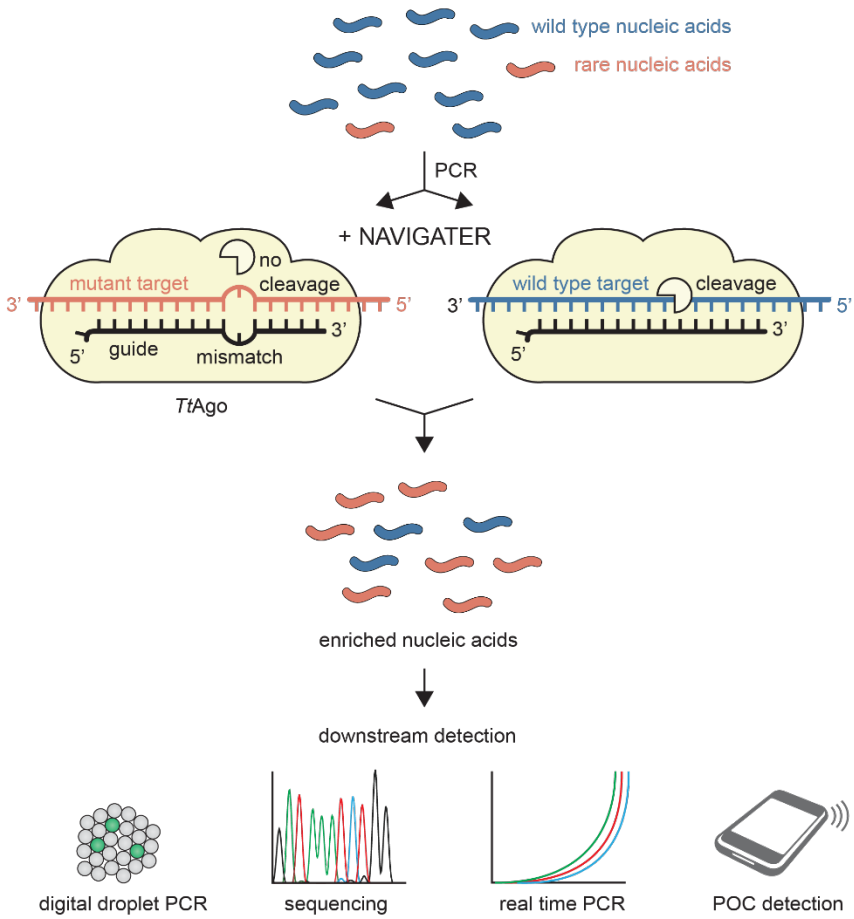



**Figure 1 | Schematic overview of NAVIGATER.** Programmed with a DNA guide, Argonaute from *Thermus thermophilus* (*Tt*Ago) specifically eliminates (cleaves) the abundant and fully complementary WT sequences (black) while sparing the rare nucleic acids (blue) that contain a single nucleotide mutation. This results in an enrichment of the disease related nucleic acids, making them easier to detect with downstream detection methods.

# Results

## Optimization of the endonucleolytic activity of *Tt*Ago

For rapid and inexpensive genotyping of rare mutant alleles (MAs) we envision using NAVIGATER in combination with the enzymatic amplification of rare nucleic acids, in either a single-stage or a two-stage process. LAMP (Loop-mediated isothermal amplification) amplification is of particular interest as it does not require temperature cycling and can therefore be implemented with simple instrumentation in a resource poor setting (Song et al., 2018). We tested the endonucleolytic activity of *Tt*Ago in three different LAMP buffers and compared it to the activity in the previously described buffer S (**Table S1**) (Swarts et al., 2014b).

We incubated a single stranded DNA and RNA target fragment (100 nt) of the human KRAS gene with *Tt*Ago complexed with a 16 nt guide. This guide is fully complementary to the wild type (WT) KRAS fragment but contains a single nucleotide mismatch at guide position 12 (g12) with the KRAS-G12D target. Comparing the endonucleolytic activity of *Tt*Ago at 80°C in different buffers reveals that in buffer 3 the activity of *Tt*Ago is high (nearly 100%) on fully complementary WT DNA and RNA targets, and low (<1%) on the MAs with a single-nucleotide mismatch at position 12 (**Figure 2a, c**).

Next, we set out to reveal why buffer 3 outperforms the other buffers tested. Compared to the other buffers, buffer 3 contains betaine, dNTPs, and a higher concentration of magnesium (8 mM vs 2 mM). To individually examine the effect of each of these compounds on the activity of *Tt*Ago, we added them separately to buffer 2. Previous studies already showed the importance of divalent cations, such as $Mg^{2+}$, on the endonucleolytic activity of *Tt*Ago (Sheng et al., 2014; Swarts et al., 2014b; Wang et al., 2008b, 2009). In line with these studies, we see that the addition of 6 mM magnesium increases the activity of *Tt*Ago on DNA targets and to a lesser extent on RNA targets (**Figure 2b, d**). Also addition of betaine increases the cleavage efficiency of *Tt*Ago on WT DNA and RNA (**Figure 2b, d**). Betaine is known for its thermal stabilizing effects on enzymes and for its ability to dissolve GC-rich DNA structures during DNA amplification (Adamczak et al., 2018; Henke et al., 1997).

Surprisingly, the biggest increase in cleavage efficiency on both DNA and RNA targets was observed after the addition of dNTPs (**Figure 2d, S1**). To make sure that this beneficial effect is not unique to KRAS targets, we also tested the effect of dNTPs on the cleavage efficiency

on EFGR target sequences (**Figure S2a**). In the absence of dNTPs, the cleavage efficiency on EGFR WT RNA is 45% while in the presence of 1.4 mM dATP, dTTP or dCTP, this increased to nearly 100%. Remarkably the addition of 1.4 mM dGTP did not affect cleavage efficiency. Among NTPs, only CTP increased the cleavage efficiency on WT RNA (**Figure S2b**). Although the molecular basis of this phenomenon remains elusive, it appears that the combination of sugar groups and nitrogenous bases of dNTPs stimulate the activity of *Tt*Ago.

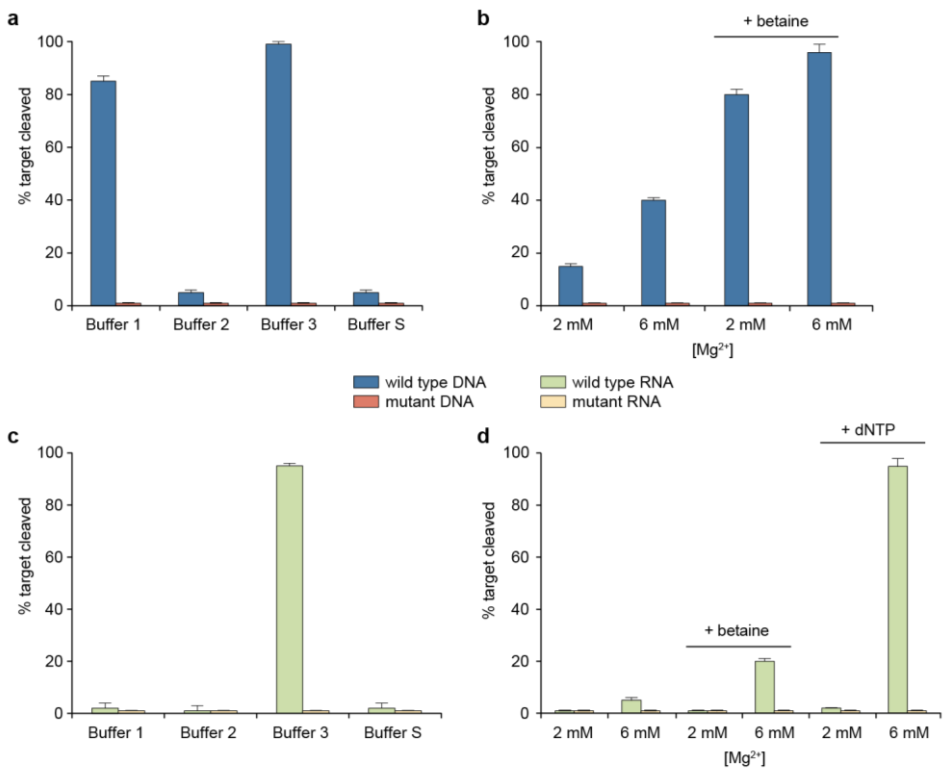



**Figure 2 | Cleavage efficiency on KRAS WT and KRAS G12D DNA and RNA in buffers 1, 2, 3, and S (Table 1) at 80°C. a |** The cleavage efficiency of *Tt*Ago on DNA targets in various buffers. **b |** Cleavage efficiency in buffer 2 supplemented with $Mg^{2+}$ and betaine. **c |** Cleavage efficiency on RNA targets in different buffers. **d |** Cleavage efficiency in buffer 2 supplemented with $Mg^{2+}$, betaine or dNTPs. All experiments were carried out with KRAS Sense (S) strand and a 16 nt KRAS-S guide with a pair mismatch at position 12 (MP12). Cleavage products were resolved on polyacrylamide gels and the cleavage efficiencies were calculated according to $I_C/(I_C+I_{UC})$, where $I_C$ and $I_{UC}$ are, respectively, the band intensities of cleaved and uncleaved alleles (N=3).

Next we tested the optimal pH and found that *Tt*Ago is most active in the pH range between pH 8-9 (**Figure S2c**). So combined, buffer 3 provides the best conditions for effective *Tt*Ago cleavage of targeted WT alleles likely due to the presence of betaine, dNTPs, and 8 mM [$Mg^{2+}$]. Notably, *Tt*Ago retained its single-nucleotide specificity in the presence of the above additives as none of the mutant allele (MA) targets, with a mutation at position 12, were cleaved.

## Single base pair-mismatch discrimination

Previous studies revealed single or dinucleotide mismatches between the guide and substrate strand can, depending on the position, severely reduce the cleavage efficiency of Argonaute proteins on target strands (Doxzen et al., 2017; Hunt et al., 2018; Kaya et al., 2016). As we aim to exploit the ability of *Tt*Ago to mediated efficient cleavage of WT alleles, while sparing rare mutant alleles with a single nucleotide mutation, we determined the location of the guide with the lowest tolerance to mutations. Hence, we screened several guide DNAs (gDNAs) of *Tt*Ago with single nucleotide mismatches at different positions along the KRAS G12D target allele (**Figure 3a, b**). Under the tested conditions, *Tt*Ago mediated cleavage of the KRAS G12D target strand was completely abolished by mutations at MP7 and MP11-13. To test whether this also extends towards other target sequences, we repeated the same screening for the EGFR and BRAF mutant alleles (**Figure 3c**). We define the differences in cleavage efficiency between WT and mutant targets as the discrimination efficiency (DE). Although in general the mismatches MP7 and MP9-MP13 that are located around the cleavage site (g10/g11), yielded the greatest discrimination (DE>80%) (**Figure 3d**), the optimal MP appears to depend on the sequence of the target. Cleavage of RNA was less tolerant to mutations than cleavage of DNA. Single mismatches between MP4-MP11 nearly completely prevented RNA cleavage (**Figure S3e, S4**).

Despite being loaded with guides that were fully complementarity to the target, WT sequences were cleaved by *Tt*Ago at variable efficiencies (**Figure S3c, e**). This suggests that besides the different buffer components, the activity of *Tt*Ago additionally depends on the sequence of the guide and target which might affect the conformation of the ternary *Tt*Ago-gDNA-DNA and *Tt*Ago-gDNA/RNA complexes (Sheng et al., 2014).

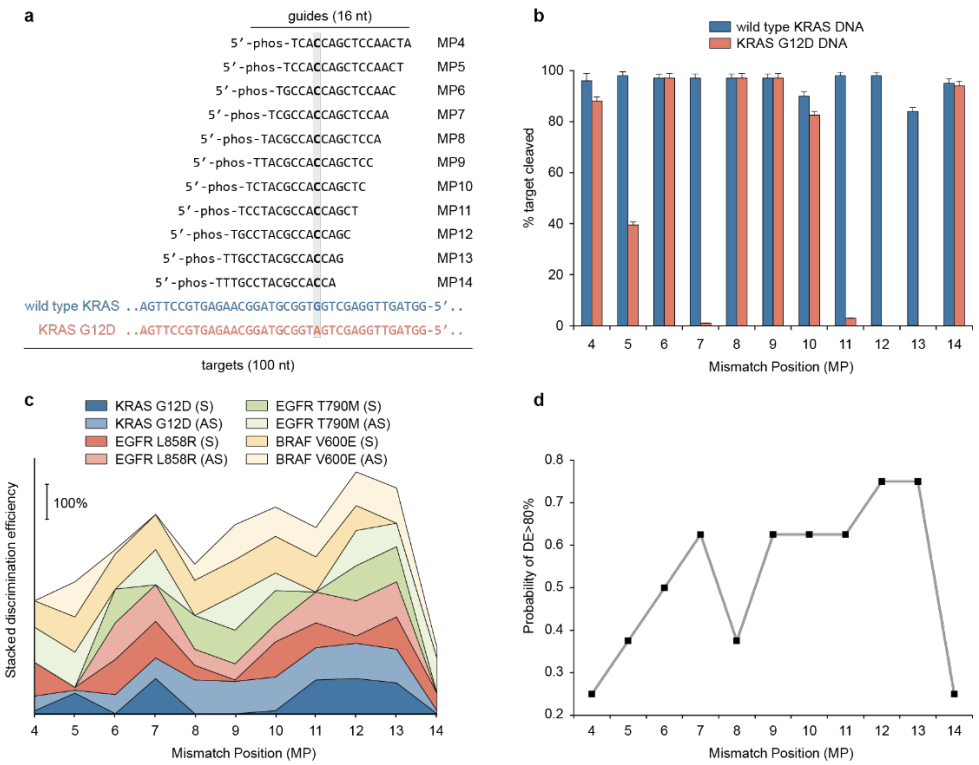



**Figure 3 | The discrimination efficiency of NAVIGATER depends on the position of the mismatch (MP). a** | Overview of the KRAS guide and target sequences used. **b** | Cleavage efficiencies are plotted for each guide. **c** | The discrimination efficiencies (DE) (difference between cleaving efficiency of WT and mutant allele) are stacked for each guide of each allele (*KRAS* G12D, *EGFR* L858R, *EGFR* T790M, and *BRAF* V600E). **d** | For each MP the probability of a DE >80% is plotted. All experiments were carried out with short guides (15/16 nt) in Buffer 3 at 80°C. *Tt*Ago: guide: target =1: 0.2: 0.2. N=3. The notation MP-X indicates the position (X) of the mismatch (MP) between the guide and target counted from 5' end of the guide.

## *Tt*Ago cleaves most specifically with short guides (15/16nt)

Heterologously expressed *Tt*Ago is typically purified with DNA guides ranging in length from 13 to 25 nt (Swarts et al., 2014b). *In vitro*, *Tt*Ago has been reported to be active with ssDNA guides ranging in length from 7 to 36 nt (Wang et al., 2008b). As little is known about the effect of guide length on the discrimination efficiency (DE) of *Tt*Ago, we examine the effect of guide length on DE in our *in vitro* assay. *Tt*Ago efficiently cleaves WT KRAS with complementary guides, ranging in length from 16 to 21 nt at both 70°C and 75°C (**Figure 4a**). Guides of 17-21 nt length with a single nucleotide mismatch at position 12 cleave MAs

at 75°C but not at 70°C (**Figure 4a**). Cleavage of MA at 75°C is completely abolished with a short 16 nt guide (**Figure 4a**). Apparently, *Tt*Ago with shorter guides form a less stable ternary complex with off-targets than with longer guides thus preventing undesired cleavage.

In contrast to MA DNA, the increase in temperature did not increase undesired cleavage of MA RNA (**Figure 4a**), likely due to differences in the effects of ssDNA and ssRNA on enzyme conformation. When operating with a short 16 nt guide, single MP12 mismatch, and LAMP Buffer 3, *Tt*Ago efficiently cleaves both KRAS RNA and DNA targets while avoiding cleavage of the KRAS G12D mutant allele between 66°C and 86°C (**Figure 4c**), providing the nucleotide specificity that is crucial for our enrichment method.

### *Tt*Ago efficiently cleaves targeted dsDNA

Guide-free (Apo-)*Tt*Ago can degrade dsDNA via a mechanisms termed 'chopping', which it exhibits to autonomously generate and selectively load functional DNA guides (Swarts et al., 2017c). However, this is a slow process that takes place only when target DNA is rich in AT-content (<17% GC) (Swarts et al., 2014b), suggesting that *Tt*Ago lacks helicase activity and depends on dsDNA thermal breathing to enable chopping (Hegge et al., 2018a; Swarts et al., 2014b). In our assays *Tt*Ago is saturated with DNA guides to suppress this Apo activity on dsDNA. The ability of *Tt*Ago to operate at high temperatures provides NAVIGATER with an advantage since *in vitro* dsDNA unwinds as the incubation temperature increases.

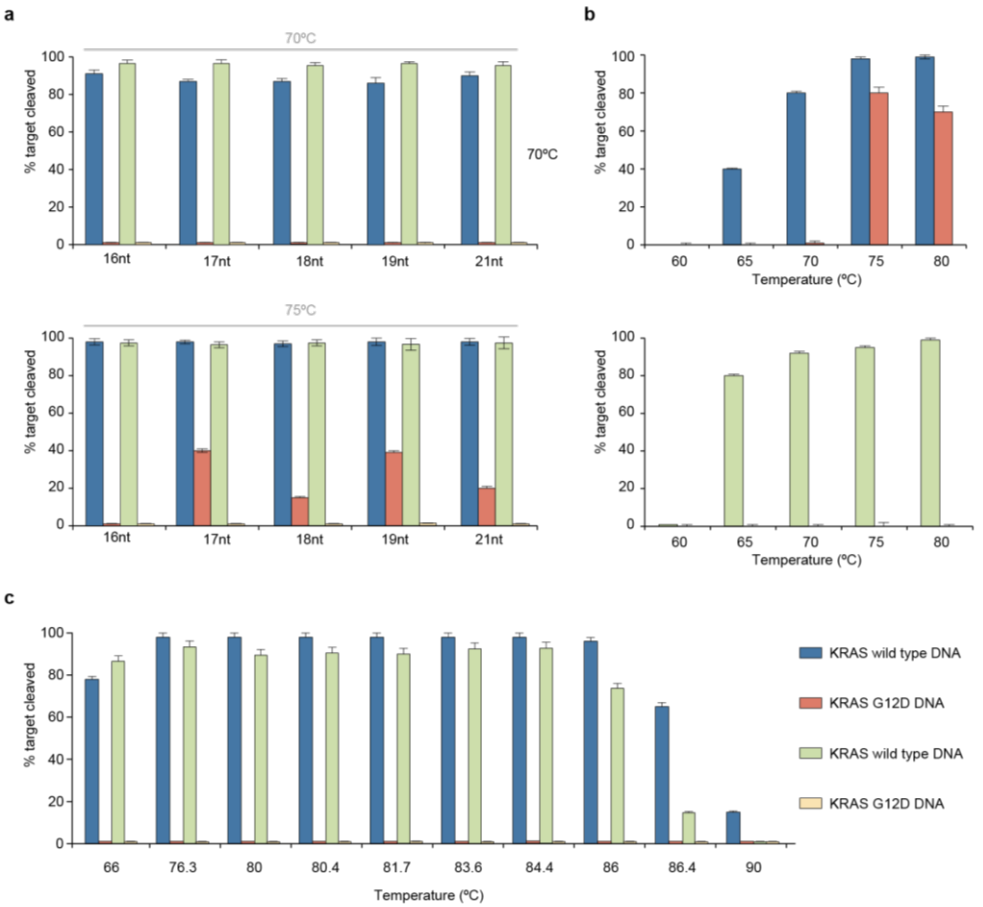



**Figure 4 | *Tt*Ago cleaves more specifically with short guides. a |** Effect of guide length and incubation temperature (70°C or 75°C) on the cleavage efficiencies of *Tt*Ago. *Tt*Ago in complex with the MP12 guide was used to target either KRAS (sense) WT or G12D. **b** | Effect of temperature on *Tt*Ago mediated cleavage of the KRAS WT or mutant alleles **c** | The effect of temperature on *Tt*Ago in complex with a 16nt MP12 guide targeting DNA or RNA. *Tt*Ago: guide: target ratio was 1:0.2:0.2 (N=3).

We determined the optimum temperature at which *Tt*Ago, saturated with guides, efficiently cleaves dsKRAS WT while sparing MA. The estimated melting temperature of the 100 bp dsKRAS in buffer 3 is 79.7°C (IDT-OligoAnalyzer). Consistent with this estimate, very little cleavage takes place at temperatures under 80°C, while dsDNA is cleaved efficiently by *Tt*Ago at temperatures above 80°C (**Figure 5a**). Cleavage efficiencies increase as the

incubation time increases and saturate after about one hour (**Figure 5c, d**). Longer incubation times are undesirable as it leads to more cleavage of the MAs.

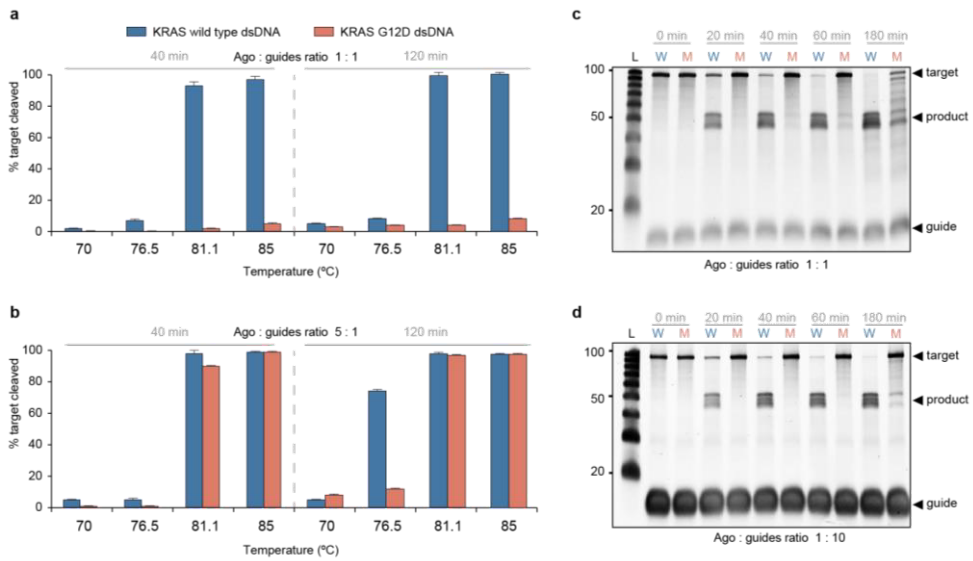



**Figure 5 | An excess of guide DNA is necessary for high discrimination efficiency on double stranded targets.** *Tt*Ago cleavage efficiencies on dsDNA *KRAS* and *KRAS* G12D targets using a *Tt*Ago:S-guide:AS-guide ratio of **a** | 1:1:1 or **b** | 1:0.2:0.2. **c** | *Tt*Ago mediated cleavage of dsDNA *KRAS* and *KRAS* G12D targets was followed in a time course at 83ºC. The products were resolved on a urea polyacrylamide gel. The *Tt*Ago:S-guide:AS-guide ratio was 1:1:1 and **d** | 1:10:10. All experiments were carried out in buffer 3 with *KRAS*-S (16nt)-MP12 and *KRAS*-AS (15nt)-MP13 guides. N=3.



## Guide saturation is necessary to avoid off-target cleavage

In case *Tt*Ago is not fully loaded with guides, undesired cleavage of dsMA occurs (**Figure 5b**). This off-target cleavage becomes more pronounced as the incubation time increases (**Figure 5b**) and can partially be attributed to the chopping ability of *Tt*Ago (Swarts et al., 2017c). Since DNA guides are tightly loaded into *Tt*Ago, an excess of DNA guides reduces undesired chopping. When guide concentrations exceed *Tt*Ago concentration, no apparent off-target cleavage takes place (**Figure 5c, d**). So for an optimal discrimination between dsWT and dsMA, it is necessary to saturate *Tt*Ago with guides and it is desired to incubate the assay at temperatures exceeding the melting temperature of the target for less than an hour.

## NAVIGATER improves the sensitivity of downstream rare allele detection

Next, we tested whether singleplex and multiplex NAVIGATER can increase the sensitivity of downstream the mutation detection methods: gel electrophoresis, ddPCR (Taly et al., 2013), PNA-PCR (Choi et al., 2010), PNA-LAMP (Tatsumi et al., 2008), XNA-PCR (Powell and Zhang, 2016) and Sanger sequencing. Moreover, we demonstrate NAVIGATER's clinical utility, by enriching rare mutant alleles in blood samples of pancreatic cancer patients, which were first analyzed with the standard ddPCR protocol (**Table S2**). These samples were pre-amplified with PCR to increase the total amount of KRAS WT and MA prior to the enrichment with NAVIGATER.

*Gel electrophoresis*

First, we resolved the enriched samples of several pancreatic cancer patients (**Table S2**) on a polyacrylamide gel (**Figure S6**). The 80 bp bands of KRAS are dark in the absence of NAVIGATER (control). After 40 minutes of NAVIGATER these bands faded, indicating a reduction of KRAS WT alleles. After 2 hours all 80 bp bands (except that of patient P6) faded, suggesting that most WT alleles were cleaved. The presence of an 80 bp band in the lane P6 can be explained by the relatively high (20%) MA fraction that was not susceptible to cleavage. We also PCR amplified products from a 2-hour NAVIGATER treatment, and subjected the amplicons to a second round of NAVIGATER (2h). The lanes P3, P4, and P6 display slightly darker bands than lanes P1, P2 and P5, indicating the presence of MAs in samples P3, P4, and P6 and demonstrating that NAVIGATER makes usually undetectable MAs, observable.

*Droplet Digital PCR (ddPCR)*

Next, to quantify our enrichment assay products, we subjected them to ddPCR (**Figure S7**). The detection limit of ddPCR relies on the number of amplifiable nucleic acids in the sample, which must be a small fraction of the total number of ddPCR droplets. The large number of WT alleles in the sample limits the number of pre-ddPCR amplification cycles that can be carried out to increase the concentration of the rare alleles. As NAVIGATER drastically reduces the number of WT alleles in the sample, it enables one to increase the number of pre-amplification cycles, increasing the number of MAs and thereby the ddPCR sensitivity. When operating with a mixture of WT and MA, NAVIGATER products include: residual uncleaved

WT ($N_{WT}$), MA ($N_{MA}$), and WT-MA hybrids ($N_H$). Hybrid alleles form during rehybridization of an ssWT with an ssMA. The allelic fraction of the mutant is fMA= ($N_{MA}$+½$N_H$) / ($N_{WT}$+$N_{MA}$+$N_H$)

We carried out ddPCR on un-enriched (control, NAVIGATER without *Tt*Ago), once-enriched, and twice-enriched samples (**Figure S7b**), increasing the fMA significantly (**Figure 6a**). For example, fMA increased from 0.5% in the un-enriched P5 (G12D) sample to ~30% in the twice enriched sample. This represents a ~60 fold increase in the fraction of droplets (fMA) containing MA (**Figure 6b**). The same assay also enriched G12R, increasing the fMA from 3% to ~66% in sample P3 and G12V, increasing the fMA from 5% to ~68% in sample P4 (**Figure 6b**).

*PNA-PCR*

PNA-PCR engages a sequence-specific PNA blocker that binds to WT alleles, suppressing WT amplification and providing a limit of detection of fMA ~1% (Choi et al., 2010). To demonstrate NAVIGATER's utility, we compared the performance of PNA-PCR when processing pancreatic cancer patient samples (**Table 2**) before and after NAVIGATER (**Figure 6c-e**). Before enrichment, PNA-PCR real-time amplification curves in the order of appearance are P6, P4, and P3, as expected (**Table 2**). Samples P1 (fMA=0), P2 (fMA=0), and P5 (fMA=0.5%) nearly overlap, consistent with a detection limit of ~1% (Choi et al., 2010). Enrichment significantly increases the threshold times of samples P1 and P2, revealing the presence of MAs in sample P5 (**Figure 6d**). PNA-PCR combined with NAVIGATER provides the linear relationship T1/2= 22.9-5 log (fMA) between threshold time (the time it takes the amplification curve to reach half its saturation value) and allele concentration (**Figure 6e**), allowing one to estimate MA concentration. The data suggests that NAVIGATER can improve PCR-PNA limit of detection to below 0.1%.

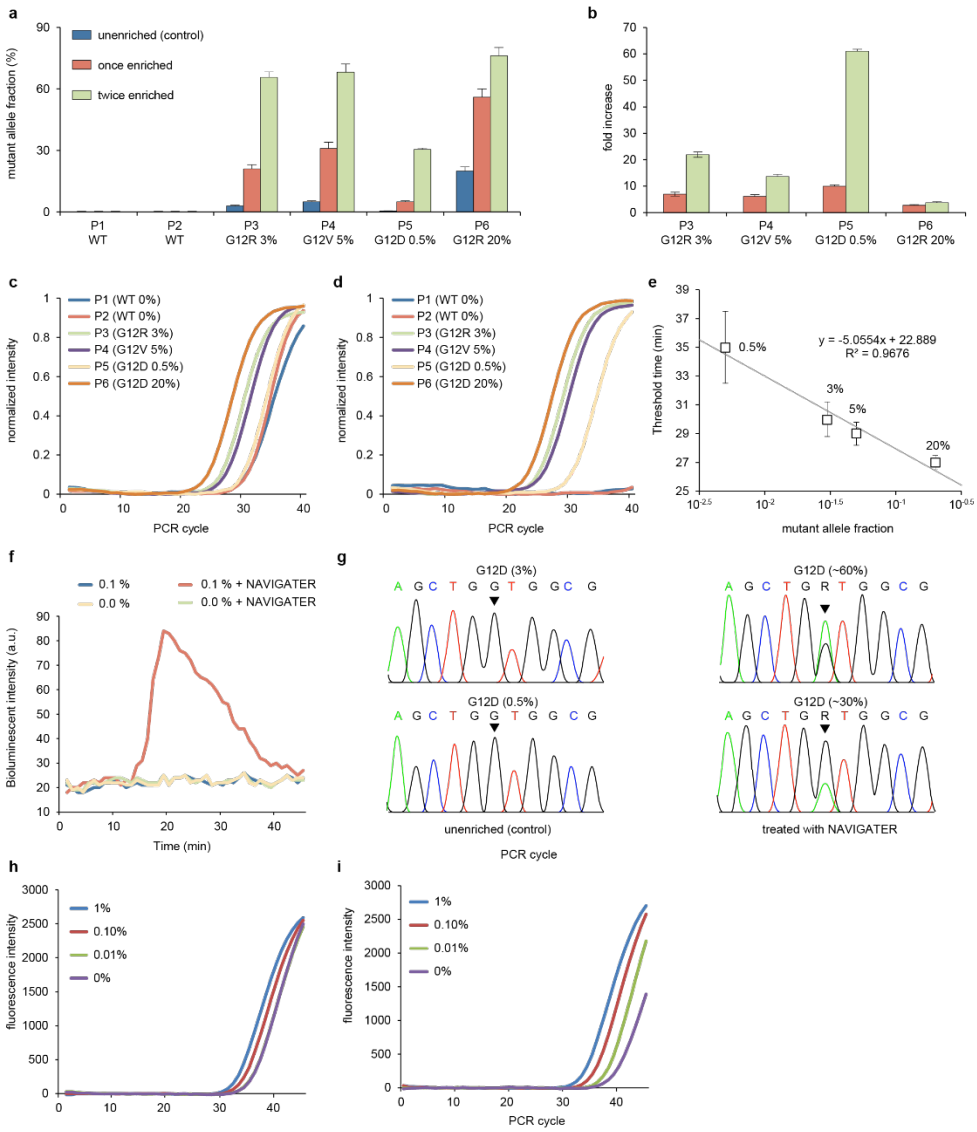



**Figure 6 | NAVIGATER enhances sensitivity of downstream detection methods. a, b** | ddPCR of samples from pancreatic cancer patients containing *KRAS* mutants (**Table 2**): **a** | Fraction of droplets containing mutant alleles. **b** | Increase in mutant allele fraction after NAVIGATER enrichment. **c, d, e** | PNA-PCR's amplification curves of pancreatic cancer patients' samples before (**c**) and after (**d**) NAVIGATER. **e** | Amplification threshold time as a function of mutant fraction. **f** | PNA-LAMP of simulated RNA samples before and after NAVIGATER carried out with a minimally instrumented, electricity-free Smart-Connected Cup (SCC) (Song et al., 2018). **g** | Sanger sequencing before and after NAVIGATER when detecting simulated RNA samples. **h, i** | XNA-PCR of samples

containing various mutant concentrations without (**h**) and with (**i**) multiplexed NAVIGATER pre-treatment. All the controls were pre-processed with NAVIGATER in the absence of *Tt*Ago.

*PNA-LAMP*

Genotyping with PNA blocking oligonucleotides can be combined with the isothermal amplification LAMP (Tatsumi et al., 2008). To demonstrate the feasibility of genotyping at the point of care and resource-poor settings, we use a minimally-instrumented, electricity-free Smart-Connected Cup (SCC) with smartphone and bioluminescent dye-based detection to incubate PNA-LAMP and detect reaction products (Song et al., 2018). To demonstrate that we can also detect RNA alleles, we used simulated samples comprised of mixtures of WT KRAS mRNA and KRAS-G12D mRNA. In the absence of pre-enrichment, SSC is unable to detect the presence of 0.1% KRAS-G12D mRNA whereas with pre-enrichment 0.1% KRAS-G12D mRNA is readily detectable (**Figure 6f**)

*Sanger Sequencing*

In the absence of enrichment, Sanger sequencers detect >5% MA fraction (Tsiatis et al., 2010). The Sanger sequencer failed to detect the presence of fMA 3% and 0.5% KRAS-G12D mRNA in our un-enriched samples, but readily detected these MAs following NAVIGATER enrichment (**Figure 6g**).

*XNA-PCR*

XNA-PCR is a clamped assay that suppresses amplification of WT alleles, enabling detection of MAs down to 0.1% fraction (Powell and Zhang, 2016). We used multiplexed NAVIGATER with 3 different guides to enrich samples of 60 ng cfDNA that included WT and various fractions of KRAS G12D, EGFR ΔE746 - A750, and EGFR L858R. Without NAVIGATER, XNA-PCR was able to detect down to 0.1% KRAS G12D (**Figure 6h**), 0.1% EGFR ΔE746 - A750 (**Figure S8e**), and 1% EGFR L858R (**Figure S8c**). With NAVIGATER pre-treatment, XNA-PCR sensitivity increased by over 10 fold to 0.01% KRAS G12D (**Figure 6i**), 0.01% EGFR ΔE746 - A750 (**Figure S8f**), and 0.1% EGFR L858R (**Figure S8d**). Here, in addition to significantly improving the sensitivity of XNA-PCR, we also demonstrate that NAVIGATER can operate as a multiplexed assay, enriching for multiple MAs.

# Discussion

Liquid biopsy is a simple, minimally invasive, rapidly developing diagnostic method to analyze cell-free nucleic acid fragments in body fluids and obtain critical diagnostic information on patient health and disease status. Currently, liquid biopsy can help personalize and monitor treatment for patients with advanced cancer, but the sensitivity of available tests is not yet sufficient for patients with early stage disease or for cancer screening (Aggarwal et al., 2018; Bettegowda et al., 2014; Oxnard et al., 2014). Detection of alleles that contain critical clinical information is challenging since they are present at very low concentrations among abundant background of nucleic acids that differ from alleles of interest by as little as a single nucleotide.

Here, we report on a novel enrichment method (NAVIGATER) for rare alleles that uses *Tt*Ago. *Tt*Ago is programmed with short ssDNA guides to specifically cleave guide-complementary alleles and stringently discriminate against off-targets with a single nucleotide precision. Sequence mismatches between guide and off-targets reduce hybridization affinity and cleavage activity by sterically hindering the formation of a cleavage-compatible state (Wang et al., 2008b, 2009). We observe that the activity and discrimination efficiency of *Tt*Ago depends on the (i) position of the mismatch, (ii) buffer composition, (iii) guide concentration, (iv) guide length, (v) incubation temperature and time, and (vi) target sequence. *Tt*Ago appears to discriminate best between target and off-target in the presence of a mismatch at or around the cleavage site located between guide nucleotides 10 and 11. Optimally, the buffer should contain $[Mg^{2+}] \geq 8$ mM, 0.8 M betaine, and 1.4 mM dNTPs. The ssDNA guides should be 15-16 nt in length with their concentration exceeding *Tt*Ago's concentration; and the incubation temperature should exceed the target dsDNA melting temperature. NAVIGATER is amenable to multiplexing and can concurrently enrich for multiple MAs while operating with different guides.

We demonstrate NAVIGATER's ability to enrich the fraction of cancer biomarkers such as KRAS, BRAF, and EGFR mutants in various samples. For example, NAVIGATER increased KRAS G12D fraction from 0.5% to 30% (60 fold) in a blood sample from a pancreatic cancer patient. The presence of 0.5% KRAS G12D could not be detected with Sanger sequencer or PNA-PCR. However after NAVIGATER pre-processing, both the Sanger sequencer and PNA-PCR readily identified the presence of KRAS G12D. Additionally, NAVIGATER

combined with PNA-LAMP detects low fraction (0.1%) mutant RNA alleles and NAVIGATER combined with PNA-LAMP enables genotyping at the point of care and in resource-poor settings. NAVIGATER improves the detection limit of XNA-PCR by more than 10 fold, enabling detection of rare alleles with frequencies as low as 0.01%.

NAVIGATER differs from previously reported rare allele enrichment methods (Aalipour et al., 2018; Bielas and Loeb, 2005; Gu et al., 2016; Kim et al., 2013; Lee et al., 2017; Li et al., 2008; Song et al., 2016a; Wu et al., 2017) in several important ways. First, NAVIGATER is versatile. In contrast to CRISPR-Cas9 (Barrangou and Doudna, 2016; Komor et al., 2017; Wu et al., 2018) and restriction enzymes (Bielas and Loeb, 2005), *Tt*Ago does not require a PAM motif or a specific recognition site. A gDNA can be designed to direct *Tt*Ago to cleave any desired target. Second, *Tt*Ago is a multi-turnover enzyme (Swarts et al., 2014b); a single *Tt*Ago-guide complex can cleave multiple targets. In contrast, CRISPR-Cas9 is a single turnover nuclease (Sternberg et al., 2014). Third, whereas CRISPR-Cas9 exclusively cleaves DNA, *Tt*Ago cleaves both DNA and RNA targets with single nucleotide precision. Hence, NAVIGATER can enrich for both rare DNA alleles and their associated exosomal RNAs (Krug et al., 2018), further increasing assay sensitivity. Fourth, *Tt*Ago is robust, operates over a broad temperature range (66-86°C) and unlike PCR-based enrichment methods, such as COLD-PCR (Li et al., 2008) and blocker-PCR (Kim et al., 2013; Wu et al., 2017) does not require tight temperature control. Moreover, NAVIGATER can complement PCR-based enrichment methods. Fifth, *Tt*Ago is more specific than thermostable duplex-specific nuclease (DSN) (Song et al., 2016a). Since DSN nonspecifically cleaves all dsDNA, DSN-based assays require tight controls of probe concentration and temperature to avoid non-specific hybridization and cleavage of the rare nucleic acids of interest. Most importantly, as we have demonstrated, NAVIGATER is compatible with many downstream genotyping analysis methods such as ddPCR, PNA-PCR, XNA-PCR, and sequencing. Last but not least, NAVIGATER can operate with isothermal amplification methods such as LAMP, enabling integration of enrichment with genotyping for use in resource poor settings.

In the future, we will design panels of DNA guides to enable NAVIGATER in combination with downstream detection methods (including next-generation sequencing) to detect MAs indicative of various types of cancer. NAVIGATER could be leveraged to help detect other rare MA such as genetic disorders in fetal DNA and drug resistant bacteria.

## Acknowledgements

Dr. Robert M. Greenberg helped with PAGE electrophoresis. Dr. Jennifer E. Phillips-Cremins provided us with access to gel imager. Dr. Changchun Liu provided helpful comments early in this project. Stephanie It is made available under a CC-BY-NC-ND 4.0 International license. (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. BioRxiv preprint first posted online Dec. 10, 2018; doi: http://dx.doi.org/10.1101/491738. The copyright holder for this preprint 12 Yee and Taylor Black assisted with ddPCR. This work is supported by the NIH NCI 1R21CA227056-01 to the University of Pennsylvania, and by grants from the Netherlands Organization of Scientific Research (NWO-ECHO 711013002 and NWO-TOP 714015001) to J.v.d.O.

## Competing interests

University of Pennsylvania and Wageningen University have applied for a patent on NAVIGATER with J.S., J.W.H., M.G.M., J.v.d.O., and H.H.B. listed as co-inventors.

## Human Subjects

This study was approved by Penn Institutional Review Board (IRB PROTOCOL #: 822028)

# Methods

## *Tt*Ago expression and purification

The *Tt*Ago gene, codon-optimized for *E. coli* Bl21 (DE3), was inserted into a pET-His6 MBP TEV cloning vector (Addgene plasmid # 29656) using ligation-independent cloning. The *Tt*Ago protein was expressed in *E. coli* Bl21(DE3) Rosetta™ 2 (Novagen). Cultures were grown at 37°C in Lysogeny broth medium containing 50 μg $ml^{-1}$ kanamycin and 34 μg $ml^{-1}$ chloramphenicol until an OD600 nm of 0.7 was reached. *Tt*Ago-expression was induced by addition of isopropyl β-D-1- thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. During the expression, cells were incubated at 18°C for 16 hours with continuous shaking. Cells were harvested by centrifugation and lysed in buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM imidazole, supplemented with EDTA-free protease inhibitor cocktail tablet (Roche). The soluble fraction of the lysate was loaded on a nickel column (HisTrap Hp, GE healthcare). The column was extensively washed with buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 30 mM imidazole. Bound proteins were eluted by increasing the concentration of imidazole in the wash buffer to 250 mM. The eluted protein was dialyzed at 4°C overnight against 20 mM HEPES pH 7.5, 250 mM KCl, and 1 mM dithiothreitol (DTT) in the presence of 1 mg TEV protease (expressed and purified as previously described Tropea et al., 2009) to cleave the His6-MBP tag. Next, the cleaved protein was diluted in 20 mM HEPES pH 7.5 to lower the final salt concentration to 125 mM KCl. The diluted protein was applied to a heparin column (HiTrap Heparin HP, GE Healthcare), washed with 20mM HEPES pH 7.5, 125 mM KCl and eluted with a linear gradient of 0.125-2 M KCl. Next, the eluted protein was loaded onto a size exclusion column (Superdex 200 16/600 column, GE Healthcare) and eluted with 20 mM HEPES pH7.5, 500 mM KCl and 1 mM DTT (**Figure S9**). Purified *Tt*Ago protein was diluted in a size exclusion buffer to a final concentration of 5 μM. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

## *Tt*Ago-based cleavage assays

5'-Phosphorylated DNA guides and Ultramer® ssDNA and ssRNA targets (100 nt) were synthesized by IDT (Coralville, IA). For ssDNA and ssRNA cleavage experiments, purified *Tt*Ago, DNA guides, and ssDNA or ssRNA targets were mixed with *Tt*Ago and guides at the ratios indicated in the buffers listed in **Table S1** and incubated at the indicated temperatures.

Reactions were terminated by adding 1 µL proteinase K (Qiagen, Cat. No. 19131) solution, followed by 15 min incubation at 56°C. Samples were then mixed with 2X loading buffer (95% (de-ionized) formamide, 5mM EDTA, 0.025% SDS, 0.025% bromophenol blue and 0.025% xylene cyanol) and heated for 10 min at 95°C before the samples were resolved on 15% denaturing polyacrylamide gels (7M Urea). Gels were stained with SYBR gold Nucleic Acid Gel Stain (Invitrogen) and nucleic acids were visualized using a BioRad Gel Doc XR+ imaging system. For dsDNA cleavage, *Tt*Ago and guides were pre-incubated in LAMP buffer 3 (**Table S1**) at 75 °C for 20 min.

## Cell-free DNA (cfDNA) and RNA Samples

Patient cfDNA samples. All six blood samples (**Table S2**) were obtained from patients with metastatic pancreatic cancer who had provided informed consent under the IRB-approved protocol (UPCC 02215, IRB# 822028). cfDNA was extracted with QIAamp® Circulating Nucleic Acid kit(Qiagen, Valencia, CA, USA). Subsequently, the extracted cfDNA was qualified and quantified with multiplex ddPCR (Raindance).

RNA samples. Total RNA was extracted with RNeasy® mini kit (Qiagen, Valencia, CA, USA) per manufacturer's protocol from Human cancer cell lines U87-MG (WT KRAS mRNA) and ASPC1 (KRAS G12D mRNA) and quantified with ddPCR.

cfDNA pre-amplification. The pre-amplification of the cfDNA was carried out in 50 µL reaction volumes using 20 ng of cfDNA, 1 × Q5 Hot Start High-Fidelity Master Mix (NEB), and 100 nM each of forward and reverse KRAS 80 bp-PCR primers (**Table S3**). Reaction mixes without DNA were included as no-template (negative) controls (NTCs). Nucleic acids were preamplified with a BioRad Thermal Cycler (BioRad, Model CFD3240) with a temperature profile of 98°C for 3 minutes, followed by 30 cycles of amplification (98°C for 10 seconds, 63°C for 3 minutes, and 72°C for 30 seconds), and a final 72 °C extension for 2 minutes.

mRNA pre-amplification. The pre-amplification of the mRNA was performed in 50 µL reactions using 30 ng of total RNA, 1 × Q5 Hot Start High-Fidelity Master Mix (New England Biolabs, Ipswich, MA), 100 nM each of forward and reverse KRAS 295 bp-PCR primers (**Table S3**), and 1 µL reverse transcriptase (Invitrogen, Carlsbad, CA). The reaction

mix was incubated at 55°C for 30 minutes and 98°C for 3 minutes, followed by 30 cycles of amplification (93°C for 15 seconds, 62°C for 30 seconds, and 72°C for 30 seconds), and a final 72°C extension for 4 minutes.

### Mutation enrichment (NAVIGATER)

The same setup as for synthetic dsDNA cleavage was used for cf-ctDNA and mutant mRNA enrichment. *Tt*Ago, S-guide, and AS-guide were mixed in 1:10:10 ratio (1.25 µM *Tt*Ago, 12.5 µM S-guide, 12.5 µM AS-guide) in the Buffer 3 and pre-incubated at 75°C for 20 min. Samples consisted of 2 µL preamplified PCR or RT-PCR products were added after pre-incubation of *Tt*Ago and guides. The reaction mixes were incubated at 83°C for 1 hour. The enriched products were diluted 104 fold before downstream mutation analysis or second-round enrichment. For second-round enrichment, the protocol outlined above was repeated.

### NAVIGATER combined with downstream mutation detection methods

Droplet digital PCR (ddPCR). ddPCR was carried out with the RainDrop Digital PCR system (RainDance Technologies, Inc.) to verify mutation abundance before and after *Tt*Ago enrichment. 2µL of the 104 fold diluted, *Tt*Ago-treated sample was added to each 30 µL dPCR. dPCRs contained 1× TaqMan Genotyping Master Mix (Life Technologies), 400 nM KRAS 80bp-PCR primers, 100 nM KRAS wild-type target probe, 100 nM KRAS mutant target probe (**Table S3**), and 1× droplet stabilizer (RainDance Technologies, Inc.). Emulsions of each reaction were prepared on the RainDrop Source instrument (RainDance Technologies, Inc.) to produce 2 to 7 million, 5 pLvolume droplets per 25 µL reaction volume. Thereafter, the emulsions were placed in a thermal cycler to amplify the target and generate signal. The temperature profile for amplification consisted of an activation step at 95°C for 10 minutes, followed by 45 cycles of amplification [95°C for 15 seconds and 60°C for 45 seconds]. Reaction products were kept at 4°C before placing them on the RainDrop Sense instrument (RainDance Technologies, Inc.) for signal detection. RainDrop Analyst (RainDance Technologies, Inc.) was used to determine positive signals for each allele type. Gates were applied to regions of clustered droplets to define positive hits for each allele, according to the manufacturer's instructions

PNA-PCR. PNA-PCR was performed in 20 µL reaction volumes, containing 4.5 µL of the 104 -fold diluted *Tt*Ago-treated products, 1 × Q5 Hot Start High-Fidelity Master Mix (New

England Biolabs, Ipswich, MA), 0.5 µL of EvaGreen fluorescent dye (Biotium, Hayward, CA), 500 nM KRAS PNA clamp (**Table S3**), and 100 nM each of forward and reverse KRAS 80 bp-PCR primers. Reactions were amplified with a BioRad Thermal Cycler (BioRad, Model CFD3240) with a temperature profile of 98°C for 3 minutes, followed by 40 cycles of amplification (98°C for 10 seconds, 63°C for 3 minutes, and 72°C for 30 seconds).

Sanger sequencing. RNA extracted from cell lines were pre-amplified by KRAS 295 bp-PCR primers as described above and treated by *Tt*Ago mutation enrichment system. 2 µL of the 104 -fold diluted, *Tt*Ago-treated sample was amplified by 295 bp PCR protocol (the same as 295 bp RT-PCR protocol without a reverse transcription step) for 30 cycles. PCR products were checked for quality and yield by running 5 µl in 2.2% agarose Lonza FlashGel DNA Cassette and processed for Sanger sequencing at Penn Genomic Analysis Core.

POC mutation detection. PNA-LAMP (SMAP-2) was prepared in 20 µL reaction volumes according to previously described protocol18. The reaction mix contained 2 µL of the 104 fold diluted *Tt*Ago-treated products (same as used for Sanger sequencing), 1 × LAMP buffer 3 (Eiken LAMP buffer), 1 µL Bst DNA polymerase (from Eiken DNA LAMP kit), 2.5 µL of BART reporter (Lot: 1434201; ERBA Molecular, UK) (Song et al., 2018), KRAS PNA clamp and LAMP primers (**Table S3**). The prepared reaction mixtures were injected into reaction chambers of our custom made multifunctional chip (Song et al., 2016c, 2016b). The inlet and outlet ports were then sealed with transparent tape (3M, Scotch brand cellophane tape, St. Paul, MN) and the chip was placed in our portable Smart-Connected Cup and processed according to previously described protocol (Song et al., 2018).

Multiplexed NAVIGATER and XNA-PCR.

Singleplex and Multiplexed preamplification: Singleplex and Triplex PCR were carried out with mutation detection kit (DiaCarta, Inc). The 10 µL reaction mixture contains 60 ng of cfDNA (reference standard that includes various MAs, Horizon Discovery, HD780), 1 × PCR Master Mix, 1 µL of either single or mixed PCR primers (1: 1: 1) for targets of interest. Nucleic acids were preamplified with a BioRad Thermal Cycler (BioRad, Model CFX96) with a temperature profile of 95°C for 5 minutes, followed by 35 cycles of amplification (95°C for 20 seconds, 70°C for 40 seconds, 60°C for 30 seconds, and 72°C for 30 seconds), and a final 72°C extension for 2 minutes.

Multiplexed enrichment: Guides (1:1:1) for targets of interest were mixed with *Tt*Ago in 10:1 ratio (12.5 µM S-guides, 12.5 µM AS-guides, 1.25 µM *Tt*Ago) in buffer 3 and pre-incubated at 75°C for 20 min. Samples consisted of 1 µL pre-amplified singleplex or triplex PCR products mixed with pre-incubated *Tt*Ago-guide complexes. The reaction mixes were incubated at 83°C for 1 hour. The products with and without *Tt*Ago treatment were resolved on 15% denaturing polyacrylamide gels (7M Urea). The products were diluted 107 fold before downstream mutation analysis.

XNA-PCR: NAVIGATER products were tested by mutation detection method XNA-PCR19 (DiaCarta, Inc.). XNA-PCR was carried out for individual mutants in 10-µL reaction volumes, containing 3 µL of the 107 -fold diluted NAVIGATER products, 1 × PCR Master Mix, 1 µL of PCR primer/probe mix, and 1 µL of XNA clamp. Reactions were amplified with a BioRad Thermal Cycler (BioRad, Model CFX96) with a temperature profile of 95°C for 5 minutes, followed by 45 cycles of amplification (95°C for 20 seconds, 70°C for 40 seconds, 60°C for 30 seconds, and 72°C for 30 seconds).

# Supplementary Tables

**Table S1 | Buffer Compositions.**

| **Buffer 1** ThermoPol reaction buffer (1x, NEB) | **Buffer 2** Isothermal amplification buffer (1x, NEB) | **Buffer 3** Eiken buffer (1x) | **Buffer S** (Swarts et al., 2014b) |
|---|---|---|---|
| 20 mM Tris-HCl | 20 mM Tris-HCl | 20 mM Tris-HCl | 10 mM Tris-HCl |
| 10 mM $(NH_4)_2SO_4$ | 10 mM $(NH_4)_2SO_4$ | 10 mM $(NH_4)_2SO_4$ | 125 mM NaCl |
| 10 mM KCl | 50 mM KCl | 10 mM KCl | 2 mM $MgCl_2$ |
| 2 mM $MgSO_4$ | 2 mM $MgSO_4$ | 8 mM $MgSO_4$ | (pH 8.0 at 25°C) |
| 0.1% Triton X-100 | 0.1% Tween 20 | 0.1% Tween 20 | |
| (pH 8.8 at 25°C) | (pH 8.8 at 25°C) | 1.4 mM dNTPs | |
| | | (pH 8.8 at 25°C) | |

**Table S2 | The genotype and mutation frequency of 6 samples from pancreatic cancer patients*.**

| Patient number | P1 | P2 | P3 | P4 | P5 | P6 |
|---|---|---|---|---|---|---|
| Genotype and mutation fraction | ND** | ND** | G12R 3% | G12V 5% | G12D 0.5% | G12D 20% |

* Samples were analyzed with standard ddPCR protocol (Taly et al., 2013).

** Not detected, possibly due to mutant allele not present.

All samples contain similar numbers of WT-KRAS (s.d. <10%).

**Table S3 | Sequences and concentrations of KRAS primers, PNA clamp oligo, and Taqman probes used in downstream mutation analysis.**

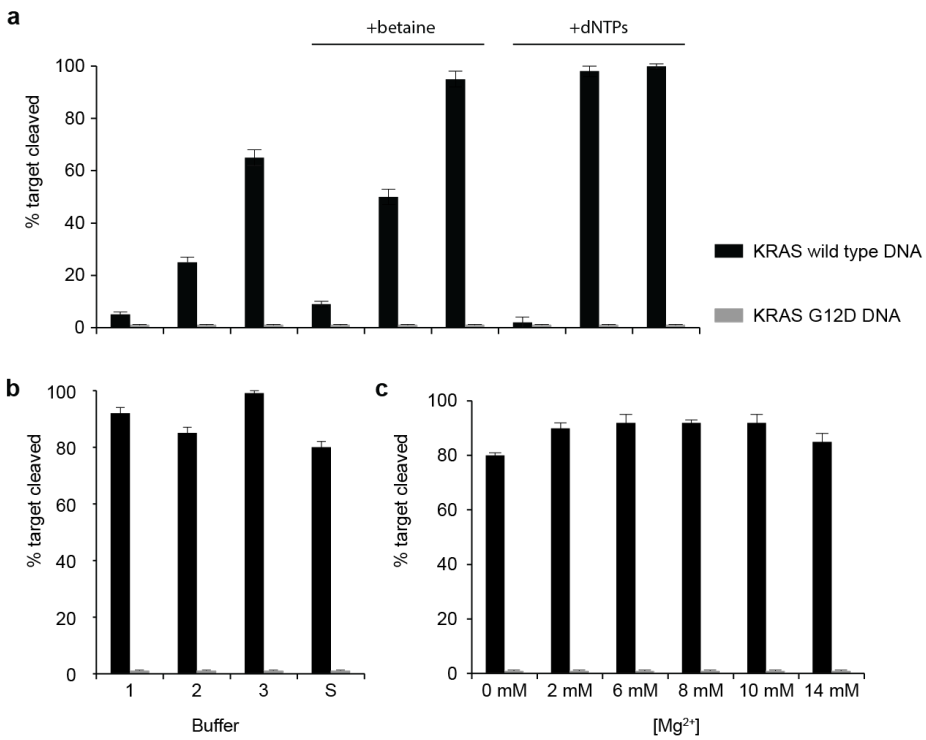
| **Name** | **Sequence (5' – 3')** | **concentration** |
|---|---|---|
| **80 bp-PCR FW**[1] | AGGCCTGCTGAAAATGACTGAATAT | |
| **80 bp-PCR-RV**[1] | GCTGTATCGTCAAGGCACTCTT | |
| **G12-WT-VIC probe**[1] | TTGGAGCTGGTGGCGT | 100 nM |
| **G12D-FAM probe**[1] | TGGAGCTGATGGCGT | 100 nM |
| **G12R-FAM probe**[1] | TTGGAGCTCGTGGCGT | 100 nM |
| **G12V-FAM probe**[1] | ACGCCAACAGCTC | 100 nM |
| **295 bp-PCR-FW**[2] | AAGGTACTGGTGGAGTATTTG | 100 nM |
| **295 bp-PCR-RV**[2] | GTACTCATGAAAATGGTCAGAG | 100 nM |
| **SMAP-2-OP1**[3] | TATTATAAGGCCTGCTG | 0.4 µM |
| **SMAP-2-OP2**[3] | TTGGATCATATTCGTCC | 0.4 µM |
| **SMAP-2-FP**[3] | ACCTTCTACCCTCAGAAGGTATAAACTTGTGGTAGTTG GAGC | 3.0 µM |
| **SMAP-2-BP**[3] | GCAAGAGTGCCTTGA | 1.5 µM |
| **SMAP-2-TP**[3] | TGGCGTAGGCATGATTCTGAATTAGCTGTAT | 3.0 µM |
| **PNA clamp**[3] | CCTACGCCACCAGCTCC | 0.7 µM |

[1] (Taly et al., 2013)

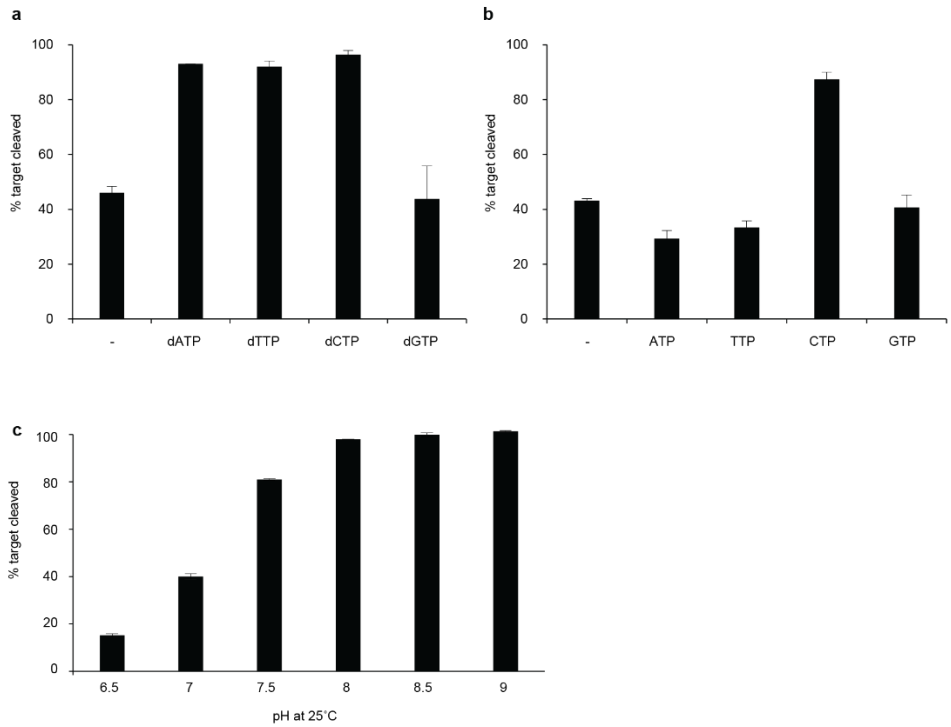
[2] (Oliner et al., 2010)

[3] (Tatsumi et al., 2008)

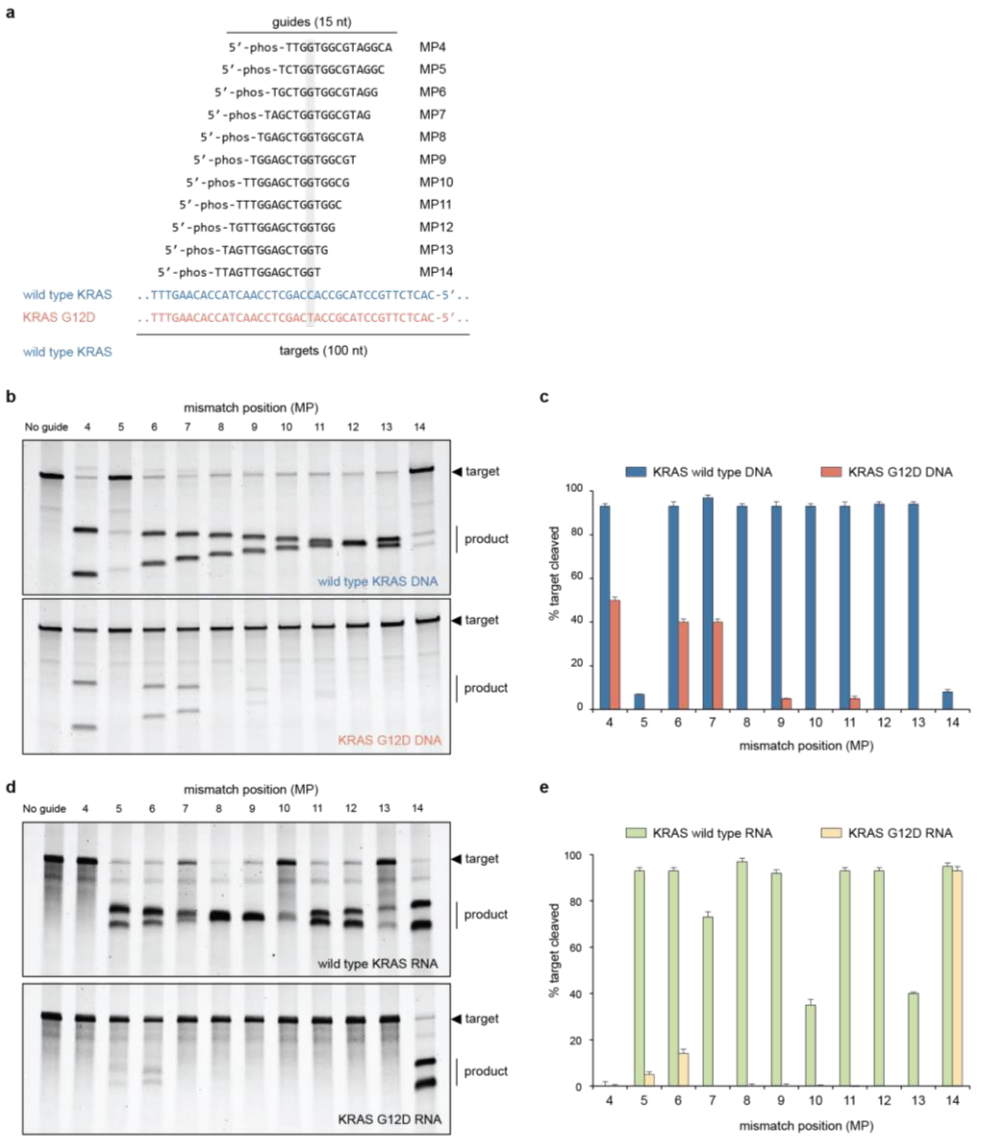
# Supplementary figures



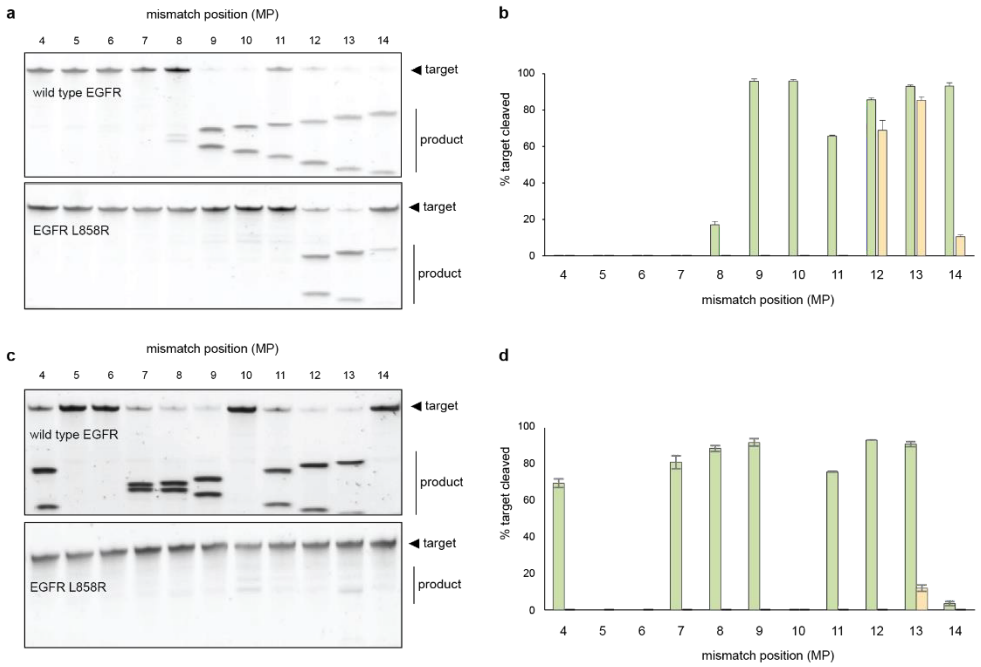

**Supplementary Figure 1 | Betaine, Mg2+, and dNTPs enhance *Tt*Ago DNA cleavage.** DNA cleavage efficiency as a function of (**a**) added [Mg2+] in buffer S in the absence and presence of betaine (0.8 M) or dNTPs (1.4 mM) at 80°C **b** | Buffer composition (**Table S1**) at 70°C; and (**c**) added [Mg2+] in buffer S at 70°C. All experiments were carried out with KRAS Sense (S) strand and 16 nt KRAS-S MP12 guide. Incubation time 20 min. *Tt*Ago: guide: target ratio =1:0.2:0.2. N=3.




**Supplementary Figure 2 |** The effects of (**a**) dNTPs and (**b**) NTPs on EGFR (L858R) sense RNA cleavage with guide EGFR L858R (16nt)-MP10 at 80°C. **c** | The effect of pH on EGFR (L858R) sense DNA cleavage with guide EGFR L858R (16nt)-MP10 at 75°C. *Tt*Ago: guide: target ratio =1:0.2:0.2. N=3.

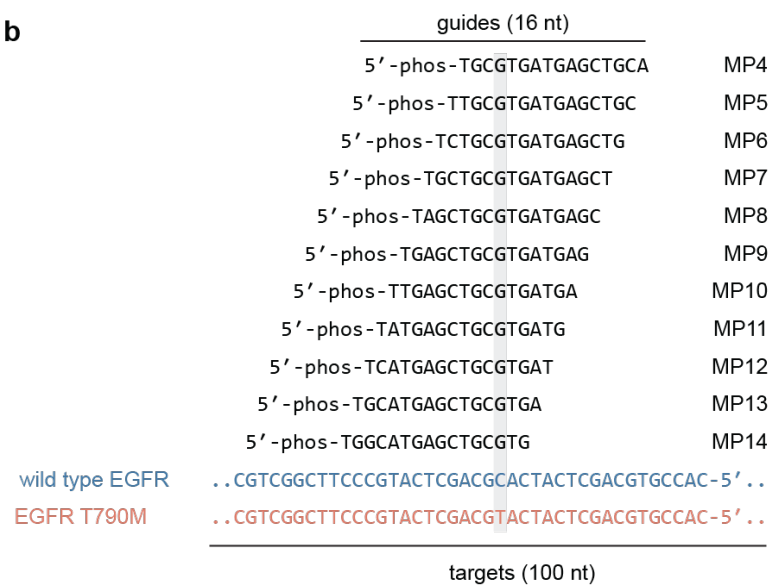


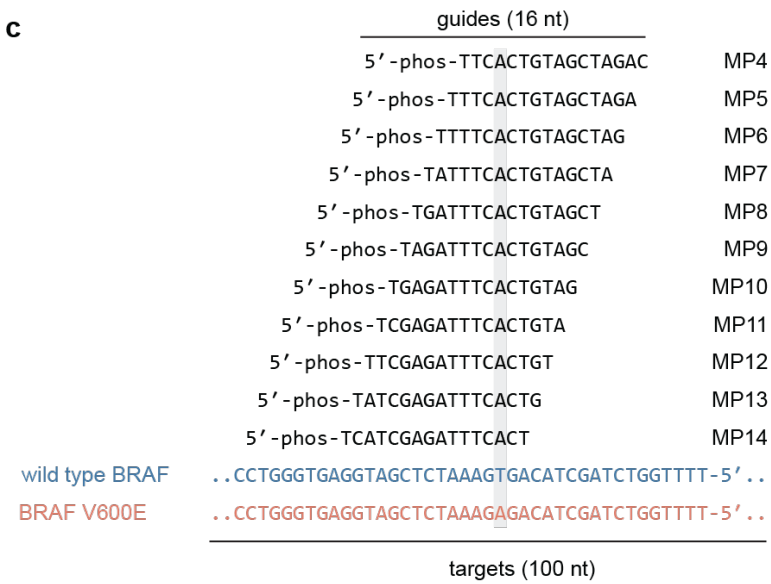

**Supplementary Figure 3 | Cleavage efficiencies on KRAS WT and KRAS G12D DNA and RNA targets using guides with varying MP. a** | KRAS – antisense (AS) guide and AS target sequences. The various guides vary in the position of the pair mismatch between AS gDNA and AS KRAS G12D. **b** | Electropherograms of cleaved AS WT KRAS and AS KRAS G12D DNA strands (80°C, 20 min). **c** | Cleaving efficiencies of AS KRAS WT and AS KRAS G12D DNA as a function of MP. **d** | Electropherograms of cleaved WT KRAS RNA and KRAS G12D RNA (80°C, 20 min). **e** | Cleaving efficiencies of WT KRAS RNA and KRAS G12D RNA as a function of MP. *Tt*Ago: guide: target =1: 0.2: 0.2. N=3.



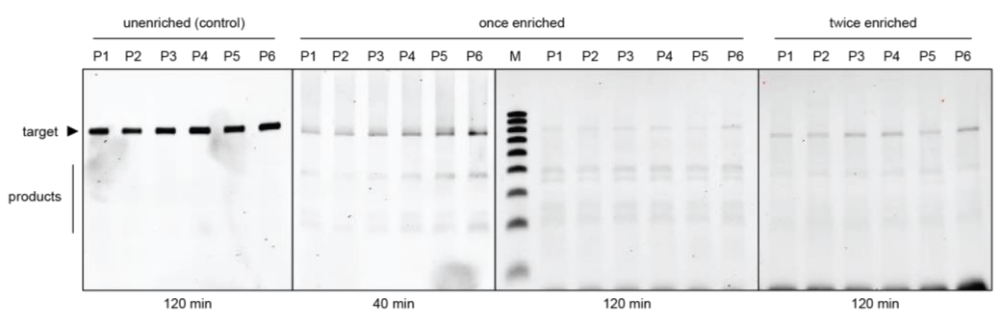

**Supplementary Figure 4 | Cleavage efficiencies on EGFR WT and EGFR L858R RNA targets using guides with varying MP's. a |** Cleavage assays using sense EGFR guides/targets. Sample were resolved on urea polyacrylamide gel. **b |** Cleaving efficiencies were determined based on the band intensities in (a) and plotted in a graph. **c |** Cleavage assays using anti-sense EGFR guides/targets. Sample were resolved on urea polyacrylamide gel. **d |** Cleaving efficiencies were determined based on the band intensities in (c) and plotted in a graph. Samples were incubated at 80ºC for 20 min. *Tt*Ago:guide:target ratio = 1:0.2:0.2. N=3.


a
guides (16 nt)
5′-phos-TAGCCCAAAATCTGTG MP2
5′-phos-TCAGCCCAAAATCTGT MP3
5′-phos-TCCAGCCCAAAATCTG MP4
5′-phos-TGCCAGCCCAAAATCT MP5
5′-phos-TGGCCAGCCCAAAATC MP6
5′-phos-TTGGCCAGCCCAAAAT MP7
5′-phos-TTTGGCCAGCCCAAAA MP8
5′-phos-TTTTGGCCAGCCCAAA MP9
5′-phos-TGTTTGGCCAGCCCAA MP10
5′-phos-TAGTTTGGCCAGCCCA MP11
5′-phos-TCAGTTTGGCCAGCCC MP12
5′-phos-TGCAGTTTGGCCAGCC MP13
5′-phos-TAGCAGTTTGGCCAGC MP14
5′-phos-TCAGCAGTTTGGCCAG MP15
wild type EGFR ..GAAGGCGTGGGTCGTCAAACCGGTCGGGTTTTAGACACTA-5′..
EGFR L858R ..GAAGGCGTGGGTCGTCAAACCGGGCGGGTTTTAGACACTA-5′..
targets (100 nt)

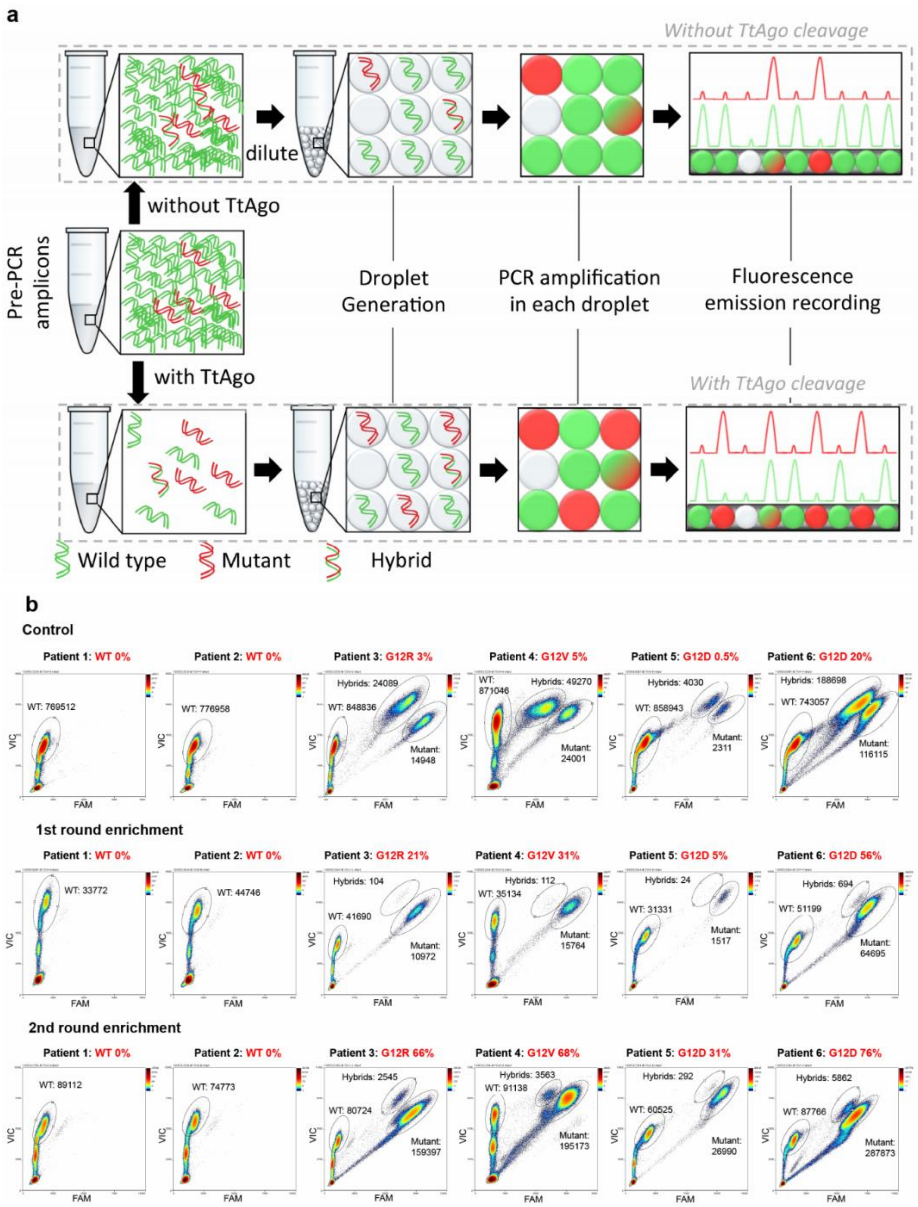

b
guides (16 nt)
5′-phos-TGCGTGATGAGCTGCA MP4
5′-phos-TTGCGTGATGAGCTGC MP5
5′-phos-TCTGCGTGATGAGCTG MP6
5′-phos-TGCTGCGTGATGAGCT MP7
5′-phos-TAGCTGCGTGATGAGC MP8
5′-phos-TGAGCTGCGTGATGAG MP9
5′-phos-TTGAGCTGCGTGATGA MP10
5′-phos-TATGAGCTGCGTGATG MP11
5′-phos-TCATGAGCTGCGTGAT MP12
5′-phos-TGCATGAGCTGCGTGA MP13
5′-phos-TGGCATGAGCTGCGTG MP14
wild type EGFR ..CGTCGGCTTCCCGTACTCGACGCACTACTCGACGTGCCAC-5′..
EGFR T790M ..CGTCGGCTTCCCGTACTCGACGTACTACTCGACGTGCCAC-5′..
targets (100 nt)



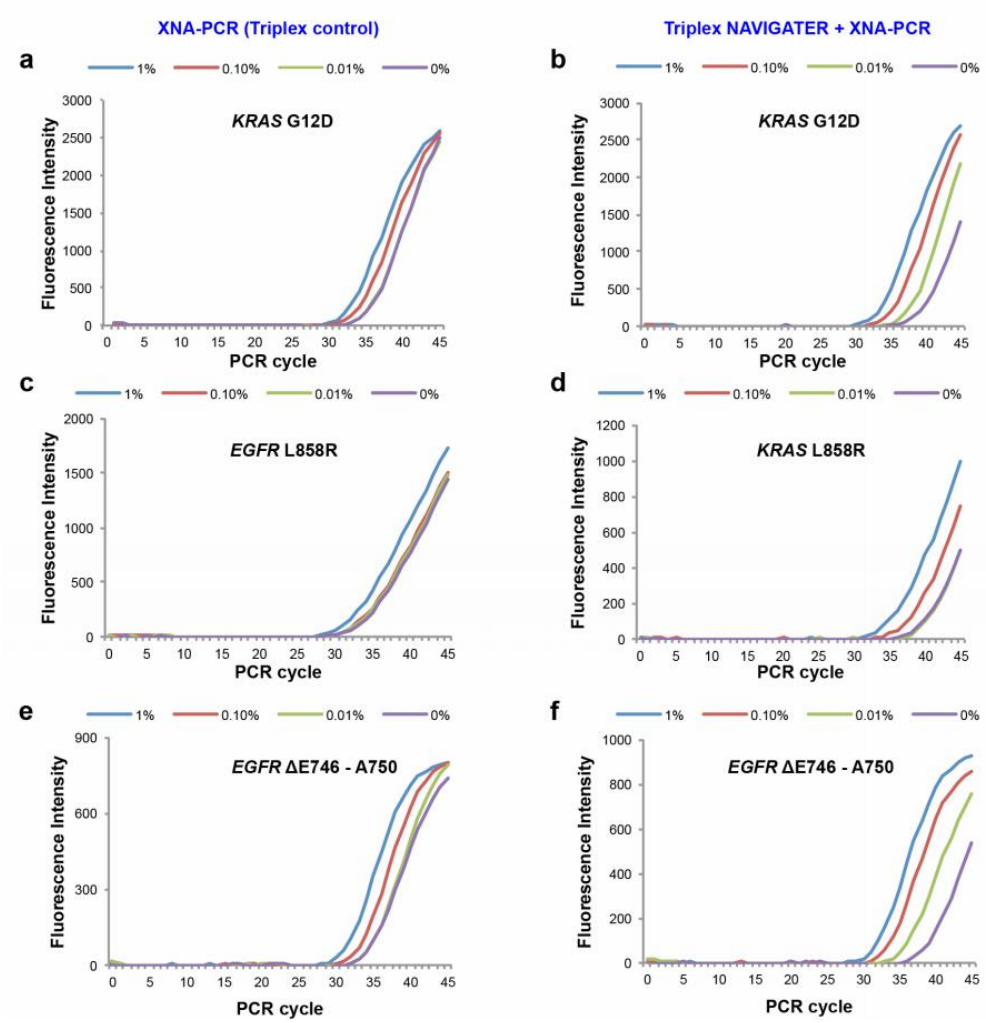

**Supplementary Figure 5 | Overview of the guide and target sequences used in the experiments. a** | EGFR WT/L858R, **b** | EGFR WT/T790M, **c** | BRAF WT/V600E.



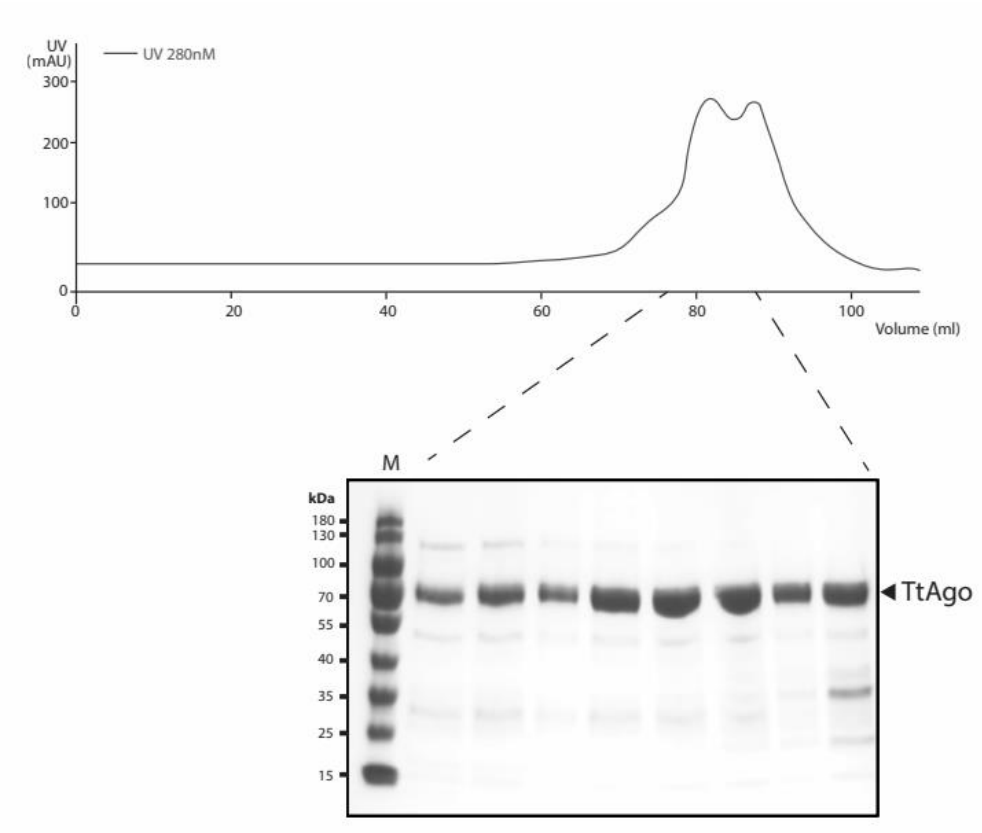

**Supplementary Figure 6 |** Urea polyacrylamide gel of six pancreatic cancer patient's samples (**Table 2**) without enrichment (control), once enriched for 40 min and 2h and twice-enriched.




**Supplementary Figure 7 | Six pancreatic cancer patient's samples processed with NAVIGATER and ddPCR**. **a |** Principle of operation of NAVIGATER combined with ddPCR. **b |** ddPCR results of non-enriched, once-enriched, and twice-enriched samples. There are three signaling droplet populations for KRAS: WT, mutant, and WT-mutant hybrid. 2~7 millions of droplets are produced for each sample




**Supplementary Figure 8 | Multiplexed enrichment of KRAS G12D, EGFR L858R, and EGFR ΔE746 - A750.** XNA-PCR amplification curves of samples of various fractions of KRAS G12D, EGFR L858R and EGFR ΔE746 - A750 in the absence (**a, c, e**) and presence (**b, d, f**) of NAVIGATER pre-treatment. Triplex NAVIGATER increases the sensitivity of XNA-PCR by at least an order of magnitude. All the controls were pre-processed with NAVIGATER in the absence of *Tt*Ago.




**Supplementary Figure 9 | Size exclusion chromatogram and SDS/PAGE.** *Tt*Ago was purified using a Superdex (200 16/60) size exclusion column.

# Chapter 6

## *Pyrococcus furiosus* Argonaute repurposed for molecular cloning

Jorrit W. Hegge*, Ismael Hinojo*, Jie Liu, Thijs Nieuwkoop, Nico Claasens, John van der Oost.

*contributed equally

# Abstract

Here we describe a novel one-tube cloning method that collectively uses the Argonaute protein of the archaeon *Pyrococcus furiosus* (*Pf*Ago) and a thermostable DNA ligase (Ampligase™). *Pf*Ago is a highly programmable endonuclease that can cleave single stranded DNA targets with nucleotide precision, when programmed with a small complementary DNA oligonucleotide guide. The method relies on a thermo-cycling reaction: (i) at 95°C two *Pf*Ago-guide complexes specifically cleave a DNA backbone vector or PCR product consequently forming 5'-end overhangs at both sides; (ii) at 45°C these overhangs anneal to overhangs of the insert (two pre-annealed DNA oligonucleotides). The two nicks that formed upon hybridization of the overhangs are covalently sealed by Ampligase™. The two-step reaction can be repeated several times to drive the reaction towards the direction of desired plasmid assembly. Although the method still requires further optimizations to increase the cloning efficiencies, this new cloning approach (called Jason cloning) has the potential to become an interesting alternative for current molecular cloning/assembly applications.

# Introduction

It is impossible to imagine molecular biological research without the numerous DNA assembly tools that are available (Engler et al., 2008, 2009, Gibson et al., 2008, 2009, 2010; Li and Elledge, 2007; Shao et al., 2009; Tsvetanova et al., 2011; Wiedmann et al., 1994). The ability to create recombinant DNA molecules led to an increased understanding of gene function, as well as to the spectacular development of synthetic biology. In the last decade multiple methods were developed for the construction of recombinant DNA sequences among which popular methods such as Gibson and Golden gate cloning (Engler et al., 2009; Gibson et al., 2009). Gibson cloning is an isothermal one-pot reaction where an exonuclease, a DNA polymerase, and a DNA ligase collaborate (**Figure 1**).

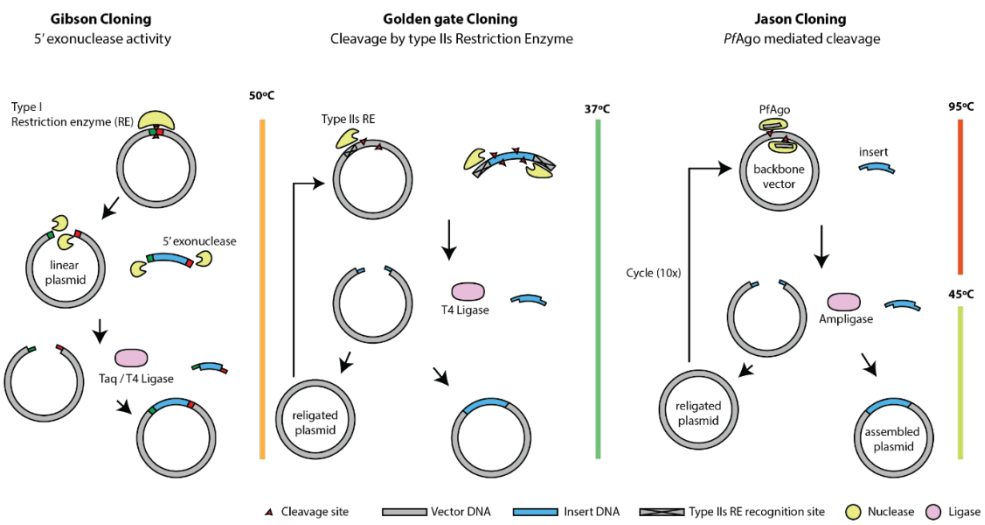



**Figure 1 | Schematic comparison of the one-pot Gibson, Golden gate and Jason Cloning.** Jason cloning requires and insert with pre-made overhangs that is incubated with a backbone vector, *Pf*Ago, ssDNA guides targeting the backbone vector, and Ampligase™. A repeated two-step thermo-cycle reaction causes an enrichment of the assembled plasmid.

The method requires a linearized backbone vector, generated by either PCR or a restriction enzyme, and an insert DNA with ends that overlap roughly 30 base pairs either with other to-be-inserted fragments or with the backbone vector. During the reaction, an exonuclease removes successive nucleotides from the 5' ends of both the backbone vector and insert DNA, allowing the overlapping regions of the different fragments to anneal. A DNA

polymerase fills in the gaps, and a ligase then seals the fragments together. In general, Golden gate cloning is also performed as a one-pot reaction but instead of using an exonuclease, it uses type IIs restriction enzymes to cleave both the insert DNA and backbone vector. Type IIs restriction enzymes cleave DNA outside their recognition site allowing the generation of customized overhangs (**Figure 1**). The popularity of such methods relates to their ability to assemble multiple insert fragments simultaneously in a one-pot reaction, which simplifies the procedure and workload. However, both technologies do have drawbacks such as, for Gibson cloning difficulties in assembling fragments smaller than 200 base pairs and for Golden Gate the type IIs recognition sites that cannot be present within any to the fragments to be assembled. Therefore, we here describe the development of Jason Cloning, an alternative one-pot, restriction site-independent DNA cloning method using Argonaute from *Pyrococcus furiosus* (*Pf*Ago) (**Figure 1**).

Prokaryotic Argonautes (pAgos) are nucleic acid-directed endonucleases that significantly differ in role and structural domain organization from their eukaryotic counterparts. Instead of being part of an RNA interference (RNAi) pathway, non-existing in prokaryotes, several pAgos have demonstrated to be involved in host defense against invading nucleic acids, such as plasmids (Olovnikov et al., 2013; Swarts et al., 2015a, 2017a; Zander et al., 2017). In contrast to eukaryotic Argonautes (eAgos), that exclusively mediate RNA-guided RNA interference, the substrate specificity of pAgos is diverse, with some pAgos being able to use DNA guides to cleave DNA targets. The highly programmable nature of these DNA-targeting pAgos paved the way for possible DNA editing applications (Hegge et al., 2018a).

The Argonaute protein of *Pyrococcus furiosus* (*Pf*Ago) is a well-studied pAgo that mediates DNA cleavage at elevated temperatures (80-100°C), using a small ssDNA (single-stranded DNA) guide of 15-21nt in length (Swarts et al., 2015a). A set of *Pf*Ago nucleases, each loaded with a guide that targets one of the two plasmid strands, can generate double stranded breaks (DSB) with designable overhangs (Swarts et al., 2015a). This already led to the development of *Pf*Ago as an universal restriction endonuclease for molecular cloning purposes (Enghiad and Zhao, 2017). Based on this we further developed the technology into a one-pot, *in vitro* cloning method that is capable of cloning a DNA insert with premade overhangs into a backbone vector. The overhangs of the insert need to be complementary to the desired cloning site of the backbone and can be as small as 18 nucleotides. The method allows assembly of fragments as small as 100 base pairs (smaller has not been tested yet) using

*Pf*Ago, a thermostable ligase (Ampligase™, Epicentre), and a set of short (21 nucleotide) ssDNA guides, via a short two-step thermo-cycling reaction. We named the method Jason cloning after the hero from the Greek mythology who, as captain of the ship Argo, leads a group of warriors called the Argonauts.

During Jason cloning the insert, backbone vector, *Pf*Ago and Ampligase™ are combined in a one-pot reaction. This reaction is cycled between 95°C and 45°C (**Figure 1**). At 95°C, while Ampligase™ is inactive but stable, *Pf*Ago nicks both plasmid strands in guide-dependent manner thereby linearizing the plasmid and generating overhangs that are complementary to the premade overhangs of the insert. On the other hand at 45°C, while *Pf*Ago is inactive, Ampligase™ ligates the annealed DNA overhangs, generating the assembled plasmid. As both overhangs of the backbone vector itself are complementary to each other, part of the cleaved backbone vector will re-circularize, thus regenerating the original plasmid. The re-circularized vector will act as the original substrate in subsequent cycles. The correctly assembled vector will be cleaved by *Pf*Ago at the same location as before, but since the insert is added in between the two nicks on both plasmid strands are thus far apart that if a double-stranded break is generated, it will have overhangs that are as long as the insert which will be annealed and sealed in the next ligation step. Hence, running multiple cycles results in an accumulation of the assembled plasmid.

### Ampligase™ ligates DNA substrates with overhangs of at least 10nt (50%-GC)

Prior to setting up the entire Jason Cloning method, we tested the effect of overhang length and temperature on the activity of Ampligase™ in a separate assay. As the optimal ligation temperature is a balance between the temperature at which the DNA overhangs anneal to each other and the temperature at which Ampligase™ is active. A ligation assay was designed using three DNA oligonucleotides (Oligo A, Oligo B, and Oligo C) (**Figure 2A**). Oligo B is has 50 complementary bases with oligo C and together they are annealed to form a double strand product. The remaining part of oligo C results in overhangs of variable lengths (6, 10, 14, and 18 nucleotides), which are complementary to part A. All four overhangs have a fixed GC-content of 50%. The mixture of the three oligonucleotides was incubated with Ampligase™ for one hour at three different temperatures (37°C, 45°C, and 60°C). An overhang of at least 10 nucleotides (50% GC-content) was necessary to successfully anneal and ligate oligo A and B together in one hour (**Figure 2B, C**). As a thermo-cycling program

with ligation steps of one hour each cycle are not feasible, we monitored the activity of Ampligase™ with the 18nt overhang during a 5 minute time course at 45°C. Despite the presence of non-ligated oligo A and B, a significant amount was ligated after 5 minutes (**Figure 2D**).

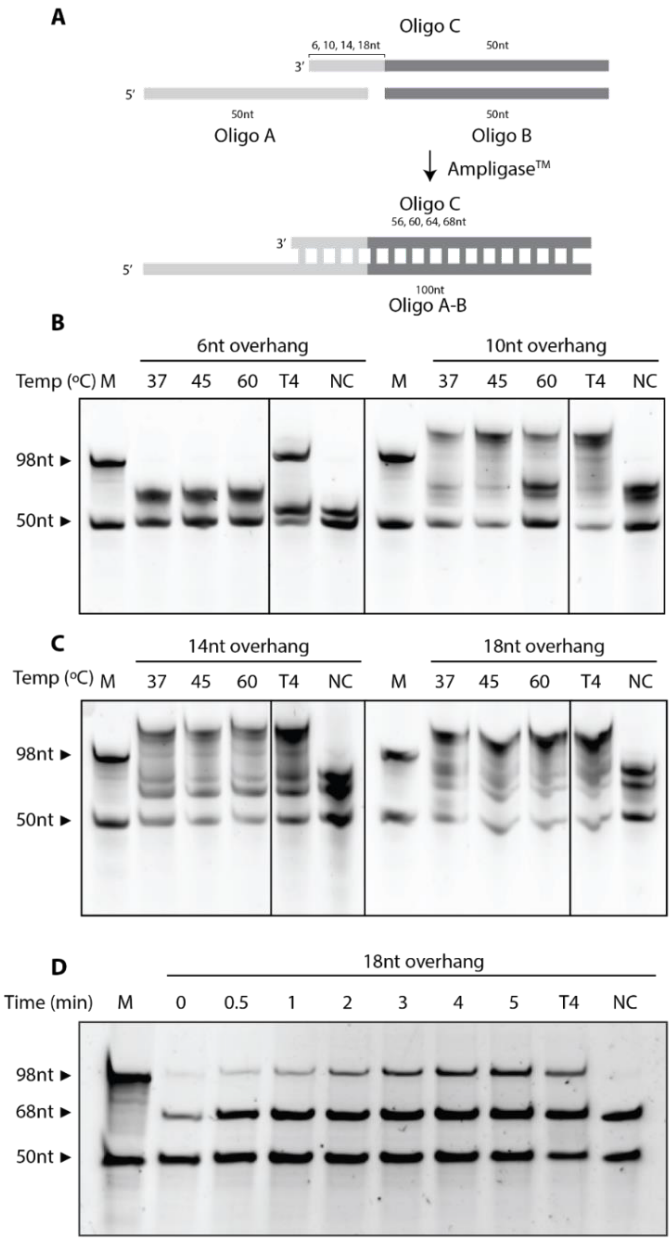


**Figure 2 | Ampligase™ successfully ligates DNA substrates with overhangs of at least 10nt (50%-GC). A |** Schematic representation of Ampligase assay using three oligonucleotides. Complementarity is indicated by matching grey shades. The length of Oligo A and Oligo B is 50 nucleotides and the length of the Oligo C varies between 56, 60, 64, and 68 nucleotides. Complementarity between oligo A and C represent overhangs of (6, 10, 14, and 18 nucleotides). **B, C, D** | Analysis on a denaturing Urea-page gel. (B and C) Ampligase was incubated with Oligo A, Oligo B, and Oligo C for 1 hour at 37°C, 45°C, or 60°C for each overhang length. **D |** Oligo A and B were ligated using an oligo C of 68 nt giving an 18nt overhang. The formation of ligated product (oligo A+B) during a time course of 5 minutes at 45°C. As a control T4 ligase was used for 1 hour at room temperature.

## *Pf*Ago cleaves double stranded DNA targets

The cleavage kinetics of *Pf*Ago on double stranded DNA targets were also tested separately on the pUC19 backbone vector. Previous studies showed that *Pf*Ago cleaves AT rich dsDNA more efficiently that GC-rich DNA (Swarts et al., 2015a). Therefore the cleavage kinetics of *Pf*Ago were tested on an AT rich region (~80%, 100bp) of the backbone vector pUC19 (**Figure 3A**). During the first two minutes *Pf*Ago rapidly linearized ~40% of the total pUC19 after which the formation of product plateaued (**Figure 3B, C**).

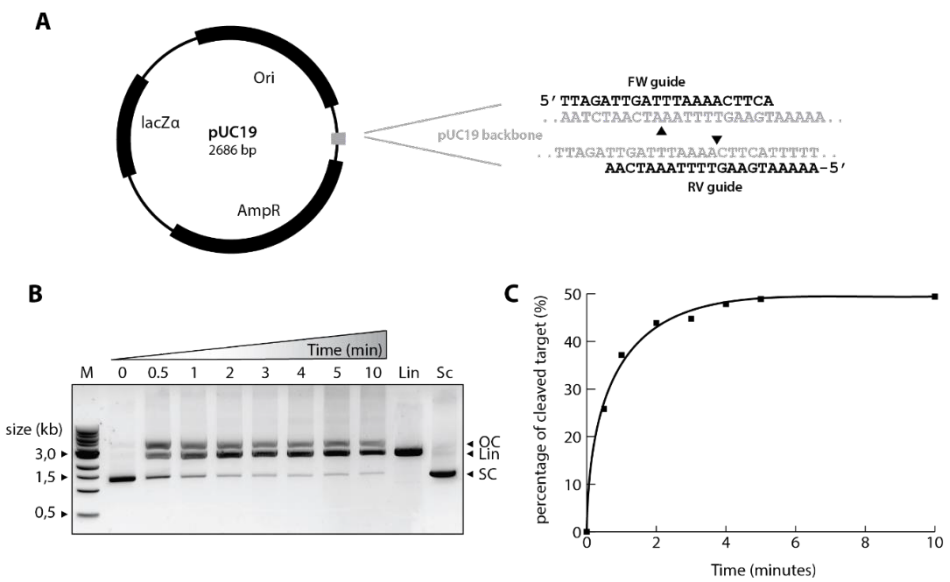



**Figure 3 | *Pf*Ago cleaved pUC19 at a high rate during the first two minutes. A |** pUC19 plasmid map and corresponding sequences of the target region (grey) and FW and RV guides (black). Cleavage sites are indicated by the black triangle. **B |** Time course of *Pf*Ago mediated pUC19 cleavage. Cleavage products were resolved in a 0.8% agarose gel stained with SYBR gold. M: ssDNA marker OC: Open circular LIN: Linear and SC: supercoiled plasmid. **C |** Time course of the cleavage efficiency of *Pf*Ago on pUC19. Efficiency is defined as the (intensity of the LIN band divided by the intensity of the SC+OC band * 100).

## Successful plasmid assembly using Jason cloning

Based on these results together with the data of the ligation assay, Jason Cloning was initially tested using a two-step thermo-cycle program of 10 cycles, each cycle starting with a 3 minute cleavage step at 95°C, followed by a 5 minute ligation step at 45°C (**Figure 4A**). While Ampligase™ favors substrates consisting of long GC-rich overhangs (more efficient annealing), *Pf*Ago generates short AT-rich overhangs more efficiently (better dsDNA accessibility; unpublished data). To find a balance between both, cloning reactions were tested using DNA inserts with overhangs that varied in length (10/18nt) and AT-content (50/80%). Accordingly, pUC19 was cleaved by *Pf*Ago using four different guide pairs; each pair was designed to generate overhangs compatible with the inserts. Successfully assembled plasmids could be detected by PCR using two primers that anneal adjacent to the insertion site (**Figure 4B**). Without the insert pUC19 plasmid generates an amplicon of 264bp, whereas the insert adds 100bp resulting in an amplicon of 364bp. Jason cloning was only shown successful using GC-rich (50%) 18 nucleotide overhangs (**Figure 4C**).

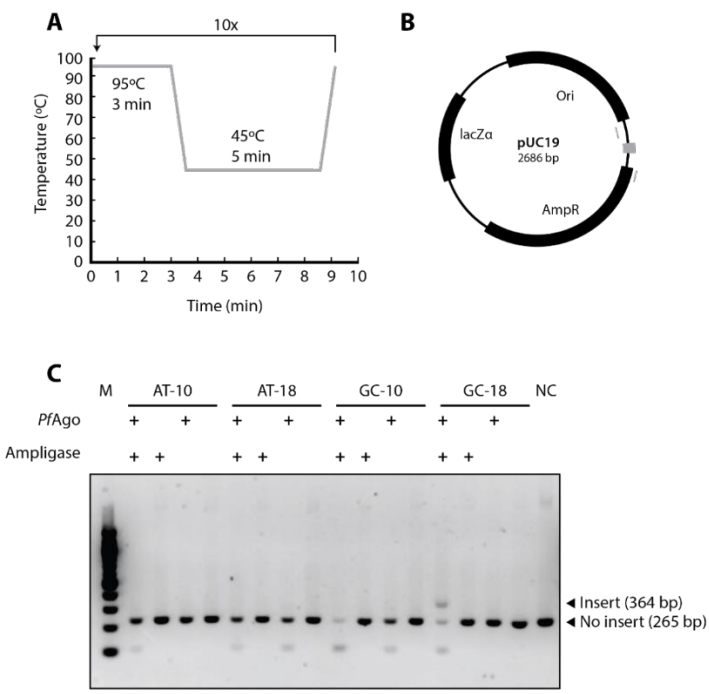


**Figure 4 | Plasmid assembly with Jason cloning was successful using overhangs of 18nt with a GC-content of 50%. A |** Schematic representation of the two step thermo-cycle. **B |** Schematic representation pUC19 target plasmid, insert indicated in grey together with the amplification primers. **C |** Correctly assembled plasmids were screened by PCR. Amplification of pUC19 gives a band of 264 bp, while the correctly assembled plasmid gives a band of 364 bp. The products of the PCR were resolved in a 2% agarose gel. AT-10/18: AT-rich overhang of 10nt or 18nt. GC-10/18, 10 or 18nt overhang with a GC content of 50%. NC: Negative control. M: ssDNA marker.

Unfortunately, prolonged incubation at 95°C (30 min in total) resulted in significant plasmid degradation by *Pf*Ago (**Figure 5A**), thereby lowering the plasmid assembly efficiency. Therefore, we designed new cycling programs in which the total incubation time at 95°C was reduced from 30 minutes, of the original cycle (now cycle 2.3), to 20 minutes (Cycle 2.2) and 10 minutes (Cycle 2.1). In addition, we decreased the amount of backbone vector used (120 ng, 180 ng, and original 300 ng).

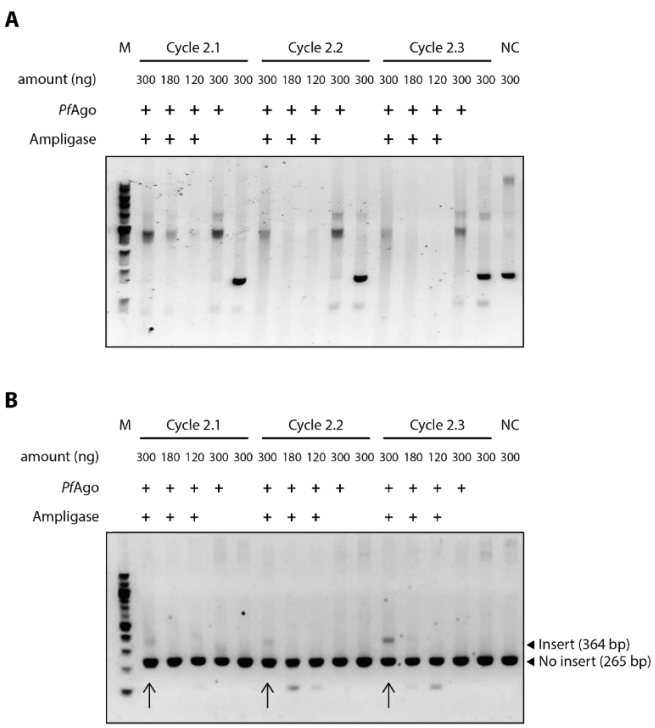



**Figure 5 | Plasmid degradation and plasmid assembly on combined cycle variation and amount of starting plasmid.** *Pf*Ago, the ligase, the premade insert and the ssDNA guides were incubated according to Cycle 2.1, 2.2, and 2.3 with different amounts of pUC19 (120 ng, 180 ng, and 300 ng). **A |** The direct products of the cycling reactions were resolved on a 2% agarose gel. NC: negative control. M: ssDNA marker. **B |** PCR was used to detect the presence of correctly assembled plasmids (band of 364 bp). NC: negative control = pUC19. M: DNA ladder.

The products of the cycling reactions were resolved on an agarose gel (**Figure 5A**) in order to assess the plasmid degradation. The amount of plasmid degradation positively correlates to an increased total incubation times at 95°C. Nevertheless, despite the degradation, successfully assembled plasmids could still be detected by PCR (364bp amplicon) after each of the three cycles (indicated by the arrows in **Figure 5B**). Assembly products of cycling program 2.1 were transformed into *Escherichia coli* DH10B cells. Colony PCR revealed that out of 6 colonies one positive clone yielded the correct amplicon (**Figure S1A**). The correct plasmid assembly was furthermore confirmed by sequencing of the plasmid-derived from this positive clone (**Figure S1B**).

Next, we tested if multiple thermo-cycles actually result in an accumulation of assembled plasmid as was hypothesized. The presence of assembled plasmid was detected by PCR (product of 265 bp) after an increasing number of thermo-cycles of a newly designed program (20 sec at 95°C followed by 2 min at 45°C), which further reduced the total incubation time at 95°C (~7 min) to avoid degradation. The amount of correctly assembled plasmids increases as the number of cycles increases. Until after 20-30 cycles the amount starts to go down, probably due to the aforementioned degradation of the plasmid by *Pf*Ago (**Figure 6B**). More optimization experiments were performed varying the cleavage and ligation temperature. By lowering the cleavage and ligation temperature to 90°C and 37°C, respectively, the yield of assembled plasmid seemed to further increase (**Figure 6C**). Additionally, we showed that yield of assembled plasmid improves further after the addition of a final ligation step up to 30 minutes (**Figure 6D**).

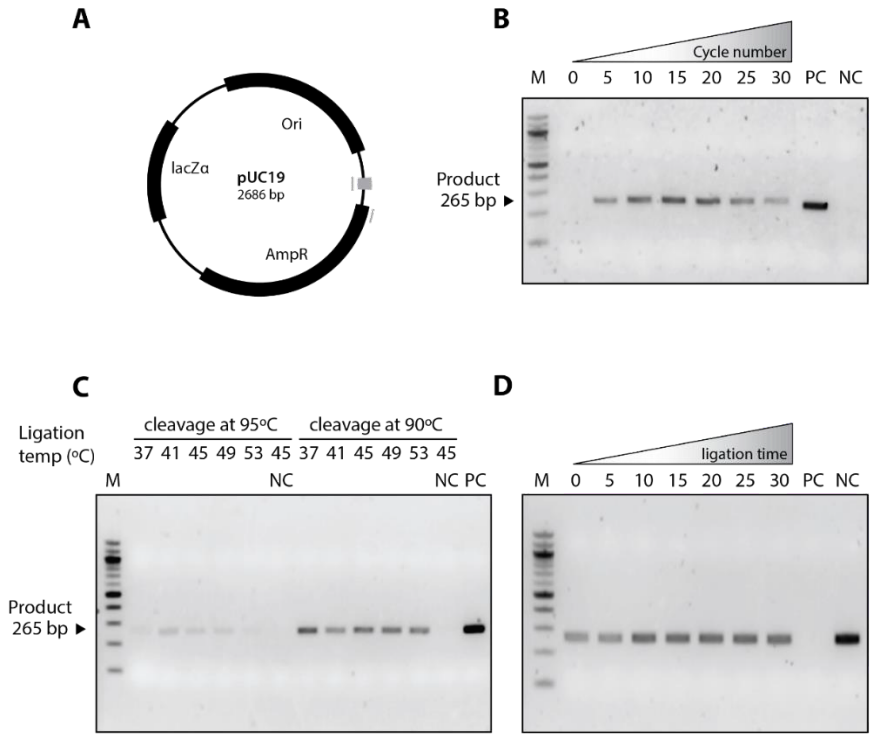



**Figure 6 | The effect of multiple thermo-cycles on the yield of assembled plasmid. A |** Plasmid map of pUC19. Primes that were used to amplify the assembled plasmid are shown in grey. In contrast to the primers indicated in **figure 4B**, one primer is now located on the insert itself. **B |** The presence of assembled plasmid was assessed with PCR. The effect of the cycle number (20 sec at 95°C followed by 2 min at 45°C) on the assembled plasmid yield. **C |** A cloning method of 30 cycles with different ligation and cleavage temperatures was tested. **D |** The effect of a final ligation step of different lengths was tested. The positive control (PC) is a pUC19 with the insert. Negative control (NC) is pUC19.

# Discussion

Taken together, we have demonstrated the basic principles of Jason cloning; a novel one-pot cloning method that collectively uses Argonaute of *P. furious* (*Pf*Ago) and the commercially available Ampligase™. A DNA insert of 100bp was successfully ligated into a pUC19 backbone vector by Jason cloning. After several optimization steps the highest cloning efficiencies were obtained with *Pf*Ago-linearized backbone vectors and inserts that had overhangs of 18nt (50% GC) at both ends and by using a two-step cycling program of 15 cycles, with each cycle a 20 seconds incubation step at 90°C followed by a 2 minute incubation step at 37°C. An additional ligation step of 10 minutes after the last cycle further improved the cloning efficiency.

Despite the success of this method, some aspects require further optimizations. The current cloning efficiencies of Jason cloning that are not yet comparable to those of other cloning methods, such as Gibson and Golden gate cloning. One of the likely causes for the low efficiency is the plasmid degradation that was observed at prolonged incubations at 95°C. In the absence of a guide some DNA targeting pAgos were shown to generate small DNA guides from plasmids via a dual activity termed "chopping" (Swarts et al., 2017a; Zander et al., 2017). Although *Pf*Ago was saturated with DNA guides throughout all the experiments, the degradation that was observed at 95°C might be due to guide free *Pf*Ago that has lost the guide in the process. This could potentially be avoided or minimized by crosslinking the DNA guide to *Pf*Ago, similar to what has been done for Argonaute from *Marinitoga piezophila* (*Mp*Ago) (Lapinaite et al., 2018).

Linearization of the backbone vector by *Pf*Ago appears to be highly dependent on the presence of AT-rich regions in the plasmid, as was reported for *Tt*Ago (Swarts et al., 2014b). The base pairing of AT-rich dsDNA is less tight than GC-rich dsDNA, consequently the individual DNA strands of AT-rich DNA are likely better accessible for *Pf*Ago. Improved cleavage efficiencies of GC-rich dsDNA would significantly enhance Jason Cloning. Currently, the DNA insert that is ligated into pUC19 consist of two pre-annealed DNA oligo's with premade overhangs. To allow the use of blunt ended inserts, such as PCR products, the feasibility of *Pf*Ago preloaded with an additional set of guides that are necessary to generate overhangs in the insert, should be explored.

Unlike Gibson cloning, we show that Jason cloning can assemble DNA fragments as small as 100nt and presumably as small as 20nt, although this has not been tested yet. Moreover, in contrast to Golden gate cloning, Jason cloning is restriction site independent and can therefore be used for insert fragments of any sequence. So, although Jason cloning requires additional optimizations to further increase its cloning efficiencies, the method has the potential to become an interesting alternative for current molecular cloning applications.

# Materials and Methods

## *Pf*Ago expression and purification

The *Pf*Ago gene was codon optimized for *E .coli* Bl21 (DE3) and inserted into a pET-His6 MBP TEV cloning vector (Addgene plasmid # 29656) using ligation-independent cloning. The *Pf*Ago protein was expressed in *E. coli* Bl21(DE3) Rosetta™ 2 (Novagen). Cultures were grown at 37°C in LB medium containing 50 µg ml-1 kanamycin and 34 µg ml-1 chloramphenicol till an OD600 nm of 0.7 was reached. *Pf*Ago expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. During the expression cells were incubated at 18°C for 16 hours with continuous shaking. Cells were harvested by centrifugation and lysed in lysis buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM imidazole, supplemented with an EDTA free protease inhibitor cocktail tablet (Roche). The soluble fraction of the lysate was loaded on a nickel column (HisTrap Hp, GE healthcare). The column was extensively washed with wash buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 30 mM imidazole. Bound protein was eluted by increasing the concentration of imidazole in the wash buffer to 250 mM. The eluted protein was dialyzed at 4°C overnight against 20 mM HEPES pH 7.5, 250 mM KCl, and 1 mM dithiothreitol (DTT) in the presence of 1mg TEV protease to cleave of the His6-MBP tag. Next, the cleaved protein was diluted in 20 mM HEPES pH 7.5 to lower the final salt concentration to 125 mM KCl. The diluted protein was applied to a heparin column (HiTrap Heparin HP, GE Healthcare), washed with 20 mM HEPES pH 7.5, 125 mM KCl and eluted with a linear gradient of 0.125-2 M KCl. Next, the eluted protein was loaded onto a size exclusion column (Superdex 200 16/600 column, GE Healthcare) and eluted with 20 mM HEPES pH 7.5, 500 mM KCl and 1 mM DTT. Purified *Pf*Ago protein was diluted in size exclusion buffer to a final concentration of 5 µM. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

## Activity assays of Ampligase

In one reaction 25U of Ampligase (Epicentre) were mixed with 5pM of Oligo A, B, and C (**Table S1**), 5µL 10X Ampligase Reaction Buffer containing 200 mM Tris-HCl pH 8.3, 250mM KCl, 100mM MgCl2, 5mM NAD and 0.1% Triton® X-100 (Epicentre), filled up with ultrapure water to a total reaction volume of 50 µL. For the positive controls the Ampligase and the Ampligase buffer were exchanged for, respectively, 400U of T4 DNA Ligase and 10X T4- ligase buffer containing 50mM Tris-HCl pH 7.5, 10mM MgCl2, 1mM ATP, 10mM DTT (New England Biolabs) in a total volume of 20µL. The incubation time and temperature varied as indicated in the figures. After incubation, Loading Buffer (95% (deionized) formamide, 5mM ethylenediaminetetraacetic acid, 0.025% sodium dodecyl sulphate, 0.025% Bromophenol blue and 0.025% xylene cyanol) was mixed with EDTA and SDS to a final concentration of 25µM EDTA and 1% SDS. The Modified Loading Buffer was added in a 1:1 ratio to the samples, followed by a 10 min incubation at 95°C. Next the samples were resolved on 15% denaturing polyacrylamide gels (7M Urea). A custom DNA marker was made by mixing BG4263 and BG13198 to a final concentration of 5uM. The gels were stained using SYBR gold Nucleic Acid Gel Stain (Invitrogen) and visualized using a G:BOX Chemi imager (Syngene).

## Activity assays of *Pf*Ago

For the cleavage kinetics assay 0.5 µl of *Pf*Ago (6 µM) was mixed with 7.25µl of 40mM Tris-HCl (pH 8), 2.25µl of 1M KCl, 1µl of 5mM MnCl2, 0.6 µl of 5µM ssDNA guide BG10650, 0.6µl of 5µM ssDNA guide BG10634, 300ng of pUC19, and ultrapure water to a final reaction volume of 20µl. The reaction was incubated at 95°C for 0, 0.5, 1, 2, 3, 4, and 5 minutes. After incubation, 2.5µl of Proteinase K (Ambion) and 2.5µL of 50mM CaCl2 were added and the mix was incubated at 65°C for 1 hour. Nucleic acids were resolved on 0.8% agarose gels pre-stained with SYBR gold Nucleic Acid Gel Stain (Invitrogen) and visualized using a G:BOX Chemi imager (Syngene). As a marker, 1 kb DNA Ladder (New England Biolabs) was used, EcoRI-HF (New England Biolabs) linearized pUC19 was used as a linear marker, and untreated pUC19 was used as a supercoiled marker.

### Insert pre-assembly

To generate each insert with the pre-made overhangs, two ssDNA sequences were ordered from IDT for each insert. The sequences were fully complementary with the exception of the flanking overhangs. The ssDNA strands were diluted to 5 µM and mixed in equimolar amounts in ultrapure water. The mix was incubated at 94°C for 5 minutes, and cooled down to 20°C at 0.5°C per second.

### Cycling assays

For the cycling assays, 0.5 µl of *Pf*Ago (6 µM) were mixed with 0.6 µL of 5 µM of the forward and reverse ssDNA guides, 1 µl of 5 U/µl Ampligase (Epicentre), 2 µl of 10X Ampligase Reaction Buffer (Epicentre), 300 ng pUC19 (unless noted otherwise), insert at a molar ratio of insert to plasmid of 7:1, and ultrapure water to a final volume of 20 µl. Note that in different experiments incubation temperature, time, and initial plasmid mass were varied (indicated in the figures of the corresponding experiments).

### Assembly screening

For screening of the correctly assembled plasmid, a PCR reaction was carried out using Q5® High-Fidelity 2X Master Mix (New England Biolabs), and the products were resolved on 2% agarose gels pre-stained with SYBR gold Nucleic Acid Gel Stain (Invitrogen) and visualized using a G:BOX Chemi imager (Syngene). As a marker, 100 bp DNA Ladder (New England Biolabs) was used, and as a negative control, PCR amplification of untreated pUC19 was used. Two pairs of primers were used for PCR. The initial pair was BG12878 and BG12879, with amplification bands of 264 bp for pUC19 and 364 bp for the correctly assembled plasmid, and the final pair was BG12878 and BG13421, which only amplified the correctly assembled plasmid with a band of 265 bp.

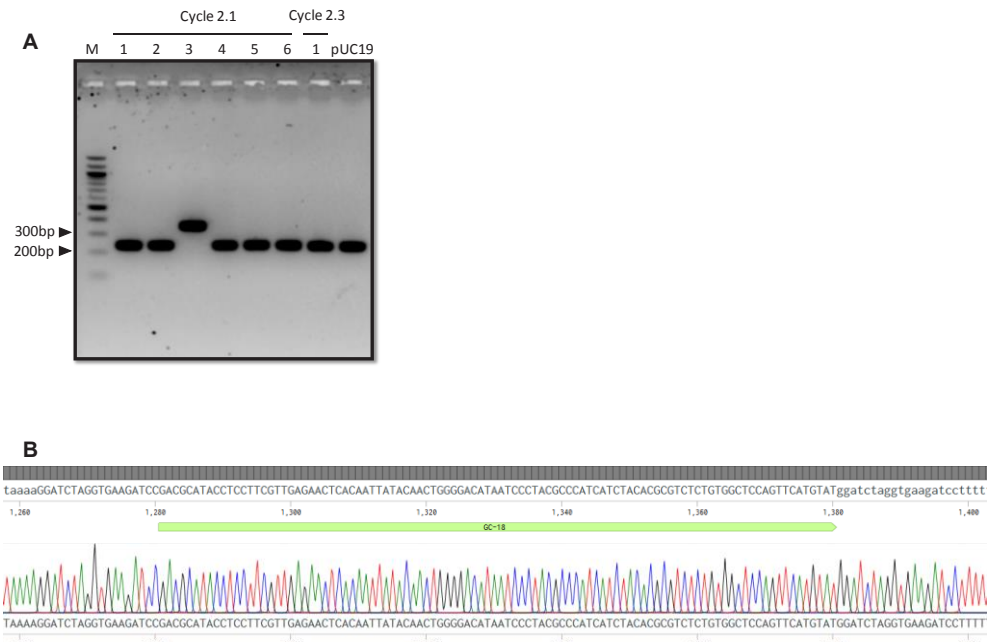


### Plasmid isolation and transformation

For pUC19, the plasmid containing *E. coli* DH10B was grown in LB medium supplemented with 1:1000 Ampicillin, overnight at 37°C with shaking, and pUC19 was isolated using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) or ZymoPURE II Plasmid Midiprep Kit and eluted in ultrapure water. For the correctly assembled plasmid from the cycling reaction, the cycle reaction was transformed into *E. coli* DH10B cells and plated in

Luria-Bertani agar medium supplemented with 1:1000 Ampicillin and incubated at 37°C for 24 hours. The obtained colonies were transferred onto a new plate and re-cultured in Luria-Bertani agar medium supplemented with 1:1000 Ampicillin and incubated at 37°C overnight. The colonies were then used to inoculate liquid Luria-Bertani medium and incubated at 37°C overnight. The colonies were assessed by PCR for correct assembly and the correct ones sent for sequencing.

# Supplementary information




**Supplementary Figure S1A |** Colony PCR on six colonies obtained after Jason cloning. Amplification of pUC19 gives a band of 264 bp, while the correctly assembled plasmid gives a band of 364 bp. **B |** Sequencing results of the correctly assembled plasmid (#4). Assembly performed using Benchling.

**Supplementary table S1 | Oligonucleotides used in the ligation feasibility assay.**

| Oligo | Sequence (5' to 3') | Length | Description |
|---|---|---|---|
| **BG13198** | TTTACTCATATATACTTTAGATTGATTTAAAAACGTA CTGAGCTAGGTAC | 50 | Oligo A, Marker for the ligation feasibility assay |
| **BG13199** | CTTCATTTTTAATTTAACTGCATCTAGGTGGAGATCC TTTTTGATAATCA | 50 | Oligo B |
| **BG13200** | TGATTATCAAAAAGGATCTCCACCTAGATGCAGTTAA ATTAAAAATGAAGGTACCT | 56 | Oligo C6 |
| **BG13201** | TGATTATCAAAAAGGATCTCCACCTAGATGCAGTTAA ATTAAAAATGAAGGTACCTAGCT | 60 | Oligo C10 |
| **BG13202** | TGATTATCAAAAAGGATCTCCACCTAGATGCAGTTAA ATTAAAAATGAAGGTACCTAGCTCAGT | 64 | Oligo C14 |
| **BG13203** | TGATTATCAAAAAGGATCTCCACCTAGATGCAGTTAA ATTAAAAATGAAGGTACCTAGCTCAGTACGT | 68 | Oligo C18 |
| **BG4263** | TCGACTTTATATTTAAATAATTTAATATACTATACAA CCTACTACCTCGTATAAATTTTTAAATAAATATTGCA TTCAAGCTTTTAATTTAATTAAAT | 98 | Marker for the ligation feasibility assay |

**Supplementary table S2 | Guides used for cleavage of pUC19 during the cleavage kinetics assay and the cycling reaction assay.**

| Guide | Sequence (5' to 3') | Length | Description |
|---|---|---|---|
| **BG10634** | TTAGATTGATTTAAAACTTCA | 21 | pUC19 RV Guide Kinetics assay |
| **BG10650** | AAAAATGAAGTTTTAAATCAA | 21 | pUC19 FW Guide Kinetics assay |
| **BG12743** | TAGATTGATTTAAAACTTCAT | 21 | pUC19 10nt FW guide (AT-rich) |
| **BG12744** | AATCAATCTAAAGTATATATG | 21 | pUC19 10nt RV guide (AT-rich) |
| **BG13276** | GTGAAGATCCTTTTTGATAAT | 21 | pUC19 10nt FW guide (GC-rich) |
| **BG13277** | ACCTAGATCCTTTTAAATTAA | 21 | pUC19 10nt RV guide (GC-rich) |
| **BG13268** | TTGATTTAAAACTTCATTTTT | 21 | pUC19 18nt FW guide (AT-rich) |
| **BG13269** | AATCTAAAGTATATATGAGTA | 21 | pUC19 18nt RV guide (AT-rich) |
| **BG13272** | GTGAAGATCCTTTTTGATAAT | 21 | pUC19 18nt FW guide (GC-rich) |
| **BG13273** | ACCTAGATCCTTTTAAATTAA | 21 | pUC19 18nt RV guide (GC-rich) |

**Supplementary table S3 | Oligonucleotides used for premade insert generation.**

| Oligo | Sequence (5' to 3') | Length | Description |
|---|---|---|---|
| **BG13266** | TAGATTGATTGACGCATACCTCCTTCGTTGAGAACTCACAA TTATACAACTGGGGACATAATCCCTACGCCCATCATCTACA CGCGTCTCTGTGGCTCCAGTTCATGTAT | 110 | Insert 10nt overhang FW strand (AT-rich) |
| **BG13267** | AATCAATCTAATACATGAACTGGAGCCACAGAGACGCGTGT AGATGATGGGCGTAGGGATTATGTCCCCAGTTGTATAATTG TGAGTTCTCAACGAAGGAGGTATGCGTC | 110 | Insert 10nt overhang RV strand (AT-rich) |
| **BG13278** | CTAGGTGAAGGACGCATACCTCCTTCGTTGAGAACTCACAA TTATACAACTGGGGACATAATCCCTACGCCCATCATCTACA CGCGTCTCTGTGGCTCCAGTTCATGTAT | 110 | Insert 10nt overhang FW strand (GC-rich) |
| **BG13279** | CTTCACCTAGATACATGAACTGGAGCCACAGAGACGCGTGT AGATGATGGGCGTAGGGATTATGTCCCCAGTTGTATAATTG TGAGTTCTCAACGAAGGAGGTATGCGTC | 110 | Insert 10nt overhang RV strand (GC-rich) |
| **BG13270** | ACTTTAGATTGATTTAAAGACGCATACCTCCTTCGTTGAGA ACTCACAATTATACAACTGGGGACATAATCCCTACGCCCAT CATCTACACGCGTCTCTGTGGCTCCAGTTCATGTAT | 118 | Insert 18nt overhang FW strand (AT-rich) |
| **BG13271** | TTTAAATCAATCTAAAGTATACATGAACTGGAGCCACAGAG ACGCGTGTAGATGATGGGCGTAGGGATTATGTCCCCAGTTG TATAATTGTGAGTTCTCAACGAAGGAGGTATGCGTC | 118 | Insert 18nt overhang RV strand (AT-rich) |
| **BG13274** | GGATCTAGGTGAAGATCCGACGCATACCTCCTTCGTTGAGA ACTCACAATTATACAACTGGGGACATAATCCCTACGCCCAT CATCTACACGCGTCTCTGTGGCTCCAGTTCATGTAT | 118 | Insert 18nt overhang FW strand (GC-rich) |
| **BG13275** | GGATCTTCACCTAGATCCATACATGAACTGGAGCCACAGAG ACGCGTGTAGATGATGGGCGTAGGGATTATGTCCCCAGTTG TATAATTGTGAGTTCTCAACGAAGGAGGTATGCGTC | 118 | Insert 18nt overhang RV strand (GC-rich) |
| **BG12878** | GTTATCTACACGACGGGGAGT | 21 | PCR FW pUC19 |
| **BG12879** | CTTTTCTACGGGGTCTGACG | 20 | PCR RV pUC19 |
| **BG13421** | TGAGTTCTCAACGAAGGAGG | 20 | PCR RV insert |

# Chapter 7

Summary and general discussion

# Thesis summary

Argonaute proteins (Ago) constitute a diverse group of nucleic acid-guided enzymes that are present in every domain of life and are important regulators of intracellular nucleic acid levels (**Chapter 1**). This thesis is divided into two parts (a fundamental and an applied part) and focuses on Argonaute proteins from prokaryotes (pAgo). The fundamental part describes the combined biochemical and biophysical characterization of a newly discovered pAgo from *Clostridium butyricum* (*Cb*Ago). The applied part describes the development of two novel molecular applications which are based on the previously characterized pAgos from *Thermus thermophilus* (*Tt*Ago) and *Pyrococcus furiosus* (*Pf*Ago).

All pAgos that have been studied experimentally thus far are summarized in **chapter 2**. Those studies mainly focused on long pAgo variants that contain the same structural domain architecture (N-PAZ-MID-PIWI) as Agos in eukaryotes (eAgos). The N-domain facilitates the dissociation of cleaved target strands, the PAZ and MID domains form binding pockets for the 3' and 5' end of the guide nucleic acids, respectively, and the PIWI domain of catalytically active pAgos contains a conserved DEDX tetrad (where X represents a D, H or N) that mediates cleavage of target nucleic acid strands. Whereas eAgos strictly target RNA, certain long pAgos, such as *Tt*Ago and *Pf*Ago, target DNA instead. This DNA-targeting behavior raised the idea to repurpose such pAgos as genome editing tools. However, currently two major drawbacks prevent them from being utilized for such applications, which are their low nuclease activity on double stranded DNA targets (due to lack of helicase activity) and their low activity at temperatures relevant for genome editing of mesophilic organisms (37°C). The main reason for the latter issue is that the catalytically active DNA-targeting pAgos that have been described thus far, are exclusively derived from thermophilic organisms. To try and overcome these hurdles we started a quest for catalytically active pAgos from mesophilic organisms.

This resulted in the discovery of *Cb*Ago, whose biochemical characterization is described in **chapter 3**. *Cb*Ago is a long pAgo that is phylogenetically closely related to *Tt*Ago, which is also reflected in the structural and functional similarities between both. Like *Tt*Ago, *Cb*Ago acquires small plasmid-derived DNA guides upon heterologous expression in *Escherichia coli. In vitro*, small DNA guides could be used to efficiently cleave complementary ssDNA targets, suggesting that similar to *Tt*Ago, *Cb*Ago may be involved in host-defense. Despite

the similarities, *Cb*Ago is highly active at moderate temperatures (37°C) and displays a slight preference for DNA guides with a deoxyadenosine at the 5'-end and for deoxythymidines in the downstream sub-seed segment (guide nucleotides 2-4), both *in vivo and in vitro*. Although *Cb*Ago could be reprogrammed with any synthetic DNA guide to cleave both ssDNA and AT-rich (AT-content ≥65%) dsDNA at 37°C, preliminarily genome editing experiments with *Cb*Ago in mammalian cells were not yet successful. Nevertheless, the characterization of a mesophilic DNA-targeting pAgo can be seen as an important step towards the development of potential pAgo-based genome editing tools.

The pool of nucleic acids that *Cb*Ago-guide complexes have to scan before finding a complementary target is much bigger inside cells than during the *in vitro* cleavage assays. This might explain why the *in vitro* assays were successful while the genome editing attempt in mammalian cells was not. To improve our understanding of the target search mechanism of *Cb*Ago, we used single molecule assays, described in **chapter 4**. While scanning the target DNA, *Cb*Ago combines a search mechanism termed 'gliding' with intersegmental jumping. During gliding, *Cb*Ago diffuses in lateral direction while maintaining loose contact with the target DNA. In order to bypass obstacles such as protein roadblocks or local double stranded DNA regions, *Cb*Ago is capable of making intersegmental jumps after which it continues gliding until it finds complementary DNA targets. The fact that *Cb*Ago jumps over small regions of double stranded DNA underscores the inability of *Cb*Ago to unwind dsDNA. This suggests that *Cb*Ago merely probes DNA that is single stranded, and thus limits its use as genome editing tool. Although single stranded DNA is rare inside the cell, it does occur, especially during infections of mobile genetic elements, such as plasmid or viruses, which would be in agreement with the hypothesis that *Cb*Ago is involved in host defense.

The second half of this thesis covers two newly developed pAgos-based molecular applications. For this *Tt*Ago and *Pf*Ago were repurposed, which both mediate DNA-guided DNA targeting and have their optimum temperature above 65°C. **Chapter 5** describes the use of *Tt*Ago in a novel diagnostic application termed, NAGIVATER (enriched **N**ucleic **A**cids **Vi**a DNA-**G**uided **A**rgonaute of ***T**hermus th**er**mophilus*)**.** NAVIGATER allows to enrich rare nucleic acids that differ from highly abundant background nucleic acids by as little as one nucleotide. *Tt*Ago, programmed with a DNA guide, eliminates abundant background nucleic acids (RNA and DNA) with nucleotide precision while sparing rare nucleic acids, allowing these to be specifically amplified with PCR. NAVIGATER acts as a

pre-treatment for established rare nucleic acid detection methods, such as digital droplet PCR (ddPCR) and Peptide Nucleic Acid-Mediated PCR (PNA-PCR). We demonstrated that NAVIGATE could improve the limit of detection of ddPCR, PNA PCR and Sanger sequencing (up to a 60 fold) for the detection of KRAS G12D mutant alleles in blood samples of several pancreatic cancer patients.

**Chapter 6** describes a novel cloning method termed 'Jason cloning' that combines *Pf*Ago and a thermostable ligase (Ampligase™) in a one tube. The method consists of a two-step reaction. The first step is performed at 95°C and requires two *Pf*Ago-guide complexes that each cleaves one of the two strands of backbone plasmid or PCR product, such that it leaves behind overhangs of 18nt. A second step at 45°C enables Ampligase™ to seal the two nicks that are formed upon the hybridization of an insert (two pre-annealed DNA oligonucleotides) with overhangs compatible to the backbone product, cleaved by *Pf*Ago. By repeating those two steps multiple times the reaction is driven towards the correct assembly of the plasmid. Although, the assembly efficiencies of Jason cloning are not as high as the efficiencies well-known previously established methods (e.g. Gibson assembly® and Golden Gate cloning), Jason cloning certainly has potential to become an interesting alternative when further improvements are realized.

To conclude, this PhD project has led to the discovery of a novel DNA-targeting pAgo from the mesophilic bacterium *Clostridium butyricum*. Research on *Cb*Ago improved our molecular understanding on both the catalytic and target search mechanisms of pAgos. Although *Cb*Ago programmed with a DNA guide could cleave target DNA at 37°C *in vitro*, preliminary genome editing experiments with *Cb*Ago in mammalian cells were not yet successful and require further experimentation. In addition, the well-studied DNA targeting *Tt*Ago and *Pf*Ago, were successfully used for the development of two novel pAgo-based diagnostic and molecular applications.

# General discussion

In contrast to the well-studied eukaryotic Argonaute proteins (eAgos), the field that studies its prokaryotic counterparts is still in its infancy. In this final chapter, the fundamental work on *Cb*Ago is put in perspective by comparing it with other characterized Agos. Updated with the findings described in this thesis, it gives an overview of our current understanding on each of the mechanistic steps that Ago proteins undertake to achieve RNA or DNA interference. The last part discusses the novel pAgo-based molecular applications that we developed and the hurdles that still need to be overcome in the near future.

## Prokaryotic Argonaute diversity

Nearly 10 years after the discovery of the existence of Argonaute homologs in bacteria and archaea (Cerutti et al., 2000), a bioinformatics study by Makarova and co-workers revealed that the diversity of pAgos highly exceeds the diversity of eAgos in terms of domain architecture (Makarova et al., 2009). Besides, pAgo variants that comprise of the same N-PAZ-MID-PIWI domains as in eAgo (long pAgos), also many truncated pAgos that lack or have substituted N and PAZ domains (short pAgos) were found (**Chapter 2**). Phylogenetic analyses on sequence alignments of the conserved MID and PIWI domains suggested that the long pAgos could be subdivided into two clades; a clade of "long pAgos" that are mainly catalytically active and a clade of long pAgos in which, during evolution, one or more catalytic residues were substituted and consequently they lost their catalytic behavior. Based on the same sequence alignments, the evolutionary journey of pAgos could be reconstructed. Currently, two possible scenarios of the journey exist (Swarts et al., 2014a), both of which start with an ancient long active pAgo precursor (**Figure 1**). In the first and most likely scenario this precursor lost one or more of its catalytic residues giving rise to inactive "non-slicing" pAgos. The loss of catalytic site residues must have occurred at multiple independent occasions as also in the clade of mainly long active pAgos some pAgos exists that have lost their catalytic behavior. These non-slicing pAgos successively lost their N-PAZ domain resulting in short pAgos (**Figure 1; Scenario 1).** This is likely because none of the short-Agos contains a conserved active site. Alternatively, short pAgos first diversified from long pAgos after which these short and some long variants lost their catalytic behavior at multiple independent occasions (**Figure 1; Scenario 2**).

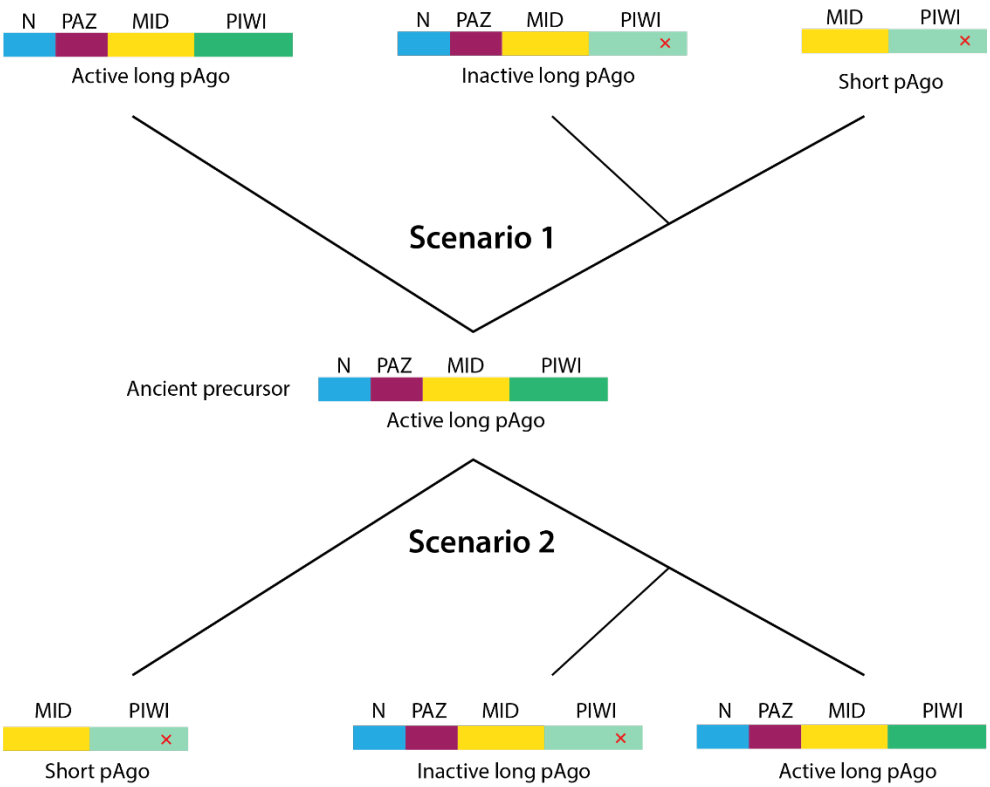



**Figure 1 | The two possible evolutionary journeys of pAgo proteins**. It should be noted that in the clade of long pAgos some truncated Agos (e.g. Argonaute of *Archaeoglobus fulgidus, AfAgo*) exist and that, similarly, the clade of catalytic active pAgos contains several inactive pAgos. The light green PIWI-domain with a red cross indicates catalytic inactive PIWI-domains, typical for non-slicing pAgos.

By 2015, a total of 261 non-redundant pAgo proteins were found in all sequenced bacterial and archaeal genomes (Swarts et al., 2014a). Combined with a study in 2013 by Burroughs and colleagues that identified a third group of pAgos, termed PIWI-RE, which are typified by two uniquely conserved arginine (R) and glutamic acid (E) residues in the PIWI domain (Burroughs et al., 2013), the pAgo family could roughly be divided into three classes: long pAgos, short pAgos and PIWI-RE (**Chapter 2**). While writing this chapter, a bioinformatics study was published that tripled the number of available pAgos sequences (721 pAgos) compared to the previous studies in 2015 (Ryazansky et al., 2018). However, despite the increase in pAgos sequences, no novel pAgo variants were discovered. In this thesis we have mainly focused on the active long pAgos variants.

## Function and mechanism of pAgos

Although the evolutionary journey of Argonaute proteins (Agos) resulted in an enormous diversity in terms of domain architecture and functionality, the fundamental and mechanistic properties of Agos are well conserved (Lisitskaya et al., 2018; Swarts et al., 2014a; Willkomm et al., 2018). In general, Agos are important regulators of intracellular nucleic acid levels by using single stranded nucleic acid guides to specifically bind and, depending on the presence of an intact catalytic site, cleave complementarity nucleic acid targets (Lisitskaya et al., 2018; Swarts et al., 2014a; Willkomm et al., 2018). In eukaryotes Ago forms the principal component of the RNA Induced Silencing Complex (or RISC) that is involved in RNA interference pathways, whereas in prokaryotes Ago is involved in host defense against DNA invaders, such as plasmids (Olovnikov et al., 2013; Swarts et al., 2015a, 2014b). Structural and biochemical studies that elucidated the role and mechanisms of pAgos focused mainly on long pAgos variants due to their structural similarities with the well-studied eAgos (**chapter 2**). In fact, the first Ago protein of which the structure was solved belongs to the long pAgos (Song et al., 2004), as eAgo proteins were difficult to obtain in the quantities required for X-ray crystallography (Parker, 2010). In this part of the chapter we outline our current view, updated with findings described in this thesis, on each of the mechanistic steps that long pAgo proteins have to go through to achieve RNA or DNA Interference.

## Guide generation

Protein homologs that are responsible for the guide RNA generation of eukaryotic Agos, such as Drosha and Dicer, are absent in prokaryotes (Shabalina and Koonin, 2008). Nonetheless, bacterial Agos of *Rhodobacter sphaeroides* (*Rs*Ago), *Thermus thermophilus* (*Tt*Ago) and *Cb*Ago co-purify with guides upon heterologous expression in *Escherichia coli* (**Chapter 3**; Olovnikov et al., 2013; Swarts et al., 2014b). This indicated that guide processing either is performed by pAgo itself or by common host factors in *E. coli*. Recently, the former hypothesis was confirmed for *Tt*Ago and *Mj*Ago (Ago from *Methanocaldococcus jannaschii*) as both pAgos were shown to autonomously generate functional DNA guides from long dsDNA precursors *in vitro* via a mechanism termed "chopping" (Swarts et al., 2017a; Zander et al., 2017). Similarly, *Cb*Ago and *Pf*Ago exhibit DNA chopping activity (**Chapter 3 and 6**; Swarts et al., 2015a). However, the severity of chopping in the individual experiments

varied. *Mj*Ago completely degraded a target plasmid via chopping (Zander et al., 2017), while *Tt*Ago and *Pf*Ago only linearized circular supercoiled target plasmids without further degradation (Swarts et al., 2015a, 2014b). Chopping by *Cb*Ago was even less severe. In the absence of guides, *Cb*Ago used the relatively open supercoiled topology of the target plasmid to presumably nick one of the two target DNA strands (**Chapter 3**). Since all these pAgos mediate DNA-guided cleavage of ssDNA *in vitro*, the most likely determinant that causes the different outcomes of chopping activity on target plasmids is the variation in incubation temperature at which the chopping experiments were performed. Due to the nature of the host of these proteins, *Cb*Ago, *Tt*Ago, *Pf*Ago and *Mj*Ago were incubated at 37°C, 65°C, 75°C and 85°C, respectively (**Chapter 3**, Swarts et al., 2014b, 2015, 2017; Zander et al., 2017). This suggests that higher incubation temperature results in more severe chopping of the target plasmid. As higher temperatures generally result in more DNA unwinding, this theory would agree with the observation that chopping by *Tt*Ago heavily depends on the presence of unwound double stranded target DNA (Swarts et al., 2017a): during *in vitro* assays, at higher temperatures there is more thermal breathing (partial unwinding) of the DNA which would result in more chopping. In thermophiles, the topology of cellular DNA is strongly positively coiled due to reverse gyrase activity (Forterre et al., 1996). As a result the accessibility of dsDNA in thermophiles is comparable to that of the dsDNA accessibility in mesophiles. Intracellular DNA unwinding in thermophilic and mesophilic organisms occurs through natural unwinding processes such as replication and transcription. Presumably due to the lack of these processes outside the cell, *Cb*Ago did not generate functional guides through chopping *in vitro,* whereas in *E. coli* it did. Hence, this would suggest that for DNA chopping *Cb*Ago, but possibly also *Mj*Ago and *Tt*Ago, rely on common host factors, such as helicases and polymerases.

Supporting evidence for the importance of DNA unwinding for guide generation was found when mapping the plasmid-derived guides that co-purified with *Cb*Ago *in vivo,* against the expression plasmid (**Figure 2**). Nearly 81% (34/42) of the guides were derived from an open reading frame (ORF), whereas only 70% of the plasmid consists of ORFS. Furthermore the average GC content of the guides mapping the plasmid was 37.2% whereas the average GC-content of the plasmid was 47.8%. Combined, this indicates that a majority of the guides acquired by *Cb*Ago are preferably acquired from ORF regions that are low in GC content.

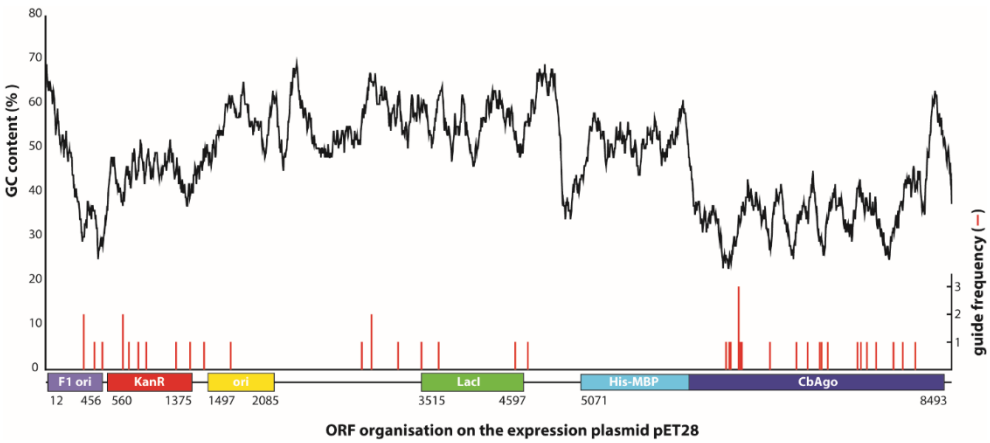



**Figure 2 | A majority of the co-purified guide sequences map at the AT-rich regions of the open reading frames of the expression plasmid.** The red lines indicate the location where the co-purified guide was mapped on the expression plasmid. The colored boxes represent the genes on the plasmid. The black line graph displays the GC-content along the plasmid.

Similar to other DNA targeting pAgos, the co-purification of plasmid-derived guides suggests the involvement of *Cb*Ago in host defense against invading DNA. Generally DNA invaders such as viruses and plasmids tend to have a GC-content that is on average 4%-10% lower than the host (Nishida, 2012; Rocha and Danchin, 2002). This difference in GC-content was proposed to possibly be one of the determinants that *Tt*Ago employs to discriminate between host and invading DNA (Swarts et al., 2014b). As the genome of the native host of *Cb*Ago, *Clostridium butyricum*, has a GC-content of only 28.5% (Dwidar et al., 2012), it might deal with DNA invaders that are low in GC-content as well. Possibly *Cb*Ago evolved as an pAgo with a specific preference for AT-rich DNA targets, which could explain that guides were mainly derived from AT-rich plasmid regions. Alternatively, the preference could also be explained by the fact that these regions are more prone to DNA unwinding. Characterizations of catalytic active pAgos derived from a mesophilic organisms that have a GC-rich genome would therefore be interesting candidates for future research.

## Guide anchoring

The numerous crystal structures of binary (Ago with a guide) and ternary (Ago with a guide and target) Ago complexes have significantly contributed to our current understanding of

their cleavage mechanism (Boland et al., 2010; Doxzen et al., 2017; Frank et al., 2010; Kaya et al., 2016; MacRae et al., 2012; Matsumoto et al., 2016; Mi et al., 2008; Miyoshi et al., 2016; Nakanishi et al., 2012; Sheng et al., 2014; Wang et al., 2008a, 2008b, 2009; Willkomm et al., 2017a). These structures revealed that Ago separates its guide sequence into different functional segments, each contributing to a specific task during the binding/cleavage process (Wee et al., 2012). The first guide nucleotide at the 5'-end, often phosphorylated, forms the anchor that is bound into the binding pocket of the MID domain on the border with the PIWI domain. Inside this pocket the 5'-end phosphate of the guide is recognized by four well-conserved residues (YKQK). Most pAgos use these conserved residues to coordinate a cation that interacts with the 5'-phoshorylated terminus of the guide (Ma et al., 2005; Parker et al., 2005; Sheng et al., 2014). In contrast, eAgos almost exclusively use a conserved positively charged lysine for the binding of the 5'-phosphate (MacRae et al., 2012; Nakanishi et al., 2012). Recently one noticeable exception was found as the PIWI-clade Ago of silkworms (SIWI) was shown to also exhibits metal-dependent 5'-end guide phosphate recognition, similar to pAgos (Matsumoto et al., 2016).

An amino acid sequence alignment of *Cb*Ago with other pAgos indicated that *Cb*Ago contains the same conserved residues (YKQK), which could also be traced back in the 5'-end binding pocket of its ternary structure (**Chapter 3**). However, despite the presence of these conserved residues that specifically interact with the 5'-end phosphate of the guide (**Figure 2**), *Cb*Ago is also able utilize guides with a hydroxylated 5'-end. This was previously also observed for *Pf*Ago, *Tt*Ago (**Chapter 5 and 6**) and *Mj*Ago (Zander et al., 2017). However, the cleavage efficiencies of these pAgos with hydroxylated guides appears to be significantly lower than with phosphorylated guides.

By now several prokaryotic Ago variants have shown specific preferences for a 5'-end nucleotide even though this nucleotide does not participate in base pairing with the target. Similarly, also eukaryotic Argonautes show specific preferences for 5'-end guide nucleotides. Human Ago 2 (hAgo2) uses a structural feature in the MID domain, termed nucleotide specificity loop (NSL), to preferably sort RNA guides with a 5'-end uracil (Frank et al., 2010). For *Tt*Ago the preferred 5'-end cytidine is a consequence of the guide loading process, during which *Tt*Ago preferentially binds dsDNA with a guanine at the 3'-end of the passenger (target) strand, opposite to the 5'-end cytosine of the guide strand (Swarts et al., 2017a). *Tt*Ago encompasses a special binding pocket formed by the PIWI domain and the L2

loop to accommodate the guanine of the passenger strand. Also the preference of *Rs*Ago for RNA guides with a 5'-end uracil could be explained by the same underlying mechanism (Liu et al., 2018). However, in the structure of *Cb*Ago the binding pocket is visible but the thymidine of the target strand that is opposed to the preferred 5'-adenine of the guide did not show any specific interactions with the pocket. In fact, the thymidine was flipped away from the binding pocket (**Chapter 3**). This might hint towards the presence of alternative guide loading mechanism and thus is more research required to elucidate the origin of the preference of *Cb*Ago for guides with a 5'-end adenine.

**Table 1 | Amino acid alignment of MID domain regions of characterized pAgos.** Structures of hAgo2, *Tt*Ago, *Rs*Ago reveal that the well conserved Y/R, K, Q, K amino acids contribute to the recognition of the 5' end phosphate in the binding pocket of the MID-domain.

| | | Y/R K Q K | |
|---|---|---|---|
| hAgo2 | -*Homo sapiens* | TPV**Y**AEV**K**RVGDTVLGMAT**Q**CVQ \| INV**K**LGC | |
| CbAgo | -*Clostridium butyricum* | ENP**Y**NPF**K**KVWAKL-NIPS**Q**MIT \| ILG**K**IGG | 5'-end $PO^{3-}$ |
| RsAgo | -*Rhodobacter spaeroides* | RNP**Y**IHT**K**SLLLTL-GVPT**Q**QVR \| TYA**K**LNG | |
| AfAgo | -*Archaeoglobus fulgidus* | TPL**Y**YKL**K**SYLI-N-SIPS**Q**FMR \| FVS**K**LGG | |
| MjAgo | -*Methanocaldococcus jannaschii* | NDY**Y**EIL**K**KQLFDL-KIIS**Q**NIL \| IMG**K**LGI | |
| PfAgo | -*Pyrococcus furiosus* | SEKFEEI**K**RRLFNL-NVIS**Q**VVN \| VLS**K**LGV | |
| TtAgo | -*Thermus thermophilus* | WED**R**NRL**K**ALLLRE-GLPS**Q**ILN \| LLA**K**AGL | |
| NgAgo | -*Natronobacterium gregoryi* | TET**Y**DEL**K**KALANM-GIYS**Q**MAY \| LLAAAGG | Extra N-terminal OB-like domain |
| HdAgo | -*Halopiger djelfamassiliensis* | ERV**Y**DEL**K**KVLATK-DLNS**Q**FAY \| LVAAAGG | |
| HbAgo | -*Halogeometricum borinquense* | DDP**Y**PEF**K**RRLGQL-GVPS**Q**MIT \| LIG**K**AGG | |
| MpAgo | -*Marinitoga piezophila* | IGNIDPLVRNFPDN--LILQPIL \| MGNFIPE | 5'-end $OH^-$ |
| TpAgo | -*Thermotoga profunda* | QESLDNLIKNAPKN--VVILPVL \| IVNFSQD | |
| MsAgo | -*Marinitoga sp.* | LGNIDPLIQKFPEN--LILQPIL \| MGNFIQE | |

The 3'-end of the guide is accommodated into a binding pocket of the PAZ domain (Wang et al., 2008b). In pAgos this is essential to prevent the guide form being degraded by nucleases but also to induce a conformational change upon target strand binding (Hur et al., 2013; Sheng et al., 2014). This conformational change ensures that a glutamate residue (which resides on a loop called the glutamate finger) (Nakanishi et al., 2012), which is located away from the catalytic site in the absence of a bound target, now completes that catalytic tetrad that is furthermore formed by three well-conserved aspartic acids (Sheng et al., 2014). Interestingly recent structural data revealed that both *Rs*Ago (Liu et al., 2018) and *Mp*Ago (Doxzen et al., 2017) lack one of the four sub regions that form the hydrophobic binding pocket that anchors the 3'end of the guide into the PAZ domain. Despite the incomplete PAZ domain, *Mp*Ago still anchors and, upon target binding releases the 3'-end of the guide

(Doxzen et al., 2017). For *Rs*Ago however, the incomplete PAZ domain failed to facilitate the anchoring of the guides 3'-end (Liu et al., 2018). As a result the conformational change that is necessary to swing the glutamate finger into the catalytic tetrad could not occur. Besides, *Rs*Ago lacks two of the four catalytic residues of the DEDX tetrad (Liu et al., 2018; Olovnikov et al., 2013). Restoring the catalytic tetrad did not restore the catalytic activity (Liu et al., 2018), indicating the importance of this conformation change that is triggered upon target binding followed by the release of the 3'-end of the guide from the binding pocket.

## Target recognition and binding

Once both ends of the guide are anchored into the canonical binding pockets of the Argonaute protein, the binary complex is primed for target recognition (Wang et al., 2008a, 2008b). The number of potential targets is enormous in the pool of cellular nucleic acids. Therefore Ago requires an efficient and fast target search mechanism. In **chapter 4** we aimed to elucidate this target mechanism for *Cb*Ago. The nucleotides downstream of the anchor nucleotide (nt 2-7) form the seed region (**Chapter 3**). The seed is the region of the guide that is essential for initial target search and binding. In all binary Ago complexes (Ago + guide) that have been crystallized to date, a kink divides the seed is into a smaller 'sub-seed' segment (nt 2-4) and a 3'-end region (reviewed by Willkomm et al., 2018). The seed region of the guide is solvent exposed and pre-ordered in an A-helical like conformation favorable for base pairing (Schirle et al., 2014; Wang et al., 2009). By pre-ordering the seed, Ago prepays the entropic penalty that is associated with base pairing to the target (Wang et al., 2009). For hAgo2 a proposed stepwise mechanism already exists in which hAgo2 first exposes the sub-seed segment of the guide (nt 2-4) towards the solvent to allow target pairing (Schirle et al., 2014), prior to target nucleation with the rest of the seed sequence (nt 5-8). This mechanism is reminiscent to the search strategy that the CRISPR-Cas systems (Clustered Regularly Interspaced Short Palindromic Repeats systems utilize CRISPR-associated genes) employ (Globyte et al., 2018b; Jiang et al., 2015; Künne et al., 2014; Sternberg et al., 2014). To find cognate targets Cas proteins utilize RNA guides, that are noncoding RNA transcripts derived from the genomic CRISPR locus (Horvath and Barrangou, 2010; Rotem Sorek, 2008). Where Ago proteins initially use the sub-seed segment of the guide to scan for potential targets, Cas-proteins use a small protein domain to scan their target for a 2-6 nucleotide long sequence

called the protospacer adjacent motif (PAM) (Marraffini and Sontheimer, 2010b; Westra et al., 2012). This mechanism allows Cas proteins to distinguish the host genome from invading nucleic acids and thereby prevent that the genomic CRISPR locus, from which their guide RNA originates, is targeted and cleaved. Target sequences that are entirely complementary to the guide RNA but that lack the exact PAM sequence are ignored by Cas proteins (Sternberg et al., 2014). For recognition of the protospacer by Cas nucleases, a perfect match between the seed region of the guide (5-8nt) and the target strand is required. Whereas for *Tt*Ago, a mutation in the sub-seed segment of the guide (nt 2-4) is tolerated *in vitro* (**Chapter 5**). In case of a mutation in the sub-seed, Ago apparently still tests the remainder of the guide sequence (nt 5-21) for complementarity with the target, indicating that the target search process by Ago is not that stringent, at least not *in vitro*. To conclude, both proteins employ a similar search strategy; the initial search for only a few target nucleotides enables them to quickly scan target chains via lateral diffusion and allows them to rapidly dissociate from the target in the case the remainder of the guide is not complementary (Globyte et al., 2018b; Klein et al., 2017; Klum et al., 2018).

Surprisingly, besides the preference of a 5'-end adenosine anchor nucleotide, a majority of guides that were pulled down together with *Cb*Ago after heterologous expression, contained a thymidine rich sub-seed region (**Chapter 3**). *In vivo* co-purified guides of *Tt*Ago and Argonaute from *Rhodobacter sphaeroides* (*Rs*Ago) also showed sequence preferences for the second nucleotide of the guide (for *Tt*Ago an adenosine and for *Rs*Ago a cytidine or uracil), although these preferences were less strong than the preferences observed for the first nucleotide (Olovnikov et al., 2013; Swarts et al., 2014b). In addition, neither *Tt*Ago nor *Rs*Ago had preferences for the third or fourth guide nucleotide that are part of the sub-seed. Our single molecule experiments showed that *Cb*Ago bind targets much stronger and more frequent with guides having a thymidine rich sub-seed than with guides having an adenosine rich sub-seed. This preference also extends to cleavage assays, where *Cb*Ago with showed higher initial cleavage efficiencies with guides having a thymidine rich sub-seed compared to guides with an adenosine, guanosine or cytidine rich sub-seed. However at this stage, the underlying mechanism that causes the observed biases in co-purified guides and target binding/cleavage kinetics remains unknown. Possibly the guide generation process plays an important factor as it might mainly yield guides with thymidine in the sub-seed region. Also,

the structure of *Cb*Ago could play an important role in this. More research is necessary to elucidate the underlying cause of these biases.

**Cleavage and target release**

All catalytically active pAgos that have been characterized thus far, including *Cb*Ago, cleave targets strands between the nucleic acids that are complementary at least up to nucleotide 10 and 11 of the guide (counted from the 5'end of the guide) (Sheng et al., 2014). The target cleavage mechanism has been well studied for *Tt*Ago, which uses the charged DEDD amino acid tetrad to coordinate a pair of divalent cations (typically $Mg^{2+}$ or $Mn^{2+}$) together with a nucleophilic water molecule, toward the cleavable phosphodiester bond of the target backbone (Sheng et al., 2014). As the positioning of the cations towards the target backbone nucleic acids is crucial, mismatches around the cleavage site significantly decrease the cleavage efficiencies of *Tt*Ago, as seen in **Chapter 5**.

Both eAgos and pAgos were demonstrated to be multiple-turnover enzymes (Haley and Zamore, 2004; Hutvágner and Zamore, 2002; Swarts et al., 2015a). This means that after slicing the target the cleaved products are released and the Ago-guide complexes are recycled for another round of cleavage. Additional evidence for Agos multiple-turnover behavior was found structural data. The ternary *Tt*Ago and *Cb*Ago structures showed several contacts between the Ago and the phosphodiester backbone of the guide whereas no hydrogen-bond contacts with the target were found. This indicated that they maintains the guide whereas the target strand is more likely to dissociate (**Chapter 3**; Parker, 2010; Wang et al., 2008b, 2009)

The turnover rate of *Cb*Ago was limited *in vitro*. A rapid initial burst (single turnover) of activity was followed by a much slower multiple turnover phase. This effect was also observed for heterologously expressed hAgo2 (Rivas et al., 2005). Interestingly, single nucleotide mismatches between the guide-target sequences can increase the turnover rate (Parker, 2010). For *Tt*Ago we observed that mismatches in nucleotide 4 and 5 of the guide did not negatively affect the cleavage efficiency. Instead, when either of these two nucleotides of the seed segment contained mismatches, *Tt*Ago showed a slightly higher cleavage efficiency (**Chapter 5**). This suggests that the mismatches improve the release of the cleaved target strands which in turn may increase the turnover rate.

## Potential cellular functions pAgos

The extent of the structural and mechanistic diversity in pAgos is rapidly becoming clear; however knowledge about the cellular function of pAgos is still limited to only a few long pAgos, which were proposed to be involved in host defense proteins. Therefore several hypotheses about other possible functions of pAgos exist.

Given the current variations in substrate specificity, it did not come as a surprise that also RNA-guided pAgos exist, that similar to eAgos target transcripts or invading RNA. A potential candidate is *Mp*Ago which was shown to mediate RNA-guided DNA and RNA cleavage *in vitro* (Kaya et al., 2016). Further experiments are necessary to elucidate the exact role of *Mp*Ago *in vivo*.

Interestingly *Mp*Ago is located on an operon with CRISPR-Cas (Kaya et al., 2016). While no direct interaction between the two systems has been demonstrated yet, it is tempting to speculate about a possible synergy. Potentially cleavage products of CRISPR-Cas systems are used as guides for Argonaute proteins, or vice versa. CRISPR-Cas showed similarities with the eukaryotic immune system that uses PIWI-clade Argonaute proteins guided by piRNAs (PIWI-interacting RNAs) to fend of invaders (**Chapter 1**). Recently it was also shown that expression of *Tt*Ago in the presence of a target plasmid stimulated the expression CRISPR-Cas loci (Swarts et al., 2015b), further indicating that both immune systems might collaborate.

## Applications of prokaryotic Argonaute

The antiviral immune systems has emerged as a wealthy source for numerous molecular tools, which have shaped and revolutionized modern biological research (Castanotto and Rossi, 2009; Engler et al., 2009; Hsu et al., 2014; Wu et al., 2018). Many proteins of the antiviral immune system are specific due to their ability to discriminate foreign genetic elements from its own genome (Marraffini and Sontheimer, 2010b). Among others, restriction enzymes, RNA interference and CRISPR are well-known examples of antiviral immune systems that have been repurposed to powerful biological tools (**Figure 4**).

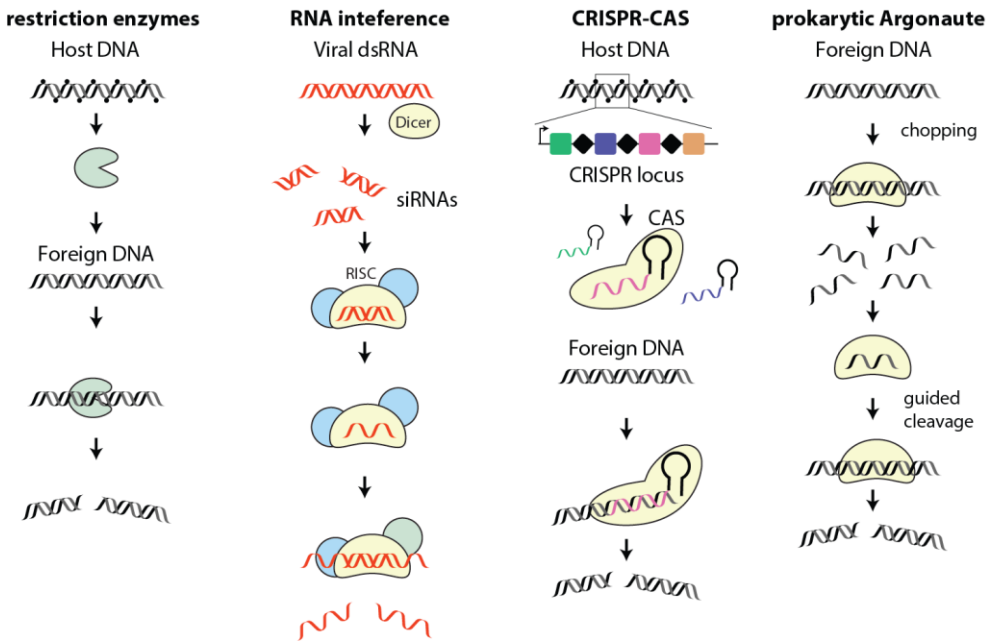



**Figure 4 | Overview of antiviral immune systems whose activities have been repurposed for molecular applications.** Restriction enzymes recognize fixes nucleic acid sequences of 4-8nt in length. Host DNA is methylated (small black dots on the DNA) and thereby protected from degradation by restriction enzymes. Argonaute and CRISPR utilize nucleic acid guides to specifically target invading nucleic acids. CRISPR distinguishes invader DNA from its host genome by the requirement of a PAM sequence next to the target sequence which is lacking in the CRISPR locus. The self-non-self-discrimination of prokaryotic Argonaute proteins is less understood.

CRISPR-nucleases emerged as the main players in genome editing technologies, due to their ability to introduce double stranded DNA breaks (DSB) in genomes in a guide-dependent manner (Jinek et al., 2012). Depending on the organism, such a DSB triggers an endogenous DNA repair response that in the presence of a DNA repair template integrates this repair template via homologous recombination (HR) (Wu et al., 2018). This way, parts of the genome can be removed, corrected or rewritten. DNA-targeting pAgos show similarities with CRISPR nucleases (nucleic acid-guided DNA cleavage activity) and have therefore also been explored for their potential use in similar molecular applications. In 2016, this resulted in a discovery by Han and colleagues, who claimed to have found a mesophilic DNA-targeting Argonaute (NgAgo), superior to CRISPR, that could be utilized for genome editing (Gao et al., 2016). Due to reproducibility issues, however, they were forced to retract their research

within a year (Burgess et al., 2016; Javidi-Parsijani et al., 2017; Khin et al., 2017; O'Geen et al., 2018). Despite the bad publicity, we were successfully able to demonstrate the use of pAgo in two promising applications, of which also CRISPR-based variants exist.

## Argonaute in diagnostic applications

We applied *Tt*Ago in a diagnostic application that is used to enrich rare oncogenic mutant DNA/RNA sequences which are massively outnumbered by "normal" wild-type (WT) nucleic acids in human plasma (liquid biopsies) (**Chapter 5**). Rare nucleic acid sequences often contain important diagnostic and biological information (i.e. cancer biomarker) but are challenging to detect due to their low abundance (Han et al., 2017; Schwarzenbach et al., 2011). Modern sequencing technologies are despite numerous advances unable to detect mutant alleles at frequencies lower than 1% (**Table 2**) (Jancik et al., 2012; Shao et al., 2016), while especially in early stages of cancer such biomarkers can be present in concentrations that are much lower (Bettegowda et al., 2014). The NAVIGATER (**N**ucleic **a**cid enrichment **Vi**a DNA-**g**uided **A**rgonaute from ***T****hermus* ***th****e****r****mophilus*) method that we developed combines *Tt*Ago mediated elimination of abundant WT nucleic acids with PCR amplification which ultimately results in the enrichment of the rare nucleic acids. The enriched nucleic acids from the sample can be detected using existing methods such as Sanger sequencing, ddPCR or qPCR (**Chapter 5**) but possibly also with the recently developed CRISPR-based detection applications (**Table 2**). In 2017 a CRISPR-based nucleic detection system called SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) was developed (Gootenberg et al., 2017), followed by a series of similar CRISPR-based detection methods (Chen et al., 2018; Gootenberg et al., 2018; Li et al., 2018). The CRISPR-based nucleic acid detection methods rely on Cas13a (SHERLOCK) or Cas12a which upon specific target recognition exhibit collateral nonspecific single stranded ribonuclease or deoxyribonuclease activity, respectively (Chen et al., 2018; Gootenberg et al., 2017; Li et al., 2018). The collateral activity, in turn, degrades the single stranded nucleic acids of a reporter molecule which links a fluorophore to a quencher. As the fluorophore is no longer quenched, the fluorescent signal that can be detected. These CRISPR-based detection systems were reported to detect pure nucleic acids at extreme low (attomolar) concentrations (~1000 copies per ml) (Gootenberg et al., 2017). However, detection of low-abundant oncogenic mutant alleles in human samples is challenged by the presence of a high amount of wild-type allele.

Yet, the CRIPR-based nucleic acid detection systems were able to detect point mutated alleles in the presence of WT DNA with a limit of detection around 0.1% (**Table 2**) (Chen et al., 2018; Gootenberg et al., 2017; Li et al., 2018).

**Table 2 | The detection limit of techniques that can be used to detect low-frequency cancer mutations in cell-free DNA.**

| Method | Nuclease involved | LOD (%) | References |
|---|---|---|---|
| Digital droplet PCR | | 0.01-0.1 | (Hindson et al., 2013) |
| Next generation sequencing (NGS) | | 1-3 | (Shao et al., 2016) |
| CUT-PCR + NGS | Cas9/12a | 0.01 | (Lee et al., 2017) |
| Sanger sequencing | | 25 | (Jancik et al., 2012) |
| NAVIGATER + Sanger Sequencing | *Tt*Ago | 0.5 | **Chapter 5** |
| Clamp-based PCR | | 0.1-1 | (Itonaga et al., 2016) |
| NAVIGATER + Clamp-based PCR | *Tt*Ago | 0.01-0.1 | **Chapter 5** |
| SHERLOCK | Cas13a | 0.1 | (Gootenberg et al., 2017) |
| HOLMES | Cas12a | 0.1 | (Li et al., 2018) |
| DETECTR | Cas12a | 0.1 | (Chen et al., 2018) |

*LOD, limit of detection (Ratio of mutant to wild-type copies)*

CRISPR nucleases have also been utilized in a molecular application that is similar to our NAVIGATER method, and used to enrich and improve detection of rare nucleic acids. CUT-PCR (**C**RISPR-mediated **U**ltra-sensitive detection of **T**arget DNA)-PCR, as it is called, employs the CRISPR-Cas 9 and Cas12a nucleases to eliminate wild-type DNA, while sparing rare mutant alleles (Lee et al., 2017). As most cancer associated mutant alleles differ by only a single mismatch from WT alleles, CUT-PCR makes use of the low tolerance to mismatches in the PAM recognition site of CRISPR nucleases to discriminate between the mutant and wild-type targets (Cho et al., 2014). Based on the COSMIC (Catalogue Of Somatic Mutations In Cancer) database, only 27% of all oncogenic point mutations are covered by the PAM sites of SpCas9 (PAM: GG) and FnCas12a (PAM: TT) combined (Lee et al., 2017). The coverage can be further increased to ~80% by using all the other CRISPR nuclease variants, with unique PAMs, that are currently available (Lee et al., 2017).

The PAM-independent targeting of pAgos make it possible to discriminate any point mutation listed in the COSMIC database, from its wild-type DNA. Although, a screening of guides is necessary because the length and the location of guide segment that has low tolerance to mismatched appears to be guide sequence dependent (**Chapter 5**).

A general downside of mutation enrichment methods that use PCR is that it is hard to trace back the initial ratio of wild-type to mutant DNA. Quantitation and measuring this ratio is extremely valuable for disease monitoring before and after treatment. The ratio can be estimated by comparing results to those obtained using wild-type DNA spiked with known mutant DNA as standards. Nevertheless, we successfully applied NAVIGATER to detect sequences with oncogenic mutations in the cell free DNA from pancreatic cancer patients, demonstrating that our technique can be used for diagnosing various types of cancer at early stages.

## Molecular cloning with prokaryotic Argonaute

For decades, restriction enzymes have been the workhorses of molecular biology (Roberts, 2005). In nature, they equip bacteria and archaea with a defense system to fend of DNA invaders (**Figure 4**). Depending on the type, restriction enzymes typically recognize specific restriction sites of ~4-8 bases in length, and produce DSB with sticky or blunt end overhangs (Roberts et al., 2003). Based on these features they have been utilized for molecular cloning purposes. However for cloning, restriction enzymes do have limitations; they do not cleave locations other than the fixed recognition site and the recognition sites should be avoided in DNA regions where cleavage is not desired. Therefore recently, a new alternative platform was developed that uses *Pf*Ago as an artificial restriction enzyme (Enghiad and Zhao, 2017). *Pf*Ago is programmable with DNA guides, which means that two *Pf*Ago-guide complexes each targeting one strand, can cleave dsDNA at virtually any location. We developed this *Pf*Ago-based cloning method further into a one-pot reaction together with a ligase, similar to Gibson and Golden Gate cloning. Although, we could successfully clone an insert into a plasmid that was linearized by *Pf*Ago, the ultimate efficiency was low and requires further optimization (**Chapter 6**). Also CRISPR-Cas12a has been used for several cloning applications (Lei et al., 2017; Li et al., 2016; Wang et al., 2015). Cas12a generates DSB with staggered ends whereas Cas9 leaves DSB with blunt ends behind (Wu et al., 2018). Compared to restriction enzymes, Cas12a has a longer, programmable, and much more specific recognition site of ~18nt (spacer), which indicates the potential of Cas12a as a powerful candidate for cloning purposes. However the main limitation of Cas12a, is that besides the PAM requirement which slightly limits its flexibility, they generate double stranded breaks with sticky ends that are not consistent in length (Wu et al., 2018). Especially

for the assembly of multiple fragments this is a problem as it results in a drop in efficiency for each extra fragment that needs to be ligated.

## Bottlenecks and future directions

As discussed in **chapter 2** pAgos still need to overcome several hurdles to unleash their full potential for molecular applications. In Argonaute-based applications, as described in **chapter 5 and 6**, non-specific DNA degradation caused by chopping is detrimental. For example during Jason cloning (**Chapter 6**) prolonged incubations at 95°C resulted in non-specific degradation of the backbone plasmid, lowering the cloning efficiencies. The degradation was likely a result of chopping by *Pf*Ago, even though *Pf*Ago was saturated with guides. Several optimizations could potentially prevent the undesired chopping activity. As chopping is caused by guide-free Ago, chopping can potentially be prevented or minimized by crosslinking the DNA guide to *Pf*Ago, similar to what has been demonstrated for *Mp*Ago (Lapinaite et al., 2018). Chopping was furthermore shown to depend on the salt concentration of the buffer (Swarts et al., 2015a, 2017a). Chopping by *Tt*Ago and *Pf*Ago took place at 250mM salt but was completely abolished at 500mM salt, while canonical cleavage by Argonaute preloaded with a guide still occurred at 500mM salt (Swarts et al., 2015a, 2017a). Based on this an increase in the salt concentration might further minimize the severity of chopping, however it should be kept in mind that the other key enzyme of Jason cloning, Ampligase, has the highest activity under low salt conditions. Alternatively the incubation temperature could be lowered as chopping was shown to be less severe at lower temperatures (**Chapter 3**; Swarts et al., 2015).

Chopping activity furthermore reduces the number of approaches that can be used to deliver Ago proteins into cells for genome editing purposes. The DNA chopping activity that pAgos mediate in the absence of guides could potentially be toxic for cells through nonspecific DNA damage. Therefore in our attempt to utilize *Cb*Ago for genome editing experiments, transfection of *in vitro* assembled *Cb*Ago–siDNA complexes seemed the only appropriate cellular delivery method (**Chapter 3**). In contrast, CRISPR-Cas nucleases which do not exhibit chopping activity, can be delivered as plasmid because the protein-guide complex also efficiently assembles *in vivo*.

From an application perspective, the efficiencies at which pAgos bind and cleave double stranded DNA should be improved, especially at low temperatures where double stranded DNA is strongly base paired. Recently a study showed that the activity of *Tt*Ago and *Pf*Ago could be slightly enhanced by adding single stranded binding proteins (SSB) and UrvD-helicases to the *in vitro* cleavage reaction (Hunt et al., 2018). In order to further increase the dsDNA accessibility of pAgos, it might be worth exploring the option to fuse such proteins or domains of it to Argonaute. Although, it should be noted that the activity of Argonaute can be affected by fusions to especially the C-terminal end. While preparing *Cb*Ago for mammalian genome editing experiments we fused a nuclear localization (NLS) signal, which is a signal peptide that allows the protein to migrate to the nucleus of a eukaryotic cell, to both ends of *Cb*Ago **(Figure 5A)**. While *Cb*Ago with the N-terminally fused NLS peptide could cleave single stranded DNA, the C-terminally fused NLS completely abolished the activity of *Cb*Ago (**Figure 5B**). This can be explained by the fact that the carboxylate group at the C-terminus is a conserved structural feature of many pAgos and is involved in the binding of the 5'-end of the guide (Parker et al., 2005). Fusions might disrupt this functionality. Recently a study showed that *Tt*Ago was active *in vitro* while having an NLS-tag at the C-terminal end (O'Geen et al., 2018), however they used a 25 amino acid long (glycine-serine) linker between the NLS and *Tt*Ago which might have rescued the activity.

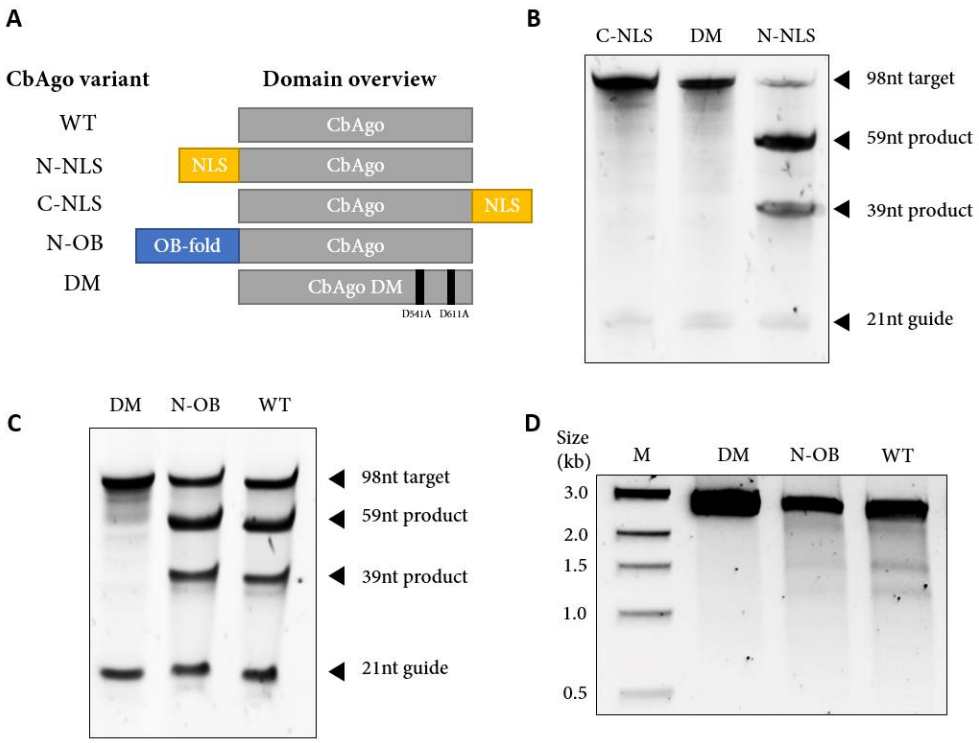



**Figure 5 | The effects of extra domains on the activity of *Cb*Ago. A |** *Cb*Ago variants with different N- and C-terminal fusions. **B |** A C-terminal NLS tag abolishes the nuclease activity of *Cb*Ago on single stranded DNA targets. **C |** An OB-fold on the N-terminal of *Cb*Ago has no effect on the single stranded cleavage activity of *Cb*Ago. **D |** An N-terminal OB-fold had no enhanced effect on the activity of *Cb*Ago on double stranded targets (plasmid)

In an attempt to improve the dsDNA accessibility of *Cb*Ago, we fused the N-terminal domain (148 amino acids) of *Ng*Ago to *Cb*Ago (in blue, **Figure 5B**). *Ng*Ago encodes this unique N-terminal domain together with other members of its phylogenetic clade (Ryazansky et al., 2018). The domain is predicted to be OB-fold domain (SSF50249) (Ryazansky et al., 2018), which are compact structural motifs that are named after their **o**ligonucleotide **b**inding properties. Although OB-folds only bind single stranded DNA, similar single stranded binding proteins (SSB) were demonstrated to enhance the activity of pAgo on double stranded DNA *in vitro* (Hunt et al., 2018). We rationalized that the fused OB-fold, once bound to the DNA, could potentially function as a wedge to promote unwinding of double stranded DNA and increase the efficiency of *Cb*Ago. However, although the OB fold-*Cb*Ago cleaved single stranded DNA at similar efficiencies as the *Cb*Ago without the fused domain,

no enhanced cleavage could be observed on dsDNA targets (**Figure 5C, D**). Further efforts are needed using other fusions to optimize the activity of *Cb*Ago on dsDNA.

Given that many different pAgo variants exist (**Chapter 2**), it is very interesting from a fundamental point of view to continue the characterization of new Argonaute variants to further increase our understanding in function, role and mechanisms of this exciting protein family. Besides, new pAgo variants, also other prokaryotic defense systems should be explored in the quest to find novel candidates for molecular applications, such as genome editing. After all the antiviral immune systems has proven to be a source of novel exciting molecular applications. All prokaryotic immune systems found to date are likely the tip of the iceberg, given the enormous pool of microbial genomes that still waits to be sequenced. With recent discovery of novel anti-phage defense islands that are, like Argonaute proteins, named after a myth (**Chapter 1**, Doron et al., 2018), we are almost certainly at the beginning of turning more myths into reality.

# References


Aalipour, A., Dudley, J.C., Park, S., Murty, S., Chabon, J.J., Boyle, E.A., Diehn, M., and Gambhir, S.S. (2018). Deactivated CRISPR Associated Protein 9 for Minor-Allele Enrichment in Cell-Free DNA. Clin. Chem. *64*, 307–316.

Adamczak, B., Kogut, M., and Czub, J. (2018). Effect of osmolytes on the thermal stability of proteins: replica exchange simulations of Trp-cage in urea and betaine solutions. Phys. Chem. Chem. Phys. *20*, 11174–11182.

Aggarwal, C., Thompson, J.C., Black, T.A., Katz, S.I., Fan, R., Yee, S.S., Chien, A.L., Evans, T.L., Bauml, J.M., Alley, E.W., et al. (2018). Clinical implications of plasma-based genotyping with the delivery of personalized therapy in metastatic non–small cell lung cancer. JAMA Oncol. 1–8.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569–573.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203–207.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. Mol. Cell *31*, 785–799.

Averhoff, B. (2009). Shuffling genes around in hot environments: The unique DNA transporter of Thermus thermophilus. FEMS Microbiol. Rev. *33*, 611–626.

Banerjee, D., and Slack, F. (2002). Control of developmental timing by small temporal rnas: A paradigm for RNA-mediated regulation of gene expression. BioEssays *24*, 119–129.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. *34*, 933.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Patrick Boyaval, Moineau, S., Romero, D., and Horvath, P. (2007). CRISPRProvides Acquired Resistance Against Viruses in Prokaryotes. Science (80-. ). *315*, 1709–1712.

Bartel, D.P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell *116*, 281–297.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215–233.

Beaudoin, J.-D., Novoa, E.M., Vejnar, C.E., Yartseva, V., Takacs, C.M., Kellis, M., and Giraldez, A.J. (2018). Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. Nat. Struct. Mol. Biol. *25*, 677–686.

Berg, O.G., Winter, R.B., and von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. Biochemistry *20*, 6929–6948.

Bergman, C.M., Quesneville, H., Anxolabéhère, D., and Ashburner, M. (2006). Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogastergenome. Genome Biol. *7*, R112.

Bernstein, E., Caudy, a a, Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., et al. (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci. Transl. Med. *6*, 224ra24.

Bevilacqua, P.C., Ritchey, L.E., Su, Z., and Assmann, S.M. (2016). Genome-Wide Analysis of RNA Secondary Structure. Annu. Rev. Genet. *50*, 235–266.

Bielas, J.H., and Loeb, L.A. (2005). Quantification of random genomic mutations. Nat. Methods *2*, 285.

Blesa, A., César, C.E., Averhoff, B., and Berenguer, J. (2015). Noncanonical cell-to-cell DNA transfer in Thermus spp. Is insensitive to argonaute-mediated interference. J. Bacteriol. *197*, 138–146.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopis

controlling leaf development. Curr. Opin. Plant Biol. *1*, 188.

Boland, A., Tritschler, F., Heimstädt, S., Izaurralde, E., and Weichenrieder, O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep. *11*, 522–527.

de Borne Dorlhac, F., Vincentz, M., Chupeau, Y., and Vaucheret, H. (1994). Co-suppression of nitrate reductase host genes and transgenes in transgenic tobacco plants. MGG Mol. Gen. Genet. *243*, 613–621.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

van den Broek, B., Lomholt, M.A., Kalisch, S.-M.J., Metzler, R., and Wuite, G.J.L. (2008). How DNA coiling enhances target localization by proteins. Proc. Natl. Acad. Sci. U. S. A. *105*, 15738–15742.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E. V., and Oost, J. van der (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science (80-. ). *322*, 949–953.

Burgess, S., Cheng, L., Gu, F., Huang, J., Huang, Z., Lin, S., Li, J., Li, W., Qin, W., Sun, Y., et al. (2016). Questions about NgAgo. Protein Cell *7*, 913–915.

Burroughs, A.M., Iyer, L.M., and Aravind, L. (2013). Two novel PIWI families: roles in inter-genomic conflicts in bacteria and Mediator-dependent modulation of transcription in eukaryotes. Biol. Direct *8*.

Cai, H., and Wind, S.J. (2016). Improved Glass Surface Passivation for Single-Molecule Nanoarrays. Langmuir *32*, 10034–10041.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642–655.

Castanotto, D., and Rossi, J.J. (2009). The promises and pitfalls of RNA-interference-based therapeutics. Nature *457*, 426.

Cerutti, L., Mian, N., and Bateman, A. (2000). Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends Biochem. Sci. *25*, 481–482.

Chandradoss, S.D., Haagsma, A.C., Lee, Y.K., Hwang, J.-H., Nam, J.-M., and Joo, C. (2014). Surface Passivation for Single-molecule Protein Studies. J. Vis. Exp. 1–8.

Chandradoss, S.D., Schirle, N.T., Szczepaniak, M., Macrae, I.J., Chandradoss, S.D., Schirle, N.T., Szczepaniak, M., Macrae, I.J., and Joo, C. (2015). A Dynamic Search Process Underlies MicroRNA Targeting. Cell *162*, 96–107.

Chen, H., Meisburger, S.P., Pabit, S. a., Sutton, J.L., Webb, W.W., and Pollack, L. (2012). Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. Proc. Natl. Acad. Sci. *109*, 799–804.

Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M., and Doudna, J.A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science (80-. ). *360*, 436–439.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 132–141.

Choi, J.-J., Cho, M.-H., Oh, M.-A., Kim, H.-S., Kil, M.-S., and Park, H.-K. (2010). PNA-mediated Real-Time PCR Clamping for Detection of EGFR Mutations. Bull. Korean Chem. Soc. *31*, 3525–3529.

Crowley, E., Di Nicolantonio, F., Loupakis, F., and Bardelli, A. (2013). Liquid biopsy: monitoring cancer-genetics in the blood. Nat. Rev. Clin. Oncol. *10*, 472.

Cyranoski, D. (2016). Replications, ridicule and a recluse: the controversy over NgAgo gene-editing intensifies. Nature *536*, 136–137.

Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem. Sci. *41*, 324–337.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science (80-. ). *4120*.

Doxzen, K.W., Doudna, J.A., Adams, P., Winn, M., Storoni, L., and Read, R. (2017). DNA recognition by an RNA-guided bacterial Argonaute. PLoS One *12*, 1–14.

Dwidar, M., Park, J.Y., Mitchell, R.J., and Sang, B.I. (2012). The future of butyric acid in industry. Sci. World J. *2012*, 4–5.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270.

Elbashir, S.M., Elbashir, S.M., Lendeckel, W., Lendeckel, W., Tuschl, T., and Tuschl, T. (2001). RNA interference is mediated 1- and 22-nucleotide RNAs. Genes Dev. *15*, 188–200.

Elbashir, S.M., Harborth, J., Weber, K., and Tuschl, T. (2002). Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods *26*, 199–213.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100–110.

Ender, C., and Meister, G. (2010). Argonaute proteins at a glance Argonaute Proteins at a Glance. J. Cell Sci. *2*, 1819–1823.

Enghiad, B., and Zhao, H. (2017). Programmable DNA-Guided Artificial Restriction Enzymes. ACS Synth. Biol. *6*, 752–757.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. PLoS One *4*.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). Getting to the Root of miRNA-Mediated Gene Silencing. Cell *132*, 9–14.

Faehnle, C.R., and Joshua-Tor, L. (2010). Argonaute MID domain takes centre stage. EMBO Rep. *11*, 564–565.

Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. *10*, 637–643.

Fellmann, C., Gowen, B.G., Lin, P.C., Doudna, J.A., and Corn, J.E. (2017). Cornerstones of CRISPR-Cas in drug discovery and therapy. Nat. Rev. Drug Discov. *16*, 89–100.

Fire, A., Xu, S.Q., K, M., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by doublestranded RNA in Caenorhabditis elegans RNAi RNA Interference or Experimental introduction of RNA into cells can be used in certain biological systems RNA silencing. Nature *806*.

Forterre, P., Bergerat, A., and Lopex-Garcia, P. (1996). The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. FEMS Microbiol. Rev. *18*, 237–248.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818–822.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. EMBO J. *31*, 3588–3595.

Fray, R.G., and Grierson, D. (1993). Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. Plant Mol. Biol. *22*, 589–602.

Friedman, R.C., Farh, K.K.H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Gao, F., Shen, X.Z., Jiang, F., Wu, Y., and Han, C. (2016). DNA-guided genome editing using the Natronobacterium gregoryi Argonaute. Nat. Biotechnol. *34*, 768–773.

Gerland, U., Moroz, J.D., and Hwa, T. (2002). Physical constraints and functional characteristics of transcription factor-DNA interaction. Proc. Natl. Acad. Sci. *99*, 12015–12020.

Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., and Hutchison, C.A. (2008). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc. Natl. Acad. Sci. *105*, 20404–20409.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R., Algire, M.A., Benders, A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *329*, 1–6.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199–202.

Globyte, V., Lee, S.H., Bae, T., Kim, J., and Joo, C. (2018b). CRISPR Cas9 searches for a protospacer adjacent motif by one-dimensional diffusion. BioRxiv.

Globyte, V., Kim, S.H., and Joo, C. (2018a). Single-Molecule View of Small RNA–Guided Target Search and Recognition. Annu. Rev. Biophys. *47*, 569–593.

Golden, D.E., Gerbasi, V.R., and Sontheimer, E.J. (2008). An Inside Job for siRNAs. Mol. Cell *31*, 309–312.

Gootenberg, J.S., Dy, A.J., Freije, Joung, J., Myhrvold, C., Deborah T. Hung, * Omar O. Abudayyeh 2, 3, 4, 6, 1, Vanessa Verdine 2, 3, 4, 1, Roby P. Bhattacharyya, 1, Pardis C. Sabeti 11, 12, 13, 1, et al. (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. Science (80-. ). *442*, 438–442.

Gootenberg, J.S., Abudayyeh, O.O., Kellner, M.J., Joung, J., Collins, J.J., and Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. Science (80-. ). *360*, 439–444.

Gorman, J., Chowdhury, A., Surtees, J.A., Shimada, J., Reichman, D.R., Alani, E., and Greene, E.C. (2007). Dynamic Basis for One-Dimensional DNA Scanning by the Mismatch Repair Complex Msh2-Msh6. Mol. Cell *28*, 359–370.

Großhans, H., and Filipowicz, W. (2008). Molecular biology: The expanding world of small RNAs. Nature *451*, 414–416.

Gu, W., Crawford, E.D., O'Donovan, B.D., Wilson, M.R., Chow, E.D., Retallack, H., and DeRisi, J.L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol. *17*, 41.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science (80-. ). *315*, 1587–1590.

Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. Nat. Struct. Mol. Biol. *11*, 599–606.

Halford, S.E., and Marko, J.F. (2004). How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res. *32*, 3040–3052.

Hamilton, A.J., and Baulcombe, D.C. (1999). A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *213*, 1997–2000.

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G., and Elf, J. (2012). The lac Repressor Displays Facilitated Diffusion in Living Cells. Science (80-. ). *336*, 1595–1598.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293–296.

Han, X., Wang, J., and Sun, Y. (2017). Circulating Tumor DNA as Biomarkers for Cancer Detection. Genomics. Proteomics Bioinformatics *15*, 59–72.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244–251.

He, L., and Hannon, G.J. (2004). MicroRNAs: Small RNAs with a big role in gene regulation. Nat. Rev. Genet. *5*, 522–531.

Hegge, J.W., Swarts, D.C., and van der Oost, J. (2018a). Prokaryotic Argonaute proteins: novel genome-editing tools? Nat. Rev. Microbiol. *16*, 5–11.

Hegge, J.W., Swarts, D.C., Chandradoss, S.D., Cui, T.J., Kneppers, J., Jinek, M., Joo, C., and Van Der Oost, J. (2018b). DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute.

Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S.A. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. *25*, 3957–3958.

Hindson, C.M., Chevillet, J.R., Briggs, H.A., Gallichotte, E.N., Ruf, I.K., Hindson, B.J., Vessella, R.L., and Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. Nat. Methods *10*, 1003–1005.

Von Hippel, P.H., and Berg, O.G. (1989). Facilitated target location in biological systems. J. Biol. Chem. *264*, 675–678.

Höck, J., and Meister, G. (2008). The Argonaute protein family. Genome Biol. *9*, 210.

Holoch, D., and Moazed, D. (2015). Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. Nat. Struct. Mol. Biol. *22*, 328–335.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167–170.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Hunt, E.A., Evans Jr, T.C., and Tanner, N.A. (2018). Single-stranded binding proteins and helicase enhance the activity of prokaryotic argonautes in vitro. PLoS One *13*, 1–20.

Hur, J.K., Zinchenko, M.K., Djuranovic, S., and Green, R. (2013). Regulation of Argonaute slicer activity by guide RNA 3' end interactions with the N-terminal lobe. J. Biol. Chem. *288*, 7829–7840.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat. Rev. Mol. Cell Biol. *9*, 22–32.

Hutvágner, G., and Zamore, P.D. (2002). A microRNA in a Multiple- Turnover RNAi Enzyme Complex. Science (80-. ). *297*, 2056–2060.

Itonaga, M., Matsuzaki, I., Warigaya, K., Tamura, T., Shimizu, Y., Fujimoto, M., Kojima, F., Ichinose, M., and Murata, S. ichi (2016). Novel methodology for rapid detection of KRAS mutation using PNA-LNA mediated loop-mediated isothermal amplification. PLoS One *11*, 1–12.

Jancik, S., Drabek, J., Berkovcova, J., Xu, Y.Z., Stankova, M., Klein, J., Kolek, V., Skarda, J., Tichy, T., Grygarkova, I., et al. (2012). A comparison of Direct sequencing, Pyrosequencing, High resolution melting analysis, TheraScreen DxS, and the K-ras StripAssay for detecting KRAS mutations in non small cell lung carcinomas. J. Exp. Clin. Cancer Res. *31*, 1.

Javidi-Parsijani, P., Niu, G., Davis, M., Lu, P., Atala, A., and Lu, B. (2017). No evidence of genome editing activity from Natronobacterium gregoryi Argonaute (NgAgo) in human cells. PLoS One *12*, 1–14.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). A Cas9–guide RNA complex preorganized for target DNA recognition. Science (80-. ). *348*, 1477 LP-1481.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA – Guided. *337*, 816–822.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat. Rev. Genet. *16*, 421–433.

Jones, D.L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M.J., and Elf, J. (2017). Kinetics of dCas9 target search in *Escherichia coli*. Science (80-. ). *357*, 1420–1424.

Joo, C., and Ha, T. (2012). Single-molecule FRET with total internal reflection microscopy. Cold Spring Harb. Protoc. *7*, 1223–1237.

Joshua-Tor, L., and Hannon, G.J. (2011). Ancestral roles of small RNAs: An ago-centric perspective. Cold Spring Harb. Perspect. Biol. *3*, 1–11.

Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' End Formation of PIWI-Interacting RNAs In Vitro. Mol. Cell *43*, 1015–1022.

Kaya, E., Doxzen, K.W., Knoll, K.R., Wilson, R.C., Strutt, S.C., Kranzusch, P.J., and Doudna, J.A. (2016). A bacterial Argonaute with noncanonical guide RNA specificity. Proc. Natl. Acad. Sci. U. S. A. *113*, 4057–4062.

Ketting, R.F. (2010). MicroRNA biogenesis and function: An overview. Adv. Exp. Med. Biol. *700*, 1–14.

Ketting, R.F. (2011). The Many Faces of RNAi. Dev. Cell *20*, 148–161.

Khin, N.C., Lowe, J.L., Jensen, L.M., and Burgio, G. (2017). No evidence for genome editing in mouse zygotes & HEK293T human cell line using the DNA-guided Natronobacterium gregoryi Argonaute (NgAgo). PLoS One *12*, 1–10.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.

Kim, D.H., and Rossi, J.J. (2007). Strategies for silencing human disease using RNA interference. Nat. Rev. Genet. *8*, 173–184.

Kim, H.S., Sung, J.S., Yang, S.-J., Kwon, N.-J., Jin, L., Kim, S.T., Park, K.H., Shin, S.W., Kim, H.K., Kang, J.-H., et al. (2013). Predictive Efficacy of Low Burden EGFR Mutation Detected by Next-Generation Sequencing on Response to EGFR Tyrosine Kinase Inhibitors in Non-Small-Cell Lung Carcinoma. PLoS One *8*, e81975.

Kim, K., Ye, S., Bae, T., Kim, K., Habib, O., , Seung Hwan Lee, Y., Kim, Y., Lee, K.-I., Kim, S., and Kim, and J.-S. (2017). DNA-dependent RNA cleavage by the Natronobacterium gregoryi. BioRxiv 1–9.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126–139.

Klein, M., Chandradoss, S.D., Depken, M., and Joo, C. (2017). Why Argonaute is needed to make microRNA target search fast and reliable. Semin. Cell Dev. Biol. *65*, 20–28.

Klum, S.M., Chandradoss, S.D., Schirle, N.T., Joo, C., and MacRae, I.J. (2018). Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition. EMBO J. *37*, 75–88.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III DCR-1 in RNA interference and germ line development in C. elegans. Science (80-. ). *293*, 2269–2271.

Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science (80-. ). *361*, 866–869.

Kolomeisky, A.B., and Veksler, A. (2012). How to accelerate protein search on DNA: Location and dissociation. J. Chem. Phys. *136*.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20–36.

Krol, A.R. van der, Mur, L.A., Beld, M., Mol, J, N., and Stuitje, A.R. (1990). Flavonoid Genes in Petunia: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. Plant Cell *133*, 291–299.

Krug, A.K., Enderle, D., Karlovich, C., Priewasser, T., Bentink, S., Spiel, A., Brinkmann, K., Emenegger, J., Grimm, D.G., Castellanos-Rizaldos, E., et al. (2018). Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. *29*, 700–706.

Kuhn, C.D., and Joshua-Tor, L. (2013). Eukaryotic Argonautes come into focus. Trends Biochem. Sci. *38*, 263–271.

Künne, T., Swarts, D.C., Brouns, S.J.J., Kunne, T., Swarts, D.C., and Brouns, S.J.J. (2014). Planting the seed: Target recognition of short guide RNAs. Trends Microbiol. *22*, 74–83.

Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A., Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y., and Nakano, T. (2001). Two mouse piwi-related genes: Miwi and mili. Mech. Dev. *108*, 121–133.

Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat. Struct. Mol. Biol. *19*, 145–151.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of Novel Genes Coding for Small Expressed ... Identification of Novel Genes Coding for Small Expressed ... Science (80-. ). *294*, 10–12.

Lapinaite, A., Doudna, J.A., and Cate, J. (2018). Programmable RNA recognition using a CRISPR-associated Argonaute. PLoS One 208041.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabidtis elegans. Science (80-. ). *294*, 862–864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). the C Elegans Heterochronic Gene Lin4 Encodes Small Rnas With Antisense Complementarity To Lin14. Cell *75*, 843–854.

Lee, S.H., Turchiano, G., Ata, H., Nowsheen, S., Romito, M., Lou, Z., Ryu, S.-M., Ekker, S.C., Cathomen, T., and Kim, J.-S. (2016). Failure to detect DNA-guided genome editing using Natronobacterium gregoryi Argonaute. Nat. Biotechnol. 3–4.

Lee, S.H., Yu, J., Hwang, G.H., Kim, S., Kim, H.S., Ye, S., Kim, K., Park, J., Park, D.Y., Cho, Y.K., et al. (2017). CUT-PCR: CRISPR-mediated, ultrasensitive detection of target DNA using PCR. Oncogene *36*, 6823–6829.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwisee processing and subcellular localization. EMBO J. *21*, 4663–4670.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. Eur. Mol. Biol. Organ. J. *23*, 4051–4060.

Lei, C., Li, S.Y., Liu, J.K., Zheng, X., Zhao, G.P., and Jin, W. (2017). The CCTL (Cpf1-assisted Cutting and Taq DNA ligase-assisted Ligation) method for efficient editing of large DNA constructs in vitro. Nucleic Acids Res. *45*, 1–7.

Li, M.Z., and Elledge, S.J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat. Methods *4*, 251–256.

Li, C., Vagin, V. V, Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009a). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell *137*, 509–521.

Li, G.-W., Berg, O.G., and Elf, J. (2009b). Effects of macromolecular crowding and DNA looping on gene regulation kinetics. Nat. Phys. *5*, 294–297.

Li, J., Wang, L., Mamon, H., Kulke, M.H., Berbeco, R., and Makrigiorgos, G.M. (2008). Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. Nat. Med. *14*, 579.

Li, S.Y., Zhao, G.P., and Wang, J. (2016). C-Brick: A New Standard for Assembly of Biological Parts Using Cpf1. ACS Synth. Biol. *5*, 1383–1388.

Li, S.Y., Cheng, Q.X., Li, X.Y., Zhang, Z.L., Gao, S., Cao, R.B., Zhao, G.P., Wang, J., and Wang, J.M. (2018). CRISPR-Cas12a-assisted nucleic acid detection. Cell Discov. *4*, 18–21.

Lindbo, J.A., Silva-rosales, L., Proebsting, W.M., and Dougherty, W.G. (1993). lnduction of a Highly Specific Antiviral State in Transgenic Plants: lmplications for Regulation of Gene Expression and Virus Resistance. Plant Cell *5*, 1749–1759.

Lisitskaya, L., Aravin, A.A., and Kulbachinskiy, A. (2018). DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. Nat. Commun. *9*, 5165.

Liu, J., Carmell, M. a, Rivas, F. V, Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437–1441.

Liu, Y., Esyunina, D., Olovnikov, I., Teplova, M., Kulbachinskiy, A., Aravin, A.A., and Patel, D.J. (2018). Accommodation of Helical Imperfections in Rhodobacter sphaeroides Argonaute Ternary Complexes with Guide RNA and Target DNA. Cell Rep. *24*, 453–462.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear Export of Microrna Precursors. Science (80-. ). *303*, 95–98.

Ma, J.-B., Yuan, Y.-R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666–670.

MacRae, N.T.S. and I.J., Schirle, N.T., MacRae, I.J., and MacRae, N.T.S. and I.J. (2012). The Crystal Structure of Human Argonaute2. Science (80-. ). *336*, 1037–1040.

Makarova, K.S., Wolf, Y.I., van der Oost, J., and Koonin, E. V (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol. Direct *4*, 29.

Marraffini, L.A., and Sontheimer, E.J. (2010a). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. *11*, 181–190.

Marraffini, L.A., and Sontheimer, E.J. (2010b). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568–571.

Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563–574.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607–620.

Matsumoto, N., Nishimasu, H., Sakakibara, K., Nishida, K.M., Hirano, T., Ishitani, R., Siomi, H., Siomi, M.C., and Nureki, O. (2016). Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell *167*, 484–497.

McCaffrey, A.P., Nakai, H., Pandey, K., Huang, Z., Salazar, F.H., Xu, H., Wieland, S.F., Marion, P.L., and Mark A Kay (2003). Inhibition of hepatitis B virus replication and expression by RNA interference in vitro. Nat. Biotechnol. *85*, 2503–2506.

Van De Meent, J.W., Bronson, J.E., Wiggins, C.H., and Gonzalez, R.L. (2014). Empirical bayes methods enable advanced population-level analyses of single-molecule FRET experiments. Biophys. J. *106*, 1327–1337.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. *14*, 447–459.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., et al. (2008). Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide. Cell *133*, 116–127.

Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J., and Kosmrlj, A. (2009). How a protein searches for its site on DNA: the mechanism of facilitated diffusion. J. Phys. A Math. Theor. *42*, 434013.

Miyoshi, T., Ito, K., Murakami, R., and Uchiumi, T. (2016). Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. Nat. Commun. *7*, 1–12.

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature *457*, 413–420.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368–374.

Nam, Y., Chen, C., Gregory, R.I., Chou, J.J., and Sliz, P. (2011). Molecular basis for interaction of let-7 MicroRNAs with Lin28. Cell *147*, 1080–1091.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279–289.

Nishida, H. (2012). Comparative Analyses of Base Compositions, DNA Sizes, and Dinucleotide Frequency Profiles in Archaeal and Bacterial Chromosomes and Plasmids. Int. J. Evol. Biol. *2012*, 1–5.

O'Geen, H., Ren, C., Coggins, N.B., Bates, S.L., and Segal, D.J. (2018). Unexpected binding behaviors of bacterial Argonautes in human cells cast doubts on their use as targetable gene regulators. PLoS One *13*, 1–12.

Oliner, K., Juan, T., Suggs, S., Wolf, M., Sarosi, I., Freeman, D.J., Gyuris, T., Baron, W., Bakker, A., Parker, A., et al. (2010). A comparability study of 5 commercial KRAS tests. Diagn. Pathol. *5*, 23.

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D., and Aravin, A. (2013). Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. Mol. Cell *51*, 594–605.

Oxnard, G.R., Paweletz, C.P., Kuang, Y., Mach, S.L., O'Connell, A., Messineo, M.M., Luke, J.J., Butaney, M., Kirschmeier, P., Jackman, D.M., et al. (2014). Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin. Cancer Res. *20*, 1698–1705.

Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2018). PIWI-interacting RNAs: small RNAs with big functions. Nat. Rev. Genet.

Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., and Milos, P.M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell *143*, 1018–1029.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (1997). Cosuppression in Drosophila: Gene Silencing of Alcohol dehydrogenase by white-Adh Transgenes Is Polycomb Dependent. Cell *90*, 479–490.

Parker, J.S. (2010). How to slice: snapshots of Argonaute in action. Silence *1*, 3.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663–666.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., et al. (2004). Identification of Virus-Encoded MicroRNAs. Science (80-. ). *304*, 734–737.

Powell, M.J., and Zhang, A. (2016). DNA mutation detection employing enrichment of mutant polynucleotide sequences and minimally invasive sampling. Patent.

Pratt, A.J., and MacRae, I.J. (2009). The RNA-induced silencing complex: A versatile gene-silencing machine. J. Biol. Chem. *284*, 17897–17901.

Qi, J., Dong, Z., Shi, Y., Wang, X., Qin, Y., Wang, Y., and Liu, D. (2016). NgAgo-based fabp11a gene knockdown causes eye developmental defects in zebrafish. Cell Res. *26*, 1349–1352.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., Manuscript, A., Houk, A.R., Jilkine, A., et al. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence- Specific Control of Gene Expression. Curr. Opin. Cell Biol. *152*, 1173–1183.

Ragunathan, K., Liu, C., and Ha, T. (2012). RecA filament sliding on DNA facilitates homology search. Elife *2012*, 1–14.

Rand, T.A., Ginalski, K., Grishin, N. V., and Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc. Natl. Acad. Sci. *101*, 14385–14389.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, and D.P. (2002). MicroRNAs in plants: Possible contributions to phenotypic diversity. Genes Dev.

Rivas, F. V, Tolia, N.H., Song, J.-J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. *12*, 340–349.

Roberts, R.J. (2005). How restriction enzymes became the workhorses of molecular biology. Proc. Natl. Acad. Sci. U. S. A. *102*, 5905–5908.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T.F.,

Dybvig, K., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. *31*, 1805–1812.

Rocha, E.P.C., and Danchin, A. (2002). Base composition bias might result from competition for metabolic resources. Trends Genet. *18*, 291–294.

Romano, N., and Macino, G. (1992). Quelling: Transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol. Microbiol. *6*, 3343–3353.

Rotem Sorek, V.K. and P.H. (2008). CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol. *6*, 181–186.

Russell, R.J.M., Ferguson, J.M.C., Hough, D.W., Danson, M.J., and Taylor, G.L. (1997). The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiosus at 1.9 Å resolution. Biochemistry *36*, 9983–9994.

Ryazansky, S., Kulbachinskiy, A., and Aravin, A.A. (2018). The Expanded Universe of Prokaryotic Argonaute Proteins. MBio *9*, e01935-18.

Salomon, W., Bulock, K., Lapierre, J., Pavco, P., Woolf, T., and Kamens, J. (2010). Modified dsRNAs that are not processed by Dicer maintain potency and are incorporated into the RISC. Nucleic Acids Res. *38*, 3771–3779.

Salomon, W.E., Jolly, S.M., Moore, M.J., Zamore, P.D., and Serebrov, V. (2015). Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. Cell *162*, 84–95.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355.

Savić, N., and Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. Transl. Res. *168*, 15–21.

Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. Science (80-. ). *346*, 608–613.

Schirle, N.T., Sheu-Gruttadauria, J., Chandradoss, S.D., Joo, C., and MacRae, I.J. (2015). Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. Elife *4*, 1–16.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.

Schwarzenbach, H., Hoon, D.S.B., and Pantel, K. (2011). Cell-free nucleic acids as biomarkers in cancer patients. Nat. Rev. Cancer *11*, 426.

Sen, G.L., and Blau, H.M. (2006). A brief history of RNAi : the silence of the genes. Faseb J. *20*, 1293–1299.

Shabalina, S.A., and Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. Trends Ecol. Evol. *23*, 578–587.

Shao, D., Lin, Y., Liu, J., Wan, L., Liu, Z., Cheng, S., Fei, L., Deng, R., Wang, J., Chen, X., et al. (2016). A targeted next-generation sequencing method for identifying clinically relevant mutation profiles in lung adenocarcinoma. Sci. Rep. *6*, 1–9.

Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. *37*, 1–10.

Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D.C., van der Oost, J., Patel, D.J., and Wang, Y. (2014). Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. Proc. Natl. Acad. Sci. U. S. A. *111*, 652–657.

Sijen, T., Steiner, F. a, Thijssen, K.L., and Plasterk, R.H. a (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244–247.

Slutsky, M., and Mirny, L.A. (2004). Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. Biophys. J. *87*, 4021–4035.

Smalheiser, N.R., and Gomes, O.L.A.A. (2015). Mammalian Argonaute-DNA binding? Biol. Direct *10*, 1–11.

Song, C., Liu, Y., Fontana, R., Makrigiorgos, A., Mamon, H., Kulke, M.H., and Makrigiorgos, G.M. (2016a). Elimination of unaltered DNA in mixed clinical samples via nuclease-assisted minor-allele enrichment. Nucleic Acids Res. *44*, e146–e146.

Song, J.-J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R. a, Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat. Struct. Biol. *10*, 1026–1032.

Song, J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. Science (80-. ). *305*, 1434–1437.

Song, J., Mauk, M.G., Hackett, B.A., Cherry, S., Bau, H.H., and Liu, C. (2016c). Instrument-Free Point-of-Care Molecular Detection of Zika Virus. Anal. Chem. *88*, 7289–7294.

Song, J., Liu, C., Bais, S., Mauk, M.G., Bau, H.H., and Greenberg, R.M. (2016b). Molecular Detection of Schistosome Infections with a Disposable Microfluidic Cassette. PLoS Negl. Trop. Dis. *9*, e0004318.

Song, J., Pandian, V., Mauk, M.G., Bau, H.H., Cherry, S., Tisi, L.C., and Liu, C. (2018). Smartphone-Based Mobile Detection Platform for Molecular Diagnostics and Spatiotemporal Disease Mapping. Anal. Chem. *90*, 4823–4831.

Song, J., Hegge, J.W., Mauk, M.G., Chen, J., Bhagwat, N., Till, J.E., Azink, L.T., Peng, J., Sen, M., Mays, J., et al. (2019). Highly Specific Enrichment of Rare Nucleic Acids using Thermus Thermophilus Argonaute. BioRxiv 491738.

Steensels, J., and Verstrepen, K.J. (2016). Stop that Noise and Turn Up the Antisense Transcription. Cell Rep. *15*, 2575–2576.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67.

Swarts, D.C., and Jinek, M. (2018). Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. Wiley Interdiscip. Rev. RNA *9*, 1–19.

Swarts, D.C.., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., and Van Der Oost, J. (2015a). Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. Nucleic Acids Res. *43*, 5120–5129.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J.J., et al. (2014b). DNA-guided DNA interference by a prokaryotic Argonaute. Nature *507*, 258–261.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F.R.F., Koonin, E. V., Patel, D.J., and Van Der Oost, J. (2014a). The evolutionary journey of Argonaute proteins. Nat. Struct. Mol. Biol. *21*, 743–753.

Swarts, D.C., Koehorst, J.J., Westra, E.R., Schaap, P.J., and Van Der Oost, J. (2015b). Effects of argonaute on gene expression in Thermus thermophilus. PLoS One *10*, 1–13.

Swarts, D.C., Szczepaniak, M., Sheng, G., Chandradoss, S.D., Zhu, Y., Wang, Y., Swarts, D.C., Szczepaniak, M., Sheng, G., Chandradoss, S.D., et al. (2017a). Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. Mol. Cell *65*, 985–998.

Swarts, D.C., van der Oost, J., and Jinek, M. (2017b). Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. Mol. Cell *66*, 221–233.e4.

Swarts, D.C., Szczepaniak, M., Sheng, G., Wang, Y., Swarts, D.C., Szczepaniak, M., Sheng, G., Chandradoss, S.D., and Zhu, Y. (2017c). Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. Mol. Cell *65*, 985--998.e6.

Taly, V., Pekin, D., Benhaim, L., Kotsopoulos, S.K., Le Corre, D., Li, X., Atochin, I., Link, D.R., Griffiths, A.D., Pallier, K., et al. (2013). Multiplex Picodroplet Digital PCR to Detect KRAS Mutations in Circulating DNA from the Plasma of Colorectal Cancer Patients. Clin. Chem. *59*, 1722 LP-1731.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz,

R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534–538.

Tatsumi, K., Mitani, Y., Watanabe, J., Takakura, H., Hoshi, K., Kawai, Y., Kikuchi, T., Kogo, Y., Oguchi-Katayama, A., Tomaru, Y., et al. (2008). Rapid screening assay for KRAS mutations by the modified smart amplification process. J. Mol. Diagn. *10*, 520–526.

Thompson, M.J., and Eisenberg, D. (1999). Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J. Mol. Biol. *290*, 595–604.

Travers, A., and Muskhelishvili, G. (2005). DNA supercoiling - a global transcriptional regulator for enterobacterial growth? Nat. Rev. Microbiol. *3*, 157–169.

Tropea, J.E., Cherry, S., and Waugh, D.S. (2009). Expression and Purification of Soluble His6-Tagged TEV Protease. In High Throughput Protein Expression and Purification: Methods and Protocols, S.A. Doyle, ed. (Totowa, NJ: Humana Press), pp. 297–307.

Tsiatis, A.C., Norris-Kirby, A., Rich, R.G., Hafez, M.J., Gocke, C.D., Eshleman, J.R., and Murphy, K.M. (2010). Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. J. Mol. Diagn. *12*, 425–432.

Tsvetanova, B., Peng, L., Liang, X., Li, K., Yang, J.P., Ho, T., Shirley, J., Xu, L., Potter, J., Kudlicki, W., et al. (2011). Genetic assembly tools for synthetic biology. Methods Enzymol. *498*, 327–348.

Vandivier, L.E., Anderson, S.J., Foley, S.W., and Gregory, B.D. (2016). The Conservation and Function of RNA Secondary Structure in Plants. Annu. Rev. Plant Biol. *67*, 463–488.

Wang, F., Redding, S., Finkelstein, I.J., Gorman, J., Reichman, D.R., and Greene, E.C. (2013). The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. Nat. Struct. Mol. Biol. *20*, 174–181.

Wang, J.W., Wang, A., Li, K., Wang, B., Jin, S., Reiser, M., and Lockey, R.F. (2015). CRISPR/Cas9 nuclease cleavage combined with Gibson assembly for seamless cloning. Biotechniques *58*, 161–170.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008a). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921–926.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008b). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209–213.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754–761.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539–543.

Wee, L.M., Flores-Jasso, C.F., Salomon, W.E., and Zamore, P.D. (2012). Argonaute divides Its RNA guide into domains with distinct functions and RNA-binding properties. Cell *151*, 1055–1067.

Werner, A., Cockell, S., Falconer, J., Carlile, M., Alnumeir, S., and John Robinson (2014). Contribution of natural antisense transcription to an endogenous siRNA signature in human cells. BMC Genomics 1–12.

Westra, E.R., van Erp, P.B.G., Künne, T., Wong, S.P., Staals, R.H.J., Seegers, C.L.C., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. Mol. Cell *46*, 595–605.

Wiedmann, M., Wilson, W.I., Luo, J., Barany, F., and Batt, A. (1994). Ligase Chain Reaction Applications. Genome Res. *3:*, S51–S64.

Wightman, B., Ha, L., and Gary, R. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern. Cell *75*, 855–862.

Willkomm, S., Oellig, C.A., Zander, A., Restle, T., Keegan, R., Grohmann, D., and Schneider, S. (2017a). Structural and mechanistic

insights into the DNA-guided DNA endonuclease activity of an archaeal Argonaute. Nat. Microbiol. *17035*, 1–7.

Willkomm, S., Oellig, C.A., Zander, A., Restle, T., Keegan, R., Grohmann, D., and Schneider, S. (2017b). Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. Nat. Microbiol. *2*, 17035.

Willkomm, S., Makarova, K., and Grohmann, D. (2018). DNA-silencing by prokaryotic Argonaute proteins adds a new layer of defence against invading nucleic acids. FEMS Microbiol. Rev.

Wu, L.R., Chen, S.X., Wu, Y., Patel, A.A., and Zhang, D.Y. (2017). Multiplexed enrichment of rare DNA variants via sequence-selective and temperature-robust amplification. Nat. Biomed. Eng. *1*, 714–723.

Wu, W.Y., Lebbink, J.H.G., Kanaar, R., Geijsen, N., and van der Oost, J. (2018). Genome editing by natural and engineered CRISPR-associated nucleases. Nat. Chem. Biol. *14*, 642–651.

Wunderlich, Z., and Mirny, L.A. (2008). Spatial effects on the speed and reliability of protein-DNA search. Nucleic Acids Res. *36*, 3570–3578.

Xie, C., Fu, L., Jin, Z., Han, L., Zhang, A., Jin, M., Tu, Z., and Xiang, Y. (2019). The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria.

Xue, C., Zhu, Y., Zhang, X., Shin, Y.K., and Sashital, D.G. (2017). Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. Cell Rep. *21*, 3717–3727.

Yamano, T., Zetsche, B., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. Mol. Cell *67*, 633–645.e3.

Yeom, K.H., Heo, I., Lee, J., Hohng, S., Kim, V.N., and Joo, C. (2011). Single-molecule approach to immunoprecipitated protein complexes: Insights into miRNA uridylation. EMBO Rep. *12*, 690–696.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. 3011–3016.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol. Cell *19*, 405–419.

Zamore, P.D., Tuschl, T., Sharp, P. a, and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25–33.

Zander, A., Holzmeister, P., Klose, D., Tinnefeld, P., and Grohmann, D. (2014). Single-molecule FRET supports the two-state model of Argonaute action. RNA Biol. *11*, 45–56.

Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., et al. (2017). Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. Nat. Microbiol. *2*, 1–10.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell *163*, 759–771.

# Acknowledgements

Although its only my name on the cover of this thesis, it has been a true team effort. Therefore, I would like to thank my colleagues, collaborators, friends and family for their support during these four years.

Allereerst wil ik jou bedanken **John** voor het vertrouwen in mij - eerst als student, toen als PhD kandidaat en nu als postdoc. Ik ben erg blij dat ik dankzij jou voor mijn afstuderen bij Caribou in Berkeley (USA) aan de slag kon, wat als ideale voorbereiding gold voor mijn promotieonderzoek. Ondanks dat in het begin de Argonaute eiwitten afkomstig uit de door jou zo geliefde cyanobacteriën niet echt mee wilden werken, hielden we hoop en met name dankzij jou positieve, inspirerend en motiverende begeleiding is alles toch nog op zijn pootjes is terecht gekomen. John bedankt voor de geweldige tijd, en ik ben enorm benieuwd naar het vervolg van "The Journey of the Argonautes". **Raymond**, leuk dat je mijn copromotor wil zijn, het was altijd fijn om ideetjes eerst even bij jou te kunnen peilen of om voor advies je kantoor binnen te lopen als John weer eens camera-crew over de vloer had.

**Daan** ik heb enorm veel van je geleerd zowel als schrijver, als op het lab. Ook de momenten buiten het lab waren altijd gezellig. Ik vond het een eer je paranimf te mogen zijn en ben blij dat jij samen met Tijn mij op de grote dag assisteren. Ik vind het erg leuk dat jij en Fieke weer terug zijn in Wageningen en dat jij twee verdiepingen lager je eigen groep aan het opstarten bent binnen de vakgroep Biochemie. Veel succes! **Tijn**, wat hebben we een hoop lol gehad. Geweldig dat je de oceaan oversteekt om mij op de belangrijke dag op het podium bij te staan als paranimf. Ik hoop ook snel een keer jou kant op te komen om te zien hoe jij en Marjolein "The American Dream" aan het beleven zijn.

**Jasper**, bedankt voor de gezellige tijd. Ik kijk uit naar de komende trip met jou en Irene. **Hugo**, dankzij jou weet ik nu van alles over paddenstoelen, arcade kasten, Bitcoins en ga nog maar even door - ik kijk uit naar ons Habanero project. De avonturen met jou en Mirelle waren onovertroffen. **Franklin**, it was fun working with you and thanks for all the memorable moments at the Javastraat. **Melvin**, hopelijk komen die startup plannen van ons ooit nog van de grond. Dank voor de leuke tijd. **Nikolas** it was fun doing many TSP courses together with you. Thanks for the good times in and outside of work, especially that last LAN-party was a blast.

**Marc** ook jij bedankt, als stagebegeleider bij AVEBE stond je aan de basis van mij carrière als wetenschapper. Jij was degene die me ervan overtuigde om een Master te gaan volgen. Ik ben trots dat ik dit heb kunnen waar maken, dank voor de mooie tijd. **Rachel** and **Andy**, thanks for your support during my stay at Caribou, it was an amazing learning experience and great to see what it is like to work in a start-up company. I appreciate that you integrated me so well into the American culture by taking me to baseball games and to the memorable Friday afternoon IPA's at Prizefighter.

I would also like to thank our collaborators, **Chirlmin**, **Thijs** and **Stanley**. Thanks to your enthusiasm and interest in *Cb*Ago it always felt like we had gold in our hands. **Jinzhao**, **Michael** and **Haim** thanks for the collaboration on the NAVIGATER project; although we did not publish our story yet I am sure it will find its way into a nice journal.

Zonder de hulp van de vaste kern van Microbiologie is het doen van een PhD onmogelijk. Daarom ontzettend bedankt **Rob**, **Wim**, **Steven**, **Philippe**, **Ton**, **Sjon**, **Ineke, Merlijn**, **Tom** en **Tom** voor jullie technische hulp en alle bestellingen door de jaren heen. **Anja**, **Heidi** en **Carolien**, jullie ook bedankt dat ik altijd bij jullie kantoor binnen kon lopen voor wat voor vraag dan ook. I would like to thank my students for all their efforts in the Argonaute project **Ismael, Nirajan, Tess, Robert, Ilona, Jeroen, Sabrina, Koen, Luuk, Lotte** and **Jie**. I really enjoyed supervising every single one you and I wish you all the best in your future careers. **Carina, Martijn, Angelina**, the CRISPR-clear dream team, het was een erg leuk om jullie te mogen begeleiden met jullie prijswinnende project.

I would also like to thank team Argonaute, **Ismael** and **Jie.** Although I was writing a majority of the time, it was a pleasure working with both of you during the last part of my PhD. **Jurre** I am happy you started your PhD on the Argonaute-project. I look forward collaborating with you on many nice projects to come. Then my office mates: **Daan, Tim, Marnix, Stan, Becca, Yifan, Prarthana, Wen, Sjoerd, Despoina, Thomas.** I really enjoyed all our scientific discussions, chats and jokes. Also a big thanks to all other (present or previous) colleagues of Bacterial Genetics **Serve, Bas**, **Teunke, Nico, Alex, Joyshree, Mihris, Yannis, Jeroen**, **Belen, Max, Thijs, Janneke, Joep and James.** Thanks for the good times and for making BacGen such a nice group to work in. Finally also thanks to **all the other members of Microbiology and Systems and Synthetic Biology** for the amazing times we had during the years.

**Tine** en **Ton** bedankt voor de vele lekkere maaltijden en gezellige avonden. Jullie interesse tijdens mijn studie en PhD was enorm. Dit heb ik altijd erg gewaardeerd, bedankt! **Pap, mam, Michiel en Martijn**, bedankt dat jullie er altijd voor me zijn. De weekenden thuis waren altijd om naar uit te kijken! **Family Ibars**, my Catalan family, spending time with you always felt like holidays. Thanks for showing me around in the beautiful Catalunya. Gràcies!

Last but not least, **Maria**. You are the cornerstone of this PhD thesis. I cannot thank you enough for coming to the Netherlands to live with me. I admire you and I thank you for your support, advice, humour and love that kept me inspired during the PhD. I am looking forward to all our adventures to come. T'estimo!

Thank you,

Gràcies,

Bedankt!

# List of publications

DC Swarts, **JW Hegge**, I Hinojo, M Shiimori, M Ellis, J Dumrongkulraksa, RM Terns, MP terns, J van der Oost (2015) *Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA*. **Nucleic acids research**.

**JW Hegge***, DC Swarts*, J van der Oost (2018) *Prokaryotic Argonaute proteins: novel genome-editing tools?* **Nature Reviews Microbiology**, 16(1), p.5.

**JW Hegge**, DC Swarts, SD Chandradoss, TJ Cui, J Kneppers, M Jinek, C Joo, J van der Oost (2019) *Argonaute of Clostridium butyricum is a DNA guided single stranded DNA* endonuclease. **Accepted for peer-review in Nucleic acids research**

**JW Hegge***, I Hinojo* J Liu, J Van der Oost. *Jason Cloning: a novel one-pot cloning method using Pyrococcus furiosus Argonaute*. Manuscript in preparation

J Song*, **JW Hegge***, MG Mauk, N Bhagwat, JE Till, M Sen, J Mays, E Carpenter, LT Azink, J van der Oost, HH Bau. *Enrichment of rare nucleic acids using Thermus thermophilus Argonaute.* **Submitted and on Biorxiv**

TJ Cui, M Klein, **JW Hegge**, SD Chandradoss, J van der Oost, C Joo (2019) *Gliding and intersegmental jumping allow Argonaute to bypass nucleic acid-associated obstacles during target search.* **Accepted for peer-review in Nature communications**

AC Nieuwenweg, MM van Galen, A Horsting, **JW Hegge**, A Velders, V Saggiomo. *CRISPR-Clear: A Fieldable Detection Procedure for Potential CRISPR-Cas9 Gene Drive Based Bioweapons.* **Submitted and on BioRxiv**

*These authors contributed equally

## Patents

J van der Oost, **JW Hegge**. *Prokaryotic Argonaute proteins and uses thereof.* PCT/EP2017/067462 (filed July 11, 2017)

J Song, J van der Oost, **JW Hegge**, M Mauk, C Liu, HH Bau. *Enrichment of nucleic acids using Thermus thermophilus Argonaute* (2018)

# Co-author affiliations

Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, 6708WE Wageningen, The Netherlands.
John van der Oost, Ismael Hinojo, Jeroen Kneppers, Jie Liu, Lotte T. Azink.

Kavli Institute of NanoScience, Department of BioNanoScience, Delft University of Technology, 2629HZ Delft, The Netherlands.
Chirlmin Joo, Stanley D. Chandradoss, Tao J. Cui, Misha Klein, Martin Depken.

Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia PA, USA.
Jinzhao Song, Michael G. Mauk. Junman Chen, Jing Peng, Haim H. Bau.

Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA
Neha Bhagwat, Jacob E. Till, Moen Sen, Jazmine Mays, Erica Carpenter.

Laboratory of BioNanoTechnology, Wageningen University and Research, Axis, Bornse Weilanden 9, 6708 WG Wageningen, The Netherlands.
Anna C. Nieuwenweg, Martijn M. van Galen, Angelina Horsting, Aldrik H. Velders, Vittorio Saggiomo.

Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland.
Daan C. Swarts, Martin Jinek.

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, USA.
Masami Shiimori, Michael Ellis, Justin Dumrongkulraksa, Rebecca M. Terns, Michael P. Terns.

# About the Author

Jorrit Hegge was born on the 15th of August, 1988, in Opende (GR), the Netherlands. After completing secondary school (HAVO) at the Drachtster lyceum in Drachten, he started a Bachelor Biology and Medical Laboratory Research at the Hanze University of Applied Sciences in Groningen. Both his BSc theses were performed at the starch manufacturer AVEBE in Veendam (GR), under the supervision of Prof. Dr Marc van der Maarel. Jorrit studied the effect of starch-branching enzymes on the physical and chemical properties of starch. After obtaining his BSc he started a master Molecular Life Sciences in Wageningen where he specialized in Biological Chemistry. During his MSc thesis at the Laboratory of Microbiology of Wageningen University, Jorrit studied *Tt*Ago-interacting proteins under the supervision of Dr. Daan Swarts. His MSc internship was performed at Caribou Biosciences in Berkeley (California) under supervision of Dr. Andrew May and Dr. Rachel Haurwitz. Here he investigated the suitability of *Tt*Ago for genome editing applications in mammalian cell lines. After obtaining his MSc degree in 2014 he started his PhD project at the laboratory of Microbiology under the supervision of promotor John van der Oost and later also under supervision of co-promotor Raymond Staals. His work on the characterization and application of pAgos is described in this thesis.

# Overview of completed training activities

**Discipline specific activities**

*Meetings & Conferences*

- ✓ NWO-ALW molecular genetics meeting. Lunteren, NL (2015)*
- ✓ NWO-CHAINS, Dutch chemistry conference. Veldhoven, NL (2016)*
- ✓ Microbiology Centennial meeting. Wageningen, NL (2017)
- ✓ IMBA, Small RNA Biology meeting. Vienna, NL (2017)*
- ✓ NWO-ALW Host microbe genetics meeting. Wageningen, NL (2017)**
- ✓ NWO-ALW Host microbe genetics meeting. Groningen, NL (2018)**

*Poster presentation, ** Oral presentation

*Courses*

- ✓ Radiation Course level 5B, Boerhaave scholing, Leiden, NL (2015)
- ✓ Advanced proteomics, (VLAG), Wageningen, NL (2017)

**General courses**

- ✓ Pitch perfect
- ✓ The essentials of scientific writing and presenting
- ✓ Brain training
- ✓ Bridging cultural differences
- ✓ Scientific Writing
- ✓ Career orientation
- ✓ Illustrator Workshop

**Optionals**

- ✓ Research proposal
- ✓ PhD Study trip to Germany, Denmark & Sweden
- ✓ Bacterial genetics group meetings (weekly)
- ✓ Microbiology PhD meetings (monthly)
- ✓ Microbiology seminars

The research described in this thesis was financially supported by an ECHO grant from the Netherlands Organization of Scientific Research (NWO) to Prof. Dr John van der Oost. (grant number 711013002)

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

**Cover design:** By Philip Patenall (Art editor Nature Reviews Microbiology)

**Printed:** GVO drukkers & vormgevers B.V. Ede