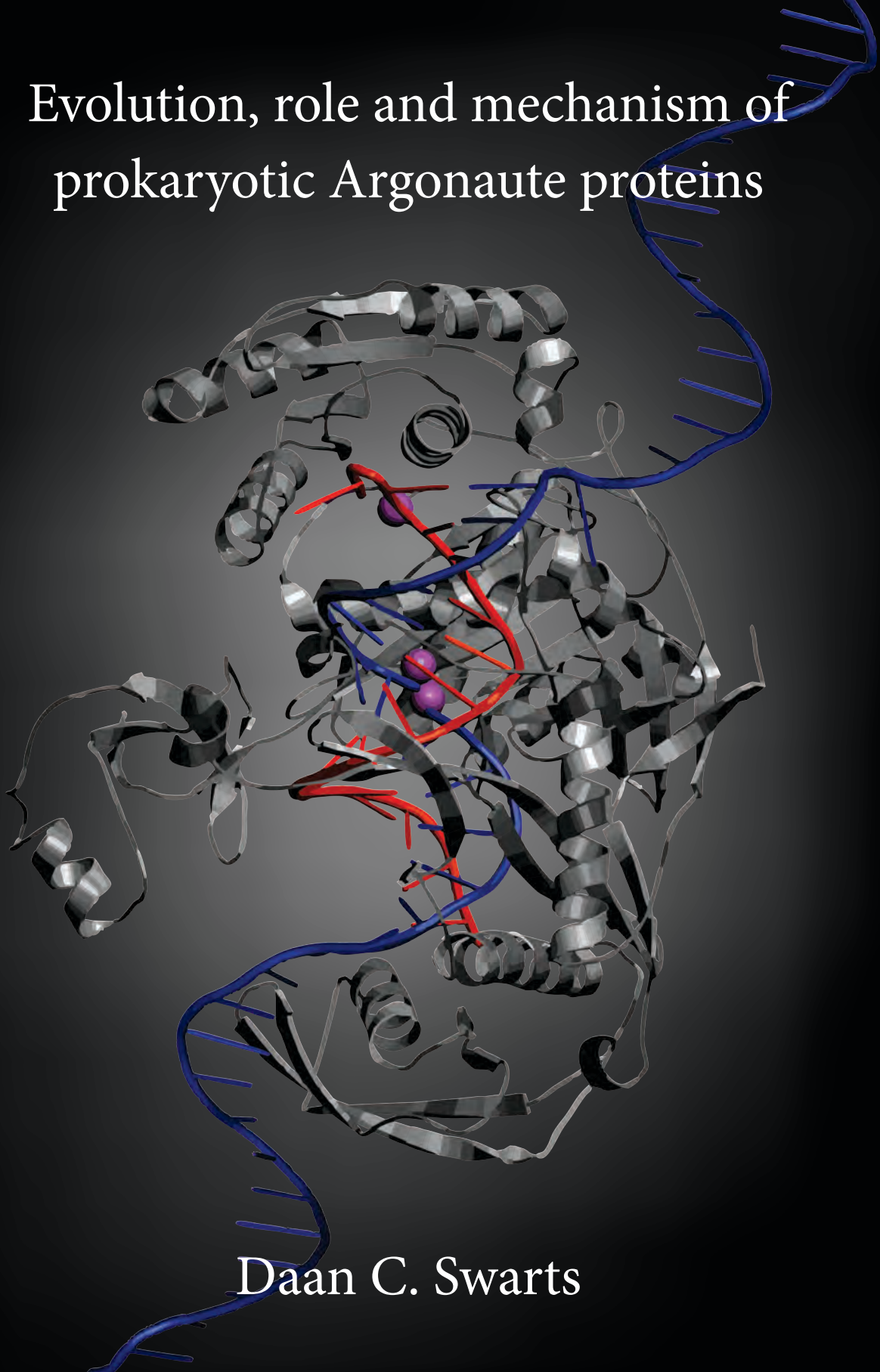
Evolution, role and mechanism of prokaryotic Argonaute proteins



Daan C. Swarts

Evolution, role and mechanism of prokaryotic Argonaute proteins

Daan C. Swarts

Thesis committee

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

Co-promotor
Dr Stan J.J. Brouns
Assistant professor, Laboratory of Microbiology
Wageningen University

Other members
Prof. Dr Willem J.H. van Berkel, Wageningen University
Dr Thijn Brummelkamp, The Netherlands Cancer Institute, Amsterdam
Prof. Dr Martin Jinek, University of Zurich, Switzerland
Dr Ronald van Rij, Radboud University Medical Centre, Nijmegen

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

Evolution, role and mechanism of prokaryotic Argonaute proteins

Daan C. Swarts

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Academic Board,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 2 June 2015
at 4 p.m. in the Aula.

Daan C. Swarts
Evolution, role and mechanism of prokaryotic Argonaute proteins
228 pages.

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summaries in Dutch and English

ISBN 978-94-6257-295-9

Table of Content

Preface and Outline

RNA interference and the Argonaute protein

The Argonaute protein (Ago) has been described for the first time in an *Arabidopsis thaliana* study in 1998 [1]. In this study, randomly mutagenized *A. thaliana* plants were visually screened for abnormal leaf morphology to identify genes involved in plant development. The *ago* gene knockout severely disturbed plant growth and leaf development. In 2001, it was demonstrated that Agos play an essential role in eukaryotic RNA interference (RNAi) pathways [2]. Since then, many proteins belonging to the Argonaute protein family and the pathways in which these proteins play a role have been identified and characterized (described below). Although research has mainly focused on eukaryotic Argonaute proteins (eAgos) and the pathways in which eAgos are involved, many prokaryotes also encode homologs of Argonaute proteins [3,4]. The roles of prokaryotic Argonaute proteins (pAgos), which do not appear to function in RNAi pathways, have long remained elusive. This chapter gives an overview of the breakthroughs made with respect to eukaryotic RNAi and – more specifically – to Argonaute proteins. It briefly touches upon the many pathways in which eAgo play roles, and reviews the little information available on pAgos. The chapter is concluded with an outline of this thesis, which describes the evolution, role, and mechanism of pAgos.

RNA interference

Since the 1980's it has been known that expression or injection of RNA complementary to specific mRNA targets (antisense RNA) can result in down-regulation of the targeted mRNA (reviewed in [5]). Although the mechanisms underlying these observations were unknown at this point, the potential of antisense RNA in regulation of gene expression was already mentioned [5]. It appeared that both antisense and sense RNA suppress gene expression, and this phenomenon was termed 'co-suppression' (in plants) [6] or 'quelling' (in fungi) [7]. It was shown that co-suppression occurs post transcription, and is the result of stimulated RNA degradation [8]. However, the mechanism for post-transcriptional gene silencing remained unclear. Later it became clear that not sense RNA but unintended production of double stranded RNA (dsRNA) is the most obvious explanation of transcriptional repression by sense RNAs [9]. In 1998, Fire *et al.* described that, in the nematode *Caenorhabditis elegans*, insertion of double-stranded RNA causes suppression of complementary mRNAs 10- to 100-fold more effective compared to inserting each strand individually, and coined the term RNA interference [10]. However, the mechanism underlying RNAi remained elusive. Fire and Mello later received the Nobel Prize for

Physiology or Medicine 'for their discovery of RNA interference – gene silencing by double-stranded RNA'.

Small RNA guides

The next step towards understanding RNAi was made in 1999, when Hamilton *et al.* reported a species of single stranded RNA (ssRNA) molecules with an estimated length of 25 nucleotides, which were involved in post-transcriptional gene silencing of complementary mRNA [11]. A year later Zamore *et al.* [12] and Hammond *et al.* [13] independently described that long dsRNA fragments are processed into small dsRNA fragments with a length of 21 to 23 base pairs, which thereafter guide mRNA cleavage. These small dsRNA molecules were termed 'small interfering RNAs' (siRNAs). Additional proof that siRNAs guide RNA interference was given by Elbashir *et al.*, who showed that the addition of synthetic siRNAs to a *Drosophila* cell extract resulted in cleavage of complementary firefly luciferase mRNA [14].

Dicer and Argonaute

Argonaute family proteins have been described before their role was linked to RNAi pathways. The first notion of PIWI, which belongs to the Argonaute family, was made in 1997 by Lin *et al.* who studied mutations in *Drosophila melanogaster* by P-element transposons [15]. P-element transposons allow the insertion of a mobile element into the genome at a random location, and are therefore commonly used in *D. melanogaster* genome mutagenesis. As mutation of a specific gene resulted in abolished germline stem cell division (resulting in small testicles), the gene was named *piwi* (P-element induced wimpy testis). The name 'Argonaute' was coined one year later, in an *A. thaliana* knockout study (**Box 1**) [1], but at that time the role of eAgos was not yet unraveled.

In 2001, Dicer (an RNase III family protein) was identified as the protein responsible for the processing (dicing) of long dsRNA and thus for siRNA generation [16]. It was also shown that subsequent target mRNA cleavage (slicing) was not performed by Dicer, but by a multiprotein complex termed the RNA-induced silencing complex (RISC). The first link between eAgos and RNAi was made by Hammond *et al.* in 2001, who identified eAgo as an essential component of RISC [2]. They showed that eAgo is indispensable for RNAi, but it remained unclear whether this protein was responsible for slicing [2,17]. Experimental proof for that was provided by Liu *et al.*, who showed that human AGO2 (hAGO2) possessed slicing activity [18]. Matranga *et al.* and Rand *et al.* demonstrated that eAgos are

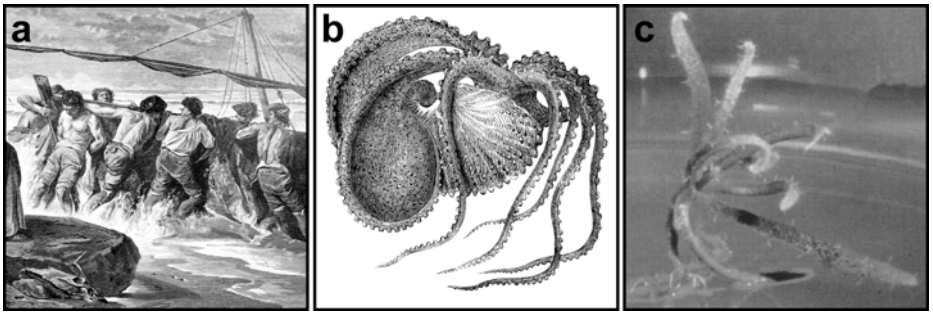
loaded with the duplex siRNAs (the cleavage products of Dicer), after which one of the strands (the passenger strand) is cleaved by eAgo, resulting in removal of the cleaved passenger strand and allowing the other strand to function as guide [19,20]. Concluding, eAgos utilize small RNA guides to bind and cleave targets complementary to the guide.

Box 1 | Origin of the name 'Argonaute'.

The origin of the name 'Argonaute' is found in Greek mythology. The Greek epic poem 'Argonautica' describes how Jason and other heroes, named the Argonauts, retrieve the golden fleece from Colchis [21]. The name Argonaut is composed from 'Argo' (the name of their ship) and 'Nautilus' (latin form of the original Greek word 'ναυτίλος', which means 'sailor'); Argonauts are sailors of the Argo.

The octopus genus *Argonauta* is the only extant genus belonging to the family of Argonautidae [22]. It encompasses different species of pelagic octopuses, of which the females generate a paper-thin shell [22]. While this shell was believed to function as boat, the shell-secreting tentacles of this octopus were assumed to function as peddles or as sails, as the octopus would float at sea surface [23]. Therefore, these octopuses were named after the heroic sailors of the Argo. More recently it has been revealed that the shell is not used to float but instead is used as a brood chamber and to trap surface air to maintain buoyancy [22,24].

It was this octopus after which the gene encoding Argonaute (*ago1*) was named after. *A. thaliana ago1* mutants showed abnormal leaf development and reduction of formation of axillary meristems [1]. As, according to Bohmert *et al*., the curled leaves of the plant gave the plant the appearance of "a small squid", the mutated gene responsible for this phenotype was named after the octopus genus *Argonauta*: Argonaute.



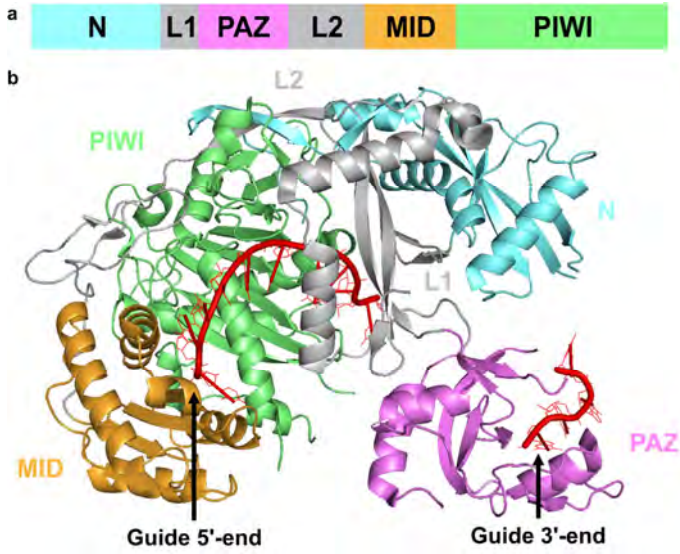
a, Victorian engraving of a scene from Jason and the Argonauts with their ship Argo (© Antiqueimages / Dollar Photo Club). **b**, Drawing of *Argonauta argo* (© Erica Guilane-Nachez / Dollar Photo Club) **c**, 35-day-old *A. thaliana* Δ*ago* seedling. Adapted from Bohmert *et al*., 1998 [1].

Roles of eukaryotic Argonaute proteins

Since the discovery of RNAi many different pathways in which eAgos function have been identified. It should be noted that, although Dicer (or a homolog of Dicer) is responsible for generation of RNA guides and RNA loading in most pathways, members of the Argonaute protein family are the only proteins conserved in all RNAi pathways. There are four families of eAgos: (I) the Ago-like family, (II) the PIWI family, (III) the WAGO family, and (IV) the Trypanosoma Ago family [25,26,27]. The domain composition and architecture of eAgos is conserved throughout these families (**Box 2**).

Box 2 | Argonaute family protein domains, architecture and activity.

Argonaute family proteins have four domains and two linker domains. The PIWI domain is named after the *piwi* gene first described in *D. melanogaster* [15], as this domain was found to be conserved in many genes essential for stem cell renewal [29]. The PIWI domain has an RNase H fold that is responsible for target strand cleavage [30]. The PAZ domain is named after the proteins in which it is found (PIWI/Argonaute/Zwille) [3], and is responsible for binding of the 3'-end of the guide [31]. The middle (MID) and N-terminal (N) domains were originally identified in structures from *Pyrococcus furiosus* pAgo [30]. The MID domain plays a role in binding the 5' end of the guide (reviewed in [32]), while the N domain plays a role in target release [33]. L1 and L2 are linker domains.



a, Schematic domain architecture of Argonaute family proteins. **b**, 2.2 Å structure of the binary complex of human Argonaute 2 (hAGO2) bound to 20 nucleotide miR-20a RNA guide (PDB 4F3T) [28]. The guide RNA is shown in red. Note that nucleotides 11 to 16 are not displayed as they could not be traced in the structure.

Furthermore, The catalytic residues required for target slicing are conserved in 90% of all eAgos [27]. This suggests basic functionality (utilizing an RNA guide to bind and cleave RNA targets) is conserved in most eAgos. eAgos that lack a complete catalytic site function by target binding alone, or recruit additional proteins upon target binding, which can have various outcomes (described below). Differences between pathways in which eAgos function are found in guide generation and loading, and in outcomes of target binding (described below) [34,35]. In general, Ago-like family proteins associate with either miRNAs or siRNAs, and PIWI family proteins utilize piRNAs. Activity of WAGO family proteins is guided by 22G small RNAs, while Trypanosoma Ago family proteins, like Ago-like proteins, bind siRNA guides. Below, the (generalized) pathways that generate the different type of small RNA guides are described.

Ago-like family - siRNA

Primary siRNAs (**Fig. 1**) are generated from long dsRNA precursors. These dsRNAs can be transcribed from the genome as long hairpin RNA or alternatively as sense and antisense transcripts [36,37,38,39,40,41,42]. Also exogenous dsRNA, such as viral RNA or dsRNA taken up from the environment, sometimes functions as siRNA precursor. Long dsRNA is a substrate for Dicer, which generates siRNA duplexes which are 20 to 25 base pairs long. These siRNA duplexes have 5' monophosphates and 2 nucleotide 3'-hydroxyl overhangs [16,43,44,45,46].

Dicer associates with other pathway-specific proteins (TRBP in *Homo sapiens*, or R2D2 in *Drosophila*) to form a loading complex [47,48]. The loading complex determines which strand is presented to eAgo as the guide strand, and which strand becomes the passenger strand. Selection of the guide strand is based on the thermodynamic stability of the ends of the siRNA duplex. In *Drosophila*, R2D2 binds the most stable 5' end of this duplex, while Dicer (Dcr-2) binds the least stable 5' end. While the siRNA duplex is loaded onto eAgo, the strand with the less stable 5' end is selected to become the guide [49]. eAgo binds this strand of the duplex siRNA, while it cleaves and removes the passenger strand [33]. Before siRNA loading, chaperones like HSP70 and/or HSP90 make eAgo available for guide loading, possibly by inducing conformational changes [50,51,52,53,54].

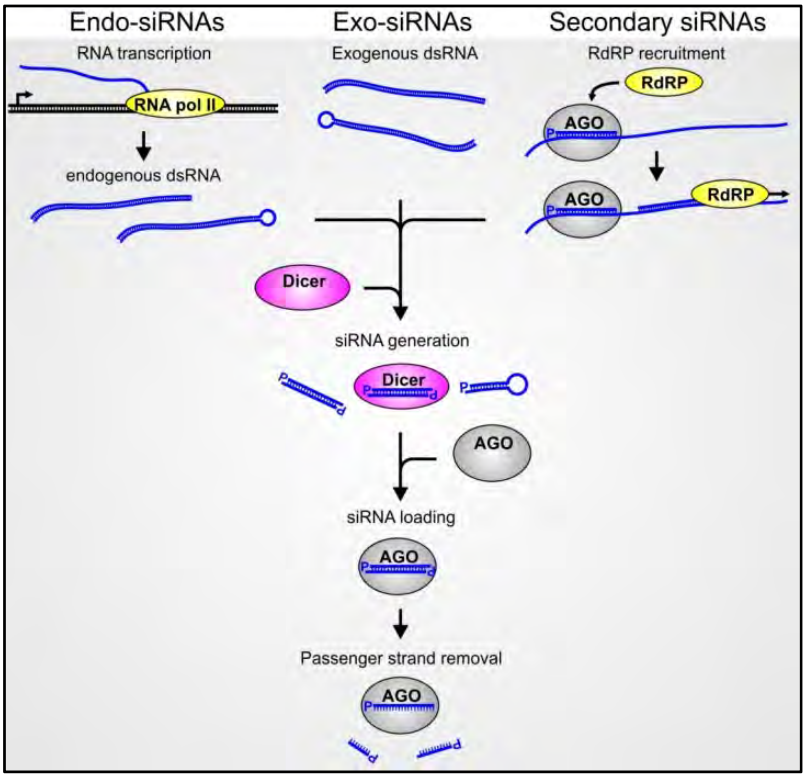


Figure 1 | Schematic overview of siRNA generation. Endo-siRNAs are generated after endogenous expression of long dsRNAs, while Exo-siRNAs have exogenous sources. Secondary siRNAs are generated upon binding of an eAgo-siRNA complex, that recruits RNA-dependent RNA polymerase (RdRP). Independent of its source, dsRNA is processed by Dicer, which generates siRNA duplexes. These duplexes are loaded onto eAgos in a strand-specific manner, after which the passenger strand is cleaved and removed. This figure displays only the most conserved proteins involved in the best studied pathways. Note that many pathways depend on additional (pathway-specific) proteins for guide processing, modification, and loading onto Ago. In addition, pathways exist that do not rely on the proteins displayed here.

Secondary siRNAs are generated via an alternative pathway that is dependent on primary siRNAs. Upon mRNA target binding by eAgo-siRNA complexes, RNA dependent RNA polymerase (RdRP) is recruited. RdRP synthesizes a strand complementary to the target mRNA, resulting in a long dsRNA. This dsRNA is subsequently processed by Dicer, forming secondary siRNAs [55,56,57,58]. eAgo-siRNA complexes generally bind RNA targets which are fully complementary to the siRNA, resulting in target cleavage. This mechanism is used to regulate endogenous gene expression [59]. Alternatively, it can be involved in silencing transposons [40,41] and in antiviral defense (described below).

Ago-like family - miRNA

Biogenesis of miRNA (**Fig. 2**) starts with the RNA polymerase II-mediated expression of endogenous encoded primary miRNA transcripts, which form hairpin structures with one or more imperfect stem loops, termed 'pri-miRNAs' [60,61]. In metazoans, RNase III-family protein Drosha processes pri-miRNAs to shorter pre-miRNAs [62], which are transported from the nucleus to the cytoplasm by Exportin-5/Ran-GTP [63,64]. Cytoplasmic Dicer processes the pre-miRNA to miRNA:miRNA* duplexes which are ~22 base pairs long and loads them onto eAgo [62,65]. In plants, a nuclear Dicer homolog (DCL-1) is responsible for processing of pri-miRNAs to miRNA:miRNA* duplexes [66]. These duplexes are transported to the cytoplasm by HASTY/Ran-GTP [67].

Like siRNAs, Dicer processing of the pre-mRNAs leaves 5' monophosphates and 2 nucleotide 3'-hydroxyl overhangs, after which the least stable 5' end of the duplex is loaded onto eAgo [68,69]. Additionally, the intrinsic structure of miRNA duplexes affects which strand is selected as guide [49,70]. In a Dicer-independent pathway, Drosha-generated pre-miRNAs are processed by the catalytic activity of Ago itself [71,72]. Additionally, some pathways rely on pre-miRNA structures (mitrons) which are formed after pre-mRNA splicing. These pre-miRNA structures are processed by the canonical miRNA biogenesis pathway [73,74].

Although miRNAs that are fully complementary to target mRNA can guide eAgo-mediated target cleavage in rare occasion [75,76], eAgo-miRNA complexes usually bind mRNA targets with imperfect complementarity, which does not allow target RNA cleavage [77]. However, it does tolerate binding to a wide range of targets, allowing regulation of multiple genes with a single miRNA. Most miRNAs are complementary to the 3' untranslated region (UTR) of mRNAs. Binding of the eAgo-miRNA complex to mRNA targets can have various outcomes [59,78,79,80,81], including: (I) Inhibition of translation elongation, by inhibiting translation initiation or premature ribosome dissociation. (II) Co-translational protein degradation, by recruitment of proteases [82]. (III) Competition for the mRNA cap structure, by displacing translation factor elF4E [83]. (IV) Inhibition of ribosomal subunit joining, by recruitment of translation factor elF6 (disrupting association of the large and small ribosomal subunits) [84]. (V) Marking mRNA for decay by recruitment of GW182, which deadenylates and decaps the mRNA [81,85].

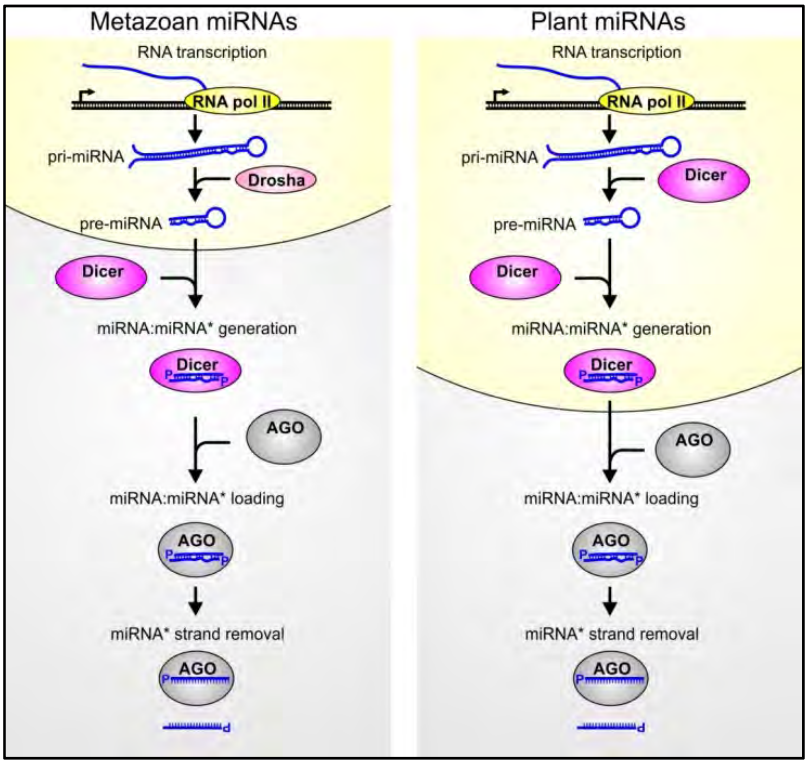


Figure 2 | Schematic overview of miRNA generation. Precursor miRNA termed pri-mRNA is transcribed by RNA polymerase II (RNA pol II) and forms structured hairpins which are processed by Drosha in metazoans or Dicer in plants, generating pre-miRNA. The formed pre-miRNA is further processed by a cytoplasmic Dicer in metazoans, or a nuclear Dicer in plants, generating miRNA:miRNA* duplexes, which are loaded onto eAgo in a strand specific manner. While binding the miRNA strand, the miRNA* strand is released without being cleaved. This figure displays only the most conserved proteins involved in the best studied pathways. Note that many pathways depend on additional (pathway-specific) proteins for guide processing, modification, and loading onto Ago. In addition, pathways exist that do not rely on the proteins displayed here.

PIWI-like family - piRNA

In contrast to siRNAs and miRNAs, piRNAs (**Fig. 3**) are generated from single stranded RNAs. Precursor piRNAs are expressed from intergenic repetitive elements termed piRNA clusters [86], consisting of transposon fragments. These transcripts are processed in a Zucchini dependent process, which produces 23 to 33 nucleotide long piRNAs with a 5' monophosphate and a 3' hydroxyl group, but is further poorly understood (for example, the 3' end is generated by an unknown 3'-5' exonuclease) [87]. Most piRNAs associated with PIWI-like proteins have a 5'-end uracil (U), which is either generated during piRNA biogenesis by Zucchini, or selected for by PIWI [86,88,89,90].

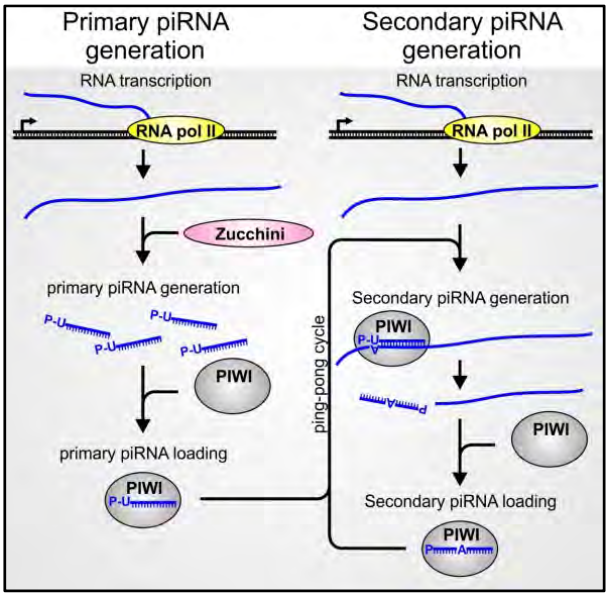


Figure 3 | Schematic overview of piRNA generation. The generation of piRNA is a poorly understood process in which long single stranded RNA transcripts generated by RNA polymerase II (RNA pol II) are processed by Zucchini and an unknown 3'-5' exonuclease, generating primary piRNAs. Primary piRNAs bound to PIWI-like proteins have a strong bias for a 5'-end monophosphorylated (P) uracil (U). Secondary piRNAs are generated in a poorly understood process termed the ping-pong cycle, during which PIWI-like proteins loaded with primary piRNAs cleave single stranded RNA targets. This process results in the generation of secondary piRNAs with an adenine (A) on the position opposite the 5'-end uracil of the primary piRNA. This figure displays only the most conserved proteins involved in the best studied pathways. Note that many pathways depend on additional (pathway-specific) proteins for guide processing, modification, and loading onto PIWI. In addition, pathways exist that do not rely on the proteins displayed here.

In some organisms, PIWI-piRNA complexes initiate a secondary piRNA biogenesis pathway termed the 'ping-pong cycle' [91,92]. In this pathway, PIWI-piRNA complexes cleave RNA targets complementary to the piRNA, generating secondary piRNAs with a strong bias for an adenine (A) at position 10. Secondary piRNAs are utilized by PIWI-like family proteins to cleave primary piRNA precursors, generating additional piRNAs [92]. Generation of piRNAs concludes by 3'-end trimming by an exonuclease [87], after which the 2'-O group of the 3' end is methylated [93]. This protects the piRNA against degradation [94]. PIWI-like proteins mostly function in germline cells, in which DNA or histone methylation is erased. PIWI-piRNA complexes are required in the germline cells to repress expression of mobile genetic elements such as transposons, which in other cells are silenced epigenetically [95,96]. Besides binding and cleaving transposon transcripts post transcription, PIWI-piRNA complexes can prevent transcription of transposable elements. In mice, PIWI-piRNA complexes guide the *de novo* methylation machinery to transposable elements, resulting in CpG methylation of the DNA and subsequent transcriptional silencing [97,98]. In *Drosophila*, non-slicing PIWI-piRNA complexes silence transcription of transposable elements by inducing heterochromatin formation [99,100,101,102,103,104].

WAGO family - G22

Worm Argonaute (WAGO) family proteins are found exclusively in worms and depend on other eAgo families for guide generation. Primary eAgos (such as RDE-1 or PRG-1) recruit RdRP upon target binding. RdRP generates '22G RNAs' (**Fig. 4**), 22 nucleotide long ssRNAs which often have a 5' triphosphate-guanosine (5'-ppp-G) [46,56,105,106]. WAGO family proteins directly utilize 22G RNAs as guide to regulate endogenous gene expression, silence exogenous dsRNA, and to affect chromosome structure and segregation [105,106,107,108,109]. Additionally, some WAGO proteins seem to counteract silencing by other WAGO proteins [110,111].

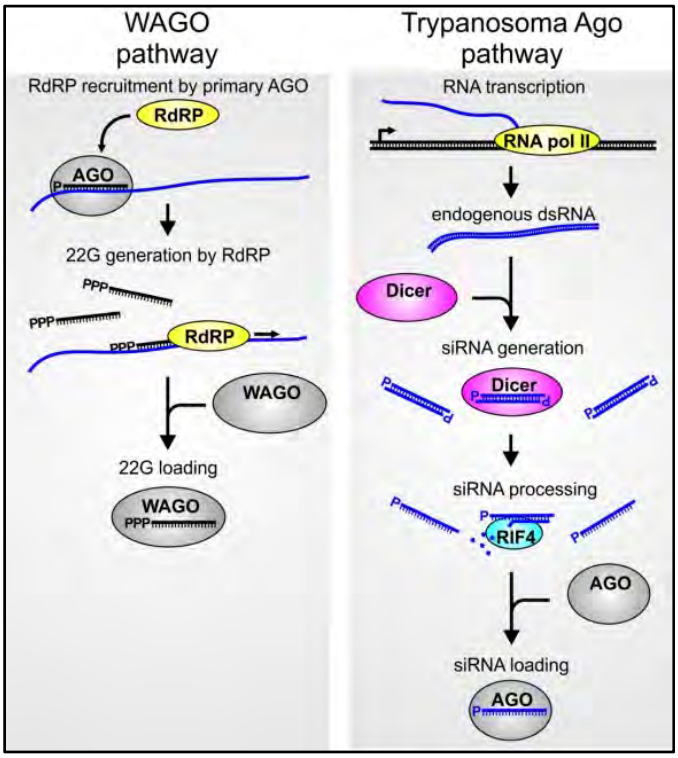


Figure 4 | Schematic overview of G22 RNA and Trypanosoma Ago siRNA generation. G22 generation in WAGO pathways is dependent on primary guide-eAgo complexes which initiate RdRP dependent guide generation. WAGO proteins can directly use the generated short 5'-end triphosphorylated RNAs. For Trypanosoma Ago family proteins, guide generation starts with transcription of endogenous tandem repeat units containing transposon fragments. These transcripts form dsRNA that is processed by Dicer, generating siRNA duplexes. Instead of directly loading the siRNA duplex onto eAgo, RIF4 degrades one of the two strands of the duplex, after which the other strand is loaded onto eAgo.

Trypanosoma Ago family - siRNA

The Trypanosoma Ago family is named after *Trypanosoma brucei*, the model organism to study this protein family. Trypanosoma Agos are guided by siRNAs (**Fig. 4**), although their generation is different from siRNAs that guide Ago-like family proteins. Transcripts from retrotransposons [112] and chromosomal 147 base pairs long tandem repeat units [113]

form dsRNA which is processed into siRNAs by a nuclear Dicer (TbDCL2) [114] or by a cytoplasmic Dicer (TbDCL1) [115], of which the latter relies on TbRIF5 for activity [116]. TbRIF4 is a predicted 3'-5' exonuclease that is essential to convert duplex siRNAs to single stranded guides [116], which are subsequently loaded into TbAGO1. Trypanosoma Agos bind and cleave retrotransposon transcripts, resulting in lowered transposon activity [117].

Antiviral defense by eukaryotic Argonaute

Besides regulating silencing endogenous gene expression and interfering with transposon activity, some eAgos play a role in host defense. In plants, RNAi functions as innate immunity system by silencing host or pathogen genes upon detection of pathogens including nematodes, fungi and protists [118,119,120]. Additionally, viral infection induces the generation of virus-derived siRNAs in plants, fungi, invertebrates and mammals (reviewed in [121,122,123,124,125]).

Upon infection of eukaryotic cells with (+)-strand and (-)-strand RNA viruses, dsRNA viruses, and even DNA viruses, viral small interfering RNAs (vsiRNAs) accumulate in the cell. Like in endogenous siRNA, vsiRNAs are a product of dsRNA processing by Dicer. The vsiRNAs that interfere with DNA viruses appear to be generated by processing of dsRNA formed by overlapping transcripts [126,127,128], while vsiRNAs targeting dsRNA viruses appear to be generated from the genome that is recognized directly by Dicer [129,130]. Single stranded (+)-strand and (-)-strand RNA viruses become Dicer substrates during their replication, when they temporarily become double stranded [131,132,133,134]. Also structured ssRNA elements (such as hairpins), which might be present in (+)-strand or (-)-strand viruses, or alternatively in viral mRNA, can function as a substrate for Dicer [122,128,133,135]. Generated vsiRNAs are fed into the canonical siRNA pathway, eventually resulting in cleavage of viral transcripts or viral genomes. eAgo knockouts in plants, nematodes, and flies, result in lower resistance against viruses [136,137,138,139,140,141,142,143,144]. This indicates that Dicer-mediated processing of viral RNA alone does not provide complete protection, but that eAgo is required to amplify antiviral defense. In addition, primary vsiRNAs can result in generation of secondary vsiRNAs via RdRP homologs, which generate dsRNA that can be processed by Dicer [145]. Virus infection in worms results in generation of secondary vsiRNAs with 5'-end di- or triphosphates, which are loaded onto WAGO-family proteins [146]. Like piRNAs and siRNAs, at least some vsiRNAs become methylated at their 3' end [147], preventing vsiRNA degradation. Upon virus infection of *Drosophila* and mosquitos cells, piRNA-sized viral

RNAs accumulate, which suggests that also PIWI proteins are involved in antiviral defense [130,148,149,150,151]. Like piRNAs, these virus-derived piRNAs (vpiRNAs) have a bias for a 5' uracil (U) and an adenine at position 10, suggesting they have been generated via the ping-pong cycle. In line with the detection of vpiRNAs, PIWI-family proteins have been shown to interfere with virus infection [136,152,153].

In the host virus arms-race, viruses usually evolve mechanisms to counteract antiviral defense systems of their host; indeed, at least plant, fly and mammalian viruses encode viral suppressors of RNAi (VSRs; reviewed in [124,125,154,155,156]). VSRs interfere with RNA silencing by sequestering siRNA duplexes, competing with Dicer by binding dsRNA genomes or dsRNA replication intermediates, or directly bind RNAi proteins inhibiting their activity. Alternatively, some viruses express highly structured RNA that competes with dsRNA substrates for Dicer processing [157].

Overlap in pathways

It should be noted that specific roles are not limited to specific eAgo families (for example, some metazoan eAgo-miRNA complexes do cleave mRNA targets) [75,76]. Also, different pathways can share proteins for guide generation (for example, PIWI-like proteins require Dicer for piRNA generation in ciliates) [158,159]. Methylation of the 2'-O group of the 3' end of the guide is not restricted to piRNAs but also occurs in Trypanosoma Ago pathways and in Ago-like family pathways [160,161,162]. Binding of an Ago-guide complex to a target can affect gene expression in many different ways. In more simple pathways, target binding results in target cleavage. In more complex pathways, which often depend on recruitment of additional proteins, eAgo stimulates translational inhibition, deadenylation of target mRNA, DNA methylation and heterochromatin formation. Thus, besides silencing gene expression post transcription, eAgo can also prevent transcription.

Prokaryotic Argonaute proteins

Cerutti *et al.* performed a bioinformatics study in which they discovered genes encoding proteins with a PIWI domain in prokaryotes in 2000 [3]. pAgos were later used in structural studies, as proteins from thermophilic prokaryotes are more rigid and generally crystalize better than proteins from eukaryotes. The first structure of a full-size Ago protein was derived from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*Ago) [30]. This

structure revealed that Ago has a bilobal structure, of which the PIWI lobe encompasses the PIWI and MID domains, while the PAZ lobe encompasses the PAZ (named after proteins in which it is present: PIWI – Argonaute – Zwille) and N-terminal domains. Additional pAgos were crystalized and characterized *in vitro* [163,164,165,166]. Interestingly, at least some of these pAgos were demonstrated to have a higher affinity for DNA guides than for RNA guides [163,164]. Structures of *Thermus thermophilus* pAgo (*Tt*Ago) were solved in complex with DNA guides and RNA targets, providing insights in guide binding, target binding, and conformational changes involved in target cleavage [167,168,169]. However, the biological significance of DNA-guided pAgos remained unknown.

Prokaryotes lack other proteins essential in RNAi pathways

In 2008, Shabalina and Koonin pointed out that, although prokaryotes encode homologs of both the RNase III and Superfamily II RNA helicase domains of Dicer, prokaryotes do not encode a combination of the two domains [170]. Additionally, eukaryotes appear to have acquired RdRP from bacteriophages rather than from prokaryotes [170]. The absence of these two proteins, which are essential in (most) eukaryotic RNAi pathways, suggests that pAgos do not function in RNAi pathways.

Prokaryotic Argonautes are predicted to be involved in host defense

In 2009, Makarova *et al*. performed a bioinformatics study that uncovered that two types of pAgos exist: long pAgo and short pAgo. Like eAgos, long pAgos encompass four domains, while short pAgos are formed by the PIWI and MID domains alone. In general, these short pAgos appear to have lost their catalytic residues, and instead are encoded by the same gene cluster as predicted nucleases. It has been hypothesized that these nucleases form the catalytic component of a short pAgo-nuclease complex, which suggests the role of pAgo depends on nucleolytic activity. Based on this observation, and in addition the frequent horizontal gene transfer of prokaryotic *ago* genes (which is common for host defense genes), occurrence of prokaryotic *ago* genes in genomic neighborhoods enriched in host defense genes, and their analogy with eAgos, Makarova *et al.* suggested that pAgos are likely to be involved in host defense [4]. Yet, there was no experimental data to back up that prediction.

Thesis outline

Chapter 1 | The evolutionary journey of Argonaute proteins

The first chapter describes the evolution, structure, mechanism and physiological roles of Agos. Ago has its origins in prokaryotes, where it most likely functions as a stand-alone protein which does not rely on other proteins for its activity. The diversity of Ago encoding genes in prokaryotes is very large, and it is not unlikely that different pAgo proteins have different mechanisms and roles. Based on Ago structures, the function of each domain and the mechanism for Ago-guide complex mediated target cleavage is explained. Additionally, it is shown that insertion segments in eAgos generate structural differences between pAgos and eAgos. These differences affect the activity of Agos and provide platforms protein-protein interactions, which allowed specialization of Ago in specific eukaryotic pathways.

Chapter 2 | Characterization of *Thermus thermophilus* Argonaute

In the next chapter the physiological role and mechanism of pAgo from the thermophilic bacterium *Thermus thermophilus* (*Tt*Ago) are described. *Tt*Ago is involved in host defense by reducing plasmid transformation efficiencies and reducing intracellular plasmid content. It is demonstrated that *Tt*Ago acquires small single stranded DNA guides from plasmid DNA *in vivo*, which it can utilize to cleave double stranded DNA (dsDNA) plasmids. These 5'-phosphorylated ssDNAs, with a length ranging from 13 to 25 nucleotides, are termed small interfering DNAs (siDNAs). *Tt*Ago-siDNA complexes are able to cleave both ssRNA and ssDNA targets, and additionally dsDNA targets if two *Tt*Ago-siDNAs complexes targeting either strand of the plasmid are provided.

Chapter 3 | Mechanism of DNA-guided DNA target cleavage by Argonaute

This chapter describes a series of *Tt*Ago structures in complex with a 21 nucleotide DNA guide and ssDNA targets of various lengths. While *Tt*Ago-siDNA complexes bound to ssDNA targets shorter than 16 nucleotides remain in a cleavage-incompatible formation, binding of targets of 16 nucleotides or more induces conformational changes leading to target cleavage. These conformational changes include release of the 3' end of the siDNA from the PAZ domain and movement of three loops in the PIWI domain. This results in insertion of PIWI loop 2 into the catalytic pocket. The glutamic acid located on this loop completes the catalytic DEDH tetrad, after which two divalent cations are bound. This conformation allows an activated water molecule to mediate cleavage of the phosphate backbone of the target strand.

Chapter 4 | Characterization of *Pyrococcus furiosus* Argonaute

The physiological role and mechanism of pAgo from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*Ago) are described in Chapter 4. Like bacterial *Tt*Ago, archaeal *Pf*Ago functions in host defense by lowering plasmid transformation efficiencies by DNA-guided DNA interference. *In vitro* characterization revealed that, in contrast to *Tt*Ago and eAgos, *Pf*Ago-siDNA complexes do not cleave ssRNA targets and cannot utilize Mg^{2+} cations for target cleavage. Instead, it uses Mn^{2+} or Co^{2+}, which mediate cleavage of ssDNA targets and dsDNA plasmids. Additionally, it is demonstrated that *Pf*Ago cleaves plasmid DNA in absence of provided siDNAs under specific conditions.

Chapter 5 | Acquisition of DNA guides by prokaryotic Argonaute

In Chapter 5 we describe *in vitro* experiments with guide-free *Tt*Ago and a variety of dsDNA targets. These experiments reveal that guide-free *Tt*Ago degrades (partially) unwound dsDNA. This process is termed 'chopping' and results in the generation of siDNAs that guide *Tt*Ago activity.

Chapter 6 | Effects of *Thermus thermophilus* Argonaute on gene expression

The effect of the presence or absence of *Tt*Ago and/or plasmid DNA on the expression of RNA is described in Chapter 6. We observed no significant changes in RNA levels in the absence or presence of *Tt*Ago or plasmid DNA alone. *Tt*Ago lowers the internal plasmid concentration ~4-fold, and lowers plasmid encoded RNA levels accordingly. This suggests that *in vivo*, *Tt*Ago interferes with plasmid DNA but not with plasmid encoded RNA. Furthermore, if both *Tt*Ago and plasmid DNA are present, various CRISPR loci and associated genes appear to be up-regulated. Some of these up-regulated genes are involved in the acquisition of new CRISPR spacers, suggesting that *Tt*Ago-mediated plasmid interference stimulates CRISPR adaptation.

Chapter 7 | CRISPR adaptation triggers plasmid curing

This chapter focusses on the adaptation stage from the CRISPR-Cas host defense system. CRISPR-Cas functions in three different stages: adaptation, expression and interference. In the adaptation stage, small pieces of DNA are acquired from invading DNA such as plasmids or phages. This DNA is integrated in genomic CRISPR loci as spacers. CRISPR loci and *cas* genes are transcribed in the expression stage. During the interference stage, the expressed RNAs guide Cas protein complexes to interfere with complementary invader

DNA. We studied the adaptation stage in plasmid-harboring *Escherichia coli* K12. During cultivation under non-selective conditions, *E. coli* acquires spacers from the plasmid DNA. *E. coli* clones that acquire spacers are cured from the plasmid and are more resistant to transformation with the same plasmid. It is demonstrated that a positive feedback loop in CRISPR adaptation exists: upon acquisition of a first spacer, secondary spacers that enhance the plasmid interfering phenotype are rapidly acquired.

Chapter 8 | Summary and General discussion

In Chapter 8 the work described in this thesis is summarized, and the (potential) physiological roles and mechanism of pAgos are discussed. Furthermore, possible interaction partners of pAgos are hypothesized upon, and the potential use of pAgos as genome editing tools is explained.

Chapter 1

The evolutionary journey of Argonaute proteins

Daan C. Swarts, Kira Makarova, Yanli Wang, Kotaro Nakanishi,
René F. Ketting, Eugene V. Koonin, Dinshaw J. Patel, John van der Oost

Adapted from:
'The evolutionary journey of Argonaute proteins'
Nature Structural & Molecular Biology. September 2014. Volume 21, number 9: 743-753

Abstract

Argonaute proteins are conserved throughout all domains of life. Recently characterized prokaryotic Argonaute proteins participate in host defense by DNA interference, whereas eukaryotic Argonautes control a wide range of processes by RNA interference. Here we review molecular mechanisms of guide and target binding by Argonaute proteins, and describe how the conformational changes induced by target binding lead to target cleavage. On the basis of structural comparisons and phylogenetic analyses of pAgos and eAgos, we reconstruct the evolutionary journey of Argonaute proteins through the three domains of life and discuss how different structural features of pAgos and eAgos relate to their distinct physiological roles.

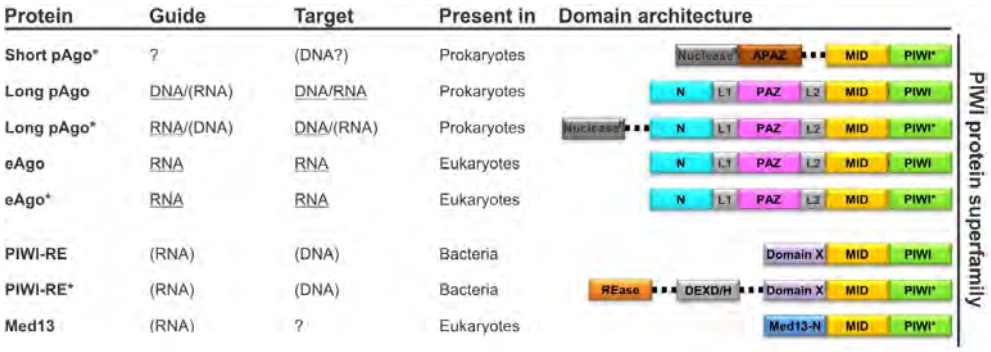


Figure 1 | Domain architectures of the PIWI superfamily proteins. Dotted lines indicate separate genes located in the same (predicted) operon. *: Ago proteins with an incomplete DEDX catalytic tetrad in the PIWI domain. Guide and target usage is based on available biochemical data (underlined) or predicted (in parentheses). [a]: predicted nucleases from Sir2, Mrr or TIR protein families. [b]: predicted nucleases from Sir2, Mrr, Cas4 or PLD protein families. REase, restriction endonuclease; DExD/H, superfamily II helicase (denoted after a signature amino acid motif).

Introduction

Argonaute (Ago) was first mentioned in a study describing a mutant in *Arabidopsis thaliana* [1]. Because the leaves of the mutant plant curled up like squid tentacles, the gene and corresponding protein were named after the octopus *Argonauta argo*. It later became clear that the Ago protein is the key player in eukaryotic RNA interference (RNAi) pathways (**Box 1**) in which Ago utilizes short 5'-phosphorylated RNA guides to target complementary RNA transcripts. The Ago proteins belong to the PIWI protein superfamily, defined by the presence of a PIWI (P-element induced wimpy testis) domain. In addition, all eAgos feature an N (N-terminal) domain, a PAZ (PIWI-Argonaute-Zwille) domain, and a MID (middle) domain, along with two domain linkers L1 and L2 (**Fig. 1**, **Box 2**).

Box 1 | RNA interference pathways.

Eukaryotic RNAi pathways (reviewed in [171,172,173,174]) include proteins with RNase III-like domains (Dicer and Drosha) that usually process dsRNA precursors into short dsRNA molecules termed small interfering (si)RNAs. With phosphorylated 5' ends and 2-nucleotide overhangs at the 3' ends, siRNAs consist of a passenger strand and a guide strand, of which the latter is selectively loaded into eAgos. After removal of the passenger strand, eAgo holds on to the guide strand, which enables eAgo to bind mRNA targets complementary to the guide. Binding of the guide to Ago results in helical preordering of the seed segments in the guide (nucleotides 2-7 or 2-8), which enhances the affinity for a matching target [175]. Target binding starts in this seed region (nucleation) and extends by zippering towards the 3' end of the guide (propagation). This results in release of the 3' end of the guide from Ago and induces conformational changes that result in target cleavage [169,176]. The cleaved target strand is released, allowing Ago to bind and cleave additional targets. In the case of imperfect targets and/or catalytically inactive eAgos, binding of eAgo, alone or with associated proteins, results in repression of mRNA translation. Both processes eventually lead to silencing of gene expression.

Many prokaryotic genomes also feature *ago* genes [3,4,170]. Long prokaryotic Agos (pAgos) encompass the same domains as eAgos, while short pAgos consist of only the MID and PIWI domains (**Fig. 1**). Like eAgos, pAgos interact with 5'-phosphorylated oligonucleotides guides, but in contrast to eAgos, some pAgos have higher affinity for DNA guides than for RNA guides [163,164,167]. Both long and short variants of pAgos (**Fig. 1**) have been proposed to function in defense against mobile genetic elements [4]. Indeed, it was recently shown that both RNA-guided [177] and DNA-guided [27] pAgos interfere with foreign DNA *in vivo*.

A major challenge in the early days of RNAi research was to uncover structure-functional relationships of Ago proteins. For practical reasons, initial efforts to obtain Ago structures focused on pAgos before their physiological role was known. These studies provided valuable mechanistic insights into guide-target pairing and guide-mediated target cleavage [167,168,169,176]. More recently, structures of eAgos have also been solved [28,178,179]. Here we review the body of structural work on pAgos and eAgos, and compare the features that determine their differential functionalities, such as guide preference (DNA versus RNA), nucleolytic activity and docking sites for partner proteins. We also discuss phylogenetic analyses that provide insight into how Agos have changed during their evolutionary journey, from relatively simple host defense proteins in prokaryotes, to key players in complex multiprotein regulatory pathways in eukaryotes.

Structures of Argonaute proteins

The first crystal structures determined were of the guide-free pAgos of *Pyrococcus furiosus* (*Pf*Ago) [30], *Aquifex aeolicus* (*Aa*Ago) [164,166] and truncated long pAgo from *Archaeoglobus fulgidus* (*Af*Ago) [180], which provided information about the overall structural organization of Agos. The long pAgos revealed a bilobal architecture, with the PAZ lobe (N, L1 and PAZ domains) connected by L2 to the PIWI lobe (MID and PIWI domains). The MID domain adopts a Rossmann-like fold with a characteristic nucleotide-binding pocket [163,165,181,182,183]. The PIWI domain adopts a typical RNase-H fold [30,164,180,183] with three catalytic aspartic acid residues, and the PAZ domain has an SH3-like barrel fold involved in nucleotide binding [31,184,185].

Binary structure of pAgo bound to guide strand

Initial attempts to produce complexes of long pAgos with 5'-phoshoryated guide RNAs were not successful. It was later found that several pAgos bind DNA guides with affinities two orders of magnitude higher than RNA guides [163,164]. Crystals of *Thermus thermophilus* pAgo (*Tt*Ago; **Fig 2a**) with a bound DNA guide (**Fig. 2b**) were eventually obtained at elevated temperatures (35 to 40 °C) [167]. The 3.0 Å structure of *Tt*Ago bound to a 5'-phosphorylated 21-mer guide DNA (**Fig. 2c**) [167] showed that the guide strand contacts all domains of *Tt*Ago, with the majority of the contacts involving interactions with the sugar-phosphate backbone of the guide DNA. The 5'-phosphorylated end was inserted into the nucleotide binding pocket in the MID domain (**Fig. 2d**), whereas the 3' end of the

guide was anchored in the PAZ domain (**Fig. 2e**), in agreement with previous structural reports on oligonucleotide complexes with the PIWI lobe [163,165] or the PAZ domain [186,187].

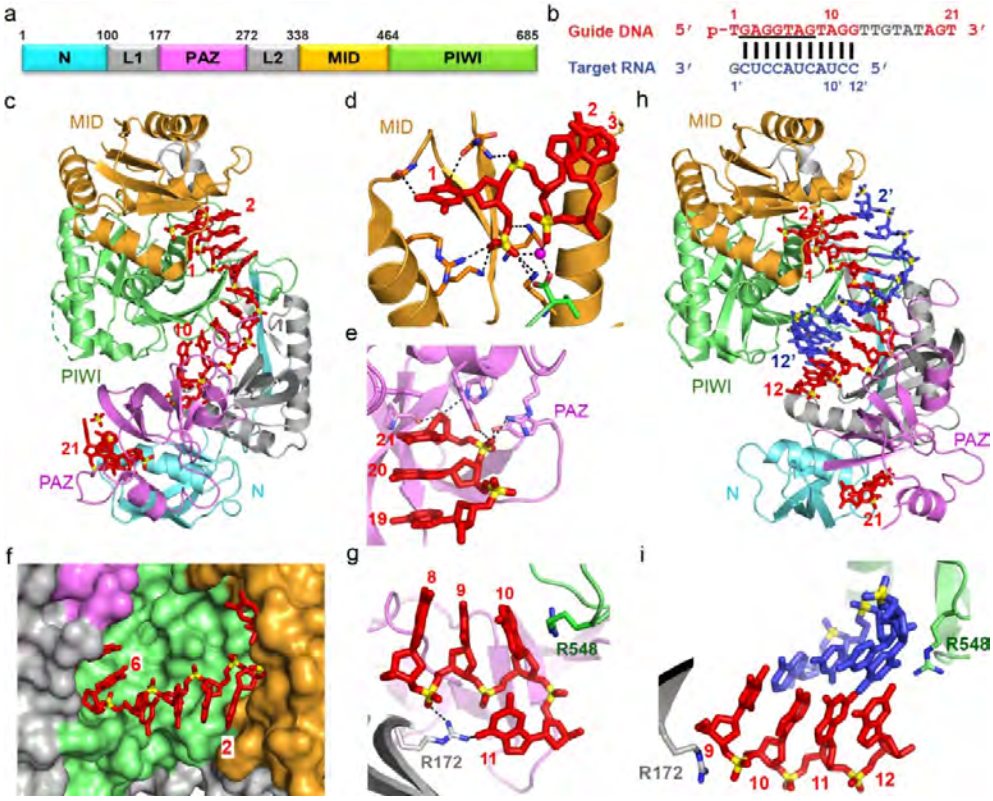


Figure 2 | *Tt*Ago with 21-mer guide DNA (binary complex) and with complementary 12-mer target RNA (ternary complex) adopt cleavage-incompatible conformations. a, Domain architecture of *Tt*Ago. **b**, Sequence of 5'-phosphorylated guide DNA (red, with disordered segment in gray and seed segment underlined) and complementary 12-mer target RNA (blue). **c-g**, 3.0 Å structure of the binary complex of *Tt*Ago bound to 5'-phosphorylated 21-mer guide DNA (PDB 3DLH). *Tt*Ago in ribbon representation, domains colored as in **a**; guide DNA in red, in stick representation. **c**, Overall view. **d**, Insertion of the 5' phosphate of the guide into the MID pocket. **e**, Insertion of the 2 nucleotide 3' end of the guide DNA into the PAZ pocket. **f**, Outward directionality of bases 2 to 6 of the guide DNA in the binary complex of *Tt*Ago with guide DNA, thereby aligning their Watson-Crick edges for pairing with target nucleic acids. **g**, Bases 10 and 11 of the guide DNA are splayed apart as a result of insertion of an arginine side chain. **h-i**, 2.6 Å structure of the ternary complex of *Tt*Ago(D546N) bound to 5'-phosphorylated 21-mer guide DNA and complementary 12-mer target RNA (PDB 3HO1). Guide DNA (red) and target RNA (blue) are in stick representation. **h**, Overall view. **i**, Bases 10 and 11 of the guide DNA are stacked.

Insertion of the 5'-phophorylated end into the MID domain binding pocket strongly bends the guide strand, precluding base pairing of nucleotide 1 (**Fig. 2d**) [163,165]. Whereas residues lining the 5' phosphate-binding pocket in the MID domain are critical for cleavage activity, substitution of the residues that compose the 3' end-binding PAZ pocket showed little effect [167]. Guide nucleotides 2 to 10 are preordered in a helical arrangement (**Fig. 2c**), with bases 2 to 6 pointing outward and thus available for pairing with the target strand (**Fig. 2f**). These observations suggest that guide-target pairing could initiate (nucleate) in the 'seed' segment (nucleotides 2-7 or 2-8; **Box 1**), with the preordered helical conformation of the guide strand reducing the entropic penalty for duplex formation. Indeed, a guide DNA strand pairs with its target RNA with much higher affinity (~300-fold increase) when its seed fragment is associated with the *A. fulgidus* PIWI lobe, compared to protein-free pairing [188]. This higher affinity could enhance the fidelity of target recognition, as well as promote and stabilize the assembly of the active silencing complex. Notably, guide-target mismatches in the seed can have a pronounced impact on the affinity of guide-target recognition (reviewed in [77,175]). There are examples of exceptions in which the seed is not essential for target binding [189], although the functional implications of these exceptions are not clear at present. In the binary *Tt*Ago structure, the preordered guide helix is interrupted by two arginine residues that lock bases 10 and 11 into a unique orthogonal arrangement (**Fig. 2g**), whereas nucleotides 12 to 17 of the DNA guide are disordered and could not be traced.

Ternary structures with pAgo bound to guide and target strands

Crystal structures of *Tt*Ago bound to a 5'-phosphorylated 21-mer DNA guide and complementary RNA targets of different lengths provided a major step in understanding Ago functionality. In order to prevent target cleavage, either mismatches were introduced in nucleotides 10 and 11 centered on the cleavage site [168], or one of the three aspartic acid residues that line the cleavage pocket were substituted [169]. The ternary complex of *Tt*Ago with a 12-mer target RNA (**Fig. 2b**) encompassed 11 Watson-Crick base pairs in an A conformation, spanning nucleotides 2 to 12 and including the seed segment and the cleavage site (**Fig. 2h**). In the guide strand, both the 5' phosphate and the 3' end remained anchored in their respective MID and PAZ pockets; in contrast, the orthogonal arrangement of bases 10 and 11 seen in the binary complex was disrupted in the ternary complex, where they appeared stacked and centered on the cleavage site (**Fig. 2i**). Pivot-like conformational transitions are observed for the N and PAZ domains on proceeding from

the binary to the ternary complex [168,169]. In structures of the ternary complexes of *Tt*Ago with a 15-mer RNA target (3.0 Å resolution; **Fig. 3a-e**) or with a 19-mer RNA target (2.8 Å resolution; **Fig. 3f-h**), the 5'-phosphorylated end of the guide remained anchored in the MID pocket, but the 3' end of the guide was released from the PAZ pocket [169].

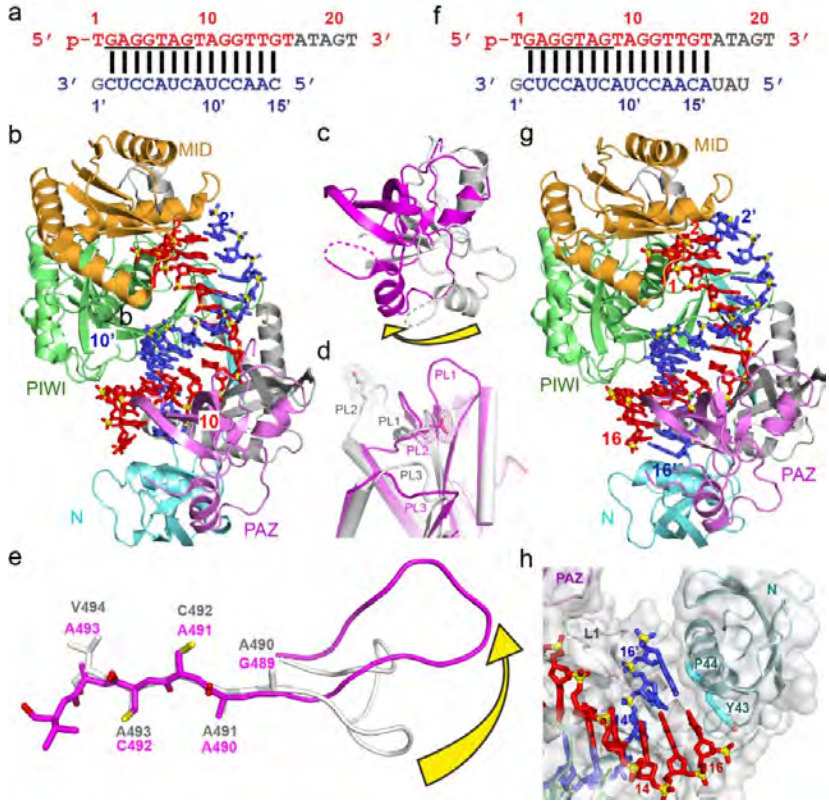


Figure 3 | *Tt*Ago with 21-mer guide DNA and complementary 15-mer and 19-mer target RNAs (ternary complex) adopt cleavage-compatible conformations. a, Sequences of 5'-phosphorylated guide DNA (red, with disordered segment in gray and seed segment underlined) and complementary 15-mer target RNA (blue). **b**, 3.0 Å ternary complex of *Tt*Ago^{D546E} bound to guide DNA and 15-mer target RNA (PDB 3HJF). The guide DNA and target RNA are in a stick representation, with same colors as in **a**. **c-e**, Conformational changes in *Tt*Ago during its transition from ternary complex with 12-mer target RNA in a cleavage-incompatible conformation (silver, PDB 3HO1) to ternary complex with 15-mer target RNA in a cleavage-compatible conformation (magenta, PDB 3HJF). **c**, Rotation of the PAZ domain. **d**, Transitions in loops PL1, PL2 and PL3. **e**, Rearrangement of the β-strand (G489 to V494) of *Tt*Ago by one residue and conformational transition in adjacent loop PL1. **f**. Sequences of 5'-phosphorylated guide DNA (red, with disordered segment in gray and seed segment underlined) and complementary 19-mer target RNA (blue). **g**, 2.8 Å Ternary complex of *Tt*Ago^{D478A} bound to guide DNA and 19-mer target RNA (PDB 3HK2). **h**, The N domain blocks base pairing of the guide and the 19-mer target RNA beyond position 16 of the target strand.

This release was required to overcome the torsional constraints that accumulate during the propagation step (**Box 1**), as longer target segments enter the Ago interior to form an uninterrupted A-form duplex with the guide strand (14 base pairs (bp) in the complex with the 15-mer RNA target; 15 bp with the 19-mer RNA target). Release of the 3' end of the guide is accompanied by rotation of the PAZ domain (**Fig. 3c**) and transitions within the nucleic acid-binding surface of the PIWI domain, namely movements in loops PL1, PL2 and PL3 (**Fig. 3d**), and a sliding and flipping of a β-strand (**Fig. 3e**). The ternary complex of *Tt*Ago with the 19-mer target RNA shows that the N domain blocks guide-target pairing beyond position 16 (**Fig. 3h**) [169]. Altogether, the structures of *Tt*Ago ternary complexes with RNA targets illustrate the conformational transitions during the progression from nucleation to propagation of guide-target duplex formation.

Structures of ternary complexes of DNA-guided *Tt*Ago with target DNAs have been solved to a substantial higher resolution (2.2 Å) [176] than those with RNA targets. A glutamic acid residue on loop PL2 (termed the 'glutamate finger') [179] is directed away from the catalytic pocket in the cleavage-incompatible conformation (i.e., in the guide-free protein, in the binary complex, and in the ternary complex with 12 nucleotide targets; **Fig. 4a**). However, in the cleavage-compatible conformation (i.e., ternary complexes with targets at least 15-mer RNA or 16-mer DNA; **Fig. 4b**), this glutamic acid is inserted into the catalytic pocket and completes the catalytic tetrad [176].

The two-state model of Ago action has been confirmed by single-molecule fluorescence resonance energy transfer studies with *Methanocaldococcus jannaschii* pAgo (*Mj*Ago) [190]. A pair of Mg^{2+} cations are known to facilitate RNA hydrolysis in RNase H family nucleases [191,192], but no metals were detectable in the cleavage-incompatible forms of *Tt*Ago (**Fig. 4c**). In contrast, two Mg^{2+} cations were identified in the cleavage-compatible conformations in the complex with the 16-mer DNA target (**Fig. 4d**), where they bridged the catalytic aspartic acid residues and the cleavage site on the target strand, i.e., the backbone phosphate between nucleotides 10 and 11 of the target strand, where cleavage takes place. In addition, a water molecule was observed at a position allowing for an in-line attack on the cleavable phosphate group (**Fig. 4d**) [176], and the catalytic glutamic acid is coordinated to the Mg^{2+} cation B through two bridging water molecules [176]; in contrast, in RNase H, the glutamic acid directly interacts with the divalent cation [192]. In *Tt*Ago, insertion of the glutamic acid in the catalytic pocket results in cleavage of the target strand between the nucleotides

that base pair with guide nucleotides 10 and 11 [176]. These structural snapshots of ternary *Tt*Ago complexes provide a model for Mg^{2+} cation-coordinated cleavage of the target strand (**Fig. 4c-f**).

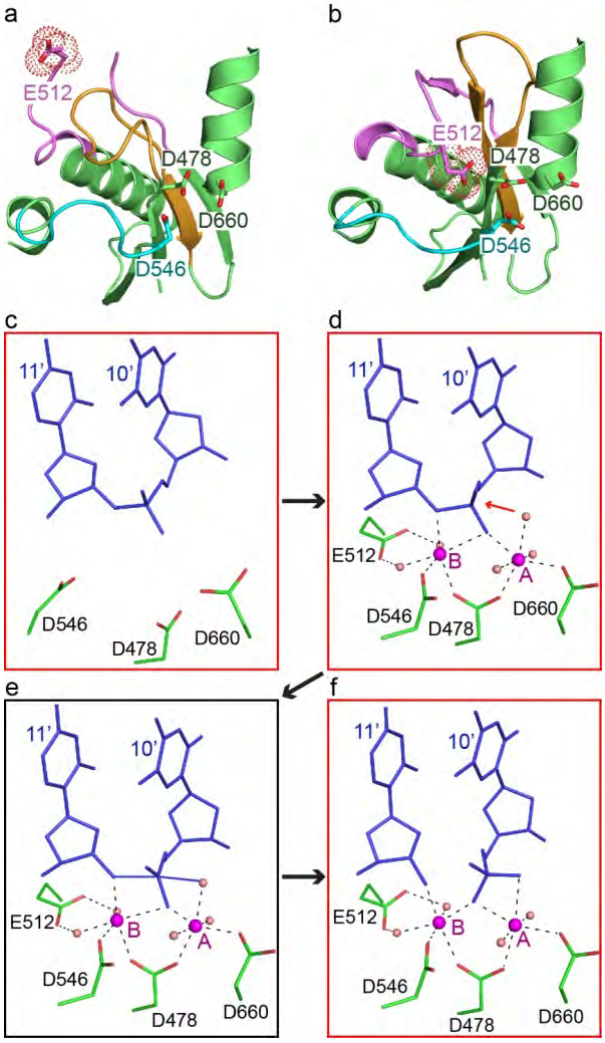


Figure 4 | Structure-based insights into the cleavage mechanism of *Tt*Ago. a, **b**, Positioning of E512 (surface shown in a dotted representation) of *Tt*Ago in the ternary complexes with 5'-phosphorylated 21-mer guide DNA and **a**, complementary 12-mer target DNA (E512 outside and directed away from the catalytic pocket) representative of a cleavage-incompatible conformation; PDB 4N47) and **b**, 19-mer target DNA (E512 inserted into the catalytic pocket) representative of a cleavage-compatible conformation; PDB 4NCB). **c-f**, Proposed mechanism for Ago-mediated Mg^{2+}-coordinated cleavage of the target strand at the 10'-11' step in the ternary complex of *Tt*Ago in with DNA guide and complementary DNA target strands. Crystal structure snapshots show cleavage-incompatible (panel c; PDB 4N47), cleavage-compatible (panel d; PDB 4NCB) and post-cleavage (panel f; PDB 4N76) states, along with a proposed model of the transition state (panel e).

Binary structures of Eukaryotic Agos bound to guide strands

Sustained efforts have resulted in the successive crystallization and structural determination of budding yeast *Kluyveromyces polysporus* Ago (*Kp*Ago, 3.2 Å; **Fig. 5a**) [179], human AGO2 (hAGO2, 2.2 Å; **Fig. 5b**) [28,178] and human AGO1 (hAGO1,1.75-2.3 Å) [193,194],

all with fortuitously loaded heterogeneous RNA guides. hAGO1 and hAGO2 were also captured as binary complexes by replacing the co-purified RNA with a defined RNA guide: hAGO1 with let-7 at 2.1 Å, and hAGO2 with miR-20a at 2.2 Å) [28,193]. Although the eAgo structures are currently restricted to binary complexes, biochemical studies have demonstrated the capacity of *Kp*Ago to load RNA duplexes, which is followed by cleavage of the passenger strand, and eventually annealing and slicing a complementary target strand [179]. In these eAgo binary complex structures, both the bases and the phosphate backbone spanning the seed segment could readily be traced, even with bound heterogeneous RNA. Similar to the DNA guides in *Tt*Ago, the seed segment of eAgo-bound RNA guides adopt an A-like conformation, which in eAgos is facilitated by hydrogen bond interactions involving the sugar ring 2'-OH group and backbone phosphate groups of the RNA guide to Ago. In all studied eAgo complexes, there is a kink between nucleotides 6 and 7 of the RNA guides, caused by the insertion of the side chain of an isoleucine residue (**Fig. 5c**). To allow guide pairing with RNA targets, this isoleucine side chain has to be displaced during ternary complex formation. Isoleucine or other hydrophobic residues are often found at this position, but they are not strictly conserved. The bases spanning the seed segment are stacked with an unusual tilting in the binary eAgo complexes [28,178,179], requiring transition to a non-tilted A-like helical state to facilitate pairing with the target strand in the ternary complex. Akin to the arginine-mediated perturbation of nucleotides 10 and 11 in *Tt*Ago with a bound DNA guide [167], base stacking at nucleotides 9 and 10 is perturbed in the complex of hAGO2 with RNA guide, with the kinked alignment stabilized by three arginine side chains (**Fig. 5d**) [28]. Yet another similarity with the binary *Tt*Ago structures concerns the disordered middle part (nucleotides 11 to 19) of the guides in the eAgo binary complexes, whereas their 3' ends are bound by the PAZ domain [28,178].

Differences between eAgo and pAgo complexes

Despite low sequence similarity between pAgos and eAgos (12% identity between various pAgos and hAGO2), their structural and functional features are remarkably similar (**Box 2**). Nevertheless, there are also notable structural differences that seem to correlate with distinct functionalities [195]. Whereas all characterized eAgos and some pAgos use RNA guides [177], other pAgos use DNA guides [27,163,164,190]. The only chemical difference between RNA and DNA nucleotides is that at the 2' position of the sugar ring, RNA has an OH group, whereas DNA has an H group. The PAZ domain does not bind the 2'-OH groups in the 3' end of the RNA guide [186,187], but in eAgos some of the 2'-OH groups

spanning the seed segment are specifically bound (either directly or via water-mediated hydrogen bonds) to the MID, L1 and PIWI domains (**Fig. 5e**) [28,178,179,193,194]. This indicates that the preference for an RNA guide is determined at the structural level, although those 2'-OH-binding residues are conserved only in a narrow group of fungal and metazoan eAgos.

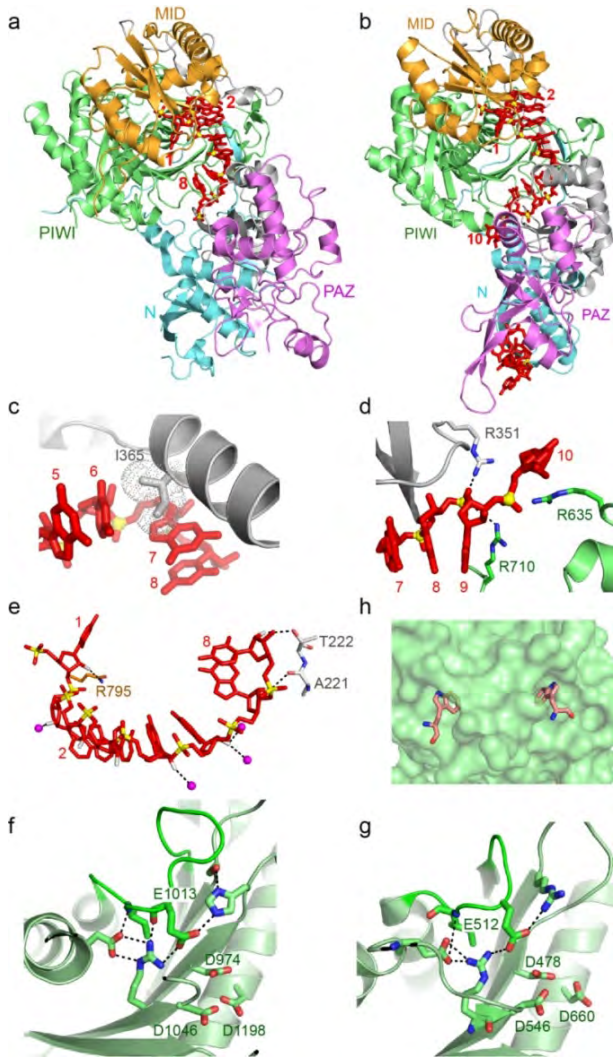


Figure 5 | Structures of binary complexes of *Kp*Ago and hAGO2 bound to 5'-phopshorylated guide RNAs. a, 3.2 Å structure of *Kp*Ago (ribbon representation) with fortuitously loaded 5'-phosphorylated guide RNA (red, stick representation; PDB 4F1N). **b**, 2.3 Å structure of hAGO2 (ribbon representation) with bound fortuitously loaded 5'-phosphorylated guide RNA (red, stick representation; PDB 4EI1). **c**-**e**, Details of the 2.2 Å structure or hAGO2 bound to a defined, mir-20a 5'-phosphorylated guide RNA (PDB 4F3T). **c**, Insertion of I365 (dotted circle), projecting from α-helix 7 of hAGO2, between bases 6 and 7 of the bound RNA guide strand. **d**, Splaying apart of bases 9 and 10 of the guide RNA by insertion of R710 side chain. **e**, Intermolecular contacts between 2'-OH groups of bound guide RNA and amino acid backbone and side chains of hAGO2; both direct and water-mediated (pink spheres) intermolecular hydrogen bonds are shown. **f**, **g**, Intermolecular hydrogen bonding interactions stabilizing the conformation of the expanded and repositioned loop PL2 that inserts the glutamic acid finger into the catalytic pocket in the structure of the *Kp*Ago binary complex with a fortuitously-loaded 5'-phosphorylated guide RNA (**f**, PDB 4F1N) and in the structure of the *Tt*Ago ternary complex with 5'-phosphorylated guide DNA and 19-mer target RNA (**g**, PDB 3HVR). **h**, A pair of tryptophan-binding pockets on the surface of hAGO2 in its binary complex with a fortuitously-loaded 5'-phosphorylated guide RNA (PDB 4EI3).

In addition, the 5' end-binding pocket of the *Tt*Ago MID domain is more hydrophobic than that of hAGO2, which might explain preference of *Tt*Ago for DNA guides [178]. In the 5' end-binding pocket of pAgos, the negative charge of two phosphates of the guide (nucleotides 1 and 3) and of the C-terminal carboxyl group of pAgo (which is inserted into the MID domain binding pocket) are neutralized by a bound divalent cation [165] (reviewed in [183]). In contrast, fungal and metazoan eAgos use the ammonium group of a conserved lysine residue to neutralize this charge [195].

In the *Kp*Ago and hAGO2 binary complexes, the glutamate finger is inserted into the catalytic pocket, even in the absence of the target strand [28,178,179]. This is in contrast with pAgos, in which insertion of the glutamate finger to complete the catalytic tetrad follows extended guide-target base pairing, leading to the cleavage-compatible state. Notably, a hydrogen bond network in eAgo stabilizes the expanded and repositioned glutamic acid-containing loop in the activate state (**Fig. 5f**), with the same alignment in cleavage compatible pAgo (**Fig. 5g**), thereby facilitating insertion of the glutamate finger into the binding pocket [179]. Further experimentation is required to define the constraints controlling cleavage activity of eAgo.

External insertion segments present in eAgos, but not in pAgos, likely provide binding surfaces for RNA-induced silencing complex (RISC) subunits or other eAgo-binding proteins [179]. *Kp*Ago contains 19 insertion segments, of which 11 are conserved segments (cSs) found in all eAgos and 8 are variable segments (vSs) found only in a subset of eAgos [179]. At least some of the cSs are essential for silencing [196] or appear to differentially affect the activity of eAgos. Although a gap between the two structural lobes is observed in *Tt*Ago, cS1, cS3 and cS10 in *Kp*Ago generate a subdomain that fills this gap [179]. The presence of this subdomain positions the N domain away from the nucleic acid-binding channel, which allows extensive guide-target pairing and accommodation [179]. The catalytically active pAgos appear to function as stand-alone proteins, but eAgos interact with a range of proteins in a variety of RNA interference pathways (described below). In hAGO1, cS7 forms a surface that could sterically hinder the positioning of a fully paired guide-target RNA duplex in the catalytic site [194]. In the catalytically activated hAGO1^{R805H} mutant, activity is further increased upon swapping specific cS7 residues with those of hAGO2 [193,194]. Structures of hAGO1 and hAGO2 have revealed that other cSs in the PIWI domain form two tryptophan-binding pockets, lined by aromatic and hydrophobic

side chains (**Fig. 5h**), which are implicated in binding glycine-tryptophan (GW) repeats of TNRC6 family proteins (for example, GW182) that promote miRNA-mediated translation regulation (deadenylation) by hAGO1 [178,194,197]. Indeed, GW-rich peptides from GW182 can target these pockets in eAgo [198], with the distance between pockets matching the pairwise arrangement of tryptophan residues in GW proteins. Thus, eAgo-specific insertion segments play a role in the binding of interacting proteins and additionally can directly influence eAgo activity.

Box 2 | Ago domains have conserved structures and functions.
The core functions of each of the four structural domains of Agos are well conserved in prokaryotes and eukaryotes.

The MID domain forms a basic nucleotide-binding pocket in which several conserved amino acids interact with the phosphate group at the 5' end of the guide [163,165,181,182,183]. In addition to the specific binding of the sugar backbone of the 5'-end terminal nucleotide, at least some Agos recognize specific 5'-end bases using a structural feature termed the 'nucleotide specificity loop' [182,199]. The MID domain also stacks the guide in a helical conformation within its seed nucleotides (2-7 or 2-8), promoting target binding (reviewed in [175]).

The PIWI domain includes the RNase H-like active site of slicing Agos [30,164,165,183]. In the cleavage-compatible conformation, two divalent cations are bound by a DDX triad (where X is usually aspartic acid or histidine; in rare cases it is lysine) [172]. The catalytic site is completed by a glutamic acid residue that resides on a mobile PIWI loop (the glutamate finger), forming the DEDX motif [176,179].

The PAZ domain. This domain binds the 3' end of the guide by interactions with the backbone of nucleotides 20 and 21 [30,167,176,186,187]. This interaction is not essential for guide binding but protects the guide from degradation [200].

The N domain is not involved in guide binding but plays a critical role in dissociation of passenger and cleaved target strands [33,193] and in target cleavage [193,201]. During duplex RNA loading, the strand with the less stable 5' end is retained as guide in Ago [68,69]. Removal of the other strand (passenger) can be slicer dependent (requires cleavage) or independent (requires mismatches or G•U wobble base pairs in the seed or in the middle of the 3' region) [70,202]. In both pathways, the N domain functions as a wedge, disrupting guide-passenger base pairs at the 3' end of the guide (active wedging) or by blocking guide-target base pairing downstream of nucleotide 16 of the guide as observed for *Tt*Ago ternary complexes (passive wedging) [33,169]. The role of the N domain in target cleavage is indicated by work on human Ago 3 (hAGO3), which is unable to cleave targets *in vitro*, even though it has an intact catalytic site; hAGO3 can be activated when its N domain is swapped for that of hAGO2 [201]. Similarly, target cleavage of activated hAGO1 is enhanced when its N domain is replaced by the counterpart of hAGO2 [193]. How the N domain facilitates slicer activity is presently unclear.

Evolution and function of Argonaute proteins

The evolutionary journey of the Argonautes has produced Ago protein families with distinct distribution patterns across the domains of life. Ago is encoded in ~65% of the sequenced eukaryotic genomes, dispersed over at least four of the five eukaryotic supergroups [170,203]. In contrast, a recent position-specific iterative basic alignment search tool (PSI-BLAST) search of the RefSeq database (November 2013) using representative PIWI domain sequences as queries shows that Agos are encoded in ~32% and ~9% of the available archaeal and bacterial genomes, respectively, and in 17 out of 37 prokaryotic phyla. Similarly to most prokaryotic defense genes [204], pAgo shows a patchy distribution, with at most 70% representation in any bacterial or archaeal phylum. Both eAgos and pAgos belong to the PIWI-protein superfamily, which is defined by the presence of a PIWI domain and in some cases a PAZ domain (**Fig. 1**) [3]. The presence of the PIWI lobe in all Agos detected so far implies that it is essential for Ago functionality [4,203]. We have thus used the sequences of only the MID and PIWI domains to build maximum-likelihood phylogenetic trees using the FastTree program (**Fig. 6**, **Fig. S1**-**4**) [205]. We discuss how this phylogeny can be linked to the structural features that are either conserved or lost in the different families.

Evolution of prokaryotic Argonautes

The topology of the phylogenetic tree of pAgos and most of its sub-trees does not follow the prokaryote phylogeny derived by analysis of ribosomal RNA and other universal genes. This pattern suggests extensive horizontal gene transfer of pAgo encoding genes, similar to the evolution of most prokaryotic defense genes [204,206]. The topology of the tree is congruent with the domain architectures of pAgo and the organization of the (predicted) operons containing pAgo genes (**Fig. 6a**, **Fig. S1**, **Fig. S2**). As shown previously [4], the tree can be confidently divided into two major branches: the short pAgo branch consists of short pAgos only, while the long pAgo branch contains all long pAgos and some short pAgos (for example, *Af*Ago). The latter variants are scattered over the long Ago branch, suggestive of multiple, independent truncation events [4]. Notably, long pAgos from several euryarchaeal species, mostly thermophiles, group with eAgos, supporting previous conclusions on the origin of eAgos from euryarchaeal pAgos [4,170].

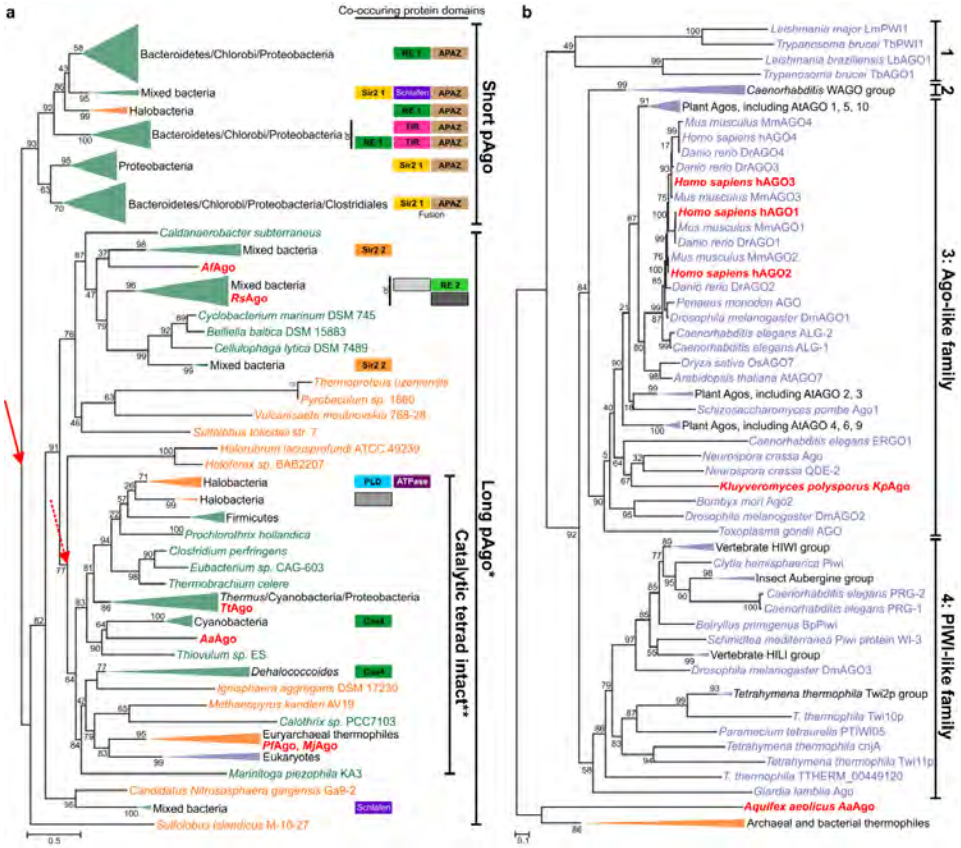


Figure 6 | Phylogenetic trees of Argonaute proteins. a, **b**, Maximum-likelihood phylogenetic unrooted trees were built using the FastTree program [205] using a multiple alignment of conserved blocks of MID and PIWI domains. The same program was also used to compute bootstrap values (percentages) that are indicated for all internal branches. Green - Bacteria; Orange - Archaea; Purple - Eukaryota. Collapsed branches are shown as triangles of the corresponding color. Organisms of which the Agos are discussed in this chapter are colored red. **a**, Phylogenetic analysis of pAgos and organization of the predicted operons. We clustered 487 pAgos identified in RefSeq by sequence similarity (Fig. S4) and selected a non-redundant representative set (261 pAgos and 8 selected eAgos). Red arrows indicate two alternative roots of the pAgo tree. *, the long pAgo clade contains several short pAgos. **, not all eukaryotic eAgos have an intact catalytic tetrad. Domains associated with pAgos are shown as boxes on the right side of the tree. Homologous domains are shown by boxes of the same color or pattern. Abbreviations: Sir2 1 and Sir2 2 are two distinct families of the predicted Sir2-like nuclease, RE1 and RE2 are two distinct families of the restriction endonuclease superfamily. TIR, predicted nuclease of TIR family; Schlafen, predicted ATPase; APAZ, 'analog of PAZ' domain; Cas4, Cas4 subfamily of restriction endonuclease superfamily; PLD, predicted nuclease of phospholipase D superfamily. Gray boxes indicate distinct families of uncharacterized proteins. Short and long pAgos are not shown but are present in all operons. Slashes denote 'and'. pAgo sequence alignment and uncollapsed phylogenetic tree are in Fig. S1 and Fig. S2, respectively. **b**, Phylogenetic analysis of a representative set of 177 eAgos. 1, Trypanosoma Ago family, 2, WAGO family. eAgo sequence alignment and uncollapsed phylogenetic tree are in Fig. S3 and S4, respectively.

On the basis of the conservation of the four catalytic residues, only 28% of the long pAgos are predicted to be catalytically active; these predicted active pAgos form a monophyletic group (**Fig. 6a**), and the encoding genes often co-occur with predicted helicases. Predicted catalytically inactive long pAgos often cluster in predicted operons with genes encoding putative nucleases (**Box 3**). Assuming that the ancestral pAgo was an active nuclease and that the primary split in the evolution of this family was the separation into short and long forms (**Fig 6a**, solid red arrow), Agos were inactivated on multiple independent occasions, which resulted in loss of activity in all short pAgos and several groups of long Agos, including a subset of eAgos. Alternatively, as the root of maximum likelihood method-generated phylogenetic trees cannot be determined, the correct root position might be between the active and inactive forms (**Fig. 6a**, dotted red arrow); in this scenario, truncation of pAgo to yield the short forms would be a relatively late evolutionary event.

Box 3 | Predicted nucleases and helicases associated with pAgos.

Genes encoding long pAgos with incomplete catalytic sites are often clustered in predicted operons with genes encoding putative nucleases of the Sir2 or Mrr families, predicted to be DNA-specific nucleases, with different catalytic motifs [207,208,209]. Genes encoding long pAgos with intact catalytic sites occasionally cluster with genes encoding Cas4-like or PLD domain nucleases. Cas4 is a clustered regularly interspaced short palindromic repeat (CRISPR)-associated nuclease/helicase, likely involved in the adaptation step of CRISPR-Cas host defense [210,211], whereas phospholipase D (PLD) family nucleases are fused to a DNA/RNA helicase domain, a combination also found in bacterial restriction-modification systems [212]. Other long pAgo encoding genes are fused to genes encoding Schlafen-like ATPases, which are putative DNA/RNA helicases [213].

All genes encoding short pAgos are associated with a gene encoding the uncharacterized APAZ (Analog of PAZ) domain (**Fig. 1**, **Fig. 6a**). APAZ lacks detectable sequence similarity with the PAZ domain and has only been detected in the context of short pAgo genes, always fused to a (predicted) nuclease domain that may belong to the Sir2 or Mrr protein families (from different subfamilies than the ones associated with long pAgos) or to TIR domains [4]. The latter are predicted to possess nuclease activity [4,214] and are involved in bacterial virulence [215] or in eukaryotic antimicrobial and antiviral response, and in apoptosis [216,217,218]. In some prokaryotic genomes, the putative Sir2 nuclease is fused not only to the APAZ domain but also to pAgo itself (**Fig. 6a**). Less commonly, Sir2-APAZ domains contain an inserted Schlafen-like ATPase domain (Sir2-Schlafen-APAZ; **Fig. 6a**). Moreover, some short pAgo genes cluster with Mrr-TIR-APAZ gene fusions.

PIWI-RE proteins are fused to an uncharacterised N-terminal domain that does not appear to be related to either PAZ or APAZ [203]. In many genomes, genes encoding PIWI-RE are clustered with two genes, encoding a DinG-like helicase and a predicted restriction endonuclease [203]. Given that DinG family helicases specifically act on R-loops [219], the PIWI-RE proteins have been hypothesised to function as part of an RNA–guided restriction system [203].

Approximately 60% of the pAgos identified lack the PAZ lobe, and most of these short pAgos have incomplete catalytic tetrads. All genes belonging to the short pAgo branch are associated with a gene encoding the uncharacterized APAZ (Analog of PAZ) domain (**Fig. 1**, **Fig. 6a**) [4]. The APAZ domain does not have detectable sequence similarity with the PAZ domain and has not been detected in any context other than the short pAgo neighborhood. The N-terminus of the APAZ domain is always fused to a (predicted) nuclease domain (**Box 3**) [4]. A highly diverged family of short pAgo derivatives, designated PIWI-RE after characteristic conserved arginine (R) and glutamic acid (E) residues (**Fig. 1**) [203], appears in a few major bacterial lineages. Notably, the set of genomes encoding PIWI-RE or pAgo show almost no overlap [203]. Similar to short pAgos, most PIWI-RE proteins feature a seemingly inactive PIWI lobe. PIWI-RE proteins are fused to an uncharacterized N-terminal domain that does not appear to be related to PAZ or APAZ [203]. In many genomes, PIWI-RE-encoding genes cluster with DinG-like helicases and predicted nucleases (**Box 3**), and thus the PIWI-RE proteins have been hypothesized to be part of an RNA–guided restriction system [203].

Function of prokaryotic Argonautes

The ability to cleave target nucleic acids *in vitro* has been investigated for four long pAgos from different branches in the Ago tree, namely *Tt*Ago, *Aa*Ago, *Mj*Ago and *Rhodobacter sphaeroides* pAgo (*Rs*Ago). *Tt*Ago utilizes DNA guides to cleave single stranded (ss)RNA, ssDNA and/or double stranded (ds)DNA plasmid targets, the latter by independently nicking the two strands [27]. *Aa*Ago utilizes ssDNA guides to cleave ssRNA strands, but its ability to cleave DNA has not been determined [164,166]. *Mj*Ago utilizes ssDNA guides to cleave ssDNA strands but cannot cleave RNA targets [190]. No catalytic activity has been observed for *Rs*Ago [177], but it co-occurs with a predicted nuclease in *R. sphaeroides*.

Although the physiological functions of *Aa*Ago and *Mj*Ago have not yet been determined, both *Tt*Ago and *Rs*Ago play a role in host defense [27,177]. *Tt*Ago lowers plasmid transformation efficiency and intracellular plasmid concentrations in *T. thermophilus* [27]. Notably, *Rs*Ago lowers intracellular plasmid concentrations in *Escherichia coli* but not in *R. sphaeroides*; however, it does interfere with plasmid-encoded RNA in *R. sphaeroides* [177]. As short DNA molecules complementary to the RNA guides associate with *Rs*Ago *in vivo*, a yet-to-be-identified nuclease has been proposed to process DNA bound by *Rs*Ago-RNA complexes [177]. *Tt*Ago and *Rs*Ago both acquire functional guides when expressed in *E. coli*

[27,177], which suggests that guide processing is performed either by pAgo itself or by common host factors. *Tt*Ago utilizes 13 to 25 nucleotide long small interfering DNA (siDNA) guides and appears to depend on its own catalytic site for guide loading [27], whereas catalytically inactive *Rs*Ago acquires 15 to 19 nucleotide long RNA guides proposed to originate from degraded mRNAs [177]. Most guides acquired by *Tt*Ago and *Rs*Ago are complementary to foreign DNA, such as plasmids or insertion elements [27,177].

The frequent association of homologous (predicted) nucleases with distinct catalytically inactive long or short pAgos (**Box 3**) suggests a modular organization of pAgo-centered defense systems, with occasional recombination between loci encoding different variants of these systems (**Fig. 6a**). In some of these pathways, the long form of pAgo is predicted to possess both target recognition and nuclease activities. In other cases, catalytically inactivated long or short pAgo might be responsible only for target recognition (using at least their MID and PIWI domains), whereas cleavage would be performed by other nucleases encoded in the same operons, which possibly physically interact with pAgo. The presence of additional non-nuclease genes near some genes encoding pAgos (**Box 3**) indicates the requirement for additional activities in those systems. Given that *Tt*Ago requires unwinding of dsDNA targets for subsequent cleavage of each strand [27], pAgo-associated helicases could play a role in enhancing the accessibility of dsDNA targets for pAgo-mediated cleavage.

Evolution and function of eukaryotic Argonautes

We also reconstructed a phylogenetic tree using a representative set of eAgos with pAgo sequences as an outgroup (**Fig. 6b**, **Fig. S3**, **Fig. S4**). In agreement with previous analyses [25,26], eAgos can be divided into four major families: the Trypanosoma Ago family [26], typified by *Trypanosoma brucei*; the WAGO family, typified by worm (*Caenorhabditis elegans*)-specific Agos; the Ago-like family, typified by *Arabidopsis thaliana* AGO1; and the PIWI family, typified by *Drosophila melanogaster* PIWI. The Ago-like and PIWI families are represented in several major groups of eukaryotes, indicating that at least one duplication of eAgo apparently antedated the last common ancestor of the extant eukaryotes. The other two families could have emerged as a result of additional, lineage-specific duplications. Another protein family belonging to the PIWI-protein superfamily was recently identified in eukaryotes [203]. These proteins have only the MID domain and

an inactive PIWI domain, and are typified by Med13, a subunit of the transcription regulatory Mediator complex in mammals [220].

The phylogenetic tree of eAgos generally follows the phylogeny of eukaryotes and, given the rarity of horizontal gene transfer in the evolution of eukaryotes, it appears that eAgos evolved solely by vertical inheritance. Thus, it has been inferred that a functional RNAi pathway, consisting of eAgo, Dicer and RNA-dependent RNA polymerase (RdRP), was present in the last eukaryotic common ancestor, where it most likely functioned in defense against viruses and transposons [170]. Dicer consists of RNase III, PAZ and DExD/H helicase domains, all with identifiable ancestors in prokaryotes, whereas RdRP apparently evolved from a group of so-far uncharacterized, predicted DNA-dependent RNA polymerases from bacteriophages [170]. All eAgos function in larger protein networks that vary substantially between and within the different families, and eAgos have evolved into distinct players in these different networks. This diversification is the result of many sequence adaptations, which allows interactions with a multitude of proteins that are involved in guide processing, guide loading, regulating eAgo activity, or recruitment of additional proteins.

Trypanosoma Ago family

This eAgo family is mainly studied in *T. brucei*, in which long dsRNA guide precursors are expressed both from retrotransposons [112] and chromosomal 147 bp tandem units [113]. These transcripts are processed either by a cytoplasmic Dicer (*Tb*DCL1) [115] which depends on *Tb*RIF5 for activity [116], or by a nuclear Dicer (*Tb*DCL2) [114]. The exonuclease *Tb*RIF4 is essential in converting the duplex siRNAs to ssRNA guides [116]. An N-terminal RGG domain allows *Tb*AGO1-guide complexes to associate with polyribosomes, which results in efficient cleavage of retrotransposon transcripts [117]. Thus, like the prokaryotic *Rs*Ago, Trypanosoma family Agos interfere with transposon activity.

WAGO family

The eAgos of the nematode-specific WAGO family generally act as so-called secondary Agos, i.e., they are loaded with guide RNAs in response to the activity of the primary Ago [105]. In *C. elegans*, a primary Ago (for example, RDE-1 or PRG-1) is believed to recruit an RdRP to the targeted mRNA, which results in the synthesis of new guide RNAs, known as 22G RNAs (22 nucleotide long guides with 5'-ppp-G), that are utilized by WAGO proteins.

As direct products from RdRP activity, 22G RNAs carry 5' tri-phosphate (GTP) groups [46,56], and it remains unclear how the WAGO proteins can accommodate this atypical guide RNA feature. The WAGO proteins execute a variety of silencing mechanisms, from target RNA destabilization [105] to transcriptional silencing [221]. Absence of secondary Agos can be enough to desilence target expression [105,221], which suggests that the action of the primary Ago is not sufficient for silencing. A notable case has been reported in which a WAGO protein seems to protect against silencing activities executed by other WAGO proteins [110,111]. Hence, apart from adapting to various mechanisms of guide RNA acquisition and target silencing, eAgos seem to play a role in counteracting or fine-tuning silencing.

Ago-like family

Guide RNAs are typically processed and loaded into Ago-like family proteins by proteins such as Dicer (reviewed in [174,222]). In some cases (such as vertebrate AGO2 proteins), Ago-like proteins themselves perform secondary processing of preprocessed RNA hairpin structures, through their endonucleolytic activity [71,72]. Many Ago-like proteins use endogenous guide RNAs, known as microRNAs (miRNAs), to regulate gene expression, mainly by affecting mRNA translation elongation, acting as a road block for the ribosome, or by affecting polyadenylation of the mRNA by extensive interactions with 3' untranslated region-processing machineries (reviewed in [59,79]). In these cases, the guide-target interactions are often characterized by limited, imperfect base-pairing that is incompatible with target RNA cleavage [77]. Thus, many eAgos act purely as sequence-specific RNA-binding proteins, whose sole function is to counteract the translation of specific mRNAs.

Once loaded with a guide, many Ago-like proteins function without involvement of other proteins. Based on the conservation of the four active site residues, ~90% of eAgos are predicted to be catalytically active. However, it should be noted that not all Agos having complete catalytic tetrads are catalytically active, as hAGO3 harbors all four residues but it cannot cleave targets *in vitro* [193,201]. Catalytically inactive hAGO1 can be activated by minimal changes in the active site, with the activity further enhanced by mutations in either the N domain [193,201] or cS7 [193,194]. These findings are compatible with a scenario in which an ancestral eukaryote inherited an active long pAgo, whose catalytic function was subsequently lost in a subset of eAgos. In plants and in some animals, Ago-like proteins use target RNA cleavage as a gene-regulatory mechanism [59] and have additionally been shown to interfere with dsRNA viruses [119,139,223,224]; the latter role is reminiscent of

the host defense functions of pAgos. However, in contrast to pAgos, eukaryotic Ago-like family proteins depend on other proteins, such as Dicer, to process guides from the viral dsRNA genome. Even when confronted with similar guide-target RNA interactions, the kinetics of binding and releasing target RNA can vary widely between different eAgos [225], which indicates that they have not only evolved to bind different protein-partners but also adapted biochemically to execute distinct functions.

PIWI-like family

The ancestral function of target cleavage is strongly conserved among the PIWI-like family proteins. Many PIWI-like family members use their guide RNAs, known as piRNAs, to control the activity of transposable elements within the germ cells of animals [95]. In contrast to the Ago-like proteins, animal PIWI-like proteins are loaded through a pathway that includes ssRNA precursors (reviewed in [173]). This process requires many different protein-protein and protein-RNA interactions and takes place in extremely protein- and RNA-rich assemblies that flank nuclear pores. In some cases, this process involves a nuclease from the PLD family (**Box 3**) [90]. In others, a member of the PIWI-like family itself catalyzes precursor processing. These endonucleases generate 5'-phosphorylated RNA fragments that are bound by a PIWI-like protein. However, not all PIWI-like proteins employ such mechanisms. For example, the PIWI-like proteins in ciliates, which are involved in sequence-specific genome rearrangements, are loaded through Dicer-dependent pathways [158,159]. These variations illustrate the high flexibility in molecular mechanisms coupled to eAgos. Some, but not all of the PIWI-like proteins display a strong preference for a uracil at the 5' end of the loaded RNA, likely reflecting the presence of a nucleotide-specificity loop [182,199], as described for some plant Agos [226]. After loading of this piRNA intermediate, the 3' end of the loaded RNA is likely trimmed by an exonuclease [87], and then 2'-O-methylated [93].

Crystal structures of the PAZ-domain of PIWI-like proteins have revealed the basis of preference for RNA guides with a 2'-O-methylation at their 3' ends over those with unmodified 2'-OH groups [227,228]. The 2'-O-methyl modification has also been demonstrated in guide RNAs of some members of the Trypanosoma Argonaute and Ago-like families, including *Tb*AGO1, *Dm*AGO2 and all the Agos in plants [160,161,162]. A common property of these eAgos is that their guides extensively pair with their target RNAs, resulting in release of the 3' end of the guide RNA from the PAZ domain, and potentially rendering the guide RNA exposed to 3' end-modifying activities. Indeed, in the

absence of 2'-O-methylation, target recognition by these Ago-like proteins results in exonuclease trimming, adenylation and uridylation of the guide RNA [94], which could all affect guide RNA stabilities [229].

Discussion

Comparison of available pAgo and eAgo structures reveals that the domain architecture and the functions of individual domains are conserved throughout the three domains of life. The MID and PIWI domains are responsible for binding and helical preordering of the RNA or DNA guide. Short pAgos, with only these two domains, most likely function as guide-mediated target binders and depend on associated nucleases (and possibly helicases) for target cleavage and/or unwinding. Long pAgos and eAgos additionally feature the PAZ domain, which binds the 3' end of the guide, and the N domain, which plays a role in unwinding of the guide-passenger duplex and interferes with guide-target base pairing towards the 3' end of the guide.

The evolutionary journey of Argonaute proteins started in prokaryotes, through a fusion of a PIWI-like RNase H domain with a MID-like nucleic-acid binding domain, yielding the first guide-dependent short pAgo (**Fig. 6**). RNase H is a nearly ubiquitous nuclease that cleaves Okazaki fragments, the RNA strand of the DNA-RNA duplexes generated during replication, in all domains of life. After the RNase H-MID fusion to generate a short Ago, there were additional associations with distinct interaction or enzymatic domains, often as N terminally-fused extensions, such as N-PAZ in long pAgos, nuclease-APAZ in short pAgos, and a unique N-terminal domain in PIWI-RE (**Fig. 1**). In different pAgo clades, these associations engendered multiple, independent variations, which resulted in active and inactive variants with different guide and target specificities. So far, two mechanistic pAgo variants have been characterized experimentally: DNA-guided DNA interference by *Tt*Ago and *Mj*Ago, and RNA-guided DNA interference by *Rs*Ago, an inactive pAgo variant associated with an uncharacterized nuclease. *Tt*Ago binds both DNA-RNA and DNA-DNA guide-target duplexes in an A-form helix, which is unusual for DNA duplexes. Notably, RNase H cleaves DNA-RNA helixes which also adopt the A conformation, suggesting that *Tt*Ago retained the ancestral preference for an A-form helix in the course of evolution. The guide and target specificity of Ago variants currently cannot be predicted from their amino acid sequence. Most of the prokaryotic MID-PIWI-containing systems likely function in

defense against invading DNA, whereby target cleavage is performed either by their PIWI domain or by co-occurring nucleases (**Box 3**). Given the variation of genes that cluster with pAgo, the functions of pAgos and partner proteins might extend beyond host defense to various regulatory pathways.

A major step in Ago evolution appears to have been the transition from stand-alone proteins to multi-protein regulatory systems. Phylogenetic analyses indicate that the last eukaryotic common ancestor possessed not only an RNA-guided RNA-interfering Ago but also all other components essential for RNAi [170]. In the course of evolution, eAgos maintained the four domains and their respective functions (although some lost catalytic activity) but additionally acquired insertion segments that allowed optimization of specific protein-protein interactions, while maintaining the basic molecular mechanism of action. Thus, various eAgos evolved to interact with pathway-specific proteins, resulting in a variety of RNAi pathways involved in a wide range of cellular processes. The functions of many insertion segments are not yet known, and both structural and biochemical research is required to reveal their roles. Elucidation of these missing links will contribute to our growing understanding of the evolution, mechanism and physiology of Argonautes, and of the diverse defense and regulatory systems of prokaryotes and eukaryotes in which these proteins play crucial roles.

Acknowledgements
This work was financially supported by grants from the Netherlands Organization for Scientific Research (NWO) to JvdO (NWO-TOP, 845.10.003) and the US National Institutes of Health to DJP (TR01 GM104962). KM and EVK are supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine). KN is supported by Precursory Research for Embryonic Science and Technology (PRESTO) from the Japan Science and Technology (JST) Agency.

Author information
Correspondence should be addressed to john.vanderoost@wur.nl or pateld@mskcc.org.

Online materials
Figures S1-4 are available as Supplementary Data 1-4 in the online version of the published manuscript [230].

Chapter 2

Characterization of *Thermus thermophilus* Argonaute

Daan C. Swarts*, Matthijs M. Jore*, Edze R. Westra, Yifan Zhu, Jorijn H. Janssen, Ambrosius P. Snijders, Yanli Wang, Dinshaw J. Patel, José Berenguer, Stan J.J. Brouns, John van der Oost
*contributed equally

Adapted from:
'DNA-guided DNA interference by a prokaryotic Argonaute'
Nature. March 2014. Vol. 507: 258-261

Abstract

RNA interference is widely distributed in eukaryotes and has a variety of functions, including antiviral defense and gene regulation [231,232]. All RNA interference pathways use small single-stranded RNA (ssRNA) molecules that guide proteins of the Argonaute family to complementary ssRNA targets: RNA-guided RNA interference [231,232]. The role of prokaryotic Argonaute variants has remained elusive, although bioinformatics analysis has suggested their involvement in host defense [4]. Here we demonstrate that Argonaute of the bacterium *Thermus thermophilus* (*Tt*Ago) acts as a barrier for the uptake and propagation of foreign DNA. *In vivo, Tt*Ago is loaded with 5'-phosphorylated DNA guides, 13-25 nucleotides in length, that are mostly plasmid derived and have a strong bias for a 5'-end deoxycytidine. These small interfering DNAs guide *Tt*Ago to cleave complementary DNA strands. Hence, despite structural homology to its eukaryotic counterparts, *Tt*Ago functions in host defense by DNA-guided DNA interference.

Results

*Tt*Ago interferes with plasmid DNA transformation and propagation

To elucidate the physiological role of Ago in prokaryotes, we studied the Argonaute protein from *T. thermophilus*. Comparison of the *ago* genes of the type strain HB27 [233,234] and a derivative with enhanced competence (HB27^{EC}; **Fig. 1a**, **Fig. S1**), revealed that an insertion sequence (IS*Tth7*) [235] disrupts *ago* in HB27^{EC}. In line with a role of *Tt*Ago in reducing competence, a generated Δ*ago* mutant (HB27Δ*ago*; **Fig. 1a**) has a natural transformation efficiency that is a factor of ten higher than the wild-type HB27 (P<0.02; **Fig. 1b**). Complementation of the knockout strain with *ago* (HB27Δ*ago*::^{s}*ago* (HB27Δ*ago* complemented with a *strep(II)-tag-ago* gene fusion insert); **Fig. 1a**, **b**) almost completely restores the wild-type phenotype.

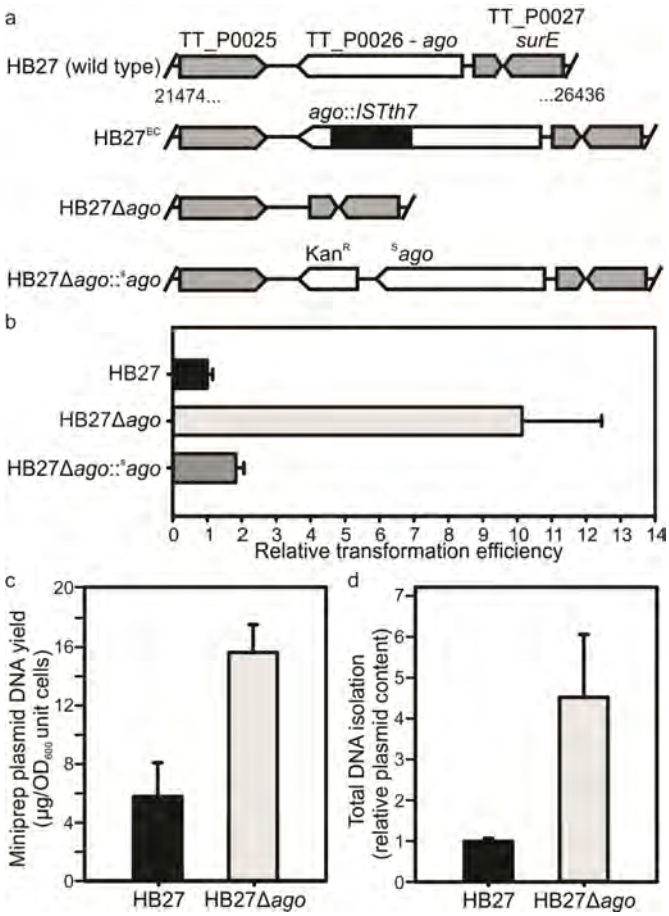


Figure 1 | *Tt*Ago interferes with plasmid DNA. a, Overview of *ago* gene loci of *T. thermophilus* strains: HB27 (wild-type), HB27^{EC} (spontaneous derivative with enhanced competence), HB27Δ*ago* (knockout), and HB27Δ*ago*::^{s}*ago* (HB27Δ*ago* complemented with a *strep(II)-tag-ago* gene fusion insert). **b**, Transformation efficiency of *T. thermophilus* strains on transformation with the plasmid pMHPnqosGFP (**Table S5**). Error bars indicate standard deviations of biological duplicates. **c**, Yield of pMHPnqosGFP plasmid mini preparation of HB27 and HB27Δ*ago*. Error bars indicate standard deviations of biological triplicates. **d**, Plasmid content of total DNA purified from HB27Δ*ago* relative to that from HB27, as quantified by Genetools (Syngene) after resolving the DNA on a 0.8% agarose gel. Error bars indicate standard deviations of biological triplicates.

Moreover, isolation of plasmid and total DNA from the wild-type and the *ago* knockout strains revealed lower plasmid yields from the wild-type strain, indicating that *Tt*Ago reduces the intracellular plasmid concentration ($P<0.02$, **Fig. 1c**; $P<0.02$, **Fig. 1d**).

We performed transcriptome analysis of HB27 and HB27Δ*ago* to determine if *Tt*Ago-mediated interference proceeds directly by targeting plasmid DNA, or indirectly by regulating gene expression. Although the comparison revealed pleiotropic changes in gene expression (**Fig. S2**), we did not observe substantial differential expression of genes involved in plasmid uptake or host defense (**Table S1**). Hence, the RNA sequencing (RNA-seq) analysis suggests that *Tt*Ago does not influence plasmid uptake and plasmid copy number at the level of transcriptional control.

5'-phosphorylated single stranded DNA guides co-purify with *Tt*Ago

We therefore studied whether *Tt*Ago interacts with plasmid DNA. In agreement with the RNA-seq analysis (**Fig. S2**), affinity-purified *Tt*Ago expressed from the chromosome of HB27Δ*ago::*^{s}*ago* could be detected by protein mass spectrometry (**Table S2**). Unfortunately, molecular analysis of *Tt*Ago expressed in *T. thermophilus* was hampered by the low *Tt*Ago yield, and attempts to overexpress *Tt*Ago in *T. thermophilus* from a plasmid were unsuccessful.

By contrast, expression of Strep(II)-tagged *Tt*Ago (**Fig. 2a**) in *Escherichia coli* was successful when performed at 20 °C. Under these conditions, *Tt*Ago has no effect on plasmid content (**Fig. S1b**). Analysis of co-purified nucleic acids revealed that *Tt*Ago-associated RNA (10 to 150 nucleotides) is preferentially ^{32}P-labeled in a polynucleotide kinase (PNK) forward reaction, indicating the presence of 5' hydroxyl groups (**Fig. S1c**). By contrast, co-purified DNA has a more defined length (13 to 25 nucleotides), and is preferentially labeled in a PNK exchange reaction, indicating phosphorylated 5' ends (**Fig. 2b**). A 5' phosphate group is a general feature of Ago guides [18,168].

Whereas eukaryotic Ago proteins exclusively use ssRNA guides, some prokaryotic Argonautes have a higher affinity for single stranded DNA (ssDNA) guides [164,236]. Moreover, the characteristics of the small DNAs that associate with *Tt*Ago *in vivo* are in agreement with previously described *in vitro* guide requirements [167,168,169]. *Tt*Ago catalyzes cleavage of ssDNA targets *in vitro* when supplied with complementary 5'-

phosphorylated 21-nucleotide ssDNA guides, but not when supplied with analogous ssRNA guides (**Fig. S3**) [167,168,169]. During isolation of an active site double mutant *Tt*AgoDM (*Tt*Ago^{D478A,D546A}; **Fig. 2a**), only RNAs co-purify (10 to 150 nucleotides; **Fig. S1c**). This suggests that active site residues are involved in processing and/or binding of the ssDNA molecules.

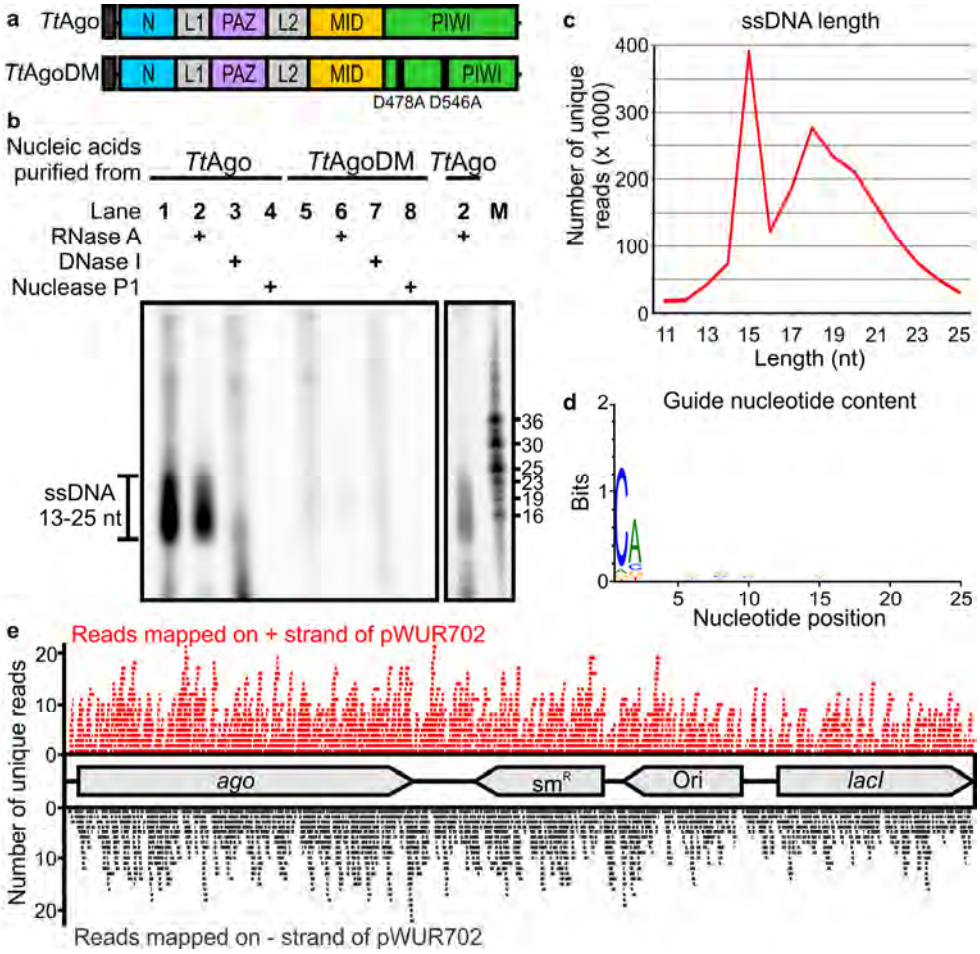


Figure 2 | *Tt*Ago guides are 5'-phosphorylated DNA molecules. a, Schematic representation of *Tt*Ago and *Tt*AgoDM proteins used for all experiments (N, PAZ, MID, and PIWI are structural domains, L1 and L2 are linkers). The amino-terminal Strep(II)-tag is indicated as a black square. **b**, Co-purified nucleic acids from *Tt*Ago and *Tt*AgoDM are labeled with [γ-^{32}P] ATP after phosphate exchange by PNK from bacteriophage T4, and treated with enzymes as indicated. M: custom ssDNA marker. nt: nucleotides. **c**, Length distribution of unique ssDNA sequences co-purified with *Tt*Ago. **d**, Nucleotide composition of unique ssDNA sequences co-purified with *Tt*Ago. **e**, Unique reads of *Tt*Ago co-purified ssDNA molecules mapped on the *Tt*Ago expression vector pWUR702.

***Tt*Ago preferentially acquires DNA guides targeting plasmid DNA**

Cloning and sequencing of *Tt*Ago-bound DNA molecules resulted in 70.6 million sequences of which 65% can be mapped on the *Tt*Ago expression plasmid pWUR702, 3% on plasmid pRARE, and 32% on the chromosome of *E. coli* K12 (**Table S3**). Remarkably, when normalized for the DNA content in each cell, *Tt*Ago predominantly co-purifies with guides complementary to pWUR702 and pRARE (approximately 54 and 8.8 times more frequently, respectively), rather than with guides complementary to the *E. coli* K12 chromosome (**Table S3**). More detailed analysis of unique guide sequences revealed two populations of DNA guides: one is 15-nucleotides long, and the other ranging from 13 to 25 nucleotides (**Fig. 2c**). No obvious bias towards specific regions of the plasmids or the chromosome was detected: the guides target coding and non-coding regions on both strands independent of GC content (**Fig. 2e**). Some guides map on one of the plasmids as well as on the chromosome of *E. coli* (for example, on *lacI* and *proL*). The fact that these guides do not seem to be under-represented compared to other plasmid targeting guides indicates that there is no selection against chromosome-targeting guides, but rather that the differential guide loading (**Table S3**) is a result of preferential acquisition of guides from plasmids. Interestingly, 89% of the DNA guides have a deoxycytidine (dC) at the first position at the 5' end and 72% have a deoxyadenosine (dA) at the second position (**Fig. 2d**). Despite this bias, identical *Tt*Ago cleavage activities are observed with DNA guides containing a 5' dC, dT, dA or dG (**Fig. S4a-d**). The 5' dC preference may result from specific guide processing, or from preferential 5' nucleoside selection by *Tt*Ago. A bias for specific 5' nucleosides also occurs in certain eukaryotic Ago proteins [182,199].

We performed activity assay to investigate whether the *in vivo* plasmid-derived ssDNA are functional guides that enable *Tt*Ago to cleave double-stranded DNA (dsDNA) targets (expression plasmid pWUR702). Purified *Tt*Ago linearizes or nicks pWUR702, resulting in linear or open circular plasmid DNA, respectively (**Fig. 3a**, lane 4). whereas *Tt*AgoDM does not show this activity (**Fig. 3a**, lane 3). The cleavage activity of *Tt*Ago is strongly temperature dependent: while ssDNA is cleaved at temperatures ≥20 °C, plasmid DNA is only cleaved at temperatures ≥65 °C (**Fig. S4e**, **f**). This agrees with the observation that during *Tt*Ago expression in *E. coli* at 20 °C, plasmid concentrations are not decreased (**Fig. S1b**). Purified *Tt*Ago is unable to cleave plasmid that have no sequence similarity to pWUR702 or pRARE (for example pWUR708; **Fig. 3b**, lane 4). However, when supplied with two synthetic 5'-phosphorylated ssDNA guides that target both strands of the plasmid

at the same locus (**Fig. 4b**), *Tt*Ago was able to linearize or nick pWUR708 (**Fig. 3b**, lane 8). These findings, together with the guide sequence data, indicate that the *in vivo* acquired DNA molecules guide *Tt*Ago to cleave complementary DNA targets. We propose to refer to these guides of *Tt*Ago as small interfering DNAs (siDNAs).

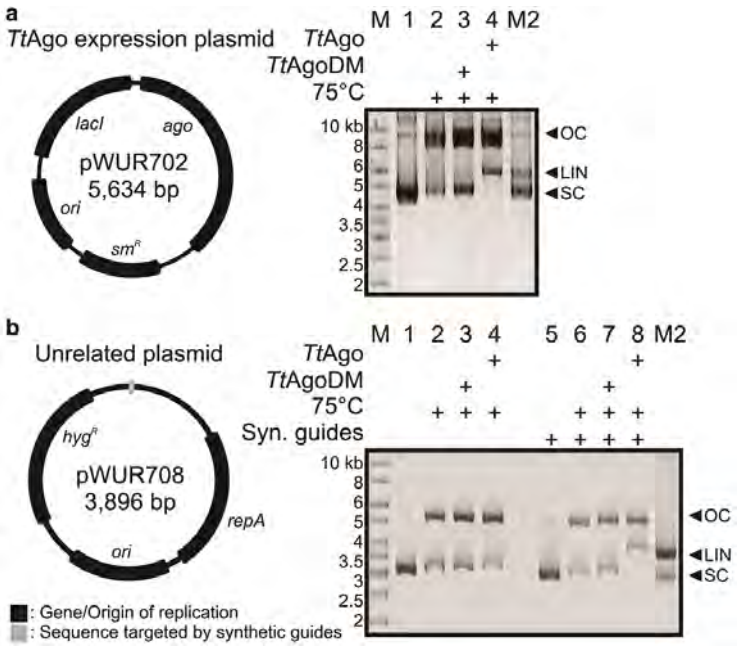


Figure 3 | *Tt*Ago cleaves plasmids complementary to its guides. Untreated target plasmid (lane 1, 5), plasmid incubated at 75 °C in absence of proteins (lane 2, 6), or in presence of *Tt*AgoDM (lane 3, 7) or *Tt*Ago (lane 4, 8) purified from *E. coli*, resolved on 0.8% agarose gels. M1: 1kb GeneRuler marker (Fermentas). M2: linearized and untreated target plasmid. OC: open circular. LIN: linear. SC: supercoiled plasmid. **a**, *Tt*Ago expression vector pWUR702. **b**, Target plasmid pWUR708, which shares no sequence identity with expression vector pWUR702 or pRARE. Additionally, synthetic (Syn.) ssDNA guides were added to the reactions with pWUR708 (lane 5-8).

To gain insight in the molecular mechanism of dsDNA cleavage by *Tt*Ago, we performed additional *in vitro* plasmid cleavage assays using purified *Tt*Ago loaded with synthetic siDNAs. Negatively supercoiled plasmids (isolated from *E. coli*) were used since at least 95% of all plasmids isolated from *T. thermophilus* have a negatively supercoiled topology [237,238]. Negative supercoiling facilitates melting of the DNA duplex, especially at elevated temperatures [239,240,241]. Target plasmids pWUR704 and pWUR705 are identical except for the flanking regions of the target site (AT-rich or GC-rich; **Fig. 4a**). Both plasmids share no sequence similarity with *Tt*Ago expression plasmid pWUR702, and they are not cleaved by *Tt*Ago unless complementary siDNAs are added (**Fig. 4c**). When supplied with a single 21-nucleotide siDNA, *Tt*Ago nicks the negatively supercoiled plasmid (**Fig. 4c**, lanes 3, 4), and when supplied with a mixture of two 21-nucleotide

siDNAs that target both DNA strands at the same locus, *Tt*Ago linearizes the plasmid (**Fig. 4b**, **c**, lane 5). Both nicking and dsDNA cleavage are more efficient when the target sequence is flanked by AT-rich regions (**Fig. 4a**, **c**, and **Fig. S5a**, **b**). Interestingly, the same *Tt*Ago-siDNA complexes are not able to cleave linearized plasmids (**Fig. S5c**, **d**). This suggests that cleavage of dsDNA by *Tt*Ago depends on the negatively supercoiled topology of the target DNA.

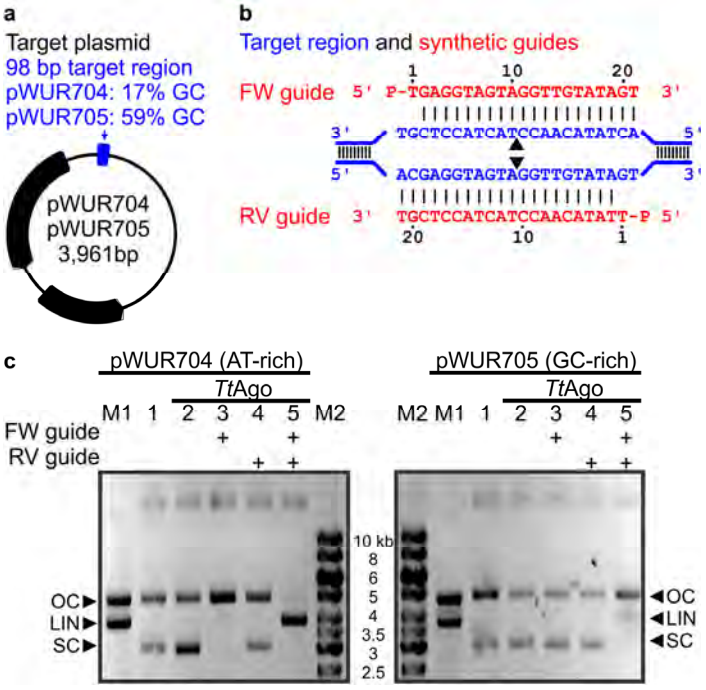


Figure 4 | *Tt*Ago cleaves plasmids by nicking two strands. **a**, Plasmids pWUR704 and pWUR705 contain a 98 bp target region with a GC content of 17% or 59%, respectively, as indicated in blue (for details, see Fig. S5a, b). **b**, Part of the pWUR704 and pWUR705 target site (indicated in blue) and complementary ssDNA guides used in this experiment (indicated in red). Black triangles indicate predicted cleavage sites. **c**, 0.8% agarose gels loaded with pWUR704 and pWUR705 plasmids which were incubated without proteins (lane 1), or with *Tt*Ago (lane 2), *Tt*Ago-forward (FW) guide complex (lane 3), *Tt*Ago-reverse (RV) guide complex (lane 4), or *Tt*Ago-FW and *Tt*Ago-RV guide complexes. M1: open circular and linear pWUR704 or pWUR705. OC: open circular. LIN: linear. SC: supercoiled plasmid. M2: 1kb GeneRuler marker (Fermentas).

Subsequent analysis revealed that the *Tt*Ago-siDNA complex is able to linearize a relaxed, nicked plasmid if its target site is directly opposite the first nick (**Fig. S5e**). If the nicked site is located further away (33 bp) from the target site, linearization of the nicked plasmid occurs only if the target region is AT-rich (**Fig. S5f**, **g**). Thus, although the negatively supercoiled topology of the plasmid is lost after the primary nick, the nick facilitates local melting of the dsDNA (especially in AT-rich DNA), which allows *Tt*Ago-siDNA complexes to nick the second strand, resulting in a dsDNA break. Like eukaryotic Ago proteins [242],

the *Tt*Ago-siDNA complex cleaves a phosphate ester bond between the target nucleotides that base pair with guide nucleotides 10 and 11 [176]. Sequence analysis of a cleaved dsDNA target (**Fig. S5h**) demonstrated that dsDNA breaks also result from nicking both strands at the canonical Ago cleavage site.

While the manuscript described in this chapter was under revision, a characterization of a prokaryotic Argonaute protein from *Rhodobacter sphaeroides* (*Rs*Ago) was published [177]. Despite similarities in the overall domain architecture of *Tt*Ago and *Rs*Ago, there are major functional differences between these proteins. *Rs*Ago acquires mRNA-derived RNA guides with a 5' uridine (U), while *Tt*Ago acquires DNA guides with a 5' dC. In both proteins, guides complementary to plasmids are overrepresented. However, *Rs*Ago lacks a functional catalytic site and functions by target-binding alone. *Tt*Ago, on the other hand, harbors a functional catalytic site allowing cleavage of both single- and double-stranded targets.

Discussion

On the basis of our findings, we propose a model for DNA interference by *Tt*Ago. Upon entry of plasmid DNA into the cell, *Tt*Ago acquires siDNA guides (13 to 25 nucleotides in length) from the invader. Although the mechanism of guide acquisition by *Tt*Ago is unknown, the requirement of an intact catalytic site suggests involvement of the nuclease itself. *Tt*Ago is loaded with siDNAs that are preferentially derived from plasmids; as such, single guides may allow for neutralization of multi-copy invaders. Combining our *in vivo* and *in vitro* data, we speculate that *Tt*Ago uses siDNA guides to specifically cleave ssDNA targets, such as DNA taken up by the natural competence system [234] or replication intermediates. The siDNA-*Tt*Ago ribonucleoprotein complex also targets negatively supercoiled dsDNA, which results in plasmid nicking. Especially in case of plasmid DNA, single strand breaks will result in loss of the supercoiled topology, and as such in decreased transcription levels [243]. Furthermore, if the nick-site is located in an AT-rich region, *Tt*Ago loaded with an siDNA that targets the opposite strand may generate a dsDNA break, potentially leading to degradation of the plasmid by other nucleases. The observation that invading DNA elements generally have a lower GC content than their host [244], may contribute to the self/non-self discrimination by *Tt*Ago. Whereas the eukaryotic Ago protein is a key component of sophisticated multi-enzyme systems for RNA-guided RNA interference, we reveal the biochemical activity and functional importance of an

evolutionary related enzyme in prokaryotes that protects its host against mobile genetic elements through DNA-guided DNA interference.

Acknowledgements

We thank A. Hidalgo, C. E. César, M. Davids and R. H. J. Staals for advice on experimental procedures. Furthermore we thank R. Engelhart, B. van Genugten, G. Göertz and R. Stolk for experimental contributions. This work was financially supported by grants from the Netherlands Organization for Scientific Research (NWO) to JvdO (NOW-TOP, 854.10.003), and to SJJB (NWO Vidi, 864.11.005), and by project BIO2010-18875 from the Spanish Ministry of Science and Innovation, and an Institutional Grant from "Fundación Ramón Areces" to CBMSO (JB).

Author contributions

MMJ and JHJ made genomic *T. thermophilus* mutants under supervision of JvdO. *T. thermophilus* experiments were performed by DCS, MMJ and JHJ under supervision of JB, SJJB and JvdO. DCS and ERW purified RNA for RNA-seq and DCS analyzed RNA-seq data under supervision of SJJB and JvdO. DCS and APS performed experiments in which *Tt*Ago expression in *T. thermophilus* was shown using mass spectrometry. DCS, MMJ and JHJ made all plasmid constructs under supervision of SJJB, JB and JvdO. DCS and ERW and YZ purified and analyzed *Tt*Ago guides. *In vitro* activity assays were designed by DCS, SJJB, YW, DJP and JvdO and performed by DCS and YZ under supervision of SJJB and JvdO. All authors read and approved the manuscript.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Data deposition

The RNA-seq data discussed in this chapter have deposited in NCBI's Gene Expression Omnibus under accession number GSE52738. The siDNA sequence data discussed in this publication have been deposited in NCBI's BioSample database and are accessible under accession number SAMN02593821.

Experimental procedures

Strains

For *in vivo* experiments, *T. thermophilus* HB27 (ATCC BAA-163 / DSM 7039 / NBRC 101085) was used, which is referred to in this chapter as HB27 or wild type. Furthermore, HB27^{EC}, and two genomic variants of the HB27 strain, HB27Δ*ago* (knockout strain) and HB27Δ*ago::*^{s}*ago* (knockout strain complemented with *strep(II)-tag-ago* fusion and kanamycin resistance marker insert), were used (**Fig. 1**, **Table S4**).

Genomic Mutants

HB27 genomic DNA including megaplasmid pTT27 was purified using the FastDNA SPIN Kit for Soil (MP biomedicals). The genomic regions directly upstream (1kb) and downstream (2.4kb) of the *ago* gene (TT_P0026) were PCR amplified from *T. thermophilus* HB27 genomic DNA. These genomic regions contained pTT27 base pair positions 26047-25061 (upstream sequence) and 22996-20583 (downstream sequence). The amplified DNA was cloned into the pUC18 vector (**Table S5**) and the insert was transferred to pK18 [245] to generate pWUR701 (**Table S5**). Strain HB27 was grown to an $OD_{600\ nm}$ of 0.4 in TTH-medium (0.8% (w/v) bacto-tryptone, 0.4% (w/v) yeast extract, 51.3 mM NaCl, pH to 7.5 with NaOH, dissolved in mineral water (Evian)). 0.5 ml of the culture was transferred to a new tube and naturally transformed by addition of 1 µg plasmid pWUR701. The culture was incubated o/n in a shaker incubator at 65 °C and plated on TTH plates with 30 µg ml^{-1} kanamycin. Cells were repetitively streaked on non-selective TTH plates and grown in non-selective TTH-medium until kanamycin^{R} was lost. Genomic DNA of kanamycin^{S} cells was purified using the FastDNA SPIN Kit for Soil (MP biomedicals) and loss of the *ago* gene was confirmed by PCR-amplification of genomic DNA and sequencing of the target region. This strain is named HB27Δ*ago* (**Fig. 1**), or knockout strain.

The genes encoding Strep(II)-tagged *Tt*Ago protein and kanamycin^{R} marker with upstream pSLPa promoter were PCR amplified from pWUR627 and pMK184 [246], respectively (**Table S5**). PCR products were cloned into pWUR676 (**Table S5**). HindIII-linearized pWUR676 was used to transform strain HB27Δ*ago* as described earlier. This strain is named HB27Δ*ago::*^{s}*ago* (**Fig. 1**). Genomic DNA was purified using the FastDNA SPIN Kit for Soil (MP biomedicals) and insertion of the ^{s}*ago*-kanamycin^{R} cassette was confirmed by PCR-amplification from genomic DNA and sequencing of the target region.

Transformations

T. thermophilus strains were cultivated in TTH-medium in a 65 °C shaker incubator until an $OD_{600\ nm}$ of 0.4 was reached. The culture was diluted 1:1 in pre-warmed TTH medium and incubated for another hour at 65 °C. 0.5 ml of the culture was transferred to a new tube which was incubated at 65 °C for 30 min. 100 ng of plasmid pMK184 or pMHPnqosGFP was added and the mixture was incubated for 4 h at 65 °C without shaking, after which it was serial diluted and plated on TTH plates (TTH-medium solidified with 1.5% agar) and on selective TTH plates (TTH plates supplied with 50 μg ml^{-1} kanamycin or 100 μg ml^{-1} hygromycin). After 48 h of incubation at 65 °C colonies were counted. Competence was determined as the amount of $kanamycin^R$ or $hygromycin^R$ colony forming units (c.f.u.; counted on selective plates) per μg DNA, divided by total c.f.u. (counted on non-selective plates). To show relative competence, HB27 wild-type transformation efficiency was set to 1 while other strain their competences were normalized against this number.

DNA purification

For plasmid purification, pMKPnqosGFP and pMHPnqosGFP (**Table S5**) harboring *T. thermophilus* HB27 and HB27Δ*ago* strains were cultivated in triplicates in TTH medium supplied with 30 ng μl^{-1} kanamycin or 100 ng μl^{-1} hygromycin. Five $OD_{600\ nm}$ units of each overnight culture were harvested and plasmids were isolated with the Fermentas GeneJET plasmid Miniprep Kit (Thermo Scientific) according to the manual provided by the manufacturer and quantified using a NanoDrop ND1000 spectrophotometer. For complete DNA (containing both genomic and plasmid DNA) purification, pMKPnqosGFP and pMHPnqosGFP (**Table S5**) harboring *T. thermophilus* HB27 and HB27Δ*ago* cultures were cultivated in triplicates to an $OD_{600\ nm}$ of 0.5. One $OD_{600\ nm}$ unit was harvested and complete DNA was isolated using the JGI 'bacterial genomic DNA isolation using CTAB' protocol. 2.5 μg DNA of each purification was resolved on 0.8% agarose gels and stained with SYBR Safe Nucleic Acid Stain (Invitrogen), visualized using a G:BOX Chemi imager and analyzed using GeneTools analysis software (Syngene).

RNA sequencing

Triplicate *T. thermophilus* strains were cultivated in 20 ml TTH medium in a 65 °C shaker incubator overnight. Cultures were diluted 1/100 and grown to an $OD_{600\ nm}$ of 0.5, after which cells were harvested by centrifugation. After harvesting, RNA was purified using mirVana RNA isolation kit (Ambion) according to the instructions provided by the

manufacturer. Biological triplicates of purified RNA were sequenced by BaseClear BV by Illumina sequencing. Reads were mapped on genomes and plasmid using Rockhopper [247], but rather than using the programs calculated expression rates and significance, the percentage of raw counts mapped on each gene were normalized against the total number of raw counts mapped on the genome. Variance in expression was calculated by dividing the average of the triplicate normalized counts mapped on single genes in strain HB27 by the average of the triplicate normalized counts mapped on the same gene in strain B27Δ*ago*.

_Tt_Ago expression and purification from _E. coli_ KRX

The *ago* gene was PCR amplified from *Thermus thermophilus* (ATCC 27634) genomic DNA (gene TTHB0068, base positions on pTT27: 61573-59516), and directionally cloned into a pET-52b(+) expression vector (pWUR627; **Table S5**). By introduction of mutations according to the QuikChange Site-Directed Mutagenesis Kit instruction manual (Stratagene), pWUR642 was generated (**Table S5**). The inserts of pWUR627 and pWUR642 were PCR amplified and ligated into pCDF-1b (pWUR702 and pWUR703; **Table S5**). These plasmids were transformed into *E. coli* KRX (Promega) simultaneously with pRARE (Novagen), purified from *E. coli* Rosetta DE3 (Novagen).

Strains were cultivated in LB medium containing the corresponding antibiotics (50 μg ml^{-1} streptomycin, 34 μg ml^{-1} chloramphenicol) in a shaker incubator at 37 °C. When the culture reached an $OD_{600\ nm}$ of 0.7-0.8, cells were cold-shocked by incubation in an ice bath for 15 min. Expression was induced by adding isopropyl-β-D-thiogalactoside (IPTG) and L-Rhamnose to a final concentration of 1 mM and 0.1% (w/v), respectively, and expression was continued for 16 h in a shaker incubator at 20 °C. Cells were harvested by centrifugation.

For plasmid quantification, plasmids were isolated from 5 $OD_{600\ nm}$ units of harvested cells using the Fermentas GeneJET plasmid Miniprep Kit (Thermo Scientific) according to the manual provided by the manufacturer and quantified using a NanoDrop ND1000 spectrophotometer.

For *Tt*Ago purification, harvested cells were resuspended in Buffer I (20 mM Tris/HCl pH 8, 1 M NaCl, supplied with either 2 mM $MnCl_2$ or 2 mM $MgCl_2$), and disrupted using a French pressure cell. Expressed proteins have an N-terminal Strep(II)-tag and were isolated using Strep-Tactin affinity chromatography (IBA) with an adapted protocol. Before loading

of the cell-free extract, columns were equilibrated in Buffer I. After loading, columns were washed with 9 column volumes of Buffer I and with 9 column volumes of Buffer II (20 mM Tris/HCl pH 8, 0.5 M NaCl, supplied with 2 mM $MnCl_2$). Proteins were eluted in Buffer III (Buffer II supplemented with 2.5 mM d-Desthiobiotin (Sigma-Aldrich)). For purification of *Tt*Ago used in Mn/Mg gradient experiments, no Mn or Mg was added to purification buffers. For other activity assays, $MnCl_2$ or $MgCl_2$ was added to all buffers at a final concentration of 0.5 mM.

_Tt_Ago purification from _T. thermophilus_

HB27Δ*ago::*^{s}*ago* was cultivated in TTH medium supplemented with 30 µg ml^{-1} kanamycin at 65 °C. After overnight growth, cells were harvested and *Tt*Ago was purified as described above. After purification, elution fractions were resolved on SDS-PAGE gels and purified proteins were stained using Coomassie brilliant blue stain. A band corresponding to the region with the molecular weight of *Tt*Ago (75-80 kDa) was excised from the gel and subjected to in-gel digestion using a Perkin Elmer Janus Automated Workstation. Peptide mixtures were injected onto a nanoACQUITY UPLC (Waters Corporation) coupled to a LTQ-Orbitap XL (Thermo Fisher Scientific) via an Advion Biosciences Nanomate. Peptides were eluted over a 30 min gradient (5-40 % ACN). MaxQuant (v1.4.1.2) and its embedded Andromeda search engine were used to search the data against a database containing *Thermus thermophilus* sequences extracted from Uniprot (8 August 2013). Methionine oxidation was used as a variable modification and a maximum of 2 missed trypsin cleavages were allowed. Peptide and protein posterior error probabilities (PEP) were calculated using the Target-Decoy using the revert scheme. The light version of intensity based absolute quantification (iBAQ) was used to rank the identified proteins by estimated relative abundance.

Guide co-purification and sequencing

Proteinase K (Ambion) and $CaCl_2$ (final concentration 5 mM) were added to purified proteins and samples were incubated for 1 h at 37 °C. Nucleic acids were separated from protein content using Roti phenol/chloroform/isoamyl alcohol pH 7.5-8.0 (Carl Roth GmbH) and further purified by ethanol precipitation. Precipitation was performed overnight at -20 °C in the presence of linear polymerized acrylamide as carrier. Purified nucleic acids were [γ-^{32}P] ATP labeled with T4 PNK (Fermentas) in exchange- or forward-labeling reactions and thereafter separated from free [γ-^{32}P] ATP using a Sephadex G-25

column (GE). Labeled nucleic acids were incubated with nucleases (DNase-free RNase A (Fermentas), RQ1 RNase-free DNase I (Promega) or P1 nuclease (Sigma)) for 1 h at 37 °C. After nuclease treatment, samples were mixed with Loading Buffer (95% (deionized) formamide, 5 mM EDTA, 0.025% SDS, 0.025% Bromophenol blue, and 0.025% xylene cyanol), heated for 5 min at 95 °C and resolved on 15% or 20% denaturing polyacrylamide gels. Radioactivity was captured from gels using phosphor screens. Nucleic acids were purified from *Tt*Ago and treated with RNase A, as described above. The small 5'-phosphorylated DNA molecules were poly-adenylated at their 3' end using recombinant terminal deoxynucleotidyl transferase (TdT, Invitrogen), according to the instructions of the manufacturer. After purification of the product using the QIAquick nucleotide removal kit (Qiagen), 5'-phosphorylated and 3'-polyadenylated products were ligated to the 3' end of oligonucleotide BG4409 (**Table S4**) using T4 RNA ligase (Ambion), according to the instructions of the manufacturer. After purification of the product using the QIAquick nucleotide removal kit (Qiagen), the product was PCR amplified using primers BG4409 and BG4436 (anchored poly-T primer (partially degenerate), **Table S4**). The PCR amplification product was gel purified using the GeneJET gel extraction kit (Fermentas) and sent for sequencing by Imagif, Plateforme de Séquençage à Haut Débit by Illumina sequencing with an adapted RNA-seq protocol. Sequences were analyzed with FastQC software (Babraham Bioinformatics). After mapping on genome and plasmids, duplicate reads were removed using SAMtools software [248], to exclude a bias for preferentially PCR amplified reads in downstream analysis. Unique read datasets were re-analyzed with FastQC software and remapped on genome and plasmid DNA using Tablet software (James Hutton Institute) [249].

DNA guides and targets

The sequence of guide BG3466 is based on *let-7* miRNA, and has been used before in experiments performed with *Tt*Ago [167,168,169], whereas the sequence of guide BG4017 is based on the reverse complementary sequence of *let-7* miRNA (**Table S4**). Both guides have a 5' phosphate, are 21-nucleotides long and have been PAGE purified after synthesis. Oligonucleotides BG4262-BG4265 (**Table S4**) were used in activity assays as ssDNA target or mixed together with 2X STE buffer (20 mM Tris-HCl pH 8, 100 mM NaCl, 2 mM EDTA) in a 1:1:2 ratio (BG4262:BG4263:2X STE or BG4264:BG4265:2X STE) and incubated at 95 °C for 5 min. Samples were cooled down to room temperature (20 °C). Annealed oligonucleotides were used as inserts for plasmid pWUR677 (generated from pFU98) [250] to generate pWUR704 and pWUR705. For experiments with nicked and

linearized targets, pWUR704 and pWUR705 were treated with Nb.BsmI or SpeI, respectively. Plasmid pWUR708 was generated as pWUR704 and pWUR705 but with annealed BG3467 and BG3468 oligonucleotides as insert.

Activity assays

Purified *Tt*Ago, ssDNA or ssRNA guides, and ssDNA targets (**Table S4**) were mixed in 5:1:1 ratio (*Tt*Ago:guide:target) in 2X Reaction Buffer (20 mM Tris-HCl pH 8, 250 mM NaCl supplied with varying concentrations of $MnCl_2$ or $MgCl_2$). Reaction mixtures were incubated 1 h at 75 °C. Reactions were stopped by the addition of Loading Buffer and heated for 5 min at 95 °C before the samples were resolved on 15% or 20% denaturing polyacrylamide gels. Gels were stained using SYBR gold Nucleic Acid Gel Stain (Invitrogen) and nucleic acids were visualized using a G:BOX Chemi imager (Syngene). Because DNA-guided cleavage of ssDNA is observed in the presence of 5-10 μM Mn^{2+} (**Fig. S5j**), but comparable cleavage levels are observed in the presence of Mg^{2+} only at 10-fold higher concentrations (**Fig. S5j**), all activity assays are performed in the presence of 0.5 mM $MnCl_2$.

Purified *Tt*Ago, ssDNA guides and plasmid targets were mixed in a 25:5:1 ratio (*Tt*Ago:guide:target) in 2X Reaction Buffer supplemented with 0.5 mM $MnCl_2$. Samples were incubated for 16 h at 75 °C. Reactions were stopped by adding Proteinase K solution (Ambion) and $CaCl_2$ (final concentration 5 mM) and samples were incubated for 1 h at 65 °C. Samples were mixed with 6X loading dye (Fermentas) before they were resolved on 0.8% agarose gels. Agarose gels were stained with SYBR safe or SYBR gold Nucleic Acid Gel Stain (Invitrogen) and nucleic acids were visualized using a G:BOX Chemi imager (Syngene).

Plasmid pWUR704 was linearized with *Tt*Ago-siDNA complexes as described earlier. The DNA was purified from the activity assay sample by PCI extraction followed by ethanol precipitation. Purified DNA was cut either by XbaI or by NheI. Restriction site overhangs were filled in with Klenow Fragment (Thermo Scientific) according to the manual provided by the manufacturer. Blunt-end linear plasmid was closed by T4 ligase ligation according to the manual provided by the manufacturer (Thermo Scientific). Ligated plasmids were treated with HindIII (in the case of the XbaI-treated plasmids) or SalI (in the case of NheI-treated plasmids) to eliminate possible background of the original plasmid. Plasmids were

transformed to NEB 5-α *E. coli* competent cells (New England Biolabs) according to the manual provided by the manufacturer. Colonies were picked, grown overnight in LB medium at 37 °C and miniprepped with the Fermentas GeneJET Plasmid Miniprep Kit (Thermo Scientific). Purified plasmids were sent to GATC Biotech (Germany) for target site sequencing.

Statistical analysis

All *P* values stated in this chapter are calculated by a two-tailed distributed two-sample *t*-test assuming equal variances. For the calculation of *P*-values of the transformation efficiencies, competence (calculated as described above) from biological duplicates of each strain was used as input. For the calculation of *P* values of plasmid purification, plasmid DNA yield of biological triplicates of each strain were used as input. For the calculation of *P* values of plasmid DNA content of complete DNA purification, plasmid DNA content of biological triplicates of each strain were used as input. For the calculation of *P* values of differences in expression levels of specific genes, normalized raw mapped counts of biological triplicates of each strain were used as input. All *P* values calculated are considered to be significant as for all calculations $P<0.02$.

Supplementary Figures and Tables

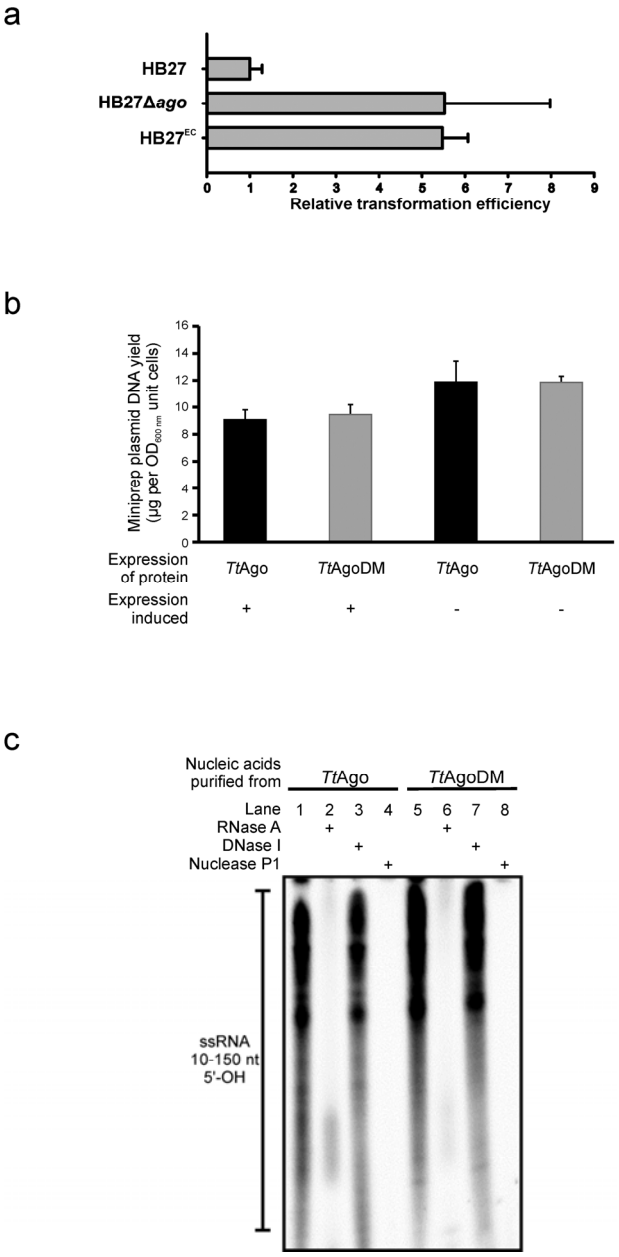
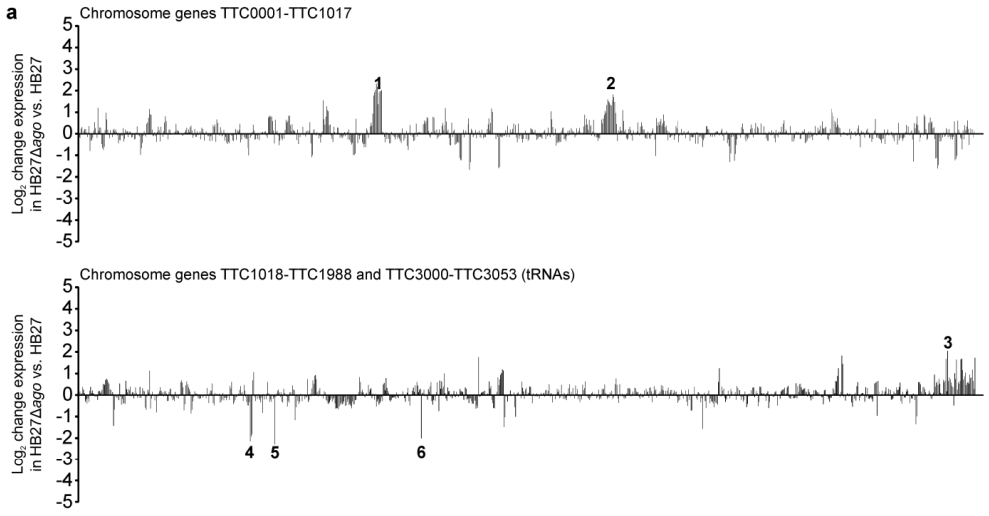
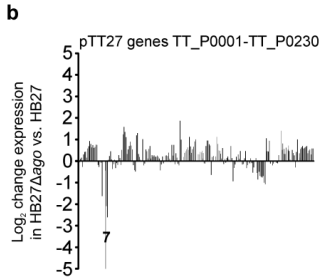


Figure S1 | Analysis of *Tt*Ago in *T. thermophilus* and *E. coli*. **a**, *Tt*Ago decreases plasmid transformation efficiency of *T. thermophilus*. Transformation efficiency of different *ago* mutant strains relative to the transformation efficiency of wild-type strain HB27. HB27^{EC} is an HB27 mutant selected for high competence, and HB27Δ*ago* is an *ago* gene knockout strain (Fig. 1a). Strains were transformed with plasmid pMK184 (Table S5). Transformations were performed in biological duplicates for each strain. Error bars indicate standard deviations**.** **b**, Effect on *Tt*Ago expression on plasmid content in *E. coli* KRX. *Tt*Ago and *Tt*AgoDM were expressed in *E. coli* KRX from plasmid pWUR702 and pWUR703. Plasmids were purified from biological triplicate cultures in which expression was induced (+) or not induced (-). Compared with *Tt*AgoDM expression, *Tt*Ago expression in *E. coli* KRX does not lead to reduced plasmid content. Changes in plasmid yield between induced and not induced cultures probably originate from protein expression energy costs. Error bars indicate standard deviations. **c**, 10-150-nucleotide (nt) RNA with 5'-OH group co-purified with *Tt*Ago and *Tt*AgoDM. Nucleic acids are phosphorylated in a T4 PNK forward reaction (5'-OH groups, and to a lesser extend 3'-P groups, are labeled) using [γ-^{32}P] ATP, and resolved on 15% denaturing polyacrylamide gels. Nucleic acids were not treated (lane 1, 5), RNase A treated lanes 2, 6), DNase I treated (lane 3, 7) or Nuclease P1 treated (lane 4, 8).





c

	Gene(s)	Encodes	Change in expression (Log_2)	p-value
1	TTC0333-0343	Branched amino acid transport system	1.85	0
2	TTC0597-0603	Branched amino acid transport system	1.43	0.01
3	TTC3000-3053	tRNAs	0.74	0.25
4	TTC1213	1-pyrroline-5-carboxylate dehydrogenase	-2.13	0
5	TTC1241	Acylamino acid releasing enzyme	-2.27	0
6	TTC1407	S-layer protein	-1.99	0
7	TT_P0024	Alkaline phosphatase	-1.64	0
	TT_P0026	*Tt*Ago	< -10.00	0
	TT_P0028	Survival protein SurE	-2.11	0.02
	TT_P0029	Hypothetical conserved protein	-2.63	0.03

Figure S2 | Change in transcription of *T. thermophilus* genes after *ago* gene knockout. a, **b**, RNA-seq analysis was performed on biological triplicates for each strain. Change in gene expression of genes encoded on the chromosome (panel a) or on the megaplasmid (panel b) is shown as the log_2 of the fold difference in expression of the average of normalized mapped reads on that gene in HB27Δ*ago* compared average of normalized mapped reads on that gene in HB27. **c**, Genes or operons containing genes with a log_2 expression change larger than 2 or -2.

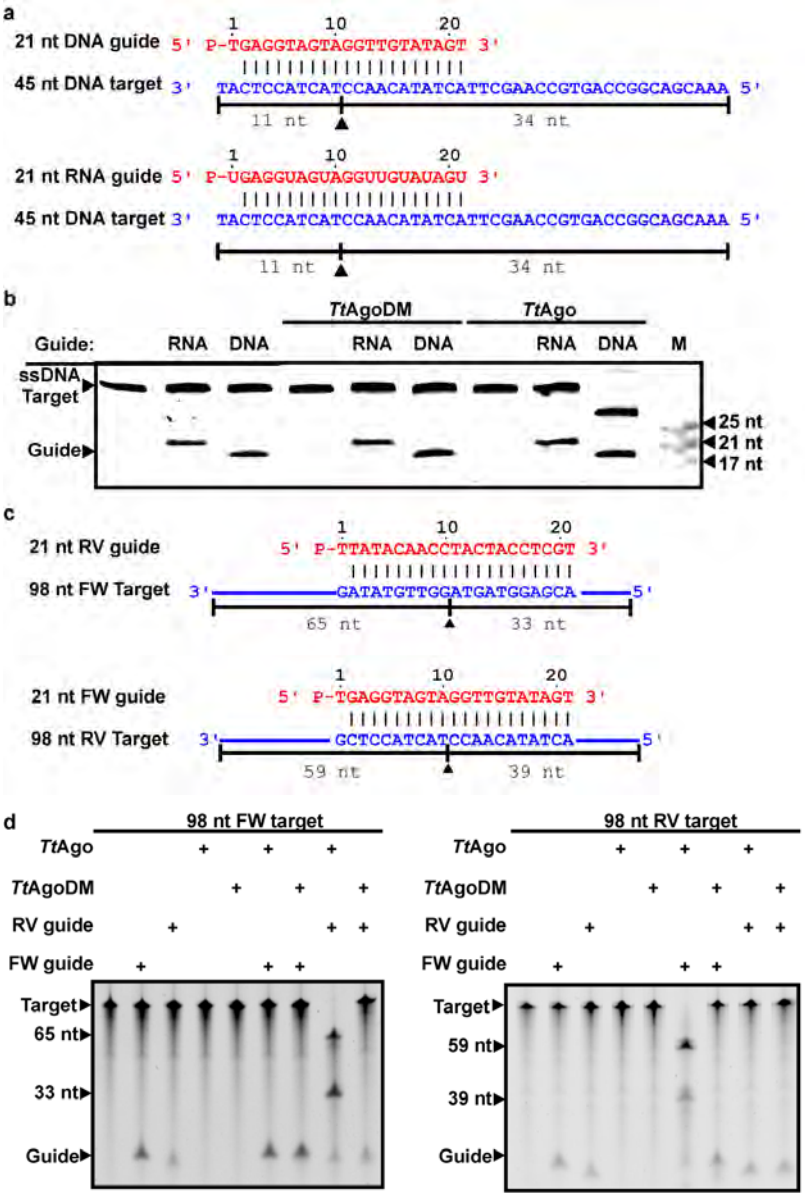


Figure S3 | *Tt*Ago cleaves ssDNA using ssDNA guides. a, 21-nucleotide (nt) DNA and RNA guides are complementary to the 45-nucleotide DNA targets. Predicted cleavage positions are indicated with a black triangle. **b**, 20% denaturing polyacrylamide gel loaded with samples in which *Tt*Ago and *Tt*AgoDM were provided with an RNA or an DNA guide to cleave a 45 nucleotide ssDNA target. **c**, 21-nucleotide RV and FW DNA guides are complementary to the 98-nucleotide ssDNA targets. Predicted cleavage positions are indicated with a black triangle. **d**, 98-nucleotide ssDNA targets are incubated with *Tt*Ago and *Tt*AgoDM provided with complementary and non-complementary DNA guides and resolved on 15% denaturing polyacrylamide gels.

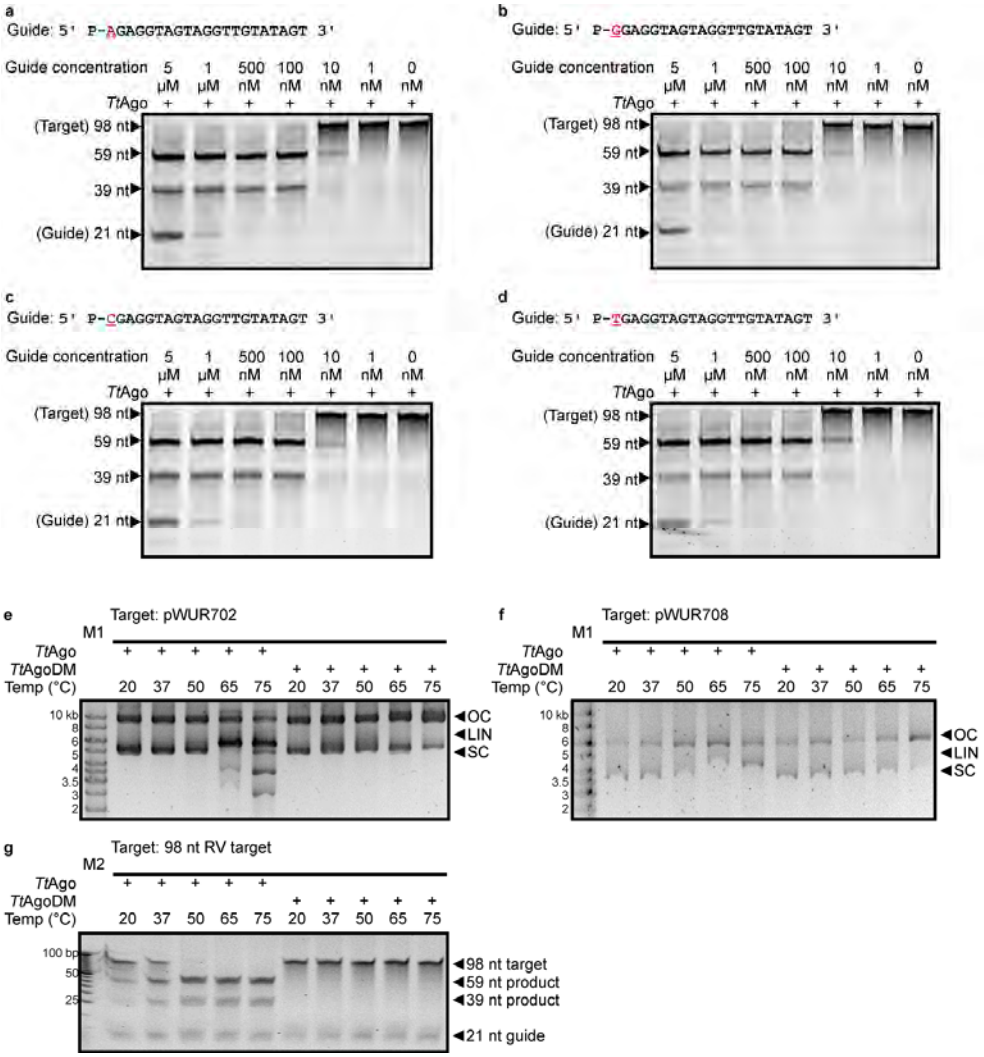


Figure S4 | Effect of variation of the 5'-end deoxynucleoside of the siDNA and the effect of temperature on *Tt*Ago cleavage efficiency. **a-d**, Cleavage of 98-nucleotide ssDNA target (Fig. S3c) by *Tt*Ago loaded with complementary siDNAs containing a different 5' deoxynucleoside, as shown in red. The concentrations of each siDNA were varied (indicated on top of the gels). Products of the reaction were resolved on 15% denaturing polyacrylamide gels. **e**, **f**, *Tt*Ago Expression plasmid pWUR702 (**e**; no guides added) and pWUR708 plasmid (**f**; FW and RV guides added), incubated with *Tt*Ago and *Tt*AgoDM at different temperatures, resolved on 0.8% agarose gels. M1: 1kb GeneRuler marker (Fermentas). OC: open circular, LIN: linear, SC: supercoiled. **g**, 98-nucleotide RV target cleavage (FW guide added) incubated with *Tt*Ago and *Tt*AgoDM at different temperatures, resolved on a 15% denaturing acrylamide gel. M2: O'RangeRuler 5 bp DNA Ladder (Thermo Scientific).

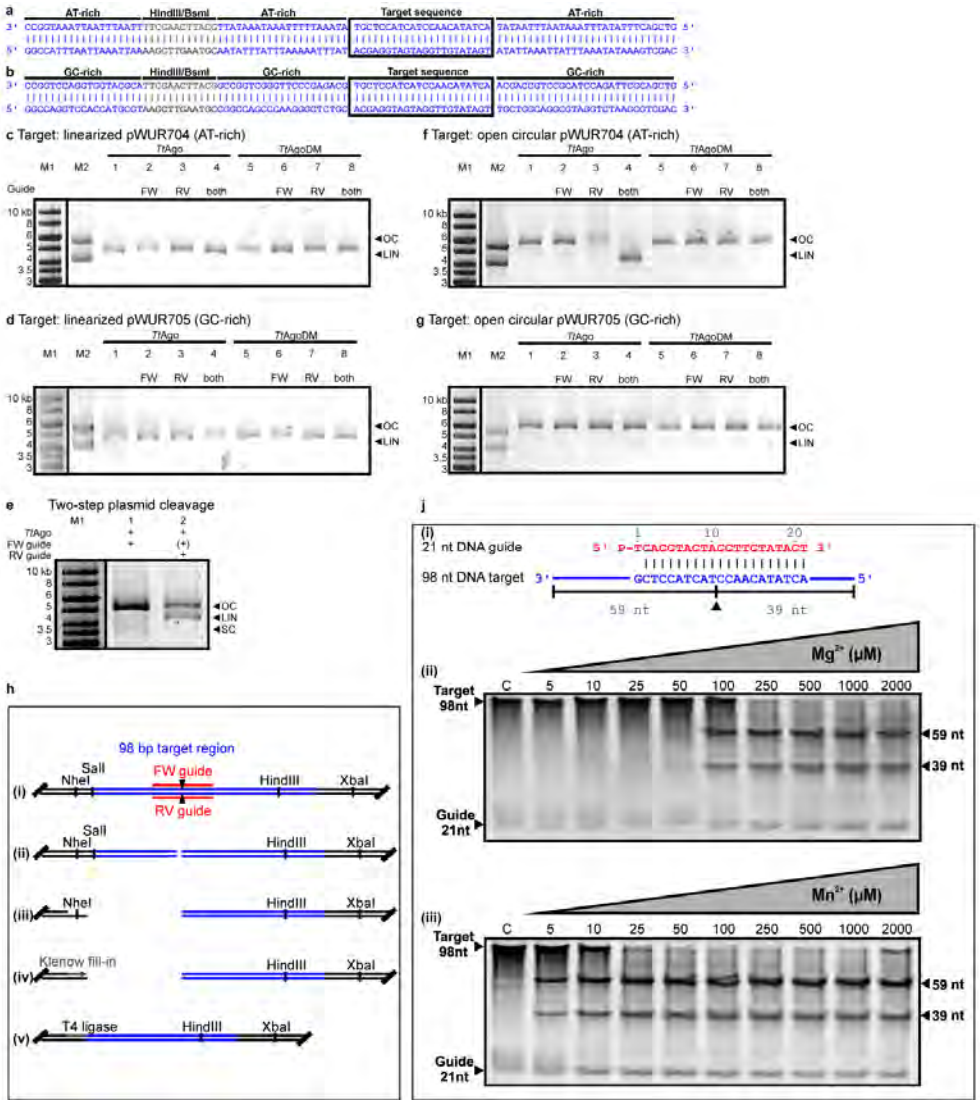


Figure S5 | Activity analyses of *Tt*Ago. a, **b**, AT-rich (17% GC) insert of pWUR704 (panel a) and GC-rich insert (59% GC) of pWUR705 (panel b). The target sequence is boxed. Restriction sites HindIII and BsmI are indicated in gray. Sequences are displayed in 3'-5' direction to allow comparison with Fig. 4b, which shows guide base pairing to this sequence. **c**, **d**, SpeI-linearized plasmid pWUR704 (panel c) and pWUR705 (panel d) incubated with *Tt*Ago-siDNA and *Tt*AgoDM-siDNA complexes targeting both strands of the plasmid, and resolved on 0.8% agarose gels. M1: 1kb GeneRuler marker (Fermentas). M2: open circular and linearized pWUR704 (panel c), or open circular and linearized pWUR705 (panel d). OC: open circular, LIN: linear. FW guide: BG3466. RV guide: BG4017. High salt concentration (250 mM NaCl) in the reaction buffer cause the *Tt*Ago treated samples to run higher in the gel than M1 and M2. **e**, Two-step plasmid cleavage. Target pWUR704 was first nicked by a *Tt*Ago-siDNA complex targeting the first strand (FW guide, lane 1), after which a *Tt*Ago-siDNA complex targeting the

other strand was added (RV guide, lane 2). FW guide is still present, and its presence is therefore indicated as (+). M1: 1kb GeneRuler marker (Fermentas). OC: open circular. LIN: linear. SC: supercoiled. **f**, **g**, Nb.BsmI-nicked plasmid pWUR704 (panel f) and pWUR705 (panel g) incubated with *Tt*Ago-siDNA and *Tt*AgoDM-siDNA complexes targeting the un-nicked strands of the plasmid (33 bp away from the nicking site), and resolved on 0.8% agarose gels. M1: 1kb GeneRuler marker (Fermentas). M2: open circular and linearized pWUR704 (panel a), or open circular and linearized pWUR705 (panel b). OC: open circular. LIN: linear. High salt concentrations (250 mM NaCl) in the reaction buffer cause the *Tt*Ago treated samples to run higher in the gel than M1 and M2. **h**, *Tt*Ago dsDNA cleavage site analysis. (i), Plasmid pWUR704 with *Tt*Ago-siDNA target sequences. Predicted cleavage sites are indicated with black triangles. (ii), pWUR704 was linearized using *Tt*Ago-siDNA complexes targeting the plasmid on both strands. (iii), The linearized plasmid was cleaved using either NheI (as shown) or XbaI (not shown). (iv), Restriction site overhangs and possible overhangs resulting from *Tt*Ago-siDNA cleavage were filled using Klenow Fragment polymerase (Fermentas). (v), Blunt-end DNA was ligated using T4 DNA ligase (Fermentas), after which the plasmid could be transformed and later sequenced to determine the cleavage site. Sequences revealed that *Tt*Ago-siDNA complexes indeed cleaved the target at the predicted locations (as shown in (i), and are shown in more detail in Fig. 4b and Fig. S5a, b. Note that in this picture target sequences are displayed in reversed order compared to Fig. 4b and Fig. S5a, b. **j**, *Tt*Ago prefers Mn^{2+} over Mg^{2+} as a divalent cation for cleavage. (i) 21-nucleotide DNA guide and 98-nucleotide ssDNA target used. The predicted cleavage site is indicated with a black triangle. (ii, iii) 98-nucleotide ssDNA target cleavage reaction with *Tt*Ago loaded with a 21-nucleotide siDNA in the presence of increasing concentrations of Mg^{2+} (ii), or Mn^{2+} (iii), as indicated on top of the gel. Samples were resolved on 15% denaturing polyacrylamide gels.

a

Genes involved in competence		
Gene	Encoded protein	Log_2 fold change (*P*-value)
TTC1603	ComEC	0.09 (0.49)
TTC1602	ComEA	-0.05 (0.74)
TTC1873	DprA	0.22 (0.19)
TTC0854	PilA1	1.15 (0.02)
TTC0855	PilA2	0.97 (<0.01)
TTC0856	PilA3	0.82 (<0.01)
TTC0858	PilA4	0.45 (0.17)
TTC1716	PilD	0.40 (0.11)
TTC1622	PilF	0.10(0.11)
TTC0440	PilC	0.43 (<0.01)
TTC1017	PilQ	0.15 (0.30)
TTC0857	ComZ	0.61 (0.05)
TTC1013	PilM	0.25 (0.06)
TTC1014	PilN	0.21 (0.05)
TTC1015	PilO	0.20 (0.43)
TTC1016	PilW	-0.17 (0.12)
TT_P0190	PilA	-0.01 (0.93)

b

Genes involved in host-defence		
Gene	Encoded protein	Log_2 fold change (*P*-value)
TT_P0026	*Tt*Ago	<-10 (<0.01)
TTC1926	Cas2	0.16 (0.03)
TTC1927	Cas6	0.02 (0.88)
TT_P0101	Cas2	0.18 (0.61)
TT_P0102	Csm1	0.39 (0.04)
TT_P0103	Csm2	0.46 (<0.01)
TT_P0104	Csm3	0.54 (<0.01)
TT_P0105	Csm4	0.64 (0.01)
TT_P0106	Csm5	0.52 (<0.01)
TT_P0107	Csx1	0.28 (0.07)
TT_P0115	Cmr2	0.32 (0.06)
TT_P0116	Cmr3	0.30 (0.16)
TT_P0117	Cmr1	0.43 (0.01)
TT_P0118	Cmr4	0.41 (<0.01)
TT_P0119	Cmr5	0.48 (<0.01)
TT_P0120	RAMP	0.38 (0.03)
TT_P0132	Cas3	0.22 (0.20)
TT_P0133	Cas4	1.14 (<0.01)
TT_P0134	Cas8C	0.80 (<0.01)
TT_P0135	Cas7	0.50 (<0.01)
TT_P0136	Cas4	0.48 (0.01)
TT_P0195	Cas2	1.42 (0.15)
TT_P0196	Cas1	0.52 (0.20)
TT_P0197	Cas4	0.34 (0.44)
TT_P0204	Cas6	0.73 (0.01)
TT_P0215	Cas1	0.16 (0.42)

Table S1 | Expression profile of *T. thermophilus* genes involved in competence and host defense. a, Genes involved in competence. **b**, Genes involved in host defense. Expression values are given log_2 values of fold expression levels of the gene in strain HB27Δ*ago* relative to strain HB27, and *P* values (*t*-test) are indicated between brackets. Changes in expression are considered substantial if the log_2 value >2 and *P*<0.02 (Fig. S2).

Table S2 | Mass-spectrometry data of identified proteins after Strep(II)-tag affinity purification.

a

Proteins with most abundant peptides

Protein ID	Name	Peptides	Sequence coverage (%)	Mol. Weight (kDa)	PEP	iBAQ
Q72G73	TT_C1975	31	42.8	76.8	1.64E-109	32756000
P61490	groL	16	28.4	57.9	1.09E-36	704330
Q746M7	*Tt*Ago	14	19	76.7	1.95E-37	2642000
Q72JL4	TT_C0758	11	26.3	49.3	1.72E-41	2238900
Q72J15	TT_C0966	9	25.3	41.6	3.57E-29	1088200
Q72H68	TT_C1627	7	16.1	48.1	1.06E-19	340990
Q72HX7	TT_C1355	7	8.7	82.5	1.05E-19	272740
Q72GH6	TT_C1872	5	15.9	33.3	4.09E-17	612660
Q72K98	TT_C0549	5	16.3	35.9	3.04E-13	585590
Q72GW4	tuf1	5	12.3	44.8	1.97E-14	497870

b

Peptides matched against *Tt*Ago

Sequence	Mass (Da)	Protein	PEP
AFGASGASLR	935.48248	*Tt*Ago	0.00054319
AQETALALLR	1084.6241	*Tt*Ago	5.97E-06
AVSKPADALR	1026.5822	*Tt*Ago	0.0052086
EGIAYDLVSVR	1220.6401	*Tt*Ago	0.0012876
EIASWIGR	930.49232	*Tt*Ago	0.0065071
LADGLYVPLEDK	1331.6973	*Tt*Ago	0.00097635
LGEEDPK	786.37595	*Tt*Ago	0.013635
LGLGTPEAVR	1011.5713	*Tt*Ago	0.00069792
LYPASGFAFPR	1224.6291	*Tt*Ago	0.021229
MGQNYAYR	1001.4389	*Tt*Ago	0.0050656
SVLSALAR	815.4865	*Tt*Ago	0.0013518
TEVFLNR	877.46577	*Tt*Ago	0.0013027
VAWVADPKDPR	1252.6564	*Tt*Ago	0.023399
VYPVQGR	817.44464	*Tt*Ago	0.033518

a, Only proteins of which 5 or more peptides were discovered are shown. The 'Peptides' column shows how many peptides are matches against a certain protein. **b**, Peptides identified that match the *Tt*Ago sequence. PEP, posterior error probability; iBAQ, intensity-based absolute quantification.

Table S3 | *Tt*Ago preferentially acquires ssDNA guides from plasmid DNA.

DNA locus	Size	Copies per cell	Total DNA per cell	Reads aligned to DNA locus	Normalized reads*	Normalized reads* corrected for DNA per cell
***E. coli* K12 chromosome**	4.64 Mb	1	4.64 Mb	$23*10^6$	1	1
pWUR702	5.6 kb	20-40	0.17 Mb	$45*10^6$	2	54
pRARE	4.7 kb	10-12	52 kb	$2*10^6$	0.1	8.8

Estimated relative quantities of guides complementary to plasmid and chromosome and DNA per cell. *Reads are normalized against the number of reads mapped against the *E. coli* K12 chromosome.

Table S4 | Strains and oligonucleotides used in this study.

a

Strain	Abbreviations	Description	Source, reference
Thermus thermophilus HB27	HB27, wild type	ATCC BAA-163 / DSM 7039 / NBRC 101085	DSMZ
T. thermophilus HB27^{EC}	HB27^{EC}	*ago*::*ago*ISTth7 and multiple mutations, selected for enhanced competence	This study
T. thermophilus HB27 Δ*ago*	HB27Δ*ago*, knockout	Δ*ago*	This study
T. thermophilus HB27 Δ*ago*::*strep(II)-ago*	HB27Δ*ago*::^{s}*ago*	HB27Δ*ago* with *strep(II)-tag-ago* fusion and kanamycin marker insert	This study

b

Experiment	Primers	Sequence (5'-3')	Description, restriction sites
Genomic mutants	BG3524	AAAAAAAAGCTTCCTCAACGGGGAGGTTCCGGA	upstream region *ago* (fw), HindIII
	BG3525	AAAAAAGTCGACGCTCAGATTTGCATAGGAGCTGC	upstream region *ago* (rv), SalI
	BG3526	AAAAAAGTCGACATGGCAAGCTGGAGCCACCCG	*strep(II)-ago* (fw), SalI
	BG3527	AAAAAATCTAGACTAAACGAAGAAGAGCTTTTCCCG	*strep(II)-ago* (rv), XbaI
	BG3528	AAAAAATCTAGATGCCCAAGCGGGGCGGAACC	downstream region *ago* (fw), XbaI
	BG3529	AAAAAAGAATTCGGTCAATCCGCCCCGCTTCCA	downstream region *ago* (rv), EcoRI
	BG3563	GGCCGTCTAGACCCGGGAGTATAACAGAAACCTT	PslpA-Kan^{R}-stop (fw), XbaI
	BG3564	GCGCGTCTAGATCAAAATGGTATGCGTTTTGACAC	PslpA-Kan^{R}-stop (rv), XbaI
Expression vectors *Tt*Ago	AgoFW	GCGCGCGGTACCAGATGAACCACCTTGGAAAAACGG	*T. thermophilus* HB8 *ago* (fw), KpnI
	AgoRV	GCGCGCGCGGCCGCGAATTCCTAAACGAAGAAGAGCTTTTCCC	*T. thermophilus* HB8 *ago* (rv), NotI
	BG4207	GCGCGCACATGTCAAGCTGGAGCCACCCGCAG	*strep(II)-ago* (fw), PciI
	BG4208	GCGCGCCCTAGGTTAATTAGTGGTGGTGATGG	*strep(II)-ago* (rv), AvrII
Site directed mutagenesis of *ago* gene	BG3454	GGCGGAGCTCGCCGTGGGCTTTGCCGCCGGCGGAAGGGAGTCCTTTCG	HB8 *ago* D478A (fw)
	BG3455	CGAAAGGACTCCCTTCCGCCGGCGGCAAAGCCCACGGCGAGCTCCGCC	HB8 *ago* D478A (rv)
	BG3456	CCCGGGTCCTCCTCCTTCGGGCCGGCCGCGTGCCCCAGGACGAG	HB8 *ago* D546A (fw)
	BG3457	CTCGTCCTGGGGCACGCGGCCGGCCCGAAGGAGGAGGACCCGGG	HB8 *ago* D546A (rv)
Guide sequencing	BG4409	GAGAGAGGATCCCGAATTGTGCA**GCT**GTCAATCAACC	5' Amplification primer, BamHI
	BG4436	GAGAGAGGATCCTTTTTTTTTTTTTTTTTTTTTTTTTTVN	3' Poly-T primer with 'VN' anchor, BamHI
Target sequences	BG4262	GGCCATTTAATTAAATTAAAAGCTTGAATGCAATATTTATTTAAAAATTTATA CGAGGTAGTAGGTTGTATAGTATATTAAATTATTTAAATATAAAG	Low GC-content (17%) target oligonucleotide 'FW-target'
	BG4263	TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACCTCGTATA AATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT	Low GC-content (17%) target oligonucleotide 'RV-target'
	BG4264	GGCCAGGTCCACCATGCGTAAGCTTGAATGCCGGCCAGCCCAAGGGCTC TGCACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG	High GC-content (59%) target oligonucleotide 'FW-target'
	BG4265	TCGACGCTTAGACCTACGCCTGCCAGCAACTATACAACCTACTACCTCGTG CAGAGCCCTTGGGCTGGCCGGCATTCAAGCTTACGCATGGTGGACCT	High GC-content (59%) target oligonucleotide 'RV-target'
	BG3467	CTAGACGAGGTAGTAGGTTGTATAGTA	Target sequence insert, XbaI, HindIII
	BG3468	AGCTTACTATACAACCTACTACCTCGT	Target sequence insert, XbaI, HindIII
siDNA and siRNA sequences	BG3466	P-TGAGGTAGTAGGTTGTATAGT	FW-guide, based on *let-7* miRNA
	BG4017	P-TTATACAACCTACTACCTCGT	RV-guide, based on reverse complement of *let-7* miRNA
	BG4500	P-AGAGGTAGTAGGTTGTATAGT	FW-guide, based on *let-7* miRNA, 5'-end deoxyadenosine
	BG4501	P-GGAGGTAGTAGGTTGTATAGT	FW-guide, based on *let-7* miRNA, 5'-end deoxyguanosine
	BG4502	P-CGAGGTAGTAGGTTGTATAGT	FW-guide, based on *let-7* miRNA, 5'-end deoxycytidine
	BG4503	P-TGAGGTAGTAGGTTGTATAGT	FW-guide, based on *let-7* miRNA, 5'-end deoxythymidine
	BG4508	P-UGAGGUAGUAGGUUGUAUAGU	FW-guide, based on *let-7* miRNA

a, *T. thermophilus* strains used. **b**, Oligonucleotides used. Restriction sites are underscored.

Table S5 | Plasmids used in this study.

Plasmid	Description	Restriction sites used	Primers	Source, reference
pRARE	*E. coli* Rosetta (DE3) plasmid, encodes rare tRNAs, Cam^R			Novagen
pET-52b(+)	T7 RNA polymerase based expression vector, Amp^R			Novagen
pWUR627	*T. thermophilus* HB8 *ago* with N-term. *strep(II)-tag* in pET-52b(+) expression vector for TtAgo	KpnI NotI	AgoFW AgoRV	This study
pWUR641	pWUR627, *ago* active site residue substituted (D546A)	-	BG3456 BG3457	This study
pWUR642	pWUR641, *ago* active site residue substituted (D478A) Expression vector for TtAgoDM(D478A,D546A)	-	BG3454 BG3455	This study
pCDF-1b	T7 RNA polymerase based expression vector, Sm^R			Novagen
pWUR702	*strep(II)-ago* insert from pWUR627 inserted in pCDF-1b Expression vector for TtAgo	AvrII NcoI	BG4207 BG4208	This study
pWUR703	*strep(II)-agodm*(D478A,D546A) insert from pWUR642 inserted in pCDF-1b Expression vector for TtAgoDM(D478A,D546A)	AvrII NcoI	BG4207 BG4208	This study
pUC18	Amp^R			Thermo scientific
pUC19	Amp^R			Thermo scientific
pWUR673	2.4kb downstream sequence of *ago* inserted in pUC18	XbaI EcoRI	BG3528 BG3529	This study
pWUR674	1kb upstream sequence of *ago* inserted in pWUR673	HindIII SalI	BG3524 BG3525	This study
pWUR675	*ago* with N-terminal *strep(II)-tag* inserted in pWUR674	SalI XbaI	BG3526 BG3527	This study
pWUR676	Kan^R marker with pSLPa promoter inserted in pWUR675	XbaI XbaI	BG3563 BG3564	This study
pK18	Recombination vector			[245]
pWUR701	Insert from pWUR674 transferred to pK18	HindIII EcoRI		This study
pMHPnqosGFP	*E. coli* / *T. thermophilus* shuttle vector, Hyg^R, sGFP under control of Pnqo promoter			[285]
pMKPnqosGFP	*E. coli* / *T. thermophilus* shuttle vector, Kan^R, sGFP under control of Pnqo promoter			[285]
pMK184	*E. coli* / *T. thermophilus* shuttle vector, Kan^R			[246]
pFU98	pSC101 ori, rbs-luxCDABE, Cam^R			[250]
pWUR677	pFU98, Cam^R marker replaced by Hyg^R marker	SacI NheI	BG3870 BG3871	This study
pWUR704	pWUR677, rbs-luxCDABE replaced by annealed BG4262-BG4263	NotI SalI	BG4262 BG4263	This study
pWUR705	pWUR677, rbs-luxCDABE replaced by annealed BG4264-BG4265	NotI SalI	BG4264 BG4265	This study
pWUR708	pWUR677, rbs-luxCDABE replaced by annealed BG3467-3468 insert	XbaI HindIII	BG3467 BG3468	This study

Chapter 3

Mechanism of DNA-guided DNA target cleavage by Argonaute

Gang Sheng*, Hongtu Zhao*, Jiuyu Wang, Yu Rao, Wenwen Tian,
Daan C. Swarts, John van der Oost, Dinshaw J. Patel, Yanli Wang
*contributed equally

Adapted from:
'Structure-based Cleavage Mechanism of *T. thermophilus*
Argonaute DNA Guide Strand-Mediated DNA Target Cleavage'
Proceedings of the National Academy of Sciences USA. January 2014. Vol. 111, 652-657

Abstract

We report on crystal structures of ternary *Thermus thermophilus* Argonaute (*Tt*Ago) complexes with 5′-phosphorylated guide DNA and a series of DNA targets. These ternary complex structures of cleavage-incompatible, cleavage-compatible, and postcleavage states solved at improved resolution up to 2.2 Å have provided molecular insights into the orchestrated positioning of catalytic residues, a pair of Mg^{2+} cations, and the putative water nucleophile positioned for in-line attack on the cleavable phosphate for *Tt*Ago-mediated target cleavage by an RNase H-like mechanism. In addition, these ternary complex structures have provided insights into protein and DNA conformational changes that facilitate transition between cleavage-incompatible and cleavage-compatible states, including the role of a glutamate finger in generating a cleavage-competent catalytic DEDD tetrad.

Introduction

Argonaute proteins (Agos), critical components of the RNA-induced silencing complex, play a key role in guide-mediated target RNA recognition, cleavage, and product release (reviewed in [171,251,252]). Agos adopt a bilobal scaffold composed of an amino terminal PAZ-containing lobe (N and PAZ domains), a carboxyl-terminal PIWI-containing lobe (MID and PIWI domains), and connecting linkers L1 and L2. Agos bind guides whose 5′-phosphorylated and 3′-hydroxyl ends are anchored within MID and PAZ binding pockets, respectively [28,168,178,179]. The anchored guide then serves as a template for pairing with the target [167,169]. The cleavage activity of Ago resides in the RNase H fold adopted by the PIWI domain [30,180], in which the DDX (where X can be D or H) catalytic triad [18,163,164,253] initially processes loaded double-stranded siRNAs by cleaving the passenger strand, which is released after cleavage. Subsequently the same catalytic tetrad processes guide-target RNA duplexes by cleaving the target strand (reviewed in [172,183,254]). Such Mg^{2+} cation-mediated endonucleolytic cleavage of the target RNA [255,256] resulting in 3′-OH and 5′-phosphate ends [257] requires Watson–Crick pairing of the guide and target spanning the seed segment (guide nucleotides 2-7 or 2-8) [175] and the cleavage site (between target nucleotide 10' and 11') [169]. Insights into target RNA recognition and cleavage have emerged from structural [169], chemical [258], and biophysical [188] experiments.

Notably, prokaryotic Ago proteins (pAgos) have recently been shown to preferentially bind 5′-phosphoryated guide DNA [163,164] and use an activated water molecule as the nucleophile (reviewed in [259]) to cleave both RNA and DNA targets [169]. Structural studies have been undertaken on pAgos in the free state [30,164] and bound to a 5′-phosphorylated guide DNA [168] and added target RNA [167,169]. The structural studies of *Thermus thermophilus* Ago (*Tt*Ago) ternary complexes have provided insights into the nucleation, propagation, and cleavage steps of target RNA cleavage [169]. These studies have highlighted the conformational transitions on proceeding from Ago in the free state to the binary complex [168] to the ternary complexes [167,169] and have emphasized the requirement for a precisely aligned DDD triad and a pair of Mg^{2+} cations for cleavage chemistry [169], typical of RNase H fold-mediated enzymes [259,260]. Structural studies have also been extended to binary complexes of both human [28,178] and yeast [179] Agos bound to 5′-phosphorylated guide RNAs.

Despite these singular advances in the structural biology of RNA silencing, further progress was hampered by the modest resolution (2.8-Å to 3.0-Å resolution) of *Tt*Ago ternary complexes with guide DNA [168] and added target RNAs [167,169]. This precluded identification of water molecules coordinated with the pair of Mg^{2+} cations, including the key water molecule that acts as a nucleophile and targets the cleavable phosphate between target nucleotides 10' and 11'. We have now extended our research to *Tt*Ago ternary complexes with guide DNA and target DNAs, which has permitted us to grow crystals of ternary complexes that diffract to higher (2.2–2.3 Å) resolution in the cleavage-incompatible, cleavage-compatible, and postcleavage stages. These high-resolution structures of *Tt*Ago ternary complexes provide snapshots of distinct key steps in the catalytic cleavage pathway, opening opportunities for experimental probing into DNA target cleavage as a defense mechanism against plasmids and possibly other mobile elements [27,177].

Results

***Tt*Ago ternary complexes with short target DNA are in a cleavage-incompatible state**

We have solved the 2.8-Å crystal structure of the ternary complex of *Tt*Ago bound to 12-mer target DNA (**Fig. 1a**, **b**, **Table S1**). The guide DNA, of which nucleotides 1 to 12 and 20 to 21 can be traced, is anchored at both ends, with the 5′ end anchored in the MID pocket [163,165] and the 3′ end in the PAZ pocket [186,187]. The target DNA, of which nucleotides 1′ to 12′ can be traced, base pairs with the guide spanning nucleotides 2–2′ to 12–12′, thereby encompassing both the seed segment (guide nucleotides 2 to 8) and the cleavage site (target nucleotides 10' and 11'). Terminal base pair 2–2′ of the guide-target duplex is stacked over the side chains of R446 and H445 with bases 1 of the guide and 1′ of the target splayed out relative to the 2–2′ base pair and positioned in separate pockets in the MID and PIWI domains, respectively (**Fig. 1c**). The 5' phosphate of the guide [163,180] and the 5'-terminal base (T1) are bound in the MID binding pocket, the latter with sequence specific interactions (**Fig. 1d**, **Fig. S1a**) [169,182,226]. The 3'-terminal base (G1') of the target is bound in the PIWI binding pocket with sequence specific interactions (**Fig. 1e**, **Fig. S1b**).

The recent structural studies on the binary complex of yeast Ago with a bound 5′-phosphorylated guide RNA established that the catalytic pocket is made up of three aspartic acid residues, with an inserted glutamic acid residue completing a catalytic tetrad reflective

of formation of a cleavage-compatible state [179]. In the current structure of the *Tt*Ago ternary complex with 12-mer target DNA, the highly conserved E512 is unplugged and far from the catalytic pocket made up of D478, D546, and D660 (**Fig. 1b**, **f**), indicative of the formation of a cleavage-incompatible conformation at the 12-mer target DNA level.

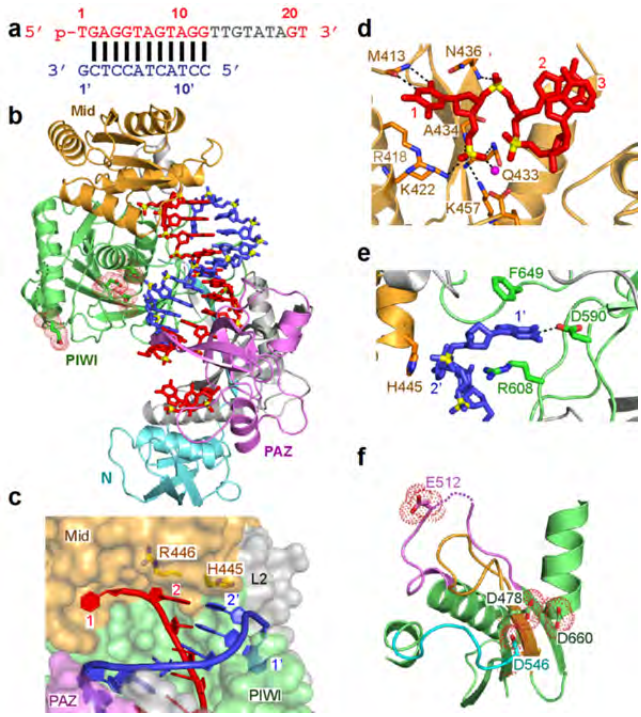


Figure 1 | Crystal structure and interactions in the *Tt*Ago ternary complex with 5′-phosphorylated 21-mer guide DNA and 12-mer target DNA. a, The sequence and pairing of guide (red) and target (blue) in the ternary complex. Disordered segments are shown in gray. **b**, 2.9-Å crystal structure of the complex. The various domains and linkers of *Tt*Ago are color-coded, as are the guide and target. The catalytic residues in a stick representation are highlighted in a red-dotted background. **c**, view of the guide-target segment highlighting splaying out of guide and target bases 1 and 1′ in their respective MID and PIWI pockets in the complex. The 2–2′ base pair stacks over side chains of R446 and H445. **d**, Positioning of the 5′-phosphate and sequence-specific recognition of splayed-out guide base T1 in the MID pocket. **e**, Positioning and sequence-specific recognition of the splayed-out target base G1′ within a pocket in the PIWI domain. **f**, E512 is unplugged and far away from the catalytic pocket composed of D478, D546, and D660 residues in the ternary complex.

The previously determined *Tt*Ago ternary complex with 12-mer target RNA [167] was compared with the current *Tt*Ago ternary complex with 12-mer target DNA. The *Tt*Ago-bound DNA–RNA (in gold) and DNA–DNA (in blue) duplexes superpose quite well (**Fig. S1c**) with both the DNA–DNA duplex and the DNA–RNA duplex adopting helical conformations closer to the canonical A form (**Fig. S2a**) than to the B form (**Fig. S2b**). Similarly, the *Tt*Ago proteins in the two complexes (same color code) also superpose reasonably well (**Fig. S1d**). Both proteins adopt the cleavage-incompatible conformation, with small differences restricted to the positions of the PAZ domain (**Fig. S1d**, indicated by a red arrow).

***Tt*Ago ternary complex with a long target DNA are in a cleavage-compatible state**

We have solved the 2.2-Å crystal structure of the ternary complex of *Tt*Ago bound to a 19-mer target DNA (**Fig. 2a**, **Table S2**). There are two molecules of the complex in the asymmetric unit with the target DNA cleaved (between target nucleotides 10′ and 11′ with retention of the duplex on both sides of the cleavage site) in one molecule, whereas the density in the other molecule could be fit only on invoking a mixture of intact and cleaved target DNA between target nucleotides 10′ and 11′. In this section, we focus on the structure of the ternary complex containing an intact 19-mer DNA target (**Fig. 2b**).

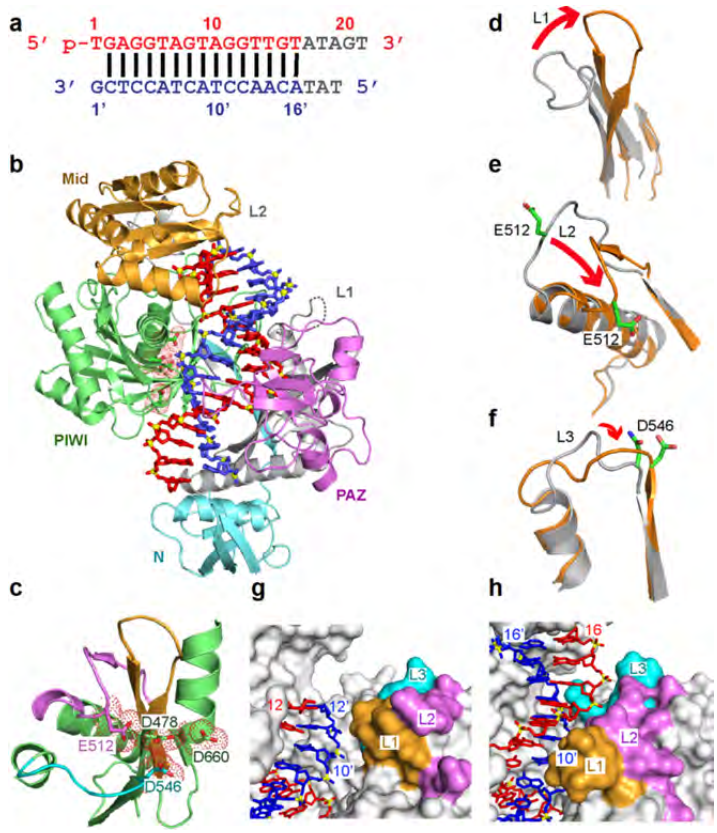


Figure 2 | Crystal structure and interactions in the *Tt*Ago ternary complex with 5′-phosphorylated 21-mer guide DNA and 19-mer target DNA. a, The sequence and pairing of guide (red) and target (blue) in the ternary complex. **b**, 2.2-Å crystal structure of the complex. **c**, Insertion of E512 residue into the catalytic pocket (D478, D546, and D660 residues) in the ternary complex. **d-f**, Conformational changes in loop PL1 (panel d), in loop PL2 that contains residue E512 (panel e), and in loop PL3 that contains residue D546 (panel f) on proceeding from the cleavage-incompatible ternary complex with 12-mer target DNA to the cleavage-compatible ternary complex with 19-mer target DNA. **g**, **h**, Relative positioning of PL1 (in gold), PL2 (in magenta), and PL3 (in cyan) in a surface representation on proceeding from the cleavage-incompatible ternary complex with 12-mer target DNA (panel g) to the cleavage-compatible ternary complex with 19-mer target DNA (panel h).

The guide DNA (red), which can be traced from nucleotides 1 to 16, is anchored at its 5′-end in the MID pocket, but the 3′ end is positioned too far from the PAZ pocket for insertion (nucleotides 17 to 21 are disordered). The target DNA (blue) can be traced from nucleotides 1′ to 16′, with guide-target pairing spanning nucleotides 2–2′ to 16–16′ within the nucleic acid-binding channel. This duplex also adopts a helical conformation closer to the canonical A form (**Fig. S2e**) than to the B form (**Fig. S2f**). The previously determined *Tt*Ago ternary complex with a 19-mer target RNA [169] was compared with the current *Tt*Ago ternary complex with a 19-mer target DNA. The DNA–RNA and DNA–DNA duplex segments superpose reasonably well (**Fig. S3a**), as do the *Tt*Ago proteins in the two complexes, except for small differences in the PAZ domain (**Fig. S3b**).

Relative to the complex with a 12-mer target DNA, an important difference in the *Tt*Ago ternary complex with a 19-mer target DNA is the transition of residue E512 over a distance of 12.8 Å into the catalytic pocket to form a tetrad with the three catalytic aspartic acid residues (**Fig. 2b**, **c**), thereby representing a cleavage-compatible conformation. We observed conformational changes on proceeding from the cleavage-incompatible ternary complex with a 12-mer target DNA to the cleavage-compatible complex with a 19-mer DNA (**Fig. S4**). Specifically, large structural transitions are observed within all three PIWI loops (PL1-3, **Fig. 2d-f**), with residue E512 positioned on PL2 which is inserted into the catalytic pocket (**Fig. 2e**). The repositioning of the three PIWI loops relative to the cleavage site between target nucleotides 10' and 11' is clearly visible in the surface representations of the three loops (PL1 in gold, PL2 in magenta, and PL3 in cyan) on proceeding from the cleavage-incompatible (**Fig. 2g**) to the cleavage-compatible (**Fig. 2h**) states. Moreover, additional protein-DNA intermolecular interactions are observed with both the guide and target on proceeding from the cleavage-incompatible state (**Fig. S5a**, **b**) to the cleavage-compatible state (**Fig. 3a**). In essence, release of the 3′ end of the guide from the PAZ pocket during the propagation step is accompanied by conformational transitions in PL1, PL2, and PL3 with PL2 residues, especially residue E512, forming stabilizing interactions on formation of the plugged-in conformation. Such stabilization provides the driving force to shift residue E512 from the unplugged to the plugged-in conformation, thereby positioning it in the catalytic pocket to complete tetrad formation.

We observe two hydrated Mg^{2+} cations (labeled A and B) that bridge between the three catalytic aspartic acid residues and the cleavable but intact phosphate at the 10′–11′ step of

the target (**Fig. 3b**) in the cleavage-compatible state of the ternary complex with 19-mer target DNA. Mg^{2+} cation A is directly coordinated with residues D478 and D660 to a non-bridging phosphate oxygen and three water molecules, one of which is poised for in-line attack on the backbone phosphate at the cleavage site (**Fig. 3b**, indicated by a red arrow). Mg^{2+} cation B is directly coordinated with residues D478 and D546, with one each of bridging and non-bridging phosphate oxygens at the cleavage site, and with two water molecules (**Fig. 3b**). Notably, the carboxylate oxygens of the inserted catalytic residue E512 are not directly coordinated to either divalent cation, but rather use a pair of bridging water molecules to coordinate with Mg^{2+} cation B (**Fig. 3b**). Given the 2.2-Å resolution of this complex, we can readily trace the Mg^{2+}-coordinated water molecules, including the proposed nucleophilic water molecule within omit maps showing the positions of water molecules.

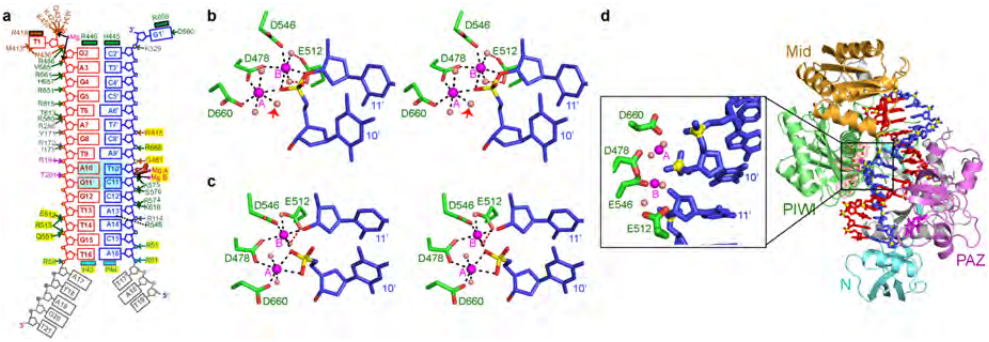


Figure 3 | *Tt*Ago ternary complexes with 5′-phosphorylated 21-mer guide DNA and added 19-mer target DNA in the presence of Mg^{2+} and Mn^{2+} containing solution. a, Intermolecular contacts in the 2.2-Å ternary complex with cleavage-compatible 19-mer target DNA in Mg^{2+}-containing solution. The interactions highlighted by a yellow background are additional contacts observed beyond those observed in the ternary complex with cleavage-incompatible 15-mer target DNA (Fig. S5b). **b**, **c**, Stereoview of the catalytic pocket in the ternary complex with an intact (panel b) and cleaved (panel c) DNA target in Mg^{2+}-containing solution. The four catalytic D478, D546, D660, and E512 are shown in stick representation. The pair of Mg^{2+} cations are labeled "A" and "B" and are shown as magenta spheres. Water molecules are shown as pink spheres. The water molecule poised for in-line attack on the backbone phosphate is indicated with a red arrow. **d**, 2.4-Å crystal structure and interactions in the *Tt*Ago ternary complex with 5′-phosphorylated 21-mer guide DNA and complementary cleaved 19-mer target DNA for crystals grown in Mn^{2+} containing solution. The inset expands the catalytic pocket segment showing the cleavage of the backbone.

After cleavage of the target between nucleotides 10' and 11', the divalent metals remain bound by *Tt*Ago (**Fig. 3c**). Mg^{2+} cation A is coordinated with D478 and D660 to a pair of

non-bridging phosphate oxygens of the newly generated 5′-phosphate and two water molecules. Mg^{2+} cation B is coordinated with D478 and D546 to a non-bridging oxygen of the newly generated 5′-phosphate and the oxygen of the newly generated 3′-OH group and two water molecules. The inserted catalytic residue E512 remains coordinated with Mg^{2+} cation B through two bridging water molecules (**Fig. 3c**). The coordination geometries around Mg^{2+} cations A and B for the intact and cleaved ternary complex are shown in **Fig. S6**. We have superposed the structures encompassing the catalytic pocket of *Tt*Ago ternary complexes containing intact (in gold) and cleaved (blue) DNA target sites with the superposition shown in stereo in **Fig. S7**. The ternary complex of *Tt*Ago bound to 5′-phosphorylated 21-mer guide DNA and cleaved 19-mer target DNA is shown in **Fig. 3d**, with the cleavage-site segment magnified in the inset. Importantly, we observed intact guide-target duplex segments on either side of the cleaved target DNA.

We also grew crystals of the *Tt*Ago ternary complex with added 19-mer target DNA in Mn^{2+}-containing solution under two different conditions. Our choice of Mn^{2+} reflected the preference for Mn^{2+} over Mg^{2+} as a divalent cation for cleavage [27,164]. The 2.4-Å crystals of the ternary complex grown in Mn^{2+}-containing solution yielded a structure (**Table S3**, **Fig. 3d**) similar to that reported above for the cleaved structure of the ternary complex in Mg^{2+}-containing solution. However, when the ternary complex in the Mn^{2+}-containing solution was incubated at 55 °C for 10 min before setting up crystallization trays, not only did the 3.0-Å crystal structure of this complex (**Table S3**) show cleavage of the target DNA between target nucleotides 10' and 11', also duplex formation was retained only on the side containing the seed segment, with the guide also disordered beyond nucleotide 10.

Comparison of *Tt*Ago ternary complexes with 15-mer target DNA and RNA

The 2.25-Å crystal structure of the ternary complex of *Tt*Ago bound to 5′-phosphorylated 21-mer guide DNA and 15-mer target DNA (**Fig. 4a**; **Table S1**) is shown in **Fig. 4b**. The guide DNA (red), of which nucleotides 1 to 14 and 20 to 21 can be traced, is anchored at both ends. The target DNA (blue), of which nucleotides 1' to 14' can be traced, base pairs with the guides spanning nucleotides 2–2′ to 14–14′. Importantly, in the *Tt*Ago ternary complex with 15-mer target DNA, 13 base pairs are formed. The 3′ end of the guide is retained within the PAZ pocket and residue E512 is unplugged and far from the catalytic pocket (**Fig. 4c**), indicative of a cleavage-incompatible conformation. By contrast, in the *Tt*Ago ternary complex with 15-mer target RNA 14 base pairs are formed. The 3′ end of the

guide is released from the PAZ pocket [169] and residue E512 is inserted into the catalytic pocket (**Fig. 4d**), indicative of a cleavage-compatible conformation. Thus, it appears that a 15-mer target represents the tipping point during the propagation step between cleavage-incompatible and cleavage-compatible conformations, with DNA targets of this length being cleavage-incompatible and their RNA counterparts adopting cleavage-compatible conformations. Both DNA-RNA and DNA-DNA duplexes bound by *Tt*Ago adopt A-like conformations (**Fig. 4e**).

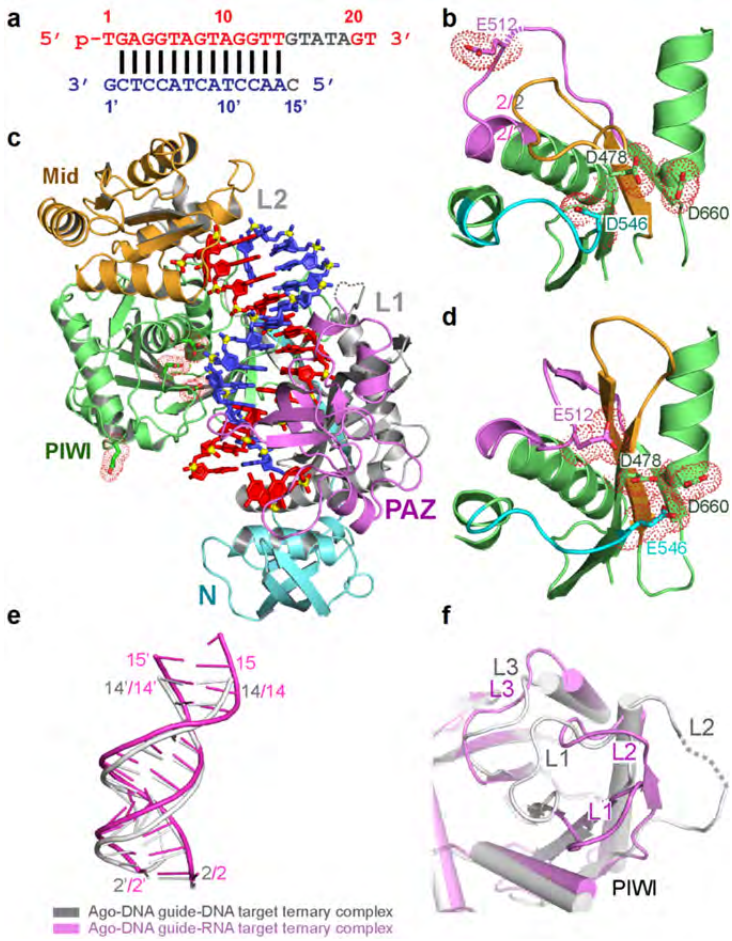


Figure 4 | Crystal structure and interactions in the *Tt*Ago ternary complex with 5′-phosphorylated 21-mer guide DNA and complementary 15-mer target DNA. **a**, The sequence and pairing of DNA guide (red) and DNA target (blue) in the ternary complex. **b**, 2.25-Å crystal structure of the complex. **c**, Positioning of residue E512 unplugged and far away from the catalytic pocket in the ternary complex with 15-mer target DNA. **d**, Insertion of residue E512 into the catalytic pocket in the ternary complex with 15-mer target RNA (PDB 3HJF). **e**, Superposition of the guide DNA-15-mer target DNA (silver) and guide DNA-15-mer target RNA (magenta) in the *Tt*Ago ternary complexes. Note that we observed one more base pair (15–15′) in the ternary complex with the target RNA. **f**, Superposition of the catalytic pockets and PL1, PL2, and PL3 in the ternary complexes with target DNA (silver) and target RNA (magenta).

The *Tt*Ago ternary complexes formed with 15-mer target DNA and 15-mer target RNA are superposed in **Fig. S8**, with the observed conformational changes in the PAZ and N-domains and the catalytic pocket reflective of transitions associated with conversion from cleavage-incompatible to cleavage-compatible conformations. The conformational changes of PL1, PL2, and PL3 within the catalytic pocket associated with this transition are shown in **Fig. 4f**. At this point we are unable the reason of difference in *Tt*Ago behavior between 15-mer DNA and RNA targets other than the one base pair extended duplex in the *Tt*Ago-DNA guide complex with the 15-mer target RNA. We also solved the 2.5-Å structure of *Tt*Ago bound to 5′-phosphorylated guide DNA and 16-mer target DNA (**Table S2**). This ternary complex formed a cleavage-compatible conformation involving formation of a 15 base pair-long guide-target duplex similar to the ternary complex observed with the 19-mer target DNA. Thus, for *Tt*Ago ternary complexes with DNA targets, the switch from cleavage-incompatible to cleavage-compatible complexes occurs on transition from 15-mer to 16-mer target DNA pairing with the guide DNA during the propagation step.

Discussion

The PIWI domain of Ago/PIWI proteins has been shown to adopt an RNase H fold as first established from structural studies of *Pyrococcus furiosus* Ago [30] and *Archaeoglobus fulgidus* PIWI [180] in the free state. The catalytic mechanism of RNase H folds has been investigated in detail (reviewed in [259]) with the emphasis on factors that contribute to substrate specificity and the key contribution of a pair of Mg^{2+} cations to cleavage chemistry [261]. The pair of Mg^{2+} cations, which are positioned to bridge the nucleic acid substrate and enzyme catalytic residues when properly aligned, have been proposed to greatly enhance substrate recognition and product release, thereby enhancing catalytic efficiency. Furthermore, the scissile phosphate group projects three coordination ligands to the pair of Mg^{2+} cations in the enzyme-substrate complex. Of the pair of Mg^{2+} cations in the catalytic pocket, cation A assists in nucleophilic attack by activating a water molecule for in-line attack on the cleavable phosphate, whereas cation B stabilizes the pentacovalent intermediate and facilitates the protonation of the 3′ oxyanion-leaving group by a water molecule (reviewed in [259]).

Previous studies of *Tt*Ago complexes with 5′-phosphorylated guide DNA and complementary target RNAs yielded structures of ternary complexes that diffracted at best

to a 2.6-Å resolution. Thus, although we were able to identify the position of Mg^{2+} cations, we were unable to identify the water molecules coordinated with these Mg^{2+} cations and hence were limited in our efforts at deducing insights into the mechanism of cleavage chemistry [169]. Nevertheless, inspired by previously reported gel-based cleavage studies of *Tt*Ago ternary complexes using guide and target DNAs [169], as well as *in vivo* demonstrated *Tt*Ago-mediated DNA interference [27], we have attempted to improve the diffraction quality of our *Tt*Ago ternary complexes by switching from target RNAs to their target DNA counterparts. This approach has yielded higher-resolution crystals of ternary complexes that diffract to 2.25 Å (with 15-mer target DNA; **Fig. 4b**) and 2.2 Å (with 19-mer target DNA; **Fig. 2b**), thereby allowing us to identify Mg^{2+}-coordinated water molecules, including the one coordinated with Mg^{2+} cation A and positioned for in-line nucleophilic attack on the cleavable phosphate (**Fig. 3b**). Furthermore, the set of structures of ternary complexes solved in this study provide snapshots of cleavage-incompatible, cleavage-compatible, and postcleavage states, as well as the conformational transitions required for formation of a cleavage-compatible pocket.

Features of the catalytic cleavage mechanism of Ago ternary complexes with guide and target DNAs are summarized in **Fig. 5**. In the cleavage-incompatible conformation of the ternary complex with 12- and 15-mer target DNAs, the three catalytic residues D478, D546, and D660 are not properly positioned relative to the cleavable phosphate, with residue E512 unplugged and far from the catalytic pocket and an absence of a pair of Mg^{2+} cations (**Fig. 5a**). The transition from a complex with 15-mer target DNA (guide-target duplex of 13 base pairs) to a complex with 16-mer target DNA (guide-target duplex of 15 base pairs) during the propagation step results in the release of the 3′ end of the guide from the PAZ pocket. The accompanying conformational changes both in the guide and in PL1, PL2, and PL3 position catalytic residue E512 for insertion into the catalytic pocket. In this process, a cleavage-compatible precleavage state with components held in optimal position for cleavage by two bound Mg^{2+} cations is formed (**Fig. 5b**). The coordination geometries of Mg^{2+} cations A and B with aspartic acid side chains, the phosphate oxygens, and coordinated waters, including the nucleophilic water poised for attack on the cleavable phosphate (**Fig. 5b**), in the precleavage structure of the complex are those characteristic of RNase H enzymes (reviewed in [259]). The pair of Mg^{2+} cations is separated by 3.7 Å with octahedral-like coordination geometries for Mg^{2+} cations A and B (**Fig. S6a**). One difference is that catalytic residue E512 is coordinated with Mg^{2+} cation B through a pair of

bridging water molecules in our *Tt*Ago ternary complexes (**Fig. 5b**), in contrast with direct coordination with Mg^{2+} cation B in canonical RNase H enzymes (reviewed in [259]). Importantly, the nucleophilic water, the cleavable phosphate, and the O3 phosphate-leaving group are positioned for in-line attack to generate the proposed pentacovalent phosphate transition state associated with an SN2 reaction with inversion of stereochemistry at the cleavable phosphate position (**Fig. 5c**).

The structure of the postcleavage state of the ternary complex is shown in **Fig. 5d**. Here again, the coordination geometries of Mg^{2+} cations A and B with aspartic acid side chains, the phosphate oxygens, and coordinated waters are those characteristic of RNase H enzymes in the postcleavage state (reviewed in [259]). The pair of Mg^{2+} cations is separated by 3.6 Å, such that Mg^{2+} cation B moves toward A, with octahedral-like coordination geometries for Mg^{2+} cations A and B (**Fig. S6b**). In our postcleavage structures of the ternary complex, catalytic residue E512 remains coordinated through two bridging water molecules to Mg^{2+} cation B.

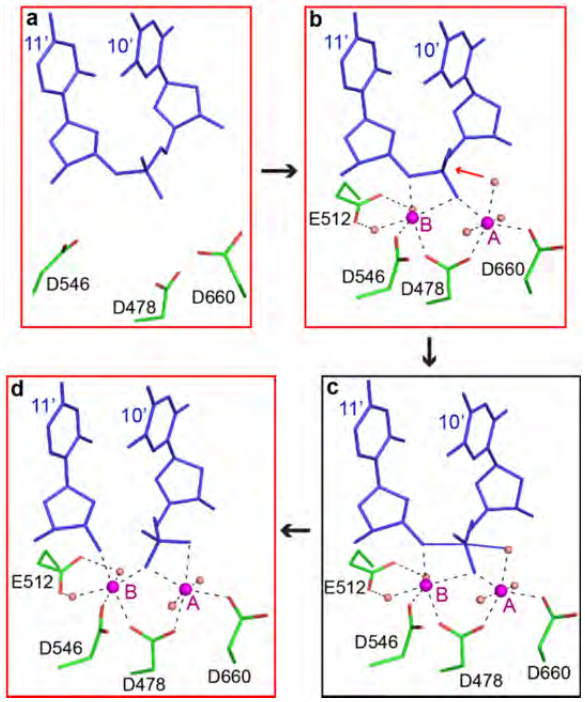


Figure 5 | Proposed mechanism for *Tt*Ago-mediated Mg^{2+} cation-dependent cleavage of targets. Crystal structure snapshots and a proposed model of transition state in the reaction pathway leading to cleavage of the target DNA between nucleotides 10′ and 11′. **a**, Structure of the catalytic pocket in the cleavage-incompatible ternary complex with residue E512 unplugged and far from the catalytic pocket as observed in ternary complexes with 12- and 15-mer target DNAs. **b**, Structure of the catalytic pocket in the cleavage-compatible ternary complex with residue E512 inserted into the catalytic pocket as observed in the ternary complex with 16- and 19-mer target DNAs. The red arrow indicates the nucleophilic water poised for attack on the cleavable phosphate. **c**, Proposed model of the transition state of the cleavage reaction in the ternary complex. **d**, Structure of the catalytic pocket of the ternary complex following cleavage of the backbone between target nucleotides 10′ and 11′ in the ternary complex with cleaved 19-mer target DNA.

Previous *in vitro* biochemical assays ([27,169], *in vivo* assays [27], and the current structural studies highlight the ability of pAgos to use guide DNA to cleave target DNA. This allows prokaryotes to utilize pAgos in DNA interference pathway as a defense mechanism against transposons and mobile genetic elements, first proposed from bioinformatic studies of pAgos [4] and recently receiving support from molecular biology experiments [27,177].

Acknowledgements

We thank the staff at beamline BL-17U at Shanghai Synchrotron Radiation Facility, beamline 3W1A at Beijing Synchrotron Radiation Facility, and beamline NE-CAT ID-24C at the Advanced Photon Source, Argonne National Laboratory. This research was funded by Chinese Ministry of Science and Technology (2011CBA01105), the Natural Science Foundation of China (31222014 and 31170705), and the Chinese Academy of Sciences (KJZD-EW-L01) (to YW); and National Institutes of Health Grant AI068776 (to DJP).

Author contributions

YW designed research; GS, HZ, JW, YR, WT, and YW performed research; GS, JW, YR, and WT undertook the crystallization; HZ and YW collected the synchrotron data; YW solved the structures; DCS and JvdO contributed insights on *Tt*Ago-mediated DNA guide-mediated DNA target cleavage; GS, DCS, JvdO, DJP, and YW analyzed data; and DJP and YW wrote the paper. All authors read and approved the manuscript.

Author information

Correspondence should be addressed to ylwang@ibp.ac.cn or pateld@mskcc.org.

Data deposition

Tables S1-S3 are available in the online version of the published manuscript [176]. The following structures have been deposited in the Protein Data Bank, www.pdb.org: *Tt*Ago ternary complexes with 5′-phosphorylated 21-mer and added 12-mer target DNA (4N47), 15-mer target DNA (4N41), 16-mer target DNA (4NCA), and 19-mer target DNA (4NCB) for crystals grown in Mg^{2+}-containing solution; *Tt*Ago ternary complexes with 5′-phosphorylated 21-mer and added 19-mer target DNA without (4KPY) and with (4N76) preheating for 10 min at 55 °C in Mn^{2+}-containing solution.

Experimental procedures

Preparation and Purification of *Tt*Ago-guide DNA-target DNA Complexes

T. thermophilus Ago (*Tt*Ago) was prepared and purified as described previously [168]. Oligodeoxyribonucleotides were purchased from Invitrogen and GenScript. For crystallization, *Tt*Ago was mixed with 5'-phosphorylated 21-mer guide DNA at 1:1.2 molar ratios, followed by addition of different length target DNAs at a 1.0 molar ratio to the binary mixture, to form the ternary complex. 20 mM $MgCl_2$ or 4 mM $MnCl_2$ were added to the complex before setting up crystallization trays.

Crystallization and Data Collection

All crystals of wild-type *Tt*Ago-guide DNA0-target DNA were grown at 33 °C using either the hanging-drop or sitting-drop vapor diffusion methods from drops containing 1 µl ternary complex and 1 µl of reservoir solution and flash-cooled in liquid nitrogen. Crystals of *Tt*Ago ternary complexes containing 21-mer guide DNA and varying length target DNAs were grown in the presence of 20 mM $MgCl_2$. Crystals of the ternary complex of *Tt*Ago in complex with 12-mer target DNA were grown in a reservoir containing 0.2 M $MgCl_2$, 0.1 M Tris pH 8.5, 25% PEG3350, and were subsequently cryoprotected with the reservoir solution, followed by flash-cooling in liquid nitrogen. The crystals belong to space group $P2_12_12_1$, and there is one *Tt*Ago ternary complex in the asymmetric unit. Crystals of the ternary complex of *Tt*Ago in complex with 15-mer target DNA were grown in a reservoir containing 3.2 M NaCl, 0.1 M Tris pH 8.5, and were subsequently cryoprotected with the reservoir solution and 15% glycerol and flash-cooled in liquid nitrogen. The crystals belong to space group $P2_1$, and there are two ternary *Tt*Ago complexes in the asymmetric unit. Crystals of the ternary complex of *Tt*Ago in complex with 19-mer target DNA were obtained in a reservoir containing 18 mM $MgCl_2$, 0.05 M Na cacodylate pH 6.5, 9% isopropanol and 2.25 mM spermine, and were subsequently cryoprotected with the reservoir solution and 25% glycerol and flash-cooled in liquid nitrogen. These crystals belong to space group $P2_1$, and there are two *Tt*Ago ternary complexes in the asymmetric unit.

Given that *Tt*Ago exhibited higher cleavage activity in the presence of Mn^{2+} compared to Mg^{2+} [27], we added 4 mM $MnCl_2$ to the ternary complex instead of 20 mM $MgCl_2$. Crystals of ternary complexes of wild-type *Tt*Ago bound to 19-mer target DNA in Mn^{2+}-containing solution were obtained with hanging-drop vapor diffusion method in a reservoir containing

2.6 M Na-acetate and 0.1 M Bis-Tris pH 7.0. In addition, the same *Tt*Ago ternary complex in Mn^{2+}-containing solution was heated for 10 min at 55 °C prior to setting up crystallization trays and the crystal was obtained in a reservoir containing 2.8 M NaCl and 0.1 M Hepes pH 7.0. The crystals under both conditions belong to space group $P2_12_12_1$, with two ternary *Tt*Ago complexes in the asymmetric unit. Diffraction data were collected at 100 K on beam line BL-17U at Shanghai Synchrotron Radiation Facility (SSRF), beam line 3W1A at Being Synchrotron Radiation Facility (BSRF), and beam line NE-CAT ID-24C at the Advanced Photon Source (APS), Argonne National Laboratory. All data sets were integrated and scaled with the HKL2000 suite (2) and data processing statistics are summarized in **Tables S1**-**S3**.

Structure Determination and Refinement

The structures of the *Tt*Ago-guide DNA target DNA ternary complexes were solved by molecular replacement with the program PHASER [262]. The domains of the ternary complex of *Tt*Ago bound to guide DNA and target RNA (PDB ID: 3HK2) without the linkers, were used as search models. Model building was done using COOT [263], and refinement was done with PHENIX [264]. The final figures were created with PyMOL (http://pymol.sourceforget.net/). The refinement statistics for the ternary complexes are summarized in **Tables S1**-**S3**. During refinement of the structure of the *Tt*Ago ternary complex with bound 19-mer target DNA, it became apparent that one of the two molecules in the asymmetric unit represented a mixture of intact and cleaved target strands. Thus, the Fo-Fc map contoured at 3.5σ based on the assumption of an intact target strand gave unaccounted for density (**Fig. S10a**), reflective of a cleaved phosphate backbone. Similarly, the Fo-Fc map contoured at 3.5σ based on the assumption of a cleaved target strand gave unaccounted for density between sugar and phosphate group (**Fig. S10b**), reflective of an intact phosphate backbone. It was only by refining using a 1:1 mixture of intact and cleaved phosphate backbones that no extra Fo-Fc density was observed (**Fig. S10c**). To identify the position of the nucleophilic water and the pair of Mg^{2+} cations in the *Tt*Ago ternary complex with bound 19-mer target DNA, we recorded Fo-Fc omit maps that identified the nucleophilic water (**Fig. S10d**) and identified the pair of Mg^{2+} cations (**Fig. S10e**).

Supplementary Figures

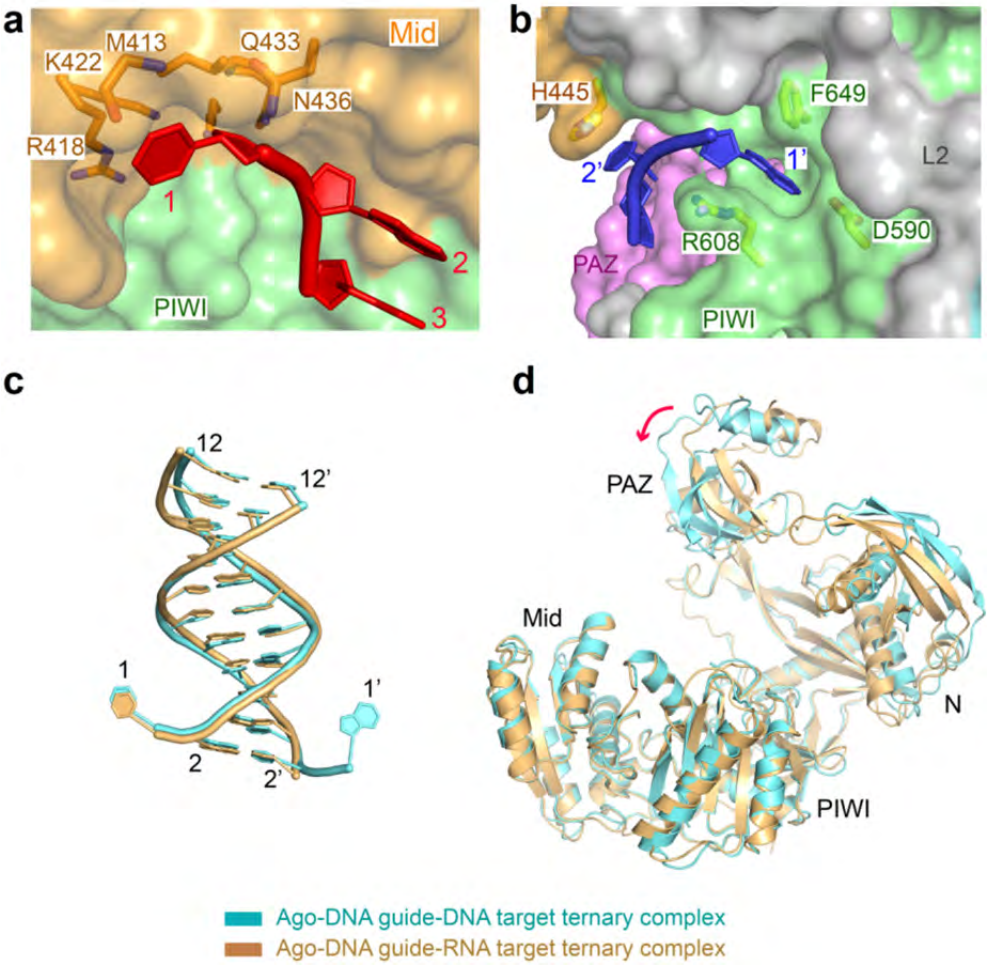


Figure S1 | Positioning of bases 1 and 1' in MID and PIWI pockets respectively in the *Tt*Ago ternary complex with 5'-phosphorylated 21-mer guide DNA and complementary 12-mer target DNA, and comparison of structures of ternary complexes of *Tt*Ago containing 5'-phosphorylated 21-mer guide DNA bound to 12-mer target DNA versus 12-mer target RNA. a, Positioning of the 5' phosphate and base T1 of the guide (stick representation) in the *Tt*Ago MID pocket (surface representation). **b**, Positioning of the base G1' of the target strand (stick representation) in the *Tt*Ago PIWI pocket (surface representation). **c**, Superposition of guide-target duplex from 2-2' to 12-12' in the ternary *Tt*Ago complexes with target RNA (gold; PDB: 3H01) and with target DNA (blue). **d**, Superposition of *Tt*Ago in ternary complexes with target RNA (gold; PDB: 3H01) and with target DNA (blue). Observed differences between the PAZ domains of the structures are indicated with the red arrow.

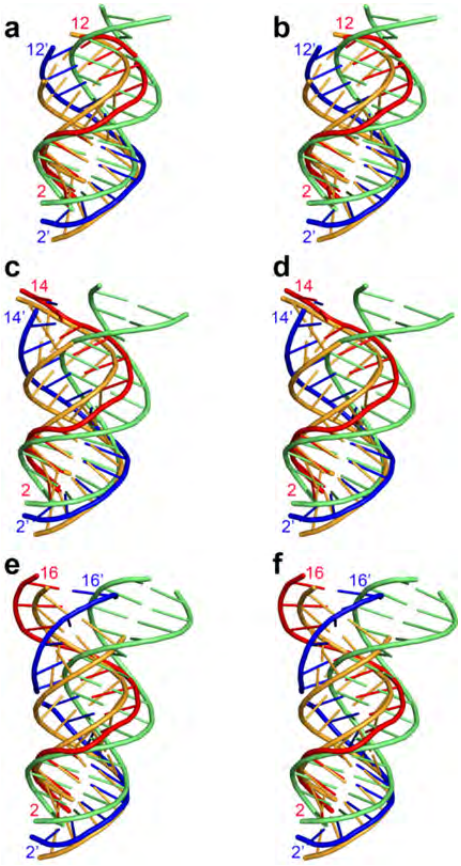


Figure S2 | Comparison of guide-target duplex structures in *Tt*Ago ternary complexes with 5'-phosphorylated 21-mer guide DNA and 12-, 15-, and 19-mer target DNAs with canonical A- and B-form helices. **a**, **b**, Superposition of the guide-target duplex (red-blue) in ternary complex with 12-mer target DNA with canonical A-form DNA (orange; panel **a**) and B-form DNA (green; panel **b**) helices. **c**, **d**, Superposition of the guide-target duplex (red-blue) in ternary complex with 15-mer target DNA with canonical A-form DNA (orange; panel **c**) and B-form DNA (green; panel **d**) helices. **e**, **f**, Superposition of the guide-target duplex (red-blue) in ternary complex with 19-mer target DNA with canonical A-form DNA (orange; panel **e**) and B-form DNA (green; panel **f**) helices. The superposition in panels **a** to **f** was done for the duplex segment spanning the seed segment (base pairs 2-2' to 8-8').

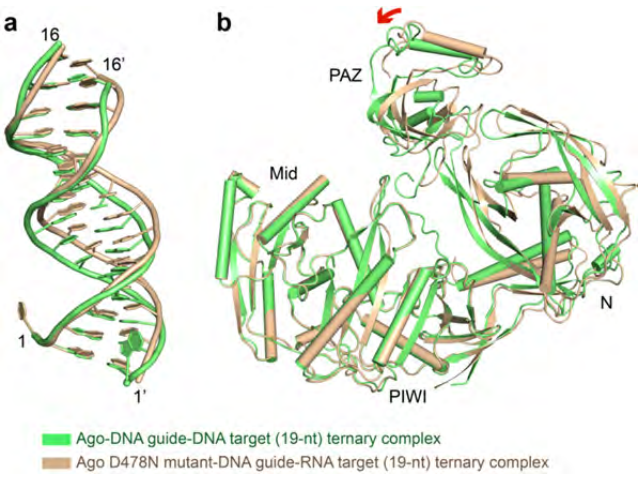


Figure S3 | Comparison of structures of ternary complexes of *Tt*Ago containing 5'-phosphorylated 21-mer guide DNA bound to 19-mer target DNA versus 19-mer target RNA. **a**, Superposition of the guide-target duplex from 2-2' to 16-16' in the ternary *Tt*Ago complexes with target RNA (gold) and with target DNA (green). **b**, Superposition of *Tt*Ago in ternary complexes with target RNA (gold) and with target DNA (green). Observed differences between the PAZ domains of the structures are indicated with the red arrow.

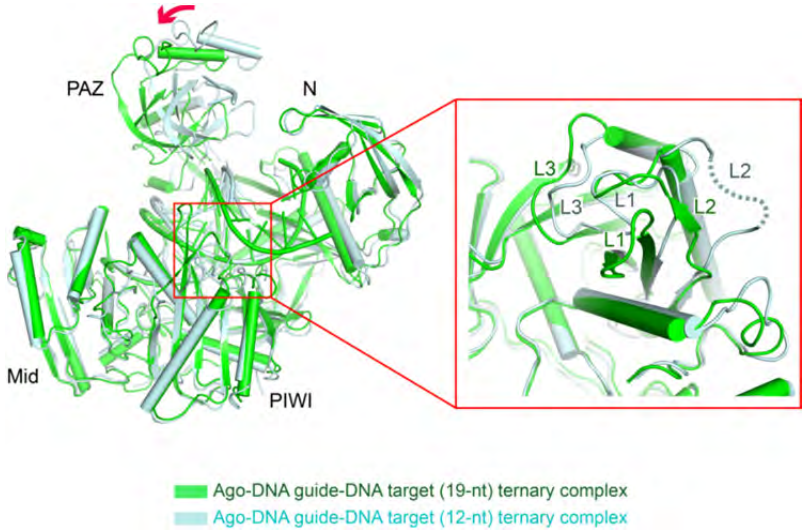


Figure S4 | Comparison of structures of ternary complexes of *Tt*Ago containing 5'-phosphorylated 21-mer guide DNA bound to 12-mer versus 19-mer target DNA. Full view of the differences following superposition of the structures of ternary complexes of *Tt*Ago containing 5'-phosphorylated 21-mer guide DNA bound to 12-mer (light blue) versus 19-mer (green) target DNAs. The boxed segment shows an expanded view of the differences following superposition of the structures of ternary complexes in the vicinity of the catalytic pocket of *Tt*Ago containing 5'-phosphorylated 21-mer guide DNA bound to 12-mer (light blue) versus 19-mer (green) target DNAs.

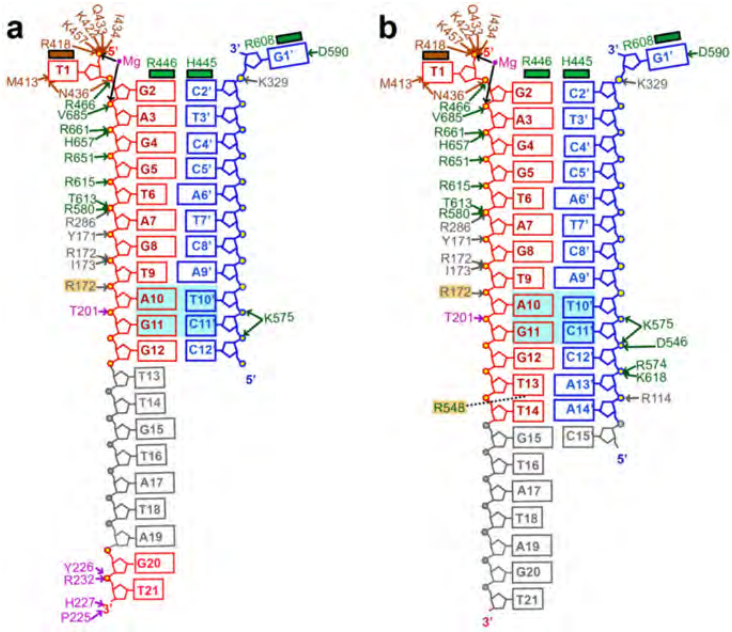


Figure S5 | Schematics showing intermolecular protein-DNA hydrogen-bonding contacts in the *Tt*Ago ternary complexes with 5'-phosphorylated 21-mer guide DNA and added 12-mer and 15-mer target DNAs. a, Intermolecular contacts in the cleavage-incompatible ternary complex with 12-mer target DNA. **b**, Intermolecular contacts in the cleavage-incompatible ternary complex with 15-mer target DNA.

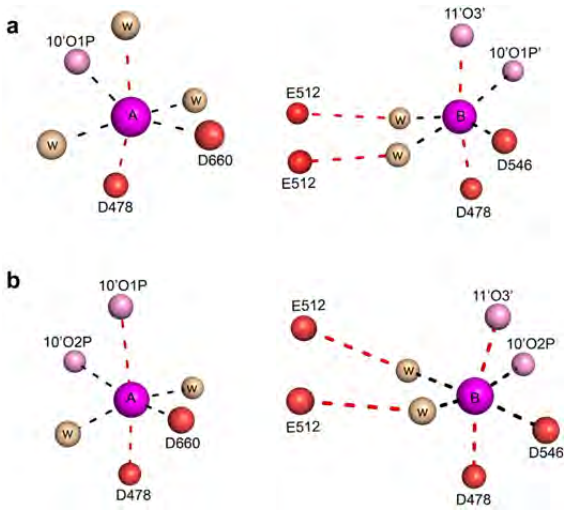


Figure S6 | Octahedral-like coordination geometries for Mg^{2+} cations A and B for intact and cleaved target strands in the ternary complex of *Tt*Ago containing 5'-phosphorylated 21-mer guide DNA bound to 19-mer target DNA. Mg^{2+} cations A and B (magenta), aspartate and glutamate carboxylate oxygens (red), phosphate oxygens (pink) and water molecules (gold). Mg^{2+} to oxygen distances are in the range of 2.1 Å, while the E512 carboxylate oxygen to water oxygen distances are between 2.7 and 2.8 Å. **a**, Coordination geometry in the ternary complex containing intact target DNA. **b**, Coordination geometry in the ternary complex containing cleaved target DNA.

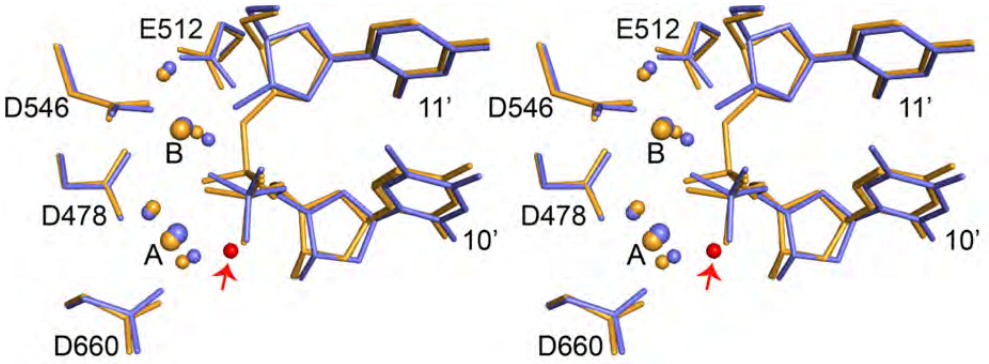


Figure S7 | Comparison of interactions within the catalytic pocket of *Tt*Ago ternary complexes with 5'-phosphorylated 21-mer guide DNA and complementary 19-mer target DNA. Stereoview of superposition of ternary complexes involving intact (gold) and cleaved (blue) phosphate backbones between nucleotides 10' and 11' of the target.

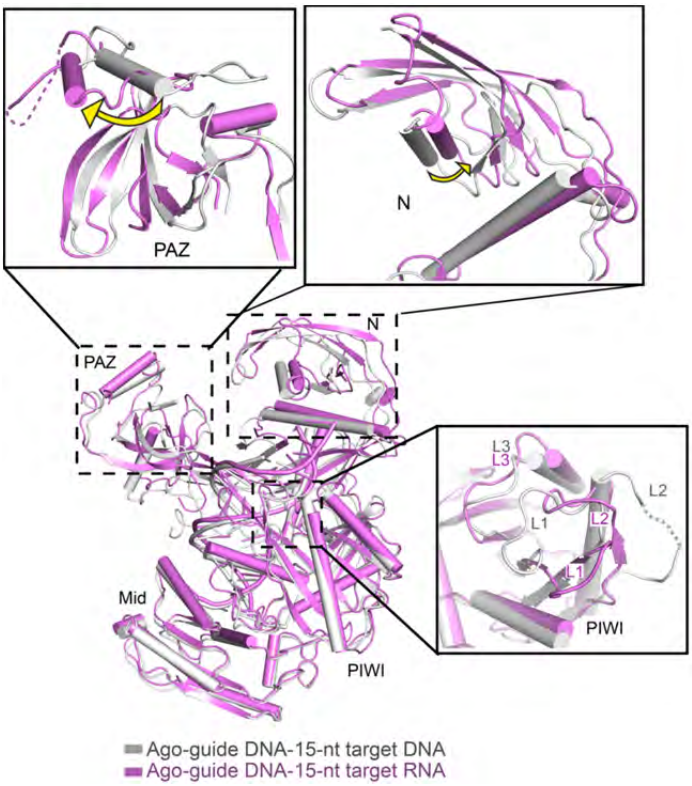


Figure S8 | Superposition of ternary complexes of *Tt*Ago with bound 5'-phosphorylated 21-mer guide DNA and 15-mer target DNA versus 15-mer target RNA. Comparison of overall structures of the ternary complexes with 15-mer target DNA (silver) and 15-mer target RNA (magenta). The expanded boxed segments show the conformational transitions within the PAZ domain, the N-domain and the loop segments within the catalytic pocket.

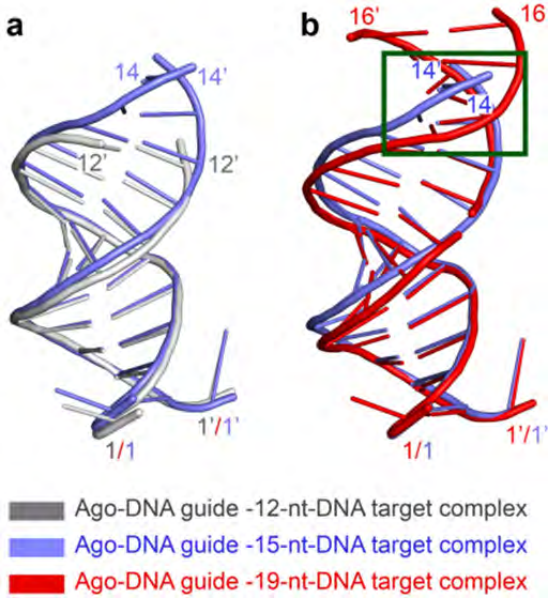


Figure S9 | Pairwise superposition of ternary complexes of *Tt*Ago bound to 5'-phosphorylated 21-mer DNA and 12-mer, 15-mer and 19-mer target DNAs. a, Pairwise superposition of ternary complexes containing 12-mer target DNA (guide-target duplex in silver) and 15-mer target DNA (guide-target duplex in blue). **b**, Pairwise superposition of ternary complexes containing 15-mer target DNA (guide-target duplex in blue) and 19-mer target DNA (guide-target duplex in red). The green box highlights the different trajectory of the guide strand beyond position 11 of the cleavage site in the two complexes. These perspectives were prepared following superposition the MID and a portion of the PIWI domains in the complexes.

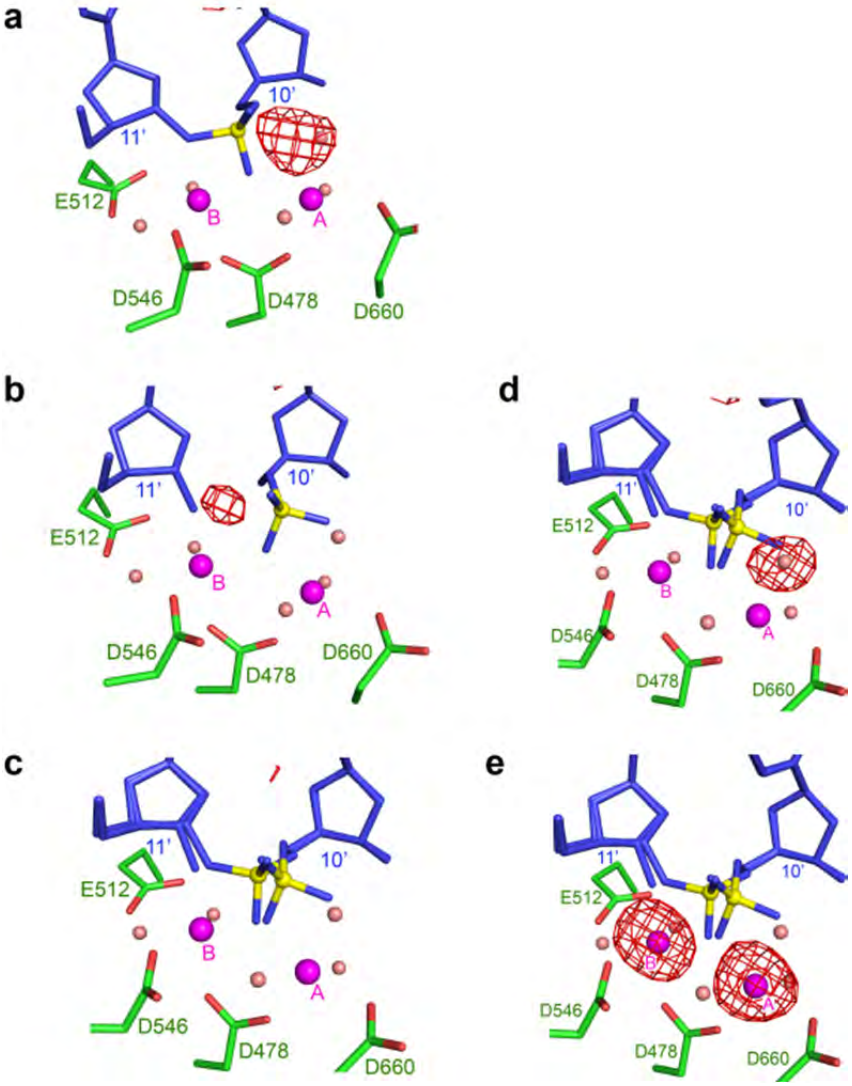


Figure S10 | Fo-Fc maps centered about nucleotides 10' and 11' of the target (cleavage site) and for identification of the position of the nucleophilic water and the pair of Mg^{2+} cations in the 2.2-Å structure of *Tt*Ago ternary complex with 5'-phosphorylated 21-mer guide DNA and complementary 19-mer target DNA. a, Fo-Fc map contoured at 3.5σ based on the assumption of an intact target strand, gave unaccounted density that was reflective of a cleaved phosphate backbone. **b**, Fo-Fc map contoured at 3.5σ based on the assumption of a cleaved target strand gave unaccounted density between sugar and phosphate group that was reflective of an intact phosphate backbone. **c**, No extra density was observed in the Fo-Fc map contoured at 3.0σ based on the assumption of a 1:1 mixture of intact and cleaved phosphate backbones. **d**, Identification of the nucleophilic water in the Fo-Fc map contoured at 3.5σ. **e**, Identification of the pair of Mg^{2+} cations in the Fo-Fc map contoured at 5.0σ.

Chapter 4

Characterization of *Pyrococcus furiosus* Argonaute

Daan C. Swarts, Jorrit W. Hegge, Ismael Hinojo, Masami Shiimori, Michael A. Ellis, Justin Dumrongkulraksa, Rebecca M. Terns, Michael P. Terns, John van der Oost

Adapted from:
'Argonaute of the Archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA'
Nucleic Acids Research, *accepted for publication*

Abstract

Functions of prokaryotic Argonautes have long remained elusive. Recently, Argonautes of the bacteria *Rhodobacter sphaeroides* and *Thermus thermophilus* were demonstrated to be involved in host defense. The Argonaute of the archaeon *Pyrococcus furiosus* (*Pf*Ago) belongs to a different branch in the phylogenetic tree, which is most closely related to that of RNA interference-mediating eukaryotic Argonautes. Here we describe a functional and mechanistic characterization of *Pf*Ago. Like the bacterial counterparts, archaeal *Pf*Ago contributes to host defense by interfering with the uptake of plasmid DNA. *Pf*Ago utilizes small 5'-phosphorylated DNA guides to cleave both single stranded and double stranded DNA targets, and does not utilize RNA as guide or target. Thus, with respect to function and specificity, the archaeal *Pf*Ago resembles bacterial Argonautes much more than eukaryotic Argonautes. These findings demonstrate that the role of Argonautes is conserved through the bacterial and archaeal domains of life, and suggests that eukaryotic Argonautes are derived from DNA-guided DNA-interfering host defense systems.

Introduction

Eukaryotic Argonaute proteins (eAgos) are the key players in RNA interference (RNAi) pathways (reviewed in [173,174,230]). During RNAi, eAgos are loaded with small 5'-phosphorylated RNAs, ranging from 20 to 30 nucleotides, in a pathway-specific ribonuclease-dependent process. The eAgos sometimes form the core of the multiprotein RNA-induced silencing complex (RISC; reviewed in [265]). Depending on which proteins associate with eAgos, Argonaute-mediated target binding can be specifically adjusted to differentially control gene expression. RNA interference generally results in silenced expression of the target gene, via functional variations that include decreased transcription (heterochromatin formation), decreased translation (mRNA binding), and decreased mRNA half-life (mRNA cleavage, mRNA de-adenylation).

Whereas prokaryotes also possess Argonaute proteins (pAgos), they appear to lack the accessory proteins involved in eukaryotic RNA interference pathways [4,170,230]. Initially, pAgo variants from *Aquifex aeolicus* and *Thermus thermophilus* were characterized biochemically, revealing that they could use DNA guides for RNA and DNA target cleavage [166,168,266]. Recently, it has become evident that at least some bacterial Agos play a role in host defense by interfering with invading nucleic acids [27,177,267]. *Rhodobacter sphaeroides* Argonaute (*Rs*Ago) acquires small RNA guides that allow interference with plasmid DNA [177]. *T. thermophilus* Argonaute (*Tt*Ago) acquires 5'-phosphorylated single stranded DNA (ssDNA) guides termed small interfering DNAs (siDNAs). These siDNAs are utilized by *Tt*Ago to cleave ssDNA and double stranded dsDNA (dsDNA) targets, the latter by cleaving each of the strands individually [27].

Archaeal pAgos initially have been explored to get insights in structural organization of Argonaute proteins [30,164,180,236]]. Along with the structure of *Archaeoglobus fulgidus* Argonaute (*Af*Ago), which is a truncated pAgo, the binding affinity of *Af*Ago for various nucleic acids has been described [164,236]. The apo-*Af*Ago has a higher affinity for ssDNA and dsDNA molecules than for ssRNA and dsRNA molecules [164,236]. The enzyme of *Methanocaldococcus jannaschii* (*Mj*Ago) is the only archaeal pAgo of which *in vitro* activity has been reported, revealing that this protein can utilize DNA guides to cleave DNA targets [190]. Like *Af*Ago, *Mj*Ago binds ssDNA and dsDNA molecules with a much higher affinity than ssRNA and dsRNA molecules [190]. Furthermore, *Mj*Ago is unable to cleave RNA target strands. The physiological role of *Mj*Ago has not been characterized. *Mj*Ago belongs

to a clade of phylogenetic clade with euryarchaeal thermophilic Argonautes, which is the pAgo clade that is most closely related to the eukaryotic Argonautes [4,230] (**Fig. 1a**). This clade also contains the *Pyrococcus furiosus* Argonaute (*Pf*Ago). *P. furiosus* is a hyperthermophilic archaeon, like *A. fulgidus* and *M. jannaschii*, and grows optimally at temperatures between 80 and 100 °C [268]. *Pf*Ago was the first Ago whose complete three-dimensional structure was determined [30]. However, the physiological role and the molecular mechanism of *Pf*Ago have not yet been reported. Despite belonging to the same clade, *Pf*Ago and *Mj*Ago only have 28% shared identity (BLASTp), which makes it difficult to predict the functionality of *Pf*Ago based on sequence homology alone. Furthermore, its closest related pAgo of which the function is determined (*Tt*Ago), has a query cover of only 38%, and only 24% shared identities with *Pf*Ago (BLASTp). In the present study, a combination of *in vivo* and *in vitro* analyses demonstrates that archaeal *Pf*Ago is involved in host defense by mediating DNA-guided DNA interference.

Results

*Pf*Ago interferes with plasmid DNA transformation

To elucidate the physiological role of *Pf*Ago in *P. furiosus* JFW02 (Pfu) [269], an *ago* knockout strain (PfuΔ*ago*) and an *ago* overexpressing strain (Pfu-*ago*-O/E) were generated (**Fig. 1b**). Immunoblots demonstrate that *Pf*Ago is expressed in Pfu, while no *Pf*Ago is detected in PfuΔ*ago* (**Fig. 1c**). Approximately 200-fold increased levels of full-length *Pf*Ago were detected in the Pfu-*ago*-O/E strain (**Fig. 1c**). Bacterial *Tt*Ago has previously been described to be involved in host defense by interfering with plasmid transformation [27,267]. These transformations were performed using the natural competence system of *T. thermophilus*, which transports extracellular DNA into the cell [234]. *P. furiosus* JFW02 also has a natural competence system [270]. This allowed comparing natural transformation efficiencies of strains Pfu, PfuΔ*ago* and Pfu-*ago*-O/E. Each strain was transformed with plasmid pJFW18 [271] or with plasmid pYS3 [272]. For both plasmids, transformation efficiency is ~1.5 to 2.5-fold higher in PfuΔ*ago* compared to wild type Pfu (**Fig. 1d**, $P<0.001$). When *Pf*Ago is overexpressed, the transformation efficiency is lowered even further (**Fig. 1d**, $P<0.001$). This demonstrates that archaeal *Pf*Ago lowers competence by interfering with plasmid transformation.

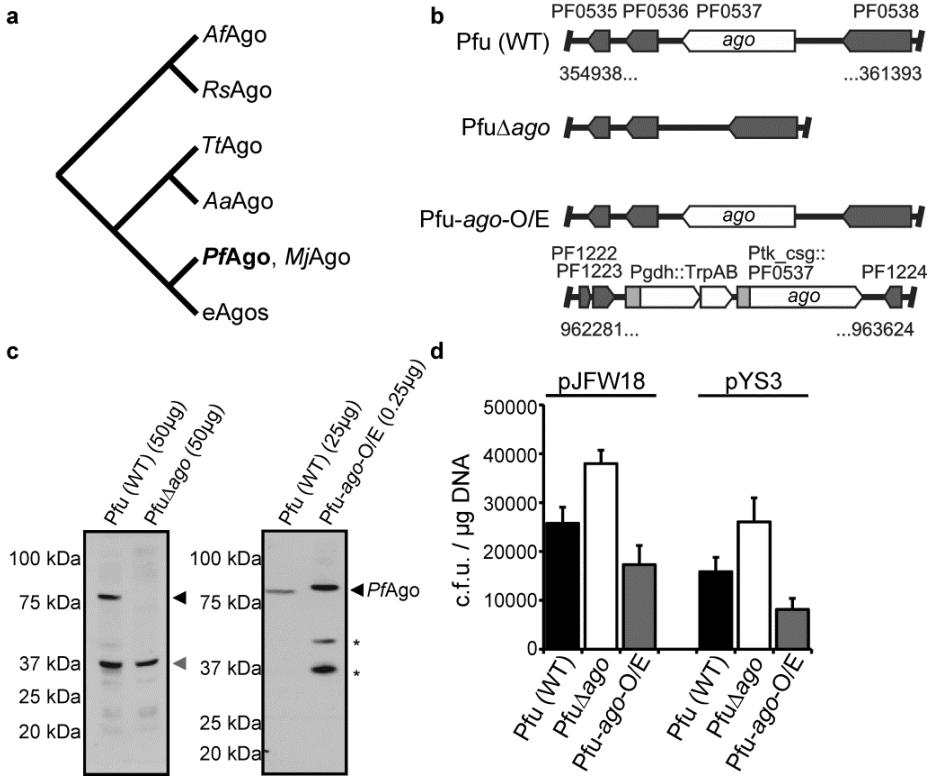


Figure 1 | *Pf*Ago interferes with plasmid transformation. a, Schematic phylogenetic tree of Argonaute proteins, adapted from [230]. *Af*Ago: *A. fulgidus* Ago. *Rs*Ago: *R. sphaeroides* Ago. *Tt*Ago: *T. thermophilus* Ago. *Aa*Ago: *A. aeolicus* Ago. *Pf*Ago: *P. furiosus* Ago. *Mj*Ago: *M. jannaschii* Ago. eAgos: eukaryotic Agos. **b**, Overview of *ago* gene loci of *P. furiosus* strains Pfu (wild type), PfuΔ*ago* (knockout) and Pfu-*ago*-O/E (*Pf*Ago overexpression strain). **c**, Immunoblot analysis of *Pf*Ago (indicated with a black triangle) content in Pfu and PfuΔ*ago* with Csa2 protein (indicated with a grey triangle) serving as the internal standard (left panel) or Pfu and Pfu-*ago*-O/E (right panel). The amount of lysate analyzed is indicated and the asterisk denotes apparent breakdown products observed when *Pf*Ago is overexpressed. **d**, Plasmid transformation efficiencies of the *P. furiosus* strains. Error bars indicate standard deviations of biological triplicates.

*Pf*Ago has a DEDH catalytic tetrad

Target cleavage by Agos is mediated by a conserved DEDX triad, of which the X can be a histidine or an aspartic acid [230] (**Fig. 2a**). Whereas the DDX residues are positioned close together in the available pAgo structures, the glutamic acid (E) is located on a structural sub-domain termed the 'glutamate finger' [179]. In *Tt*Ago, this finger is located at a distance of 12.8 Å from the catalytic site when no target nucleic acid is bound (unplugged conformation), but is inserted into the catalytic site (plugged-in conformation) upon target binding [176]. To identify the catalytic residues of *Pf*Ago, a sequence alignment [230] and a

structural alignment of *Tt*Ago and *Pf*Ago were analyzed (**Fig. 2a**, **b**). Previously, it was predicted that the catalytic site of *Pf*Ago includes residues D558, D628 and E635 [30]. However, E635 is located away from the catalytic site residues of other Agos both in sequence and structural alignments (**Fig. 2**). In contrast, H745 perfectly aligns with catalytic residues from other Agos in both sequences and structures. Like in the structure of unplugged *Tt*Ago, the glutamate finger of *Pf*Ago is unplugged in the structure of *Pf*Ago (**Fig. 2b**). However, the *Pf*Ago glutamate finger encompasses two glutamic acids (E592 and E596). Which of these two glutamate residues is involved in target cleavage is impossible to deduce from the sequence alignment or from the available *Pf*Ago structure (unplugged conformation).

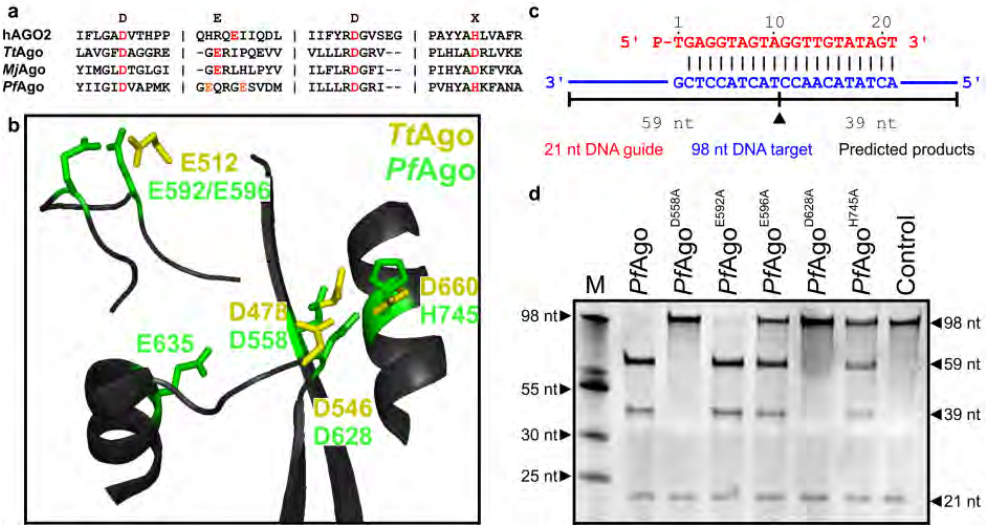


Figure 2 | DEDH catalytic site of *Pf*Ago. a, Sequence alignment of human AGO2 (hAGO2), *Tt*Ago, *Mj*Ago and *Pf*Ago, adapted from [230]. Only regions containing the DEDX catalytic residues (indicated in red) are shown. *Pf*Ago catalytic residues E592 and E596 are colored orange. **b**, *Tt*Ago catalytic residues DEDD (yellow; PDB: 4N47) aligned to *Pf*Ago catalytic site (black; PDB: 1Z25). Predicted catalytic residues of *Pf*Ago are colored green. **c**, Synthetic 21 nucleotide siDNA (red) and 98 nucleotide ssDNA target (blue) used for *in vitro* activity assays. The black triangle indicates the predicted cleavage site, black lines indicate the predicted 59 and 39 nucleotide cleavage products. **d**, *Pf*Ago and mutants were loaded with a 21 nucleotide long siDNA and were incubated with a 98 nucleotide ssDNA target in a 5:1:1 molar ratio (*Pf*Ago:guide:target). Products were resolved on a 15% denaturing polyacrylamide gel. M: ssDNA marker. nt: nucleotide. The 'Control' sample contains no protein.

To experimentally identify the residues involved in *Pf*Ago activity, Strep(II)-tagged *Pf*Ago and five predicted catalytic mutants were heterologously produced in *E. coli* KRX. After

affinity purification, expressed *Pf*Ago and mutants were tested for activity. As both *Tt*Ago and *Mj*Ago (which is closely related to *Pf*Ago) can utilize DNA guides to cleave ssDNA targets, we incubated *Pf*Ago with synthetic 5'-phosphorylated, 21 nucleotide long siDNAs. Subsequently, a synthetic 98 nucleotide ssDNA target was added (**Fig. 2c**, **d**). After 1 h incubation at 95 °C with *Pf*Ago and with *Pf*Ago^{E592A}, all ssDNA target is cleaved (**Fig. 2d**). This indicates that *Pf*Ago can utilize siDNAs to cleave DNA targets. In contrast, *Pf*Ago mutants D558A, E596A, D628A and H745A show impaired activity (**Fig. 2d**), indicating that the latter residues form the DEDH catalytic tetrad of *Pf*Ago.

Requirements for target cleavage by *Pf*Ago

To further determine the prerequisites for *Pf*Ago-mediated target cleavage, we tested the influence of temperature, salt concentration and divalent cation type on siDNA-guided ssDNA cleavage. As a negative control, we used a catalytic double mutant, *Pf*AgoDM (*Pf*Ago^{D558A,D628A}). For all assays, the guide and target shown in **Fig. 2c** were used. *Pf*Ago is most active in the range from 87 to 99 °C (**Fig. 3a**; higher temperatures were not tested). Longer (16 h) incubations show that *Pf*Ago exhibits some activity at 37 °C, but not at 20 °C (**Fig. 3b**). *Pf*Ago is active in reactions with a NaCl concentration of 50 to 250 mM, whereas at NaCl concentrations of 500 to 1000 mM the activity is lowered or absent (**Fig. 3c**, **Fig. S1**). To investigate if *Pf*Ago functions as a multi-turnover protein *Pf*Ago, siDNAs and ssDNA targets were incubated in a 2.5:1:20 ratio (*Pf*Ago:siDNA:target). *Pf*Ago-siDNA complexes cleave >95% of the 20-fold excess of target DNA within 30 minutes in buffer with 250 mM NaCl (**Fig. S1**). To investigate if higher NaCl concentrations inhibit multi-turnover reactions (for example by restricting the release of cleaved target strands) we additionally performed the same experiment at 500 mM and 1 M NaCl (**Fig. S1**). Even after 16 h, no activity is observed at 1 M NaCl. At 500 mM NaCl the activity of *Pf*Ago is severely lowered, but it still depletes the 20-fold excess of target DNA after 16 h, indicating that high NaCl concentrations do not inhibit the multi-turnover characteristic of *Pf*Ago.

Next, we investigated which divalent cations *Pf*Ago can utilize to mediate siDNA-guided DNA target cleavage. *Pf*Ago is able to utilize Mn^{2+} and Co^{2+} as cation, with Mn^{2+} being a better cation than Co^{2+} (**Fig. 3d**, **e**). In contrast to the phylogenetically closely related *Mj*Ago [190], *Tt*Ago [27,168,176] and eAgos [256], *Pf*Ago is unable to use Mg^{2+} as cation for its activity (**Fig. 3d**). Also Fe^{2+}, Cu^{2+}, Ni^{2+} and Ca^{2+} do not allow siDNA-guided *Pf*Ago cleavage of ssDNA (**Fig. 3d**). As eAgos have been shown to preferentially bind guides with

specific 5'-end nucleotides [182,199], we tested if *Pf*Ago has a preference for a specific 5'-end nucleotide on the siDNA. Like *Tt*Ago, *Pf*Ago is able to utilize siDNAs with different 5'-end nucleotides equally well (**Fig. S2**).

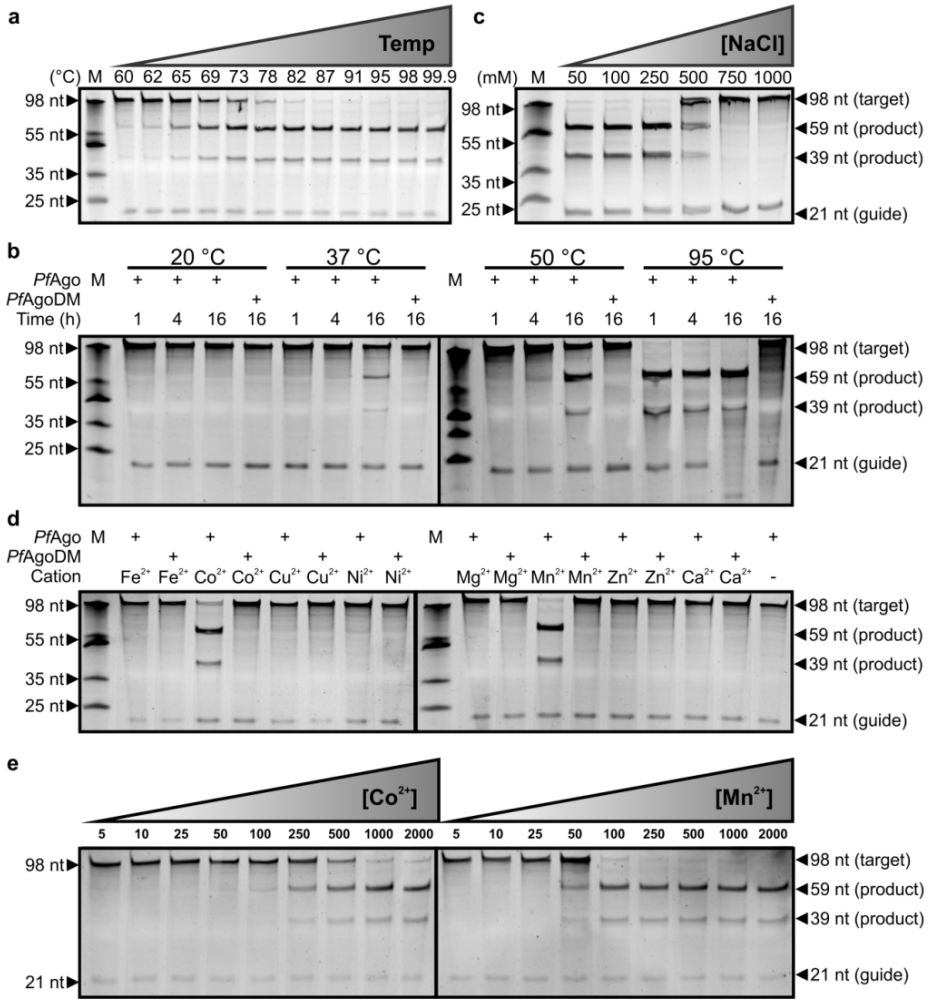


Figure 3 | Effect of temperature, salt concentration and cation on *Pf*Ago activity. *Pf*Ago loaded with a 21 nucleotide long siDNA was incubated with a 98 nucleotide ssDNA target (see Fig. 1c) in a 5:1:1 molar ratio (*Pf*Ago:guide:target) under various conditions. Unless otherwise indicated, target cleavage took place at 95 °C for 1 h, with 0.5 mM Mn^{2+} as cation. Nucleic acids are resolved on denaturing polyacrylamide gels. M: ssDNA marker. nt: nucleotide. **a**, *Pf*Ago activity is highest at temperatures between 90 °C and 99.9 °C. **b**, *Pf*Ago shows activity at temperatures ≥37 °C if incubation is extended. **c**, NaCl concentrations ≥500 mM interfere with *Pf*Ago activity. **d**, *Pf*Ago-guide complexes show Co^{2+} and Mn^{2+} mediated ssDNA target cleavage. **e**, Mn^{2+} is preferred above Co^{2+} as cation for *Pf*Ago-guide mediated ssDNA target cleavage.

*Pf*Ago is a DNA-guided protein that cleaves DNA targets

To investigate whether *Pf*Ago acquires specific guides *in vivo*, we analyzed nucleic acids that co-purify with *Pf*Ago expressed in *E. coli*. No DNA was observed in RNase A-treated samples, whereas in DNase I-treated samples it was observed that RNAs of undefined length co-purify with *Pf*Ago and *Pf*AgoDM (**Fig. S3**). The latter is suspected to be non-specifically bound RNA, as has previously been described for purification of *Tt*Ago and *Tt*AgoDM [230]. This is supported by the observation that RNAs associate with *Pf*Ago purified in presence and absence Mn^{2+}, whereas a divalent cation is required for specific binding of the 5' end of the guide by pAgos [183]. Unfortunately, attempts to identify guides associated with *Pf*Ago expressed in *P. furiosus* were not successful (data not shown)

To investigate if besides ssDNA, ssRNA can guide *Pf*Ago activity, *Pf*Ago was incubated with 21 nucleotide DNA or RNA guides after which 45 nucleotide ssDNA or ssRNA targets were added (**Fig. 4a**, **b**). As expected, *Pf*Ago catalyzes siDNA-guided cleavage of complementary ssDNA targets (**Fig. 4b**). In contrast to *Tt*Ago, *Pf*Ago does not show siDNA-guided cleavage of ssRNA targets (**Fig. 4b**).

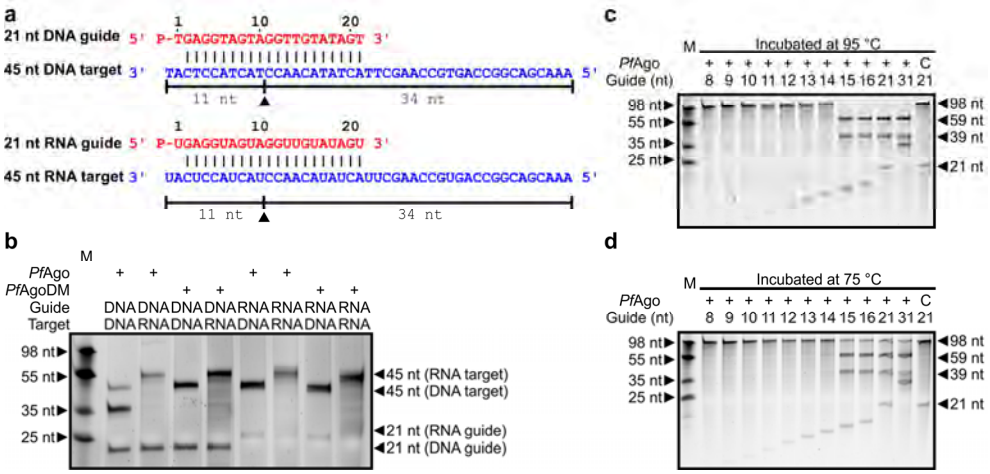


Figure 4 | *Pf*Ago utilizes 15 to 31 nucleotide long siDNAs for ssDNA target cleavage. *Pf*Ago was incubated with various guides and targets in a 5:1:1 ratio (*Pf*Ago:guide:target). Unless otherwise indicated, target cleavage took place at 75 °C for 1h, with 0.5 mM Mn^{2+} as cation. Nucleic acids are resolved on denaturing polyacrylamide gels. M: ssDNA marker. **a**, Synthetic DNA and RNA guides (red) and targets (blue). The black triangle indicates the predicted cleavage site, black lines indicate the predicted 11 and 34 nucleotide cleavage products. **b**, *Pf*Ago shows only DNA-guided DNA cleavage. **c**, **d**, Both at 95 °C and 75 °C, *Pf*Ago-mediated target cleavage is facilitated by siDNAs which are at least 15 nucleotides long. C: control reaction with *Pf*AgoDM and 21 nucleotide long siDNA.

With the RNA guide, we did not observe any *Pf*Ago-mediated cleavage, neither of ssDNA nor of ssRNA targets. This implies that *Pf*Ago only mediates DNA-guided DNA interference. In addition, the DNA guide length range of *Pf*Ago was tested. *Pf*Ago-mediated cleavage of DNA targets is only observed with siDNAs with a length ranging from 15 to at least 31 nucleotides (longer not tested; **Fig. 4c**, **d**). *Tt*Ago utilizes siDNAs with a length ranging from 9 to at least 36 nucleotides (longer not tested) [167]. As guides with a length of 31 nucleotides are too long for canonical guide binding by Ago, it has been hypothesized that the guide adopts an alternative trajectory to allow 3' end insertion into the PAZ binding pocket [167]. Alternatively, the 3' end of guide sticks out of the protein, as the nucleic acid-binding channel of *Tt*Ago is open to the outside [167]. Most likely, the same is true for *Pf*Ago.

***Pf*Ago mediates guide-free and siDNA-guided cleavage of dsDNA plasmids**

To test whether *Pf*Ago cleaves plasmid DNA, *Pf*Ago was incubated with its expression plasmid pWUR790 (**Fig. 5a**). As incubation of this plasmid at 95 °C in the presence of Mn^{2+} results in degradation of the plasmid (even in the absence of *Pf*Ago), we incubated the reaction mixtures at 75 °C for 16 h. Still, the majority of the plasmid DNA is turned to the open circular conformation under these conditions, even in absence of *Pf*Ago (**Fig. 5b**). Strikingly, pWUR790 is linearized when incubated with *Pf*Ago in absence of guides (**Fig. 5b**). We observed this suspected guide-free *Pf*Ago-mediated cleavage of pWUR790 in buffer with 250 mM NaCl, but not in buffer with 500 mM NaCl (**Fig. 5b**). These findings suggest either that *Pf*Ago cleaves plasmids independently of co-purified DNAs, or that DNA guides co-purified, but that their concentration of such co-purified DNA is below the detection limit of the assay. It has previously been demonstrated that *in vivo*, *Tt*Ago acquires guides targeting its expression vector [27]. To rule out that this *Pf*Ago activity was guided by co-purified RNA or by co-purified DNA guides that we were unable to detect, we used pWUR704 as target plasmid (**Fig. 5c**). pWUR704 has minimal sequence similarity to the *Pf*Ago expression vector (pWUR790; no sequences longer than 13 consecutive identical base pairs between the two plasmids). After incubation without *Pf*Ago, the plasmid is present both in open circular and supercoiled configuration (**Fig. 5e**, lane 1). When *Pf*Ago is added, supercoiled pWUR704 is linearized even in absence of guides (**Fig. 5e**, lane 2). Like pWUR790 cleavage, this activity is more pronounced at 250 mM NaCl compared to 500 mM NaCl. In contrast, *Tt*Ago, which co-purified with detectable levels of siDNA, was unable to cleave pWUR704, unless synthetic siDNAs targeting pWUR790 were added [27].

Low levels of guide-free *Pf*Ago-mediated nicking of pWUR704 takes place within 1 h incubation at 75 °C (**Fig. S4**). As *Pf*Ago does not mediate RNA-guided activity within the same time span (**Fig. 4**), it is unlikely that the activity on plasmid DNA is mediated by RNA that co-purified with *Pf*Ago. Instead, these findings suggest that *Pf*Ago has the potential to cleave the plasmid independently of guides.

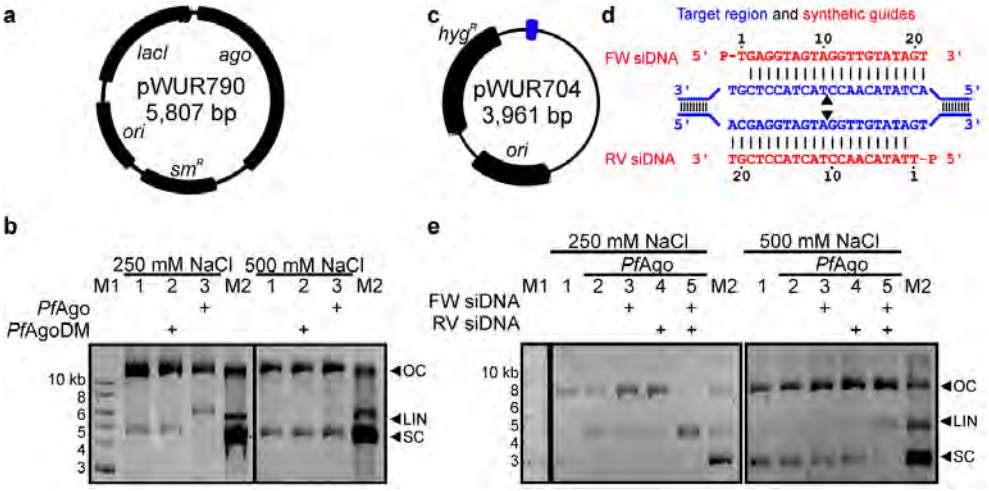


Figure 5 | Plasmid cleavage by *Pf*Ago. a, pWUR790 expression plasmid. **b**, *Pf*Ago expressed at 20 °C and purified in absence of Mn^{2+} cleaves expression plasmid pWUR790. Agarose gels with plasmid targets incubated without protein (lane 1), with *Pf*AgoDM (lane 2), and with *Pf*Ago (lane 3). M1: 1 kb GeneRuler marker (Thermo Scientific). M2: pWUR790 marker with open circular (OC), linearized (LIN) and supercoiled (SC) pWUR790. **c**, pWUR704 target plasmid, target site indicated in blue. **d**, Target region (blue) and FW and RV siDNA guides (red). Predicted cleavage sites are indicated with a black triangle. **e**, Agarose gels with plasmid targets incubated without *Pf*Ago (lane 1), with guide free *Pf*Ago (lane 2), and with *Pf*Ago loaded with FW siDNA, RV siDNA, or both (lane 3-5) in reaction buffer with 250 mM NaCl (left panel) or 500 mM NaCl (right panel). M1: 1 kb DNA ladder (New England Biolabs). M2: pWUR704 marker with open circular (OC), linearized (LIN) and supercoiled (SC) pWUR704.

We next investigated whether siDNAs can guide *Pf*Ago-mediated plasmid cleavage. We loaded *Pf*Ago either with a FW siDNA targeting the (-) strand of pWUR704, or with a RV siDNA targeting the (+) strand of pWUR704 (**Fig. 5d**). *Pf*Ago and *Pf*Ago-siDNA complexes were incubated with plasmid pWUR704 as target, in buffers containing 250 or 500 mM NaCl. Additionally, both *Pf*Agos complexes with FW guides and with RV guides were mixed and incubated with pWUR704. *Pf*Ago loaded with a single siDNA nicks the plasmid DNA, generating open circular plasmids (**Fig. 5e**, lane 3 and 4). Interestingly, reactions to

which FW or RV siDNAs are added contain more open circular plasmid compared to the sample to which no siDNA is added, in which plasmids are mostly turned to the linear state. This suggests siDNA loading hinders guide-free *Pf*Ago activity, or alternatively it suggests that guide-free cleavage of plasmid DNA by *Pf*Ago requires plasmid supercoiling. Incubation of the plasmids with *Pf*Ago-siDNA complexes targeting both strands of the plasmid results in plasmid linearization (**Fig. 5e**, lane 5). Both processes are performed more efficiently in reactions with 250 mM NaCl as compared to reactions with 500 mM NaCl. More efficient cleavage at 250 mM NaCl has also been observed for ssDNA targets (**Fig. 3c**). These findings demonstrate that *Pf*Ago-siDNA complexes can target dsDNA plasmids, resulting in a dsDNA break if both strands of the plasmid are targeted.

Discussion

This chapter describes the first combined *in vivo* and *in vitro* characterization of an archaeal Argonaute. *Pf*Ago functions as a DNA-guided DNA endonuclease that requires divalent cations such as Mn^{2+} or Co^{2+} for its activity. It utilizes 5'-phosphorylated ssDNA guides that are at least 15 and up to at least 31 nucleotides long. As such long guides cannot be bound canonically, we predict that the 3'-end of the guide 'sticks out' a gap in the PAZ domain. Whether this gap exists in the PAZ domain of *Pf*Ago, and the function of this predicted feature remain to be determined. Like the bacterial *Tt*Ago, archaeal *Pf*Agos can utilize a single siDNA to nick plasmid DNA, while two *Pf*Ago-siDNA complexes (each targeting a single strand of the plasmid) together can generate dsDNA breaks. Both pAgos mediate host defense by interfering with invading plasmid DNA (**Fig. 1**). As *Tt*Ago and *Pf*Ago do not belong to the same branch in the phylogenetic tree, these findings suggest a broad conservation of pAgo functions. Nevertheless, *Pf*Ago has some interesting characteristics: *Pf*Ago cannot use Mg^{2+} as cation (while *Tt*Ago and *Mj*Ago can [27,190]) and *Pf*Ago-siDNA complexes do not mediate RNA target cleavage (while *Tt*Ago-siDNA complexes do [168]). The preference for DNA or RNA guides or targets of eAgos, *Tt*Ago and *Pf*Ago most likely is determined on a structural level. The guide-target duplex binding channel and guide 5'-end binding pockets of eAgos are strongly positively charged, allowing RNA guide and target binding. In contrast, the duplex binding channels of *Tt*Ago and *Pf*Ago are much less positively charged, and their guide 5'-end binding pockets are hydrophobic [273], possibly excluding the possibility to bind RNA guides. However, this hypothesis requires further investigation before any claims can be made.

We were unable to co-purify siDNA guides from *Pf*Ago heterologously expressed in *E. coli* KRX. Possibly *Pf*Ago is not able to acquire guides at the low temperatures at which it is expressed (20 °C or 37 °C), or it requires host factors for guide generation and/or loading. Future research should focus on guides associated with *Pf*Ago expressed in *P. furiosus*. Interestingly, under some conditions, *Pf*Ago appears to cleave dsDNA plasmids in absence of siDNAs, suggesting that *Pf*Ago shows unguided, non-specific nuclease activity (**Fig. 5**). We have observed similar activity for *Tt*Ago under specific conditions [274], and hypothesize that unguided dsDNA cleavage might be related to the generation of siDNAs.

We chose to characterize *Pf*Ago as it belongs to the clade of euryarchaeal pAgos, which is the clade of pAgos that is most closely related to eAgos (**Fig. 1a**). eAgos are best known for their RNA-guided RNA cleaving role in RNA interference. Therefore, we predicted that *Pf*Ago possibly was involved in both RNA interference and DNA interference. In contrast, *Pf*Ago solely mediates DNA cleavage in a DNA guide-dependent manner. Strikingly, a recent comment paper cites numerous papers in which DNA binding by eAgo has been described [275]. It was suggested that some eAgos might be involved in DNA binding, a hypothesis that requires further investigation. As *Pf*Ago is one of the pAgos most closely related to eAgos, these findings suggest that eAgos were derived from the same ancestors as *Pf*Ago, which might have utilized DNA guides and/or DNA targets as well. Alternatively, the pAgos changed their guide and target specificity multiple times during evolution. Combined, these findings indicate that bacterial and archaeal Argonautes, and possibly even ancient eukaryotic Argonautes, mediate host defense by DNA-guided DNA interference.

Author contributions

MS made *ago* deletion and overexpression strains and performed plasmid transformation analysis under supervision of MPT and RMT. ME and JD carried out *Pf*Ago immunoblot analysis under supervision of MPT and RMT. *E. coli* expression plasmids were constructed by JWH and IH under supervision of DCS and by DCS under supervision of JvdO. Protein purifications and *in vitro* assays were performed by IH and JWH under supervision of DCS and by DCS under supervision of JvdO. The manuscript was written by DCS with help of all authors. All authors read and approved the manuscript.

Acknowledgements

We thank Stan J.J. Brouns for critical reading of the manuscript. This work was financially supported by grants from the Netherlands Organization of Scientific Research (NWO) to JvdO (NWO-TOP, 854.10.003) and National Institutes of Health Grant R01 GM54682 to MPT and RMT.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Experimental procedures

Strains and cultivation

For *in vivo* experiments, *P. furiosus* strain JFW02 [269] was used, which is referred to in this chapter as Pfu or wild type. Furthermore two genomic variants of strain JFW02, Pfu∆*ago* (*ago* knockout strain) and Pfu-*ago*-O/E (*ago* overexpression strain) were used (**Fig. 1; Table S1**).

Strains were cultivated anaerobically in a defined medium with cellobiose as the carbon source [270] at 90 °C in anaerobic culture bottles or on medium solidified with 1% (w/v) Gelrite (Research product international). For growth of uracil auxotrophic strains, the defined medium was supplemented uracil to a final concentration of 20 µM.

Genomic mutants

Pfu-*ago*-O/E was generated by transforming an NruI-linearized pHSG298 *ago* plasmid (**Table S2**) into the wild type Pfu strain. The plasmid contains the gene encoding *Pf*Ago with an upstream *Thermococcus kodakaraensis* csg promoter, flanked by PF1223 and PF1224 gene sequences for homologous recombination. Furthermore, the plasmid encodes TrpA and TrpB with an upstream gdh promoter. The plasmid was constructed by overlap PCR and oligonucleotides used to generate this plasmid are shown in **Table S3**.

Two rounds of colony purification was performed by plating 10^{-3} dilutions of transformant cultures onto selective plate medium (without tryptophan) and picking isolated colonies into selective liquid medium. The Pfu∆*ago* was created by pop-out marker replacement strategy as described previously [269]. The sequences of oligonucleotide used are shown in **Table S3**.

Transformation experiments

Plasmid transformation was performed as described previously using 2.5 ng plasmid DNA μl^{-1} of culture [270]. pJFW18 plasmid [271] and a modified pYS3 plasmid [272] were used (**Table S2**). The modified pYS3 plasmid was generated by replacement of the Sim^R cassette of pYS3 by the Pgdh pyrF cassette from pJFW18. The sequences of oligonucleotides used are shown in **Table S3**. The transformation efficiencies reported were calculated as the number of transformed colonies per µg of DNA added.

*Pf*Ago antibodies and *Pf*Ago immunodetection

Polyclonal antibodies were raised in chickens against purified (nickel chromatography) N-terminal, 6X His-tagged recombinant *Pf*Ago or *Pf*Csa2 (loading control) proteins as previously described [276]. Western blotting was performed by standard procedures. The blots were incubated with polyclonal IgY immune antibodies and HRP-conjugated Anti-IgY secondary antibody (Gallus Immunotech). The protein bands on the blots were detected using an enhanced chemiluminescent substrate for HRP (horse radish peroxidase) activity (GE Life Sciences) and exposure to autoradiography film.

Sequence and structural alignments

For the active site residue sequence alignment, previously published alignments were used [230]. For structural alignments, structures from *Pf*Ago (PDB: 1Z25) and *Tt*Ago (PDB: 4N47) were loaded in PyMOL. All residues but the MID and PIWI domains were deleted and the remaining residues were aligned with the PyMOL software. For clarity, only residues forming the active site of *Pf*Ago, and active site residues of *Tt*Ago were displayed.

*Pf*Ago expression and purification

A synthetic codon-optimized gene encoding *Pf*Ago with an N-terminal Strep(II)-tag was ordered from GenScript USA Inc. and was directionally cloned into expression vector pCDF-1b as indicated in **Table S2** (pWUR790). Plasmids pWUR791-pWUR796 (**Table S2**) were generated by introducing mutations according to an adapted QuikChange Site-Directed mutagenesis Kit instruction manual (Stratagene) using primers described in **Table S3**. These plasmids were transformed into *E. coli* KRX (Promega) according to the protocol provided by the manufacturer. Strains were cultivated in LB medium containing 50 μg ml^{-1} streptomycin and 0.4% (w/v) glucose in a shaker incubator at 37 °C. After overnight incubation, cultures were centrifuged for 5 min at 4700 x g, after which the supernatant was removed. Cell pellets were resuspended in LB medium containing 50 μg ml^{-1} streptomycin and incubated in shaker incubator at 37 °C until an $OD_{600\ nm}$ of 0.6-0.8 was reached. Cultures were cold-shocked by incubation in an ice bath for 15 min. *Pf*Ago expression was induced by adding isopropyl-b-D-thiogalactoside (IPTG) and L-Rhamnose to a final concentration of 1 mM and 0.1% (w/v), respectively. Expression was continued in a shaker incubator at 20 °C or 37 °C for 16 h. Cells were collected by centrifugation for 15 min at 6000 x g, after which the supernatant was removed. Cells were resuspended in Buffer I (20 mM Tris/HCl pH 8, 1 M NaCl, 2 mM $MnCl_2$) and disrupted by sonication with a Branson

Sonifier B-12 and Branson Converter with a 5 mm tip (ten 30 sec pulses at 30% power with 30 sec pause between pulses). The solution was centrifuged for 30 min at 32,913 x g at 4 °C, after which the supernatant was used for purification by Strep-Tactin affinity chromatography (IBA, Germany) with an adapted protocol. Before loading of the supernatant, the Strep-Tactin Sepharose column was equilibrated with Buffer I. After loading, the column was washed 16 CV (column volumes) Buffer I. N-terminally Strep(II)-tagged *Pf*Ago was eluted in Buffer II (buffer I supplemented with 2.5 mM biotin (Sigma-Aldrich)). For purification of *Pf*Ago used in cation preference and cation gradient experiments, no $MnCl_2$ was added to the purification buffers.

Activity assays

For activity assays, elution fractions containing *Pf*Ago or *Pf*Ago mutants were diluted with Buffer II to a final protein concentration of 5 µM. 5 µl protein sample was mixed with synthetic ssDNA or ssRNA guides (**Table S3**) in a 5:1 ratio (protein:guide) in reaction buffer (20 mM Tris/HCl, pH 8) and incubated for 15 min at 95 °C. After pre-incubation, ssDNA or ssRNA targets (**Table S3**) were added to a final 5:1:1 ratio (protein:guide:target) and incubated for 1 h at 95 °C. Final reaction concentrations were 15 mM Tris/HCl pH 8, 250 mM NaCl, 0.5 µM $MnCl_2$, 1.25 µM protein, 0.25 µM guide, and 0.25 µM target. Note that in different experiments incubation temperature, incubation time, and salt concentration were varied (indicated in figures of corresponding experiments). For cation preference experiments, *Pf*Ago or *Pf*AgoDM and ssDNA guides (**Table S3**) were mixed in a 5:1 ratio (protein:guide) in reaction buffer to which different cations were added, and incubated for 15 min at 95 °C. After pre-incubation, ssDNA targets (**Table S3**) were added to a final 5:1:1 ratio (protein:guide:target) and incubated for 1 h at 95 °C. Final reaction concentrations were 15 mM Tris/HCl pH 8, 250 mM NaCl, 0.5 mM metal-Cl_2 ($FeCl_2$, $CoCl_2$, $CuCl_2$, $NiCl_2$, $MgCl_2$, $MnCl_2$, $ZnCl_2$, or $CaCl_2$), 1.25 µM protein, 0.25 µM guide,0.25 µM target). After incubation Loading Buffer (95% (deionized) formamide, 5 mM EDTA, 0.025% SDS, 0.025% Bromophenol blue and 0.025% xylene cyanol) was added in a 1:1 ratio and samples were incubated for 10 minutes at 95 °C before resolving on 15% or 20% denaturing polyacrylamide gels. Nucleic acids were stained using SYBR gold Nucleic Acid Gel Stain (Invitrogen) and visualized using a G:BOX Chemi imager (Syngene). For plasmid assays, elution fractions containing *Pf*Ago or *Pf*Ago mutants were diluted with Buffer II to a final protein concentration of 5 µM. 5 µl protein sample was mixed with synthetic ssDNA guides (**Table S3**) in a 5:1 ratio (protein:guide) in reaction buffer (20 mM Tris/HCl, pH 8, and varying NaCl concentrations) and incubated for 15 min at 95 °C. After pre-incubation,

~200-300 ng plasmid pWUR704 or pWUR790 (**Table S2**) was added, and the samples (final reaction concentrations were 15 mM Tris/HCl pH 8, 0.5 µM $MnCl_2$, and 250 mM or 500 mM NaCl) were incubated for 16 h at 75 °C. Reactions were stopped by adding Proteinase K solution (Ambion) and $CaCl_2$ (final concentration 5 mM) and samples were incubated for 1h at 65 °C. Samples were mixed with 6x loading dye (Thermo Scientific) before they were resolved on 0.8% agarose gels. As marker, either a 1 kb Generuler Marker (Thermo Scientific) or 1 kb DNA ladder (New England Biolabs), and additionally a custom plasmid marker, were used. The custom plasmid marker consisted of non-treated pWUR704 (mostly in supercoiled conformation), Nb.BSMI (New England Biolabs) nicked pWUR704 (open circular conformation) and BcuI (Thermo Scientific) linearized pWUR704. Agarose gels were stained with SYBR-gold Nucleic Acid Gel Stain (Invitrogen) and visualized using a G:BOX Chemi imager (Syngene).

Guide co-purification

1/10 volume Proteinase K (Ambion) and 1/10 volume $CaCl_2$ (50 mM) were added to 500 pmol purified proteins in Buffer II and samples were incubated for 4 h at 65 °C. Nucleic acids were separated from protein content using Roti(R) Phenol/Chloroform/Isoamyl alcohol (Carl Roth GmbH). 1/10 volume 3 M NaAc (pH 5.2) and 10 µl 0.5% linear polyacrylamide was added, before addition of 96 % ethanol in a 2:1 ratio (v/v, ethanol:sample). Samples were incubated at -20 °C for 2 days and centrifuged for 30 min at 4 °C at 20,000 x rpm in a table top centrifuge. Supernatant was removed and the pellet was dried for 10 min at 50 °C. The pellet was washed in 70% ethanol, followed by centrifugation for 30 minutes at 4 °C at 20,000 x rpm in a table top centrifuge. Supernatant was removed and the pellet was dried for 10 min at 50 °C. The pellet was resuspended in 50 µl MilliQ H_2O. Purified nucleic acids were [γ-^{32}P] ATP labeled with T4 PNK (Thermo Scientific) in exchange or forward labeling reactions, in which 5' phosphates or 5' OH groups are labeled, respectively. Nucleic acids were separated from free [γ-^{32}P] ATP using a Sephadex G-25 column (GE). Nucleic acids were incubated with DNase-free RNase A (Thermo Scientific) or RQ1 RNase-free DNase I (Promega) for 1 h at 37 °C. Samples were mixed with Loading Buffer in a 1:1 ratio and samples were incubated for 10 min at 95 °C. [γ-^{32}P] ATP labeled nucleic acids were resolved on 20% denaturing polyacrylamide gels. Radioactivity was captured from gels using phosphor screens.

Supplementary Figures and Tables

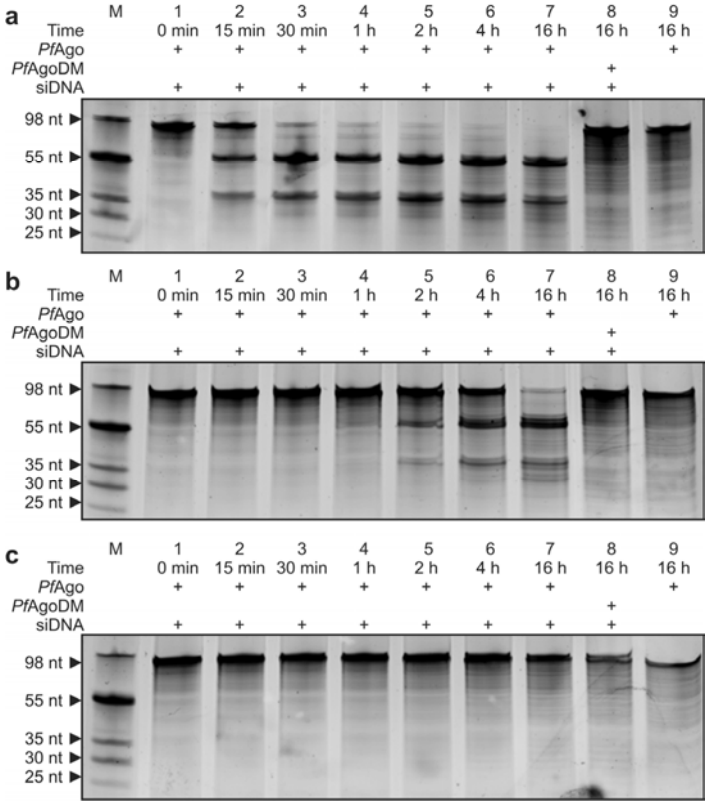


Figure S1 | *Pf*Ago-siDNA complexes show multi-turnover activity. *Pf*Ago, siDNAs and target ssDNAs (Fig. 2C) were incubated in a 2.5:1:20 ratio (*Pf*Ago:siDNA:target) and incubated at 95 °C in buffer with **a**, 250 mM NaCl or **b**, 500 mM NaCl or (**C**) 1 M NaCl. Nucleic acids are resolved on denaturing polyacrylamide gels. siDNA concentrations are below detection limits. M: ssDNA marker. nt: nucleotide.

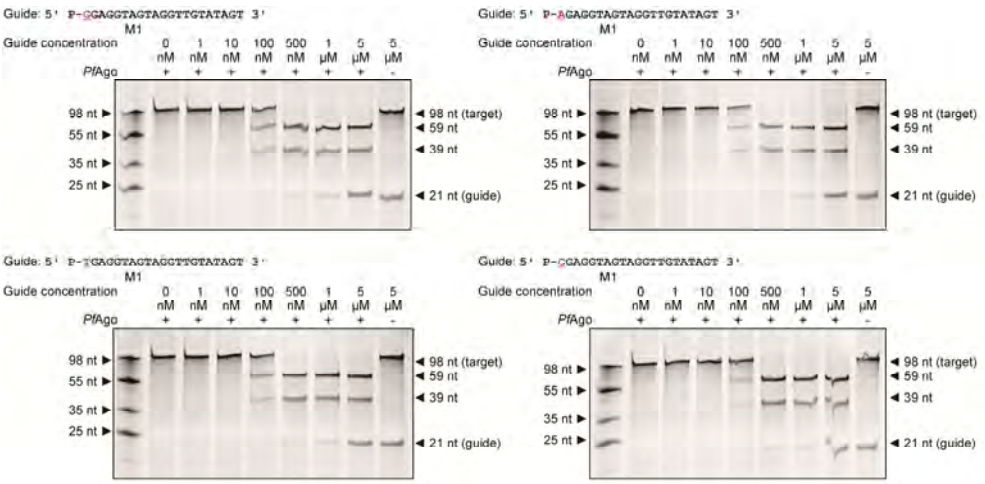


Figure S2 | Effect of the variation of the 5'-end deoxynucleoside of the siDNA on *Pf*Ago cleavage efficiency. Cleavage of 98 nucleotide ssDNA targets by *Pf*Ago loaded with complementary siDNAs containing a different 5'-end deoxynucleoside, as shown in red above each gel. The concentrations of each siDNA were varied (indicated on top of the gels). Products of the reactions were resolved on denaturing polyacrylamide gels. M1: ssDNA marker. Samples without *Pf*Ago (indicated with '-') contain *Pf*AgoDM as control.

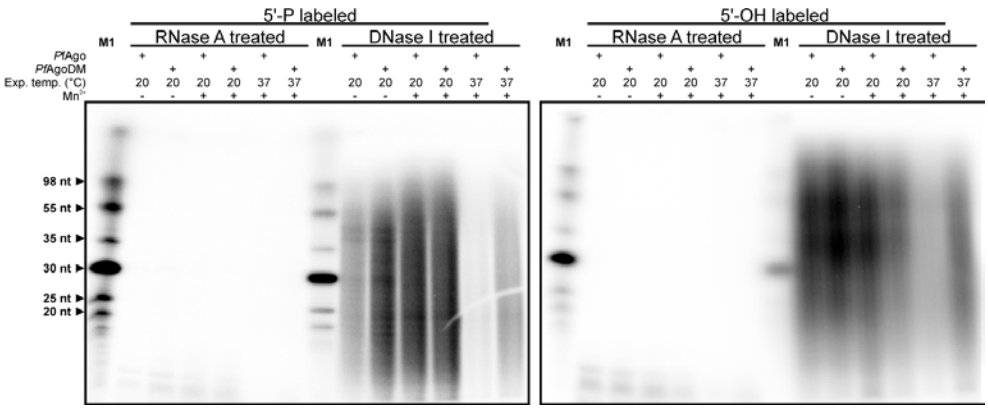


Figure S3 | Nucleic acids co-purifying with *Pf*Ago and *Pf*AgoDM. Co-purified nucleic acids are 5' phosphorylated in a T4 PNK exchange reaction (left panel; 5'-P groups, and to a lesser extend 5'-OH groups are labeled) or in a T4 PNK forward reaction (right panel; 5'-OH groups, and to a lesser extend 5'-P groups are labeled) using [γ-^{32}p] ATP and resolved on 20% denaturing polyacrylamide gels. Nucleic acids were not treated, RNase A treated or DNase I treated. M1: ssDNA marker, labeled in a T4 PNK forward reaction. Exp. temp: Expression temperature. Mn^{2+}: indicates whether the protein was purified in absence (-) or presence (+) of Mn^{2+}.

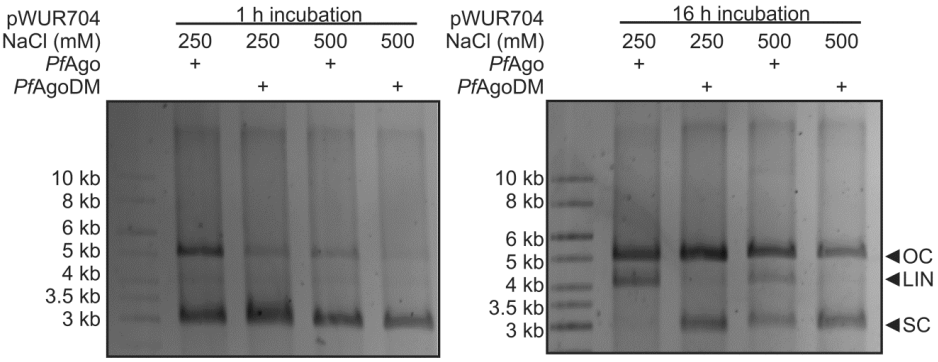


Figure S4 | Guide-free *Pf*Ago activity on pWUR704. pWUR704 was incubated with *Pf*Ago and *Pf*AgoDM for 1 h or 16 h at 75 °C at different final NaCl concentrations, and resolved on 0.75% agarose gels. Left lane: 1 kb generuler (Thermo Scientific). OC: open circular. LIN: linear. SC: Supercoiled.

Table S1 | Strains

Strain	Abbreviations	Description	Source, reference
Pyrococcus furiosus JFW02	Pfu, wild type	*Pyrococcus furiosus* JFW02	[269]
Pyrococcus furiosus Δ*ago*	PfuΔ*ago*, knockout	*Pyrococcus furiosus ago* gene (PF0537) knockout	This chapter
Pyrococcus furiosus ago overexpression strain	Pfu-*ago*-O/E, overexpression strain	*Pyrococcus furiosus Pf*Ago overexpressing strain	This chapter
Escherichia coli KRX	*E. coli* KRX	*E. coli* expression strain	Promega

Table S2 | Plasmids

Plasmid	Description	Restriction sites used	Primers	Source, reference
pHSG298	*Pf*Ago overexpression cassette genome insertion vector		PF1223_1224_1-8 PfAgo_fow PfAgo_rev	This study
pYS3	*E. coli*/*P. furiosus* shuttle vector, Amp^R, pyrF under control of gdh promoter	BamHI	Pgdh_pyrF_fow Pgdh_pyrF_rev	[272]
pJFW18	*E. coli*/*P. furiosus* shuttle vector, Apr^R, pyrF under control of gdh promoter		-	[271]
pCDF-1b	Expression vector		-	Novagen
pWUR790	Synthetic codon optimized *P. furiosus ago* with N-terminal *strep(II)-tag* in pCDF-1b vector, expression vector for *Pf*Ago.	NcoI, AvrII	-	GenScript USA Inc.
pWUR791	pWUR790, *ago* active site residue substituted (D558A)		BG5480 BG5481	This study
pWUR792	pWUR790, *ago* active site residue substituted (E592A)		BG5482 BG5483	This study
pWUR793	pWUR790, *ago* active site residue substituted (E596A)		BG5484 BG5485	This study
pWUR794	pWUR790, *ago* active site residue substituted (D628A)		BG5486 BG5487	This study
pWUR795	pWUR790, *ago* active site residue substituted (H745A)		BG5488 BG5489	This study
pWUR796	pWUR791, two *ago* active site residue substituted (D558A,D628A)		BG5486 BG5487	This study
pWUR704	Target plasmid		-	[27]

Table S3 | Oligonucleotides

Experiment	Primers	Sequence (5'-3')	Description, restriction sites
Pfu-*ago*-O/E strain	PF1223_1224_1	CCAGTGCCAAGCTTGCATGCACCG GTGCGGCCGCACGCGTTTAAACGG TTCTCAAGCGTATTTTTGG	PF1223 (FW)
	PF1223_1224_2	CTCAGCTCACTCCATTTTCAATCAT CCATCCACTGAGAATATTGAAG	PF1233 (RV)
	Pgdh_PyrF_F	GATTGAAAATGGAGTGAGCTGAG	trpAB (FW)
	TrpAB_Csg_4	TAATTCGCCTTTTGCCGATAGTCG ATTGGCTGAGCTCATG	trpAB (RV)
	Ptk_csg_fow	TATCGGCAAAAGGCGAATTATG	csg promoter (FW)
	Ptk_csg_rev	GAGGAAGCGGAGGTTCCA	csg promoter (RV)
	PF1223_1224_7	TGGAACCTCCGCTTCCTCTTCTTCT CTACTAGATCCGTTATC	PF1224 (FW)
	PF1223_1224_8	TATGACATGATTACGAATTCTACG TATCCGGATTAATTAATTTAAATG GTTAGATAAGAATCCGCTGAA	PF1224 (RV)
	PfAgo_fow	CAACCCAAGGAGGTGTTGTCATAT GAAAGCGAAAGTTGTTATTAATCT	*ago* (FW)
	PfAgo_rev	AAAGAGGAGAAGAGAGGGGGGAT CCTCAAACAAAATACAAAAATCCC TCA	*ago* (FW)
PfuΔ*ago*	PfAgo_del_1	TCTCTTCTGGCCGTTGATCT	Upstream region ago (FW)
	PfAgo_del_2	CTCAGCTCACTCCATTTTCAATCCA TTAACTTTTTCTATTTAAACAATTT CTAACCT	Upstream region ago (RV)
	PfAgo_del_5	GGTGAAAGAATGGAGCTCAAGATA ATTGAGGAATTTAGTTCTAGTTCTC AGG	Upstream region ago (FW)
	PfAgo_del_6	CATTAACTTTTTCTATTTAAACAAT TTCTAACCT	Upstream region ago (RV)
	PfAgo_del_7	AGGTTAGAAATTGTTTAAATAGAA AAAGTTAATGATTATCACCAAGTG ATTACAATTAATATCA	Downstream region ago (FW)
	PfAgo_del_8	TGTTCATCAACAGGGAGGAA	Downstream region ago (RV)
	Pgdh_PyrF_F	GATTGAAAATGGAGTGAGCTGAG	Pgdh pyrF cassette (FW)
	Pgdh_PyrF_R	TTATCTTGAGCTCCATTCTTTCACC	Pgdh pyrF cassette (RV)
pYS3	Pgdh_pyrF_fow	AGCTTCTCTGCAGGATATCTGGAT CCGATTGAAAATGGAGTGAGCTG	Pgdh pyrF cassette (FW)
	Pgdh_pyrF_rev	GCCGAAGCTAGCGAATTCGTGGAT CCGTCGATTGGCTGAGCTCATG	Pgdh pyrF cassette (RV)
Site directed mutagenesis of *ago* gene	BG5480	GATTATATCATTGGCATTGCTGTG GCACCGATGAAACG	*P. furiosus ago* D558A (FW)
	BG5481	CGTTTCATCGGTGCCACAGCAATG CCAATGATATAATC	*P. furiosus ago* D558A (RV)
	BG5482	CCGATTAAAATCGGTGCACAGCGT GGTGAAAGCG	*P. furiosus ago* E592A (FW)
	BG5483	CGCTTTCACCACGCTGTGCACCGA TTTTAATCGG	*P. furiosus ago* E592A (RV)
	BG5484	GGTGAACAGCGTGGTGCAAGCGTT GATATGAACG	*P. furiosus ago* E596A (FW)

	BG5485	CGTTCATATCAACGCTTGCACCAC GCTGTTCACC	*P. furiosus ago* E596A (RV)
	BG5486	CCTGCTGCTGCGTGCTGGTCGCAT TACCAATAATG	*P. furiosus ago* D628A (FW)
	BG5487	CATTATTGGTAATGCGACCAGCAC GCAGCAGCAGG	*P. furiosus ago* D628A (RV)
	BG5488	CCTGCACCGGTTCATTATGCAGCT AAATTTGCCAATGCCATTCG	*P. furiosus ago* H745A (FW)
	BG5489	CGAATGGCATTGGCAAATTTAGCT GCATAATGAACCGGTGCAGG	*P. furiosus ago* H745A (RV)
siDNA and siRNA sequences	BG3466	P-TGAGGTAGTAGGTTGTATAGT	21 nt DNA guide (FW), based on let-7 miRNA
	BG4017	P-TTATACAACCTACTACCTCGT	DNA guide (RV), based on reverse complement of let-7 miRNA
	BG4500	P-AGAGGTAGTAGGTTGTATAGT	DNA guide (FW), based on let-7 miRNA, 5'-end deoxyadenosine
	BG4501	P-GGAGGTAGTAGGTTGTATAGT	DNA guide (FW), based on let-7 miRNA, 5'-end deoxyguanosine
	BG4502	P-CGAGGTAGTAGGTTGTATAGT	DNA guide (FW), based on let-7 miRNA, 5'-end deoxycytidine
	BG4503	P-TGAGGTAGTAGGTTGTATAGT	DNA guide (FW), based on let-7 miRNA, 5'-end deoxythymidine
	BG4508	P-UGAGGUAGUAGGUUGUAUAGU	RNA guide (FW), based on let-7 miRNA
	BG5599	P-TGAGGTAG	8 nt DNA guide based on BG3466
	BG5600	P-TGAGGTAGT	9 nt DNA guide based on BG3466
	BG5601	P-TGAGGTAGTA	10 nt DNA guide based on BG3466
	BG5602	P-TGAGGTAGTAG	11 nt DNA guide based on BG3466
	BG5641	P-TGAGGTAGTAGG	12 nt DNA guide based on BG3466
	BG5713	P-TGAGGTAGTAGGT	13 nt DNA guide based on BG3466
	BG5714	P-TGAGGTAGTAGGTT	14 nt DNA guide based on BG3466
	BG5713	P-TGAGGTAGTAGGTTG	15 nt DNA guide based on BG3466
	BG5640	P-TGAGGTAGTAGGTTGT	16 nt DNA guide based on BG3466
	BG5603	P-TGAGGTAGTAGGTTGTATAGTATA TTAAATT	31 nt DNA guide based on BG3466
Target sequences	BG4263	TCGACTTTATATTTAAATAATTTAA TATACTATACAACCTACTACCTCGT ATAAATTTTTAAATAAATATTGCAT TCAAGCTTTTAATTTAATTAAAT	98 nt RV ssDNA target
	BG3678	AAACGACGGCCAGTGCCAAGCTTA CTATACAACCTACTACCTCAT	45 nt RV ssDNA target
	BG4427	AAACGACGGCCAGUGCCAAGCUU ACUAUACAACCUACUACCUCAU	45 nt RV ssRNA target

Chapter 5

Acquisition of DNA guides by prokaryotic Argonaute

Daan C. Swarts, Elizabeth M. Timmers, Yifan Zhu, John van der Oost

Adapted from:
'Acquisition of DNA guides by prokaryotic Argonaute'
Manuscript in preparation

Abstract

Eukaryotic Argonaute proteins mediate RNA-guided RNA interference (RNAi) to regulate gene expression. Prokaryotic Argonautes have recently been demonstrated to provide host defense by DNA-guided DNA interference. Whereas eukaryotic Argonautes rely on other proteins for generation of their RNA guides, homologs of these proteins are missing in prokaryotes. To investigate whether *Tt*Ago can generate siDNAs itself, we performed a series of *in vitro* experiments in which guide-free *Tt*Ago was incubated with a variety of DNA fragments. Whereas guide-free *Tt*Ago is unable to cleave single stranded DNA, it does degrade double stranded DNA under specific conditions. *In vitro* experiments and sequencing of the fragments revealed that the guide-free degradation of double stranded DNA does not take place at specific sequences, but rather depends on (partial) unwinding of the target. This degradation, which we termed 'DNA chopping', generates 5'-phosphorylated products which are 13 to 25 nucleotide long. This suggests that these products are siDNAs. Indeed, the generated fragments guide *Tt*Ago to complementary DNA targets resulting in target cleavage. Hence, *Tt*Ago can chop double stranded DNA to generate small interfering DNA guides.

Introduction

Argonaute proteins are key components of eukaryotic RNA interference pathways (RNAi; reviewed in [173,174,230]). In these pathways, eukaryotic Argonaute proteins (eAgos) utilize small RNA guides to bind and cleave RNA targets. There is great variation in RNAi pathways, which besides Argonaute rely on different pathway-specific proteins for guide generation, modification, loading and often also for processes that occur after target binding. Prokaryotes also possess Argonaute proteins (pAgos) [4,230], which have been demonstrated to be involved in host defense by interfering with plasmid DNA [27,177,267,277]. In contrast to eAgos, at least some of the pAgos utilize DNA guides to directly target DNA [27,163,164,167,168,169,176,190,277]. Prokaryotes do not encode orthologs of the accessory proteins involved in eukaryotic RNAi pathways [170]. Although some pAgo genes appear to co-localize with specific other genes [4,230], many pAgo genes are not clustered, suggesting they function as stand-alone proteins or require common-host factors for guide processing.

Argonaute of the bacterium *Thermus thermophilus* (*Tt*Ago) acquires small interfering DNA guides (siDNA) to interfere with plasmid transformation and propagation [27,267,278]. These siDNAs are 5' phosphorylated and are 13 to 25 nucleotides in length. *Tt*Ago acquires these guides upon heterologous expression in *Escherichia coli*, indicating either that guide acquisition is dependent on common host factors (shared between *T. thermophilus* and *E. coli*), or that *Tt*Ago itself is responsible for processing of its guides. The latter explanation is supported by the observation that a catalytic double mutant *Tt*Ago^{D478A,D546A} (*Tt*AgoDM) is unable to acquire guides *in vivo* [27]. The majority of siDNAs that associate with *Tt*Ago *in vivo* are acquired from plasmid DNA, suggesting that plasmid DNA is preferentially selected for generation of guides [27]. Argonaute of the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*Ago) also interferes with plasmid DNA transformation [277]. *Pf*Ago-siDNA complexes have been demonstrated to cleave DNA targets *in vitro*, like *Tt*Ago. However, *Pf*Ago additionally appears to linearize plasmid targets in absence of guides under specific conditions [277].

Here we demonstrate that guide-free *Tt*Ago also cleaves double stranded DNA (dsDNA) plasmids under specific conditions. Additionally, we show that *Tt*Ago can generate siDNAs from dsDNA precursors, a process that we termed 'DNA chopping'. DNA chopping generates products (siDNAs) that guide *Tt*Ago to cleave complementary target DNA.

Results

Guide-free *Tt*Ago linearizes AT-rich plasmid DNA

First, we investigated if, like *Pf*Ago [277], guide-free *Tt*Ago cleaves plasmid DNA under specific conditions. Earlier plasmid cleavage experiments with *Tt*Ago were performed in reaction buffer containing 500 mM NaCl, and resulted in linearization of plasmid DNA only when siDNAs were added [27]. As plasmid linearization by guide-free *Pf*Ago was observed at a final NaCl concentration of 250 mM but not at 500 mM [277], we incubated *Tt*Ago with plasmids pWUR704 (containing a 98 bp target region with 17% GC content; **Fig. 1a**) and pWUR705 (containing a 98 bp target region with 59% GC content; **Fig. 1a**) at various NaCl concentrations. As observed for *Pf*Ago, *Tt*Ago is able to linearize pWUR704 in the absence of guides in a buffer containing 250 mM NaCl (**Fig. 1b**, lane 3), but not in a buffer containing 500 mM NaCl (**Fig. 1b**, lane 1). This could possibly be explained by the fact that the negative charge of DNA backbones is masked by positively charged Na^+ ions. This reduces the repulsion between the two DNA backbones [279]. The resulting increased dsDNA stability might hamper guide-free plasmid degradation by *Tt*Ago and *Pf*Ago.

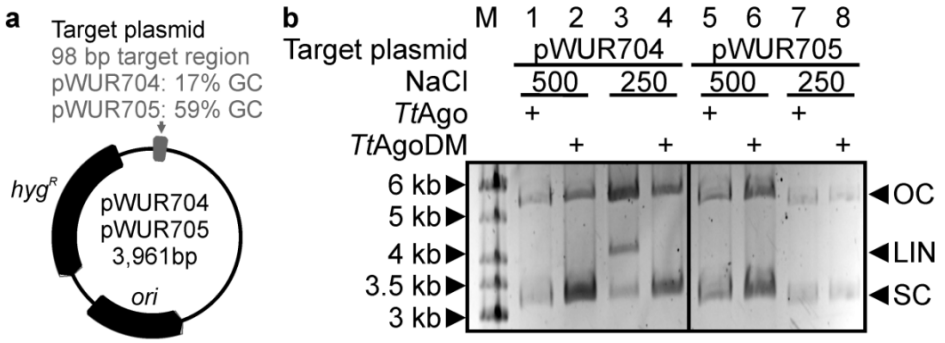


Figure 1 | Guide-free *Tt*Ago linearizes AT-rich pWUR704 at low NaCl concentrations. a, Schematic representation of target plasmids pWUR704 and pWUR705. *ori*: Origin of replication. hyg^R: Hygromycin resistance marker. **b**, Plasmids pWUR704 and pWUR705 were incubated with guide-free *Tt*Ago in buffer containing 500 or 250 mM NaCl and were resolved on 0.8% agarose gels. M: GeneRuler 1 kb DNA Ladder (Thermo Scientific). OC: Open circular. LIN: Linear. SC: Supercoiled.

Additionally, higher ionic strengths lower DNA-protein interactions [280]. Under the same conditions, pWUR705 is not linearized (**Fig. 1b**). This can be explained by the difference of the inserts of these plasmids: pWUR704 has an AT-rich insert whereas pWUR705 has a stable GC-rich insert. This AT-rich insert has a lower melting temperature than the GC-rich inserts, making pWUR704 more prone to unwinding than pWUR705. Combined,

these findings suggest that *Tt*Ago requires a certain degree of DNA unwinding for guide-free cleavage of the plasmid DNA.

Guide-free *Tt*Ago chops (partially) unwound dsDNA

To further investigate the activity of guide-free *Tt*Ago, we performed experiments with various 98 bp long dsDNA fragments (**Table S1**). When guide-free *Tt*Ago was incubated with AT-rich dsDNA fragments identical to the pWUR704 insert, the DNA appears to be degraded to smaller DNA fragments (**Fig. 2a**). *Tt*Ago-mediated degradation of AT-rich dsDNA does not rely on the presence of 5' overhangs or 3' overhangs, as both these targets are chopped as well as blunt ended targets (**Fig. 2a**). In contrast, when incubated with GC-rich dsDNA fragments identical to the pWUR705 insert, no degradation is observed (**Fig. 2b**, **Fig. S1**). These findings are in line with the observation that *Tt*Ago can linearize pWUR704 but not pWUR705 (**Fig. 1**). Furthermore, the degradation is dependent on the canonical catalytic site of *Tt*Ago, as incubation of dsDNAs with *Tt*AgoDM does not result in target degradation. Next, we tested the influence of the NaCl concentration on guide-free *Tt*Ago degradation of dsDNAs and ssDNAs. The AT-rich dsDNA is degraded by *Tt*Ago at NaCl concentrations of 250 mM, but not at 500 mM or 750 mM NaCl (**Fig. S1**). We did not observe degradation of the ssDNA strands that make up the AT-rich dsDNA (**Fig. S1**), which indicates that two complementary DNA strands are required for guide-free *Tt*Ago-mediated degradation. To distinguish unguided degradation from canonical guided-Ago target cleavage, we coin the term 'chopping' for unguided Ago-mediated degradation of dsDNAs.

To determine whether DNA instability indeed promotes chopping, GC-rich dsDNAs with internal mismatched regions (1 to 36 bp mismatches) were incubated with guide-free *Tt*Ago. In contrast to perfectly base-paired or single-mismatched GC-rich DNA, GC-rich DNAs with an internal mismatch of 6 bp or larger are chopped (**Fig. 2c**, **Fig. S1**). These observations confirm that a certain degree of dsDNA unwinding is required to allow *Tt*Ago-mediated chopping.

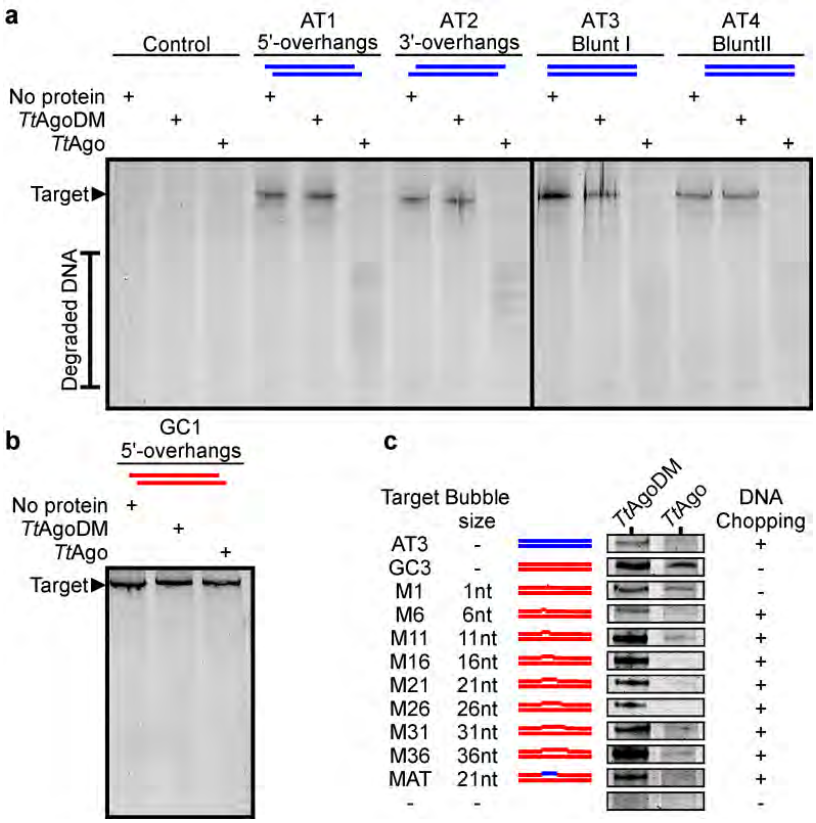


Figure 2 | Chopping of short dsDNA by *Tt*Ago. **a-c**, 20% denaturing polyacrylamide gel loaded with 98 bp AT-rich dsDNAs (panel a), a 98 bp GC-rich dsDNA (panel b) or 98 bp GC-rich dsDNAs with internal mismatches (panel c), incubated with *Tt*Ago or *Tt*AgoDM. AT-rich and GC-rich DNA is colored blue and red, respectively. 'Control' samples include no target DNA, 'No protein' samples do not contain any protein. Intact DNA is indicated with the '►' mark, while degraded DNA is visible when *Tt*Ago is incubated with AT-rich dsDNA or GC-rich DNA with internal mismatches, but not when incubated with fully complementary GC-rich dsDNA. For a detailed overview of the dsDNA targets, see Table S1. For the uncropped gels of panel c, see Fig. S1.

Products of DNA chopping guide *Tt*Ago activity

For canonical target cleavage, *Tt*Ago utilizes an ssDNA guide with a 5'-end phosphate [27,167,168,169]. To investigate whether DNA chopping generates potential guides, chopped DNA was [γ-^{32}P] ATP labeled with T4 Polynucleotide Kinase (PNK). 5'-hydroxyl (OH) and 5'-phosphate (P) groups of chopped DNA were specifically labeled in independent assays. The results confirm that AT-rich dsDNA targets are chopped by *Tt*Ago but not by *Tt*AgoDM (**Fig. 3a**). The generated DNA fragments have a size of ~13 to 25 nucleotides, which matches the size of siDNAs that *Tt*Ago acquires *in vivo* [27]. In contrast

to earlier results in which no chopping of GC-rich targets was observed (**Fig. 1b**), this detection method shows that small amounts of GC-rich targets are chopped (**Fig. 3a**, lane 3 and 9). *Tt*Ago mediated chopping generates products either with 5'-OH groups (most likely the original 5' ends of the dsDNA) or products with 5'-P groups. The 5'-P ends are most likely generated by the catalytic site of *Tt*Ago, as the catalytic site is required for chopping [27] and generates 5'-P groups during canonical target cleavage [176]. Although products with 5'-OH groups appear to be more abundant than products with 5'-P groups, it should be noted that according to the T4 PNK manufacturer (Thermo Scientific), 5'-OH labeling generally occurs more efficiently than 5'-P labeling. In any case, at least a fraction of the chopped DNA contains 5'-P groups, which implies these products could be acquired by *Tt*Ago to guide its activity.

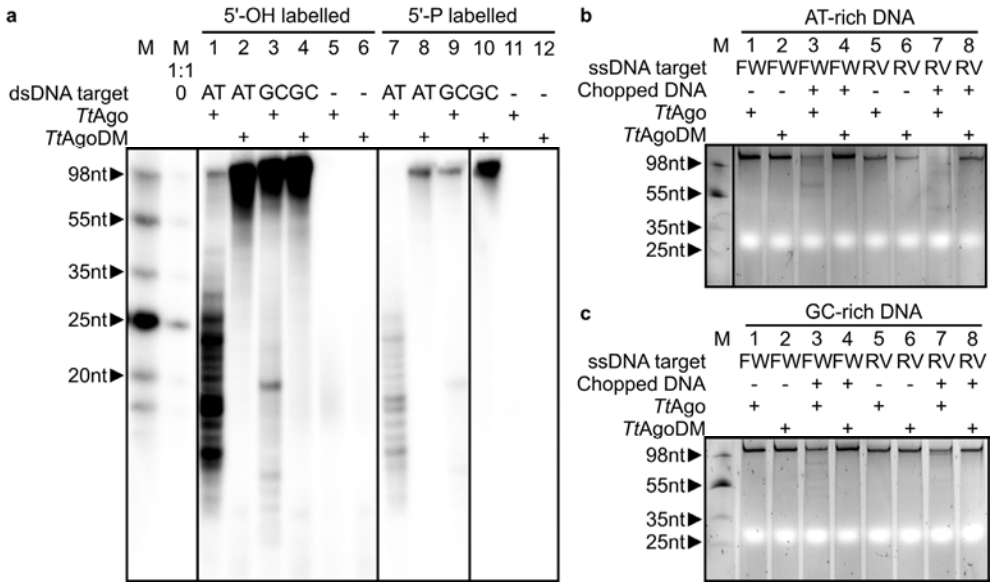


Figure 3 | DNA chopping generates functional siDNA guides. a, 20% denaturing polyacrylamide gel loaded with [γ-^{32}P] ATP labeled chopped DNA. AT-rich dsDNA C or GC-rich dsDNA G (Extended data table 1) were incubated with *Tt*Ago or *Tt*AgoDM. M: ssDNA marker. M 1:10: 10 times diluted ssDNA marker. Lanes 1 to 6 contain DNA fragments labeled in a T4 PNK forward reaction (5'-OH groups are preferentially labeled) while lanes 7 to 12 contain the same DNA fragments labeled in a T4 PNK exchange reaction (5'-P groups are preferentially labeled). **b**, *Tt*Ago incubated with chopped AT-rich dsDNA (as in panel a, lane 1 and 7) and ssDNA targets. The ssDNA targets have the same sequence as the forward (FW, BG4262) or reverse (RV, BG4724) strand of the chopped dsDNA. **c**, *Tt*Ago incubated with chopped GC-rich dsDNA (as in panel a, lane 3 and 9) and ssDNA targets. The ssDNA targets have the same sequence as forward (FW, BG4264) or reverse (RV, BG4726) strand of the chopped dsDNA. M: ssDNA marker.

To investigate if products of chopped DNA can guide *Tt*Ago activity, *Tt*Ago was pre-incubated with purified chopped DNA, after which ssDNA targets complementary to the chopped DNA were added. Indeed, chopped AT-rich DNA (**Fig. 3b**, lane 3 and 7) and, to a lesser extent, chopped GC-rich DNA (**Fig. 3c**, lane 3 and 7) is utilized by *Tt*Ago as guide to cleave complementary ssDNA targets. As both the FW and the RV target strands are cleaved, it can be concluded that functional guides are generated from both strands of chopped dsDNA. These results indicate that chopping of dsDNA results in generation of 5'-phosphorylated ssDNA molecules that guide *Tt*Ago activity.

DNA chopping takes place at the border of instable regions

In order to identify the sites at which *Tt*Ago chopped the DNA, the chopped DNA was cloned and sequenced. As the DNA fragments are generated from two complementary strands, we predicted that generated DNA fragments form short duplex DNAs. As overhangs of these duplexes might interfere with cloning, chopped DNA was modified with Klenow polymerase or Blunting Enzyme prior to cloning. These proteins remove 3' overhangs and fill in 5' overhangs, which results in loss of information on the 3' ends of the duplex DNA, while 5' ends remain unaltered. Cloning of Klenow-treated chopped DNA by TOPO cloning was not successful (data not shown). As 5'-P groups inhibit TOPO cloning, this finding supports the earlier observation that 5' ends of chopped DNA are phosphorylated. Therefore we attempted to clone Blunting-enzyme treated chopped DNA using T4 DNA ligase-dependent pJET1.2 cloning. T4 DNA ligase covalently links 5'-P to 3'-OH groups, which successfully generated vectors with chopped DNA inserts.

A total 120 recombinant vectors (some with multiple inserts), were sequenced. AT-rich DNA chopping generated considerably shorter inserts (average length 17 bp (+- 11.6)) compared to GC-rich DNA chopping (average length 59 bp (+-33.6)). This finding agrees with the difference in chopping efficiency of both targets. The shortest inserts were only 8 bp long, while the longest were intact 98 bp targets. The 5' ends of the inserts allowed the identification of a total of 202 and 48 chopping sites which for AT-rich and GC-rich target DNAs, respectively. The 5' ends generated during AT-rich DNA and GC-rich DNA chopping are indicated in **Fig. 4** and **Fig. S2**, respectively.

To investigate the chopping mechanism, we compared relative positions of identified 5' ends. If chopping would generate dsDNA breaks, this would be reflected in **Fig. 4a** by

generation of the same amount of adjacent 5' ends on opposite strands. If DNA chopping would take place in a ping-pong cycle-like process, like generation of secondary piRNAs [91,92], this would be reflected in **Fig. 4a** by generation of similar amounts of 5' ends exactly 10 bp apart on opposite strands. As we observe no such patterns, DNA degradation via such a process does not appear likely. In contrast, the DNA appears to be chopped independent of other chopping sites.

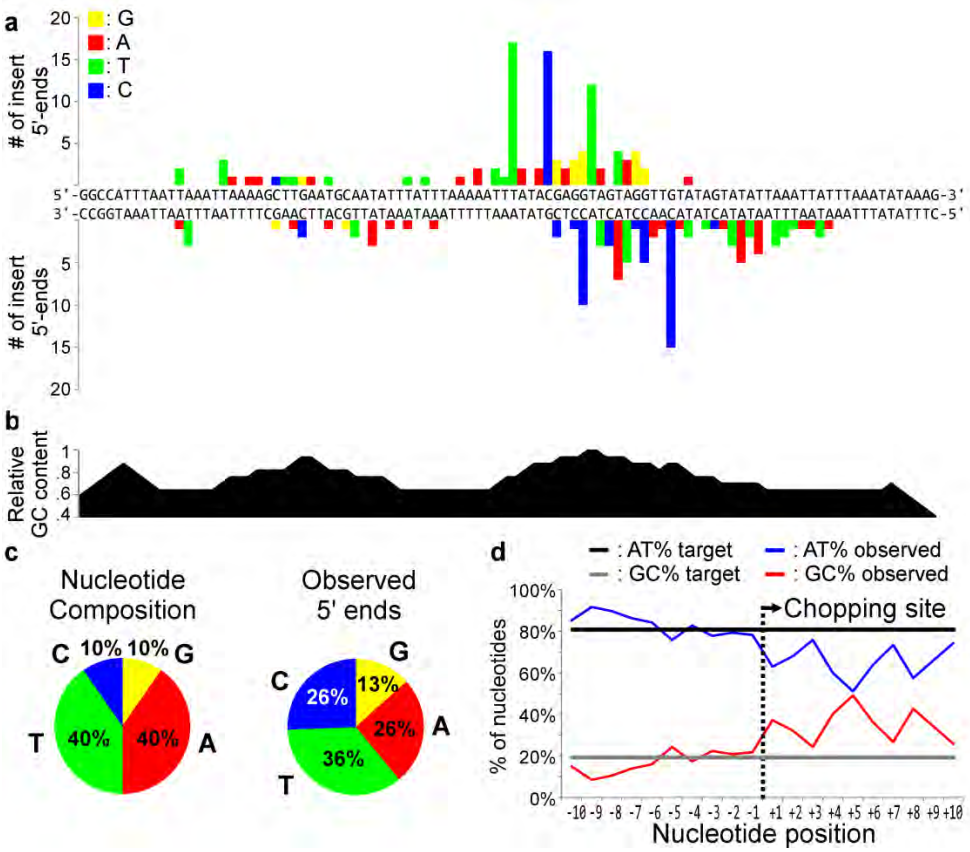


Figure 4 | Sequencing of chopped DNA reveals preferential chopping location. a, 5' ends of DNA chopping products. The sequence of the chopped dsDNA is displayed in the middle, while 5' ends of chopping products are counted in the graphs above (FW strand) and below (RV strand) the dsDNA sequence. Each target strand was chopped directly upstream of each indicated 5' end. **b**, Relative GC-content of the target displayed in panel a. Note that chopping occurred mainly in the 5' ends of GC-rich regions. **c**, Nucleotide composition of the target DNA (%) and 5-end nucleotides observed for chopped DNA (%). **d**, Average AT and GC content 10 bp upstream and 10 bp downstream chopping sites. The chopping site is indicated with a dashed line. Note that the GC content downstream the chopping site is higher than would be expected from the random GC content.

Although there seems to be a small preference for generating 5'-end deoxycytidines (26% observed versus 10% randomly expected; **Fig. 4c**), the chopping position is not strictly determined by recognition of specific sequences. Nevertheless, chopping positions appear enriched at the 5' border of GC-rich stretches within the AT-rich DNA target (**Fig. 4a**, **b**). When the sequences around each individual chopping site are aligned, it becomes clear that sequences upstream the chopping site are considerably lower in GC-content compared to sequences downstream the cleavage site (**Fig. 4d**). In contrast, such bias is not observed for the chopped GC-rich DNA target which has no low GC-content regions (**Fig. S2**). Possibly the requirement for AT-rich dsDNA upstream the chopping site reflects the aforementioned requirement for DNA unwinding for DNA chopping.

Model for DNA chopping by *Tt*Ago

Based on the observations described in this chapter, we can propose a hypothetical model for DNA chopping. This model is based on the factors required for DNA chopping: (I) double stranded DNA targets; (II) catalytic activity of *Tt*Ago; and (III) intrinsic (partial) instability of the target, preferentially downstream the chopping site.

The PIWI domains of Agos are phylogenetically related to RNase H, which can cleave the RNA strand of an RNA/DNA duplex. RNase H does not require 5'-phosphorylated guides for target cleavage, and cannot cleave ssRNA targets. It binds DNA/RNA duplexes by interacting with the backbone phosphates of the DNA and RNA strands and with 2'-OH groups of the RNA strand. By specific 2'-OH interactions and 2'-OH exclusions, it binds the RNA and DNA strands, respectively, in two separate binding grooves. The catalytic DEDD tetrad is located in the RNA binding groove and cleaves the RNA strand upon binding (**Fig. 5a**; reviewed in [281,282]).

Long pAgos like *Tt*Ago encompass four domains (N-PAZ-MID-PIWI; **Fig. 5b**; reviewed in [230]). During canonical target cleavage [176], *Tt*Ago binds the 5' end of an siDNA by a binding pocket in the MID domain, while the 3' end of the guide is bound by the PAZ domain (**Fig. 5b**). Much like RNase H, multiple amino acid residues form hydrogen bonds and salt bridges with backbone phosphates in the seed of the siDNA [168]. Target binding initiates at the seed, located at the 5' end of the guide, and zippers towards the 3' end of the guide [167,168,169,176]. As the N domain prevents duplex formation beyond nucleotide 16 of the guide [169], the guide is released from the PAZ domain when binding targets longer

than 15 nucleotides. This release is associated with conformational changes that result in the correct assembly of the active site residues cleavage of the target strand [176].

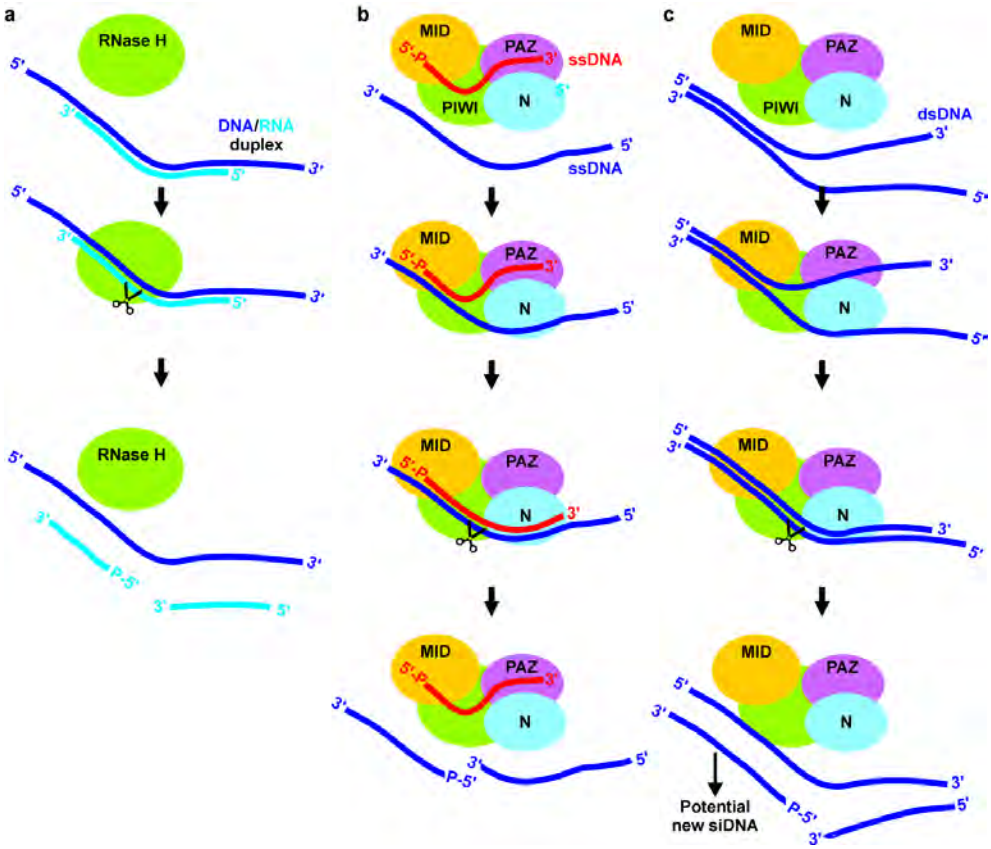


Figure 5 | Models for target cleavage by RNase H and *Tt*Ago. a, RNA degradation by RNase H. RNase H binds DNA/RNA duplexes by interacting with backbone phosphates and RNA 2'-OH groups of the bound duplex, and eventually cleaves the RNA strand. **b**, Canonical target cleavage by *Tt*Ago. *Tt*Ago binds a 5'-P guide, of which the 5' end and the 3' end are fixed in the MID and PAZ domains, respectively. Target binding initiates at the seed segment of the guide, after which zippering in the direction of the 3' end of the guide takes place. The 3' end of the guide is released from the PAZ domain and at the same time conformational changes take place, resulting in correct positioning of the catalytic residues and eventually in target cleavage. **c**, Proposed model for DNA chopping. *Tt*Ago binds the stable duplex part of the partially unwound dsDNA target with at least its PIWI domain, similar to seed-duplex binding during canonical target cleavage. Nucleation of the duplex DNA results in conformational changes that result in chopping of the secondary target strand.

Taken into account that chopping preferentially takes place at the 5' end of GC-rich regions, it is possible to predict a hypothetical model for DNA chopping (**Fig. 5c**). Like

RNase H, guide-free *Tt*Ago is unable to cleave a single stranded target, but does cleave a single strand from a duplex nucleic acid (DNA/DNA). Stable duplexes are not chopped by *Tt*Ago, possibly because duplex binding is prevented by steric hindrance of the N domain [176]. Upon unwinding of the DNA duplex (due to instability, or possibly after DNA unwinding by other proteins *in vivo*), the PIWI domain of *Tt*Ago binds the more stable part of the duplex DNA independent of the MID binding pocket. Most likely, it does this by interacting with the backbone phosphates in the seed segment of the DNA strands. The PAZ domain might contribute to stabilization of the unwound DNA conformation by binding the backbone phosphates of the primary strand. This strand can stick out the PAZ domain as the 3' end binding pocket is accessible to the outside [167]. In line with that theory, *Tt*Ago and *Pf*Ago have been demonstrated to cleave targets with guides too long to allow canonical binding of the 3' end of the guide by the PAZ domain [167,277].

As the catalytic tetrad of *Tt*Ago is essential for DNA chopping, it is likely that the canonical mechanism for *Tt*Ago target cleavage is used for chopping. Due to steric hindrance by the N domain, DNA duplex re-hybridization requires release of the primary strand from the PAZ domain, which in the canonical cleavage mechanism is associated with the conformational changes required for target cleavage [176]. We predict that These conformational changes result in nicking of the secondary strand of the DNA duplex: DNA chopping. Multiple chopping events eventually result in the complete degradation of a dsDNA target, forming new siDNA guides.

Discussion

*Tt*Ago and *Pf*Ago play a role in host defense by interfering with invading nucleic acids [27,267,277,278]. An important remaining question was how these pAgos acquire their siDNA guides. It has previously been suggested that for guide generation, pAgos rely on common host factors or that pAgos can generate their guides themselves [27]. In the present study we demonstrated that *Tt*Ago is actively involved in generating its own siDNAs from dsDNA molecules. Although the exact mechanism of guide acquisition remains to be uncovered, it is clear that dsDNA targets should be (partially) unwound to allow *Tt*Ago to chop them in fragments. We have demonstrated that *in vitro*, unwinding can be facilitated by low salt concentrations and by low GC-content and/or partially mismatched DNA. *In vivo*, other proteins that transiently unwind DNA (such as RNA and

DNA polymerases) also may stimulate DNA chopping. The observed chopped DNA fragments have the characteristics of *in vivo* acquired siDNAs: 5'-phosphorylated small DNAs that are 13 to 25 nucleotides in length.

We hypothesize that by using DNA chopping alone, guide-free pAgos can mediate a certain degree of DNA interference. The interfering effect is enhanced by the siDNAs generated during dsDNA chopping, as these fragments were demonstrated to have the potential to guide *Tt*Ago to bind and cleave complementary DNA targets. It remains unknown how pAgos are able to discriminate invader DNA from genomic DNA, but it is not unlikely that this discrimination takes place in the DNA-chopping stage, as *in vivo* it has been demonstrated that *Tt*Ago preferentially acquires siDNAs from invader plasmid DNA.

Although there are some uncertainties in the presented working model (**Fig. 5**), the research presented here provides the first proof that *Tt*Ago independently of other proteins chops dsDNA targets to generate siDNA guides. This implies that at least some pAgos are functional as stand-alone proteins, and it indicates that the requirement for interactions with other proteins, which are essential in RNAi pathways in which eAgos participate, have developed later in evolution.

Acknowledgements
We thank Stan J.J. Brouns for critical reading of the manuscript. This work was financially supported by a grant from the Netherlands Organization of Scientific Research (NWO) to JvdO (NWO-TOP, 854.10.003).

Author Contributions
In vitro chopping experiments were designed by DCS under supervision of JvdO. Proteins were purified and chopping experiments were performed by DCS under supervision of JvdO and by EMT and YZ under supervision of DCS. Preparation of samples for sequencing was performed by EMT and DCS under supervision of JvdO. DCS wrote the manuscript. All authors read and approved the manuscript.

Author information
Correspondence should be addressed to john.vanderoost@wur.nl.

Experimental procedures

Protein purification

Guide-free *Tt*Ago and *Tt*AgoDM were expressed in *E. coli* KRX and purified by Strep(II)-tag affinity purification as described previously [27]. Affinity columns were washed with 9 column volumes of washing buffer I (20 mM Tris/HCl pH 8, 1 M NaCl, 2 mM $MgCl_2$), and subsequently with 9 column volumes of washing buffer II (20 mM Tris/HCl pH 8, 1 M NaCl, 2 mM $MnCl_2$). Proteins were eluted in elution buffer (20 mM Tris/HCl pH 8, 1 M NaCl, 2.5 mM biotin, 2 mM $MnCl_2$) and diluted to 5 μM in the same buffer.

Plasmid DNA chopping

For plasmid cleavage assays, 50 ng μl^{-1} plasmid DNA, 5 μM *Tt*Ago and reaction buffer (20 mM Tris/HCl pH 8 and various concentrations of NaCl) were mixed in a 1:1:2 ratio (plasmid DNA:*Tt*Ago:reaction buffer). Samples were incubated for 16 h at 65 °C. 50 mM $CaCl_2$ and proteinase K solution (Ambion) were mixed 1:1 and added 1:4 to the reaction samples (proteinase K/CaCl2 solution:reaction sample), and the samples were incubated for 4 h at 65 °C. Plasmid DNA was mixed with 6X DNA Loading Dye (Thermo Scientific) and resolved on 0.8% agarose gels. Gels were stained using SYBR gold Nucleic Acid Gel Stain (Invitrogen) and nucleic acids were visualized using a G:BOX Chemi imager (Syngene).

dsDNA target generation

Short dsDNA targets were generated by annealing two 98 nucleotide long synthetic oligonucleotides. Two (partially) complementary oligonucleotides (see Table S1 and S2) were mixed in 2X STE buffer (0.1 M NaCl, 10 mM Tris/HCl pH 8, 1 mM EDTA) in a 1:1:2 ratio (oligo 1:oligo 2:STE). Samples were incubated at 95 °C for 5 minutes in a heat block. Then, the heat block was switched of and let to cool to room temperature. Samples were stored at -20 °C. To remove un-annealed oligonucleotides, the samples were resolved on 15% native polyacrylamide gels and visualized by UV shadowing with a TLC aluminum Silica gel 60 F_{254} plate (Merck). Bands containing the dsDNA were cut out the gel and ground. Elution Buffer (0.5 M NaAc pH 5.2, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% SDS) was added and samples were incubated for 16 h at room temperature. After diffusion, dsDNA was cleaned by ethanol precipitation and resolved in MQ H_2O to a final concentration of 2 ng μl^{-1}.

dsDNA chopping

For dsDNA chopping, 2 ng μl^{-1} dsDNA, 5 μM *Tt*Ago and reaction buffer (20 mM Tris/HCl pH 8 and 250 mM NaCl (unless indicated otherwise)) were mixed in a 1:1:2 ratio (plasmid DNA:*Tt*Ago:reaction buffer). Samples were incubated for 16 h at 65 °C. 50 mM $CaCl_2$ and proteinase K solution (Ambion) were mixed 1:1 and added 1:4 to the reaction samples (proteinase K/CaCl2 solution:reaction sample), and the samples were incubated for 4 h at 65 °C. Incubated samples were mixed in a 1:1 ratio with Loading Buffer (95% (deionized) formamide, 5 mM EDTA, 0.025% SDS, 0.025% Bromophenol blue and 0.025% xylene cyanol), heated for 10 min at 95 °C and resolved on 20% denaturing polyacrylamide gels. Gels were stained using SYBR gold Nucleic Acid Gel Stain (Invitrogen) and nucleic acids were visualized using a G:BOX Chemi imager (Syngene).

[γ-^{32}P] ATP labeling of chopped DNA

2X reaction buffer, 5 μM *Tt*Ago or 5 μM *Tt*AgoDM and 2 ng μl^{-1} dsDNA target were mixed in a 2:1:1 ratio (buffer:protein:target). Reaction samples were incubated for 16 h at 65 °C. 50 mM $CaCl_2$ and proteinase K solution (Ambion) were mixed 1:1 and added 1:4 to the reaction samples (proteinase K/CaCl2 solution:reaction sample), and the samples were incubated for 4 h at 65 °C. Samples were diluted 1:1 with MQ H_2O. PCI solution (Roth) was added 1:1 and samples were vortexed. Samples were centrifuged for 15 min in a table top centrifuge at max speed. The upper phase (containing the chopped DNA) was transferred to new tube. 3 M sodium acetate pH 5.2 was added in a 1:9 (sodium acetate:sample) and 10 μl 0.5% linear polyacrylamide was added as carrier. Nucleic acids were precipitated by adding 99% ethanol in a 2:1 ratio (ethanol:sample) and incubation for two nights at -20 °C. Samples were centrifuged in a table top centrifuge for 20 min at 4 °C at maximum speed. 500 μl 70% -20 °C ethanol was added and samples were centrifuged in a table top centrifuge for 20 min at 4 °C at maximum speed. The supernatant was removed and samples were dried 5 min in a 50 °C heat block. Precipitated nucleic acids were resolved in 50 μl MQ H2O and used for [γ-^{32}P] labeling or used in sequencing reactions. Nucleic acids were [γ-^{32}P] ATP labeled with T4 PNK (Thermo Scientific) according to the manual provided by the manufacturer. Free [γ-^{32}P] ATP label was removed using G-25 Sephadex spin columns. Samples were treated with RNase A (Thermo Scientific) or RQ1 RNase-free DNase I (Promega) for 1 h at 37 °C. After nuclease treatment, samples were mixed in a 1:1 ratio with Loading Buffer, heated for 10 min at 95 °C and resolved on 20% denaturing polyacrylamide gels. Radioactivity was captured from gels using phosphor screens. For TOPO cloning, 17 μl purified chopped DNA was mixed with 2 μl 10X Taq buffer with KCl and $MgCl_2$ (Thermo

Scientific), 0.5 µl 10 mM dNTP mix and 0.5 µl Klenow fragment (Thermo Scientific). Samples were incubated at 37 °C for 10 min. The reaction was stopped by heating at 75 °C for 10 min. 1 µl Taq polymerase was added and the sample was incubated for 10 min at 72 °C. This sample was used for TOPO cloning according to the manual provided by the manufacturer. Vectors were transformed to NEB5α *E. coli* cells according to the protocol provided by the manufacturer. For pJET1.2 cloning, 7 µl purified chopped DNA was mixed with 10 µl 2X reaction buffer (Thermo Scientific) and 1 µl DNA blunting enzyme (removes 3' overhangs and fills in 5' overhangs; Thermo Scientific). The sample was incubated for 5 min at 70 °C. For ligation into pJET1.2, 1 µl pJET1.2 (Thermo Scientific) and 1 µl T4 DNA Ligase (Thermo Scientific) was added. The sample was incubated for 30 min at 22 °C, and transformed to NEB5α *E. coli* cells according to the protocol provided by the manufacturer. For sequencing, colonies were picked and used to inoculate LB-Amp filled 96-well plates provided by GATC Biotech. Colonies were sent for sequencing by GATC Biotech using primer BG5493.

Utilizing chopped DNA as guide

*Tt*Ago or *Tt*AgoDM and purified chopped dsDNA were mixed in a 5:1 ratio in 2x reaction buffer. Reaction samples were incubated for 30 min at 65 °C, after which a ssDNA target was added in a 5:1:1 ratio (*Tt*Ago:chopped dsDNA:ssDNA target). Targets for AT-rich chopped DNA: BG4262 (FW) and BG4263 (RV). Targets for GC-rich chopped DNA: BG4264 (FW) and BG4265 (RV). See Table S2 for sequences. Reaction samples were incubated for 60 min at 65 °C after which Loading Buffer was added in a 2:1 ratio (loading dye:sample). Samples were heated for 10 min at 95 °C and resolved on 20% denaturing acrylamide gels. Gels were stained using SYBR gold Nucleic Acid Gel Stain (Invitrogen) and nucleic acids were visualized using a G:BOX Chemi imager (Syngene).

Supplementary Figures and Tables

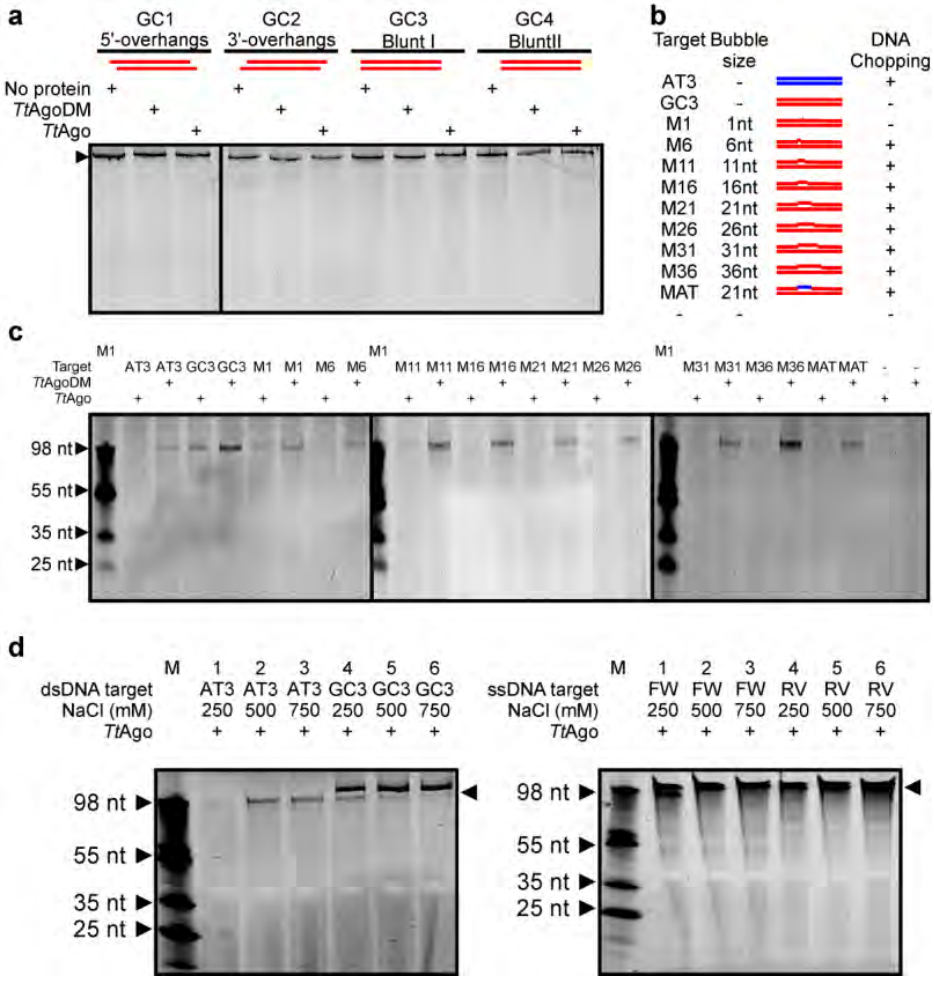


Figure S1 | Chopping of various targets by guide-free *Tt*Ago. a, GC-rich DNA is not chopped. 15% denaturing polyacrylamide gel loaded with GC-rich dsDNA targets incubated without protein (Blanco), with *Tt*Ago or with *Tt*AgoDM. b, Overview of targets with internal mismatched regions used in chopping assays. -: no target added. c, dsDNA with internal mismatches is chopped. 15% denaturing polyacrylamide gel loaded with GC-rich dsDNA targets with internal mismatches incubated with *Tt*Ago or with *Tt*AgoDM. M1: ssDNA marker. -: no target added. d, Effect of NaCl concentration on dsDNA target chopping. 15% denaturing polyacrylamide gel loaded with targets C (AT-rich) and G (GC-rich) incubated with *Tt*Ago at various NaCl concentrations. M: ssDNA marker. The black triangle right of the gel indicates the intact target. e, ssDNA is not chopped. 15% denaturing polyacrylamide gel loaded with AT-rich ssDNA targets BG4262 (FW) and BG4263 (RV) incubated with *Tt*Ago at various NaCl concentrations. M: ssDNA marker. The black triangle right of the gel indicates the intact target. For a detailed overview of the targets, see Table S1 and S2.

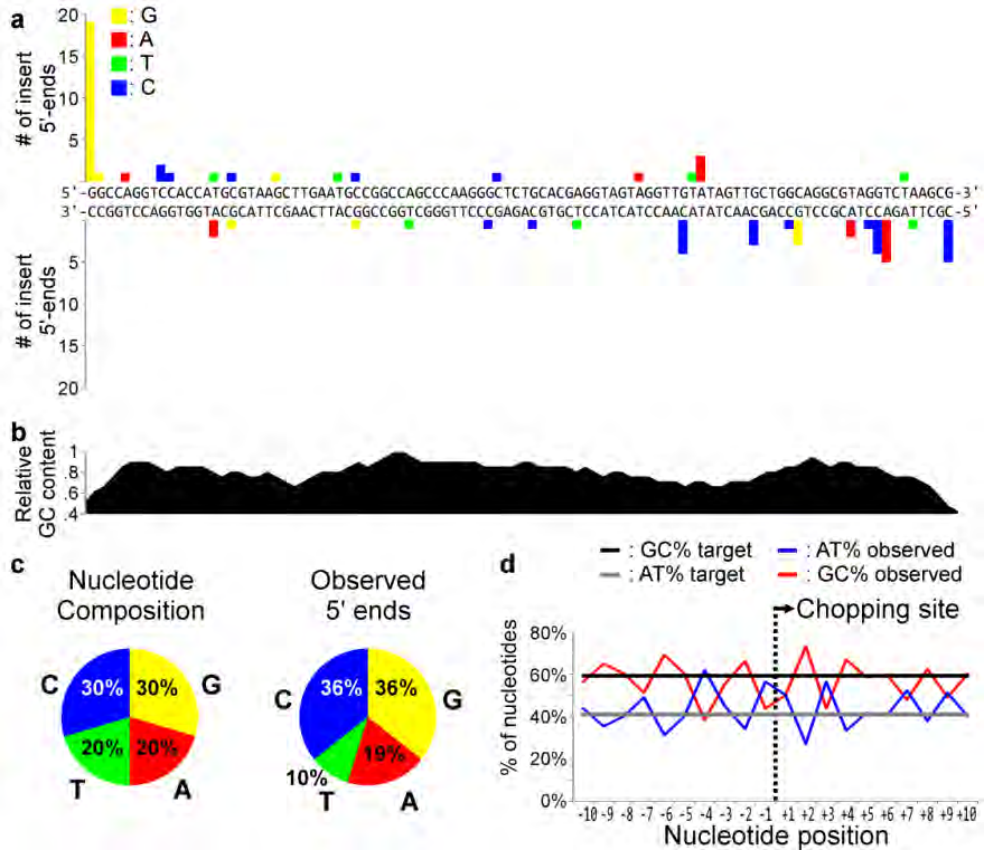


Figure S2 | Sequencing of chopped GC-rich DNA. a, 5' ends of DNA chopping products. The sequence of the chopped dsDNA is displayed in the middle, while 5' ends of chopping products are counted in the graphs above (FW strand) and below (RV strand) the dsDNA sequence. Each target strand was chopped directly upstream of each indicated 5' end. **b**, Relative GC-content of the target displayed in panel a. Note that chopping occurred mainly in the 5' ends of GC-rich regions. **c**, Nucleotide composition in the target DNA (%) and 5'-end nucleotides observed for chopped DNA (%). **d**, Average AT and GC content 10 bp upstream and 10 bp downstream chopping sites. The chopping site is indicated with a dashed line.

Table S1 | Plasmids and dsDNA targets

Target	Primers*	T_M (°C)**	Description
pWUR704	Insert: BG4262, BG4263	N/A	Target plasmid with AT-rich insert [27]
pWUR705	Insert: BG4264, BG4265	N/A	Target plasmid with GC-rich insert [27]
AT1	BG4262, BG4263	64.3/88.3	AT-rich, 5' overhangs
AT2	BG4723, BG4724	64.3/88.3	AT-rich, 3' overhangs
T3	BG4262, BG4724	66/89.9	AT-rich, Blunt end I
AT4	BG4723, BG4263	65.2/89.1	AT-rich, Blunt end II
GC1	BG4264, BG4265	81.3/105.2	GC-rich, 5' overhangs
GC2	BG4725, BG4726	81.3/105.2	GC-rich, 3' overhangs
GC3	BG4264, BG4726	82.3/106.3	GC-rich, Blunt end I
GC4	BG4725, BG4265	81.5/105.5	GC-rich, Blunt end II
GC5	BG4264, BG4727	79.7/103.5	GC-rich, Flaps
M1	BG4264, BG4729	82.1/106.2	GC-rich, 1 bp mismatch
M6	BG5268, BG4726	81.2/105.4	GC-rich, 6 bp mismatch
M11	BG5269, BG4726	80.3/104	GC-rich, 11 bp mismatch
M16	BG5270, BG4726	79.2/103.1	GC-rich, 16 bp mismatch
M21	BG4728, BG4726	78.5/102.4	GC-rich, 21 bp mismatch
M26	BG5271, BG4726	77.8/101.3	GC-rich, 26 bp mismatch
M31	BG5272, BG4726	77.5/100.8	GC-rich, 31 bp mismatch
M36	BG5273, BG4726	77.2/100.2	GC-rich, 36 bp mismatch
MAT	BG5274, BG4726	78.5/102.4	GC-rich, AT middle, 21 bp mismatch

*: For complete oligonucleotide sequence, see Table S2. **: Melting temperature (T_M) calculated for the complementary region(s) of each target using OligoCalc [283]. T_M is calculated for Basic and for Salt adjusted (250 mM NaCl) conditions.

Table S2 | Oligonucleotides

	Primer	Sequence
Targets	BG4262	GGCCATTTAATTAAATTAAAAGCTTGAATGCAATATTTATTTAAAAATTTA TACGAGGTAGTAGGTTGTATAGTATATTAAATTATTTAAATATAAAG
	BG4263	TCGACTTTATATTTAAATAATTTAATATACTATACAACCTACTACCTCGTA TAAATTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAAT
	BG4264	GGCCAGGTCCACCATGCGTAAGCTTGAATGCCGGCCAGCCCAAGGGCTCT GCACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG4265	TCGACGCTTAGACCTACGCCTGCCAGCAACTATACAACCTACTACCTCGT GCAGAGCCCTTGGGCTGGCCGGCATTCAAGCTTACGCATGGTGGACCT
	BG4723	ATTTAATTAAATTAAAAGCTTGAATGCAATATTTATTTAAAAATTTATACG AGGTAGTAGGTTGTATAGTATATTAAATTATTTAAATATAAAGTCGA
	BG4724	CTTTATATTTAAATAATTTAATATACTATACAACCTACTACCTCGTATAAA TTTTTAAATAAATATTGCATTCAAGCTTTTAATTTAATTAAATGGCC
	BG4725	AGGTCCACCATGCGTAAGCTTGAATGCCGGCCAGCCCAAGGGCTCTGCAC GAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCGTCGA
	BG4726	CGCTTAGACCTACGCCTGCCAGCAACTATACAACCTACTACCTCGTGCAG AGCCCTTGGGCTGGCCGGCATTCAAGCTTACGCATGGTGGACCTGGCC
	BG4727	CGCTTAGACCTACGCCTGCCAGCAACTATACAACCTACTACCTCGTGCAG AGCCCTTGGGCTGGCCGGCATTCAAGCTTTGCGTACCACCTGGACCGG
	BG4728	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTCGCGTTGCCGAGA CGACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG4729	CGCTTAGACCTACGCCTGCCAGCAACTATACAACCTACTACCTCGTGCAG AGCCCTTGCGCTGGCCGGCATTCAAGCTTACGCATGGTGGACCTGGCC
	BG5268	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTGCCCAAGGGCTCT GCACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG5269	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTCGGGTAGGGCTCT GCACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG5270	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTCGGGTTCCCGTCT GCACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG5271	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTCGCGTTGCCGAGA CGTGCTCGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG5272	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTCGCGTTGCCGAGA CGTGCTCCATCAAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG5273	GGCCAGGTCCACCATGCGTAAGCTTGAATGCGCCGGTCGCGTTGCCGAGA CGTGCTCCATCATCCAAGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
	BG5274	GGCCAGGTCCACCATGCGTAAGCTTGAATGCATTAACTAAACCTTTAGAG TAACGAGGTAGTAGGTTGTATAGTTGCTGGCAGGCGTAGGTCTAAGCG
Sequence	BG5493	CTGCTTTAACACTTGTGCC

Chapter 6

Effect of *Thermus thermophilus* Argonaute on gene expression

Daan C. Swarts, Jasper J. Koehorst, Edze R. Westra, Peter J. Schaap, John van der Oost

Adapted from:
'Effect of Argonaute on gene expression in *Thermus thermophilus*'
PLoS ONE. April 2015. Volume 10, Issue 4: e0124880

Abstract

Eukaryotic Argonaute proteins mediate RNA-guided RNA interference, allowing both regulation of host gene expression and defense against invading mobile genetic elements. Recently, it has become evident that prokaryotic Argonaute homologs mediate DNA-guided DNA interference, and play a role in host defense. Argonaute of the bacterium *Thermus thermophilus* (*Tt*Ago) targets invading plasmid DNA during and after transformation. Using small interfering DNA guides, *Tt*Ago can cleave single and double stranded DNAs. Although *Tt*Ago additionally has been demonstrated to cleave RNA targets complementary to its DNA guide *in vitro*, RNA targeting by *Tt*Ago has not been demonstrated *in vivo*.

To investigate if *Tt*Ago also has the potential to control RNA levels, we analyzed RNA-seq data derived from cultures of four *T. thermophilus* strain HB27 variants: wild type, *Tt*Ago knockout (Δ*ago*), and either strain transformed with a plasmid. Additionally we determined the effect of *Tt*Ago on expression of plasmid-encoded RNA and plasmid DNA levels.

In the absence of exogenous DNA (plasmid), presence or absence of *Tt*Ago had no effect on gene expression levels. When plasmid DNA is present, *Tt*Ago reduces plasmid DNA levels 4-fold, which results in 4-fold lower plasmid-encoded gene transcript levels. As *Tt*Ago does not lower these RNA levels further, we conclude that *Tt*Ago interferes with plasmid DNA, but not with plasmid-encoded RNA. Interestingly, the presence of *Tt*Ago stimulates expression of specific endogenous genes, but only when exogenous plasmid DNA was present. Specifically, the presence of *Tt*Ago directly or indirectly stimulates expression of CRISPR loci and associated genes, some of which are involved in CRISPR adaptation. This suggests that *Tt*Ago-mediated interference with plasmid DNA stimulates CRISPR adaptation.

Introduction

Argonaute proteins (Agos) have long been known as key players in eukaryotic RNA interference (RNAi) pathways, in which eukaryotic Ago (eAgo) uses a small single-stranded (ss)RNA guide to target ssRNA molecules (reviewed in [171,251,252]). While many RNAi pathways regulate host gene expression by targeting mRNAs, some RNAi pathways are involved in host defense (reviewed in [95,121,124]). In these pathways, Agos interfere with RNA transcripts from viruses or transposons, or with RNA viruses directly.

Prokaryotes also encode Agos (pAgos), but none of the additional proteins involved in canonical RNAi pathways [3,4,170,230]. Recently, it has become clear that pAgos are involved in mediating host defense, but in contrast to eAgos, they target DNA rather than RNA [27,177]. One of the best studied pAgos is that of *Thermus thermophilus* (*Tt*Ago), which has been characterized structurally and biochemically [27,167,168,169,176]. *T. thermophilus* is a gram-negative thermophilic bacterium that is used as model organism for genetic transformation, biotechnological applications and structural biology. *T. thermophilus* strain HB27 has a 1.9 Mb chromosome encoding 1988 genes (designated TTC0001-TTC1988) and harbors a 232 Kb mega-plasmid designated pTT27, encoding 230 genes (TT_P0001-TT_P0230).

In contrast to RNA-guided eAgos, *Tt*Ago has been demonstrated to utilize DNA guides in order to cleave single stranded (ss)DNA and double stranded (ds)DNA targets *in vitro* [27,167,168,169,176]. This allows *Tt*Ago to directly interfere with invading DNAs, lowering plasmid transformation efficiencies and intracellular plasmid content [27,267]. As *Tt*Ago preferentially acquires guides from plasmid DNA [27], and it is able to cleave ssRNA targets *in vitro*, it was predicted that *Tt*Ago also interferes with plasmid transcripts [27]. This would suggest a dual-function of *Tt*Ago, both in host defense and in gene regulation, which is akin to eAgos [171,251,252] and prokaryotic CRISPR-Cas [284], However, gene expression of *T. thermophilus* has not yet been investigated in strains in which invading DNA in the form of a plasmid was present. Here, we describe the analysis of a new RNA-seq dataset derived from *T. thermophilus* strains HB27 and HB27Δ*ago* harboring plasmid pMKPnqosGFP [285]. Although the presence of *Tt*Ago or plasmid DNA itself does not strongly affect gene expression, the presence of both results in decreased quantities of plasmid-encoded RNA transcripts and increased expression of specific genomic genes.

Results

We included previously obtained RNA-seq data from HB27 and HB27Δ*ago* [27] in our analysis in order to compare them with the new data from HB27 + plasmid (HB27+P) and (4) HB27Δ*ago* + plasmid (HB27Δ*ago*+P). The latter two strains were grown in medium containing kanamycin, selecting for plasmid maintenance. For each condition, RNA from biological triplicates was purified, sequenced and mapped, and for each gene the abundance was calculated as Fragments Per Kilobase of exon per Million fragments mapped (FPKM). RNA levels are considered to be changed significantly when the FPKM value of a set of biological triplicates differed from the FPKM value of another set of biological triplicates with $P<0.05$. We considered changes in RNA levels biologically relevant if FPKM averages of biological triplicates differed at least >4-fold from FPKM averages of another set of biological triplicates, while smaller changes were considered stochastic. RNA was purified from triplicate log phase ($OD_{600\ nm}$ of 0.5) cultures HB27, HB27Δ*ago*, HB27+P and HB27Δ*ago*+P (**Fig. 1a**). Using Prodigal 2.6 [286], 35 new open reading frames were identified of which 22 were located on the HB27 chromosome (tagged TTCX01-TTCX22) and 13 on the mega-plasmid pTT27 (TTPX01-TTPX13). Of these new genes, 15 encode proteins of which the function can be predicted based on (partial) homology to other proteins. Furthermore, 16 encode proteins that show (partial) similarity to hypothetical proteins, whereas four encode proteins that share no significant similarity to other proteins in the current NCBI database. The open reading frames and predicted functions of the proteins they encode are listed in **Table S1.**

Absence of *Tt*Ago results in small stochastic changes in *T. thermophilus* gene expression

The *ago* knockout in *T. thermophilus* strain HB27 has previously been demonstrated to result in small pleiotropic changes in gene expression (<4-fold change for most genes) [27], and this was confirmed in our new analyses of the same dataset (**Table S2**). Stochastic changes in gene expression include 59 genes which are >2-fold up-regulated and 35 genes which are >2-fold down-regulated in HB27Δ*ago* compared to HB27 (**Table S2**). Besides these small differences, >4-fold change in expression was observed for specific genes (**Table S2**). As expected, we observe no expression of the gene encoding *Tt*Ago (TT_P0026) in Δ*ago* strains, and low levels of *Tt*Ago expression in wild type strains (FPKM<150; **Fig. 1b**). In agreement with this observation, evidence for expression of (Strep(II)-tagged) *Tt*Ago protein encoded by the knock-in gene at the same genomic location has previously been

demonstrated [27]. The *ago* knockout resulted in 3 to 6-fold lower RNA levels mapped against genes located near and on the same strand as *ago* (**Fig. 1b**). These changes are most likely polar effects caused by *ago* deletion.

RNA mapped against two other genes is lowered 3-fold and 5-fold in HB27Δ*ago*: TTC1213 (1-pyrroline-5-carboxylate dehydrogenase) and TTC1241 (predicted acyl-amino acid-releasing enzyme). These genes, as well as the genes located near *ago* on the genome, are also down-regulated in HB27Δ*ago*+P compared to HB27+P. In addition, a predicted operon encoding a branched-chain amino acid transport system (TTC0333-TTC0343) is up-regulated (3 to 5-fold increase in RNA levels) in HB27Δ*ago* compared to HB27. This operon encodes a system homologous to the Liv ABC transporter system, which transports the amino acids leucine, isoleucine, valine, threonine and alanine in an ATP-dependent manner. The same set of genes is only moderately up-regulated (most genes <2-fold change) in HB27+P compared to HB27Δ*ago*+P. A functional link between these genes and the *ago* knockout is not obvious. The levels of these RNAs are affected in both HB27Δ*ago* and HB27Δ*ago*+P, suggesting that *Tt*Ago affects these RNA levels directly or indirectly.

Comparison of *T. thermophilus* HB8Δ*ago* and HB27Δ*ago*

A recent publication describes the differences in RNA expression between *T. thermophilus* strains HB8 (**Fig. 1c**) and HB8Δ*ago* [287]. The chromosomes of *T. thermophilus* strains HB8 and HB27 are highly conserved, while their mega-plasmids pTT8 and pTT27, which encode *ago* and most Cas genes, show a higher degree of divergence [288]. RNA was purified from log-phase cultures in both studies, but the growth medium used for HB8 cultivation [287] is slightly different from the medium we used for HB27 cultivation (**Table S3**). We compared the genes from HB27Δ*ago* and HB8Δ*ago* of which corresponding RNA levels changed >2-fold compared to the corresponding wild type strains. We found no clear correlation between the affected genes in both strains (**Table S3**, **Fig. 1d**). None of the genes of which expression changed >4-fold in HB8Δ*ago* were found to be differentially expressed in HB27Δ*ago* (**Table S3**).

Given that Agos interact with guides to bind specific complementary targets [174,230], it would be expected that *Tt*Ago strongly affects levels of specific RNAs. As the chromosomes of both HB8 and HB27 *Tt*Ago are very similar, and specific RNA levels affected by the knockout of *ago* vary greatly in these strains, it seems unlikely that *Tt*Ago targets specific

RNAs. Instead, our analysis suggests that observed differences in RNA levels are stochastic, and thus unlikely to be caused by guided *Tt*Ago activity. The observation that *Tt*Ago does not influence the transcription of genes involved in competence or host defense, suggests that *Tt*Ago only interferes with the invading DNA directly. In a recent study, the competence of HB27 and HB27Δ*ago* has been compared during natural transformation experiments and during cell-to-cell conjugation experiments with genomic *T. thermophilus* DNA. It was found that *Tt*Ago interferes with natural transformation, but not with cell-to-cell conjugation [267]. As the genes required for natural transformation are also essential for cell-to-cell conjugation, this excludes a possible indirect effect of *Tt*Ago via regulation of competence genes expression.

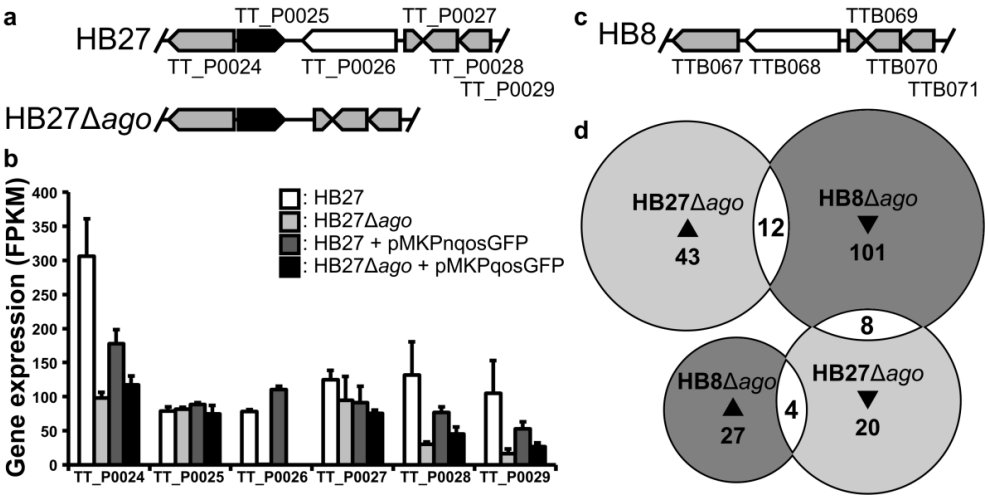


Figure 1 | Deletion of the *ago* gene results in stochastic changes in gene expression. a, Schematic representation of the gene regions encoding *Tt*Ago (TT_P0026) of *T. thermophilus* strain HB27 and HB27Δ*ago*. **b**, Schematic representation of the gene regions encoding TTB068 in *T. thermophilus* strain HB8. As no information on how the HB8 *ago* knockout was generated is available [287], HB8Δ*ago* is not displayed. HB8 genes colored grey and white are homologous to the HB27 genes indicated in Fig. 1a. **c**, Expression of genes located near *ago* (TT_P0026) on the genome. Expression values are given in Fragments Per Kilobase of exon per Million fragments mapped (FPKM). **d**, Overlap in >2-fold up-regulated (▲) and >2-fold down-regulated (▼) homologous genes in HB27Δ*ago* relative to HB27, and HB8Δ*ago* relative to HB8.

Presence of plasmid DNA results in changes in gene expression only if *Tt*Ago is present

To investigate the effect of the presence of plasmid DNA on gene expression, we compared RNA isolated from HB27Δ*ago* to that from HB27Δ*ago*+P. No significant (P<0.05) >4-fold changes in RNA levels were observed. Furthermore, the presence of plasmid DNA did not

result in significant (P<0.05) >2-fold change in expression of host defense genes (**Table S2**). Combined, these data suggest that presence of invading nucleic acids in the form of plasmid DNA does not result in differentiated gene expression in HB27Δ*ago*. This contrasts with the presence of another invader, lytic phage ϕYS40 [289] in HB8, which results in up-regulation of a plethora of host defense genes. These genes encode (amongst others) *Tt*Ago, the *T. thermophilus* Type I-E (not encoded in HB27), Type III-A and Type III-B CRISPR-Cas systems, as well as multiple other Cas genes scattered over the HB8 genome [289]. In summary, although phage infection triggers host defense-response pathways in *T. thermophilus*, the presence and replication of plasmid DNA does not trigger host defense responses. This is presumably because defense pathways are costly to induce [290], and are most beneficial in the context of parasitic infections [291], such as by lytic phages. In contrast, plasmids are far less detrimental to the host and often confer a fitness benefit [292], making it unnecessary to induce these pathways during plasmid invasion.

In contrast, when comparing WT (Ago-encoding) strains with and without plasmid DNA, we observed significant (P<0.02) >4-fold increase of RNA levels mapped to specific genes (HB27+P compared to HB27; **Table 1**, **Table S2**). Corresponding genes, difference in RNA levels, motifs and predicted functions of the proteins they encode are listed in **Table 1** and **Fig. 2**.

There seems to be no clear link between the functions of the up-regulated genes. Interestingly however, many of these genes, especially genes up-regulated >5-fold, are located directly downstream and in the same orientation as various CRISPR loci (**Table 1**, **Figure 2a**). Predicted gene TTPX09 is located in a CRISPR locus, and is unlikely to encode a functional protein. Furthermore two putative transposases (TT_P0099 and TTC1169) of which expression appears up-regulated, are located near CRISPR loci on the genome, but in reverse orientation. TT_P0211 and TTPX12 are located directly downstream each other in a predicted operon. These and three other genes located on the chromosome (TTC0310, TTC0311 and TTC0399) appear to have no link with CRISPR loci. As we observe only elevated RNA levels under these conditions, it is highly unlikely that *Tt*Ago interferes with RNA, as this would lower RNA levels. Nevertheless, the fact that these genes are up-regulated only under conditions where both *Tt*Ago and plasmid DNA are present, suggests that *Tt*Ago directly or indirectly influences expression of these genes.

Table 1: Genes differentially expressed in HB27+P compared to HB27

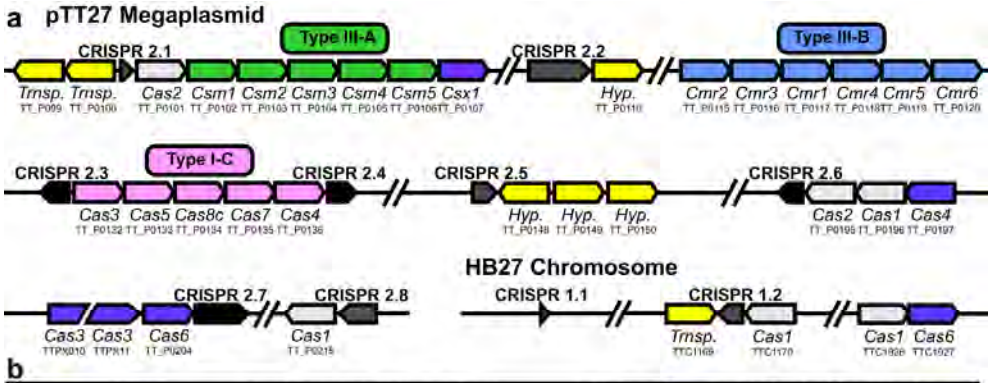
Gene	Fold change*	Motif(s) (derived from KEGG)	(Predicted) function	Located near CRISPR locus?
TT_P0099	4.3	DDE 3	**Transposase	Upstream CRISPR 2.1 (reverse orientation)
TT_P0101	27.9	CRISPR Cas2	Cas2; involved in CRISPR-adaptation	Directly downstream CRISPR 2.1
TT_P0110	8.1	DUF1887	Hypothetical protein **Csx1	Directly downstream CRISPR 2.2
TT_P0149	27.1	C-terminal AAA-associated	Hypothetical protein	Directly downstream CRISPR 2.5
TT_P0150	4.0	ABM DUF1330 Dehydratase-heme	Hypothetical protein	Directly downstream CRISPR 2.5
TT_P0211	6.6	AAA 14 DUF4143 Many HTH motifs MopB	**ATPase **Cas1 **Transposase **Transcription regulation	Downstream CRISPR 2.7 (reverse orientation)
TTPX09	5.9	-	Putative gene	Located in CRISPR 2.6
TTPX12	6.6	-	Putative gene	(Far) downstream CRISPR 2.7 (reverse orientation)
TTC0310	4.2	PAPS reduct	**Phosphoadenosine phosphosulfate reductase	No
TTC0311	4.3	NAD binding 7 CysG dimeriser Sirohm synth M	**Uroporphyrin-III C-methyltransferase	No
TTC0399	4.1	-	Hypothetical protein	No
TTC1169	12.6	DEDD Tnp IS110 Transposase 20 Helix-Hairpin-Helix (HHH)	**Transposase **DNA binding	Directly downstream CRISPR 1.2 (reverse orientation)

*: Fold-change increase in RNA levels in HB27+P compared to HB27. For all changes $P<0.02$. **: Function predicted based on domains and similarity to other genes.

Presence of both plasmid DNA and *Tt*Ago results in up-regulation of crRNA expression

As many genes that are up-regulated in HB27+P are located on the genome near CRISPR loci, we further investigated expression of *cas* genes and CRISPR loci. Mega-plasmid pTT27 encodes complete Type I-C, III-A and III-B CRISPR-Cas systems (**Fig. 2a**), and multiple scattered *cas* genes (two *cas1*, two *cas2*, one *cas4*, one *cas6* and a *cas3* gene with an internal

frameshift (TTPX10 and TTPX11; **Fig. 2a**, **Table S1**). pTT27 additionally encodes eight CRISPR arrays (**Fig. 2a**, **b**). The HB27 chromosome encodes two *cas1* genes and a *cas6* gene, as well as two CRISPR loci (**Fig. 2a**, **b**). Besides TT_P0101 (encoding Cas2), no *cas* genes appear differentially expressed (**Table S2**). This is striking, as the up-regulated *cas2* is located directly upstream of the predicted operon encoding the Type III-A CRISPR-Cas system (**Fig. 2a**).



b

Locus	Location	Representative repeat sequence (5' to 3')	Repeat type	# of Spacers	Fold change in expression HB27+P/HB27
Megaplasmid pTT27					
CRISPR 2.1	92158-92424	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	3	
CRISPR 2.2	100454-101620	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	15	
CRISPR 2.3	134658-132212	GTTGCACCGGCCCGAAAGGGCCGGTGAGGATTGAAAC	III	7	
CRISPR 2.4	138645-139200	GTTGCACCGGCCCGAAAGGGCCGGTGAGGATTGAAAC	III	7	
CRISPR 2.5	153106-153599	GTTGCAAGGGATTGAACCCCGTAAGGGGATTGCGAC	I	6	
CRISPR 2.6	192358-191888	GTTGCAAACCTCGTTAGCCTCGTAGAGGATTGAAAC	III	6	
CRISPR 2.7	203222-204195	GTTGCAAACCTCGTTAGCCTCGTAGAGGATTGAAAC	III	13	
CRISPR 2.8	213370-212645	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	9	
HB27 Chromosome					
CRISPR 1.1	541152-541341	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	2	
CRISPR 1.2	1137118-1136637	GTTGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAC	I	6	

Figure 2 | CRISPR loci and *cas* genes encoded by *T. thermophilus* HB27. a, Schematic representation of CRISPR loci and *cas* genes encoded on mega-plasmid pTT27 and the *T. thermophilus* HB27 chromosome. Encoded protein and KEGG annotation are given below each gene. Note that the size of the illustrated genes does not correspond to their actual size. CRISPR loci with type I and III repeats are colored gray and black, respectively. Repeat types are based on [293] and should not be confused with CRISPR-Cas Types [294]. Transp: Transposase. Hyp: Hypothetical protein. **b**, Characteristics of CRISPR loci encoded by *T. thermophilus* HB27. Fold change in CRISPR RNA levels is shown for HB27+P compared to HB27.

To investigate expression of CRISPR RNA (crRNA) from CRISPR loci, we used a dataset containing only reads that are partially complementary to CRISPR repeats (**Table S4**). For most CRISPR loci, expression of crRNA is highest at the leader-proximal end of the CRISPR locus, and gradually lowered towards the leader-distal end of the CRISPR locus (**Table S4**). This observation agrees with the leader harboring the promoter for crRNA expression [295,296]. When comparing crRNA expression in the different strains,

expression of crRNAs encoded by eight CRISPR loci is strongly up-regulated in HB27+P compared to HB27 (**Fig. 2b**, **Table S4**). As some of the genes mentioned in **Table 1** are located directly downstream CRISPR loci, it appears that the expression of these genes and the presence of a CRISPR locus are linked. This suggests that either these genes are expressed from the same promoter (read-through), or alternatively that they, and possibly other up-regulated genes that are not located directly downstream CRISPR loci, are under control of the same transcriptional regulator as the up-regulated CRISPR loci. As Cas2 and CRISPR leader sequences play essential roles in the acquisition of CRISPR-Cas-mediated immunity (reviewed in [295,296]), increased expression of Cas2 and crRNAs could imply that CRISPR adaptation is activated. To investigate if *Tt*Ago enhances CRISPR adaptation, we analyzed CRISPR loci for integration of new spacers. We used a PCR-based method that previously has been demonstrated to identify spacer integration in *E. coli* cultures, if at least 0.4% of the culture integrated a spacer in the amplified CRISPR locus [297]. However, no new spacers were detected (**Fig. S1**), even when cultures were grown in absence of antibiotics, in which there is no selection for plasmid maintenance. This suggests either that under the tested conditions CRISPR adaptation is not stimulated, or alternatively that CRISPR adaptation does not confer a benefit to the host (i.e. clones with novel spacers do not increase in frequency and therefore remain undetectable), which is supported by theoretical predictions that costly acquired immunity is not likely to evolve against parasites with low virulence [291].

***Tt*Ago interferes with invader DNA but not with invader-encoded RNA**

Besides the effect on genome-encoded gene expression, *Tt*Ago has a clear effect on plasmid DNA and plasmid-encoded gene expression (**Fig. 3**, **Table S5**). It has previously been shown that *Tt*Ago interferes with intracellular plasmids, resulting in 3 to 5-fold higher plasmid contents in HB27Δago compared to HB27, even when the cultures were grown under conditions selecting for plasmid maintenance [27]. To confirm these results, we determined plasmid pMKPnqosGFP content at the time at which the RNA was isolated ($OD_{600\ nm}$ of 0.5) for strains HB27+P and HB27Δ*ago*+P (**Fig. 3a**; **Table S5**). These cultures were grown in presence of kanamycin selecting for pMKPnqosGFP maintenance. In line with previous observations [27], intracellular plasmid content was significantly ($P<0.05$) lowered ~4-fold in wild type HB27 compared to HB27Δ*ago* (**Fig. 3b**), confirming that *Tt*Ago interferes with intracellular plasmid DNA. Furthermore, we observed 2.4 to 3.8-fold lower levels of plasmid-encoded RNA in the HB27+P strain compared to the HB27Δ*ago*+P

strain (**Fig. 3c**; **Table S5**). Thus, in contrast to genomic encoded RNAs, plasmid encoded RNAs are lowered in the presence of *Tt*Ago. The ~4-fold lower plasmid content itself can explain the 2.4 to 3.8-fold decrease of plasmid encoded RNA, as there are fewer plasmid copies available for RNA expression. Thus, although DNA-guided *Tt*Ago has been shown to cleave both ssDNA and ssRNA targets *in vitro* [27,167,168,169], the data suggests that *in vivo Tt*Ago solely interferes with plasmid DNA and not with plasmid-encoded RNA.

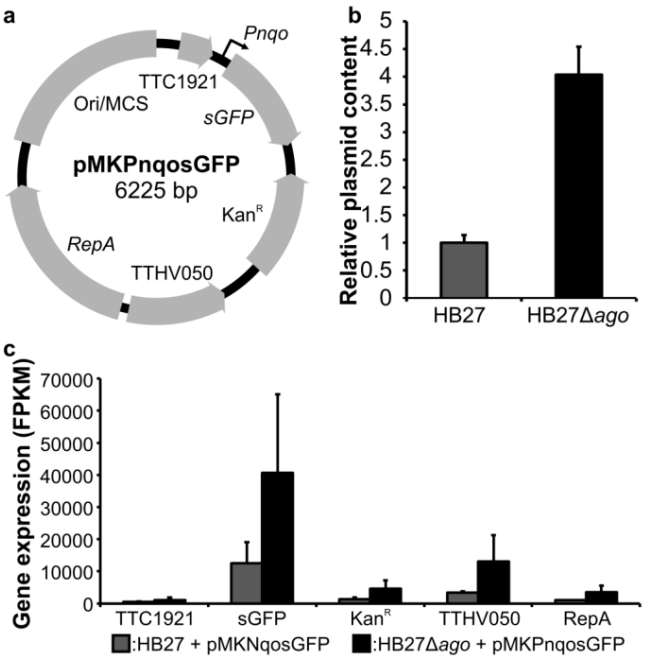


Figure 3 | Effects of *Tt*Ago on plasmid DNA and plasmid encoded RNA. a, Schematic representation of the *Escherichia coli-T. thermophilus* shuttle vector pMKPnqosGFP. Ori/MCS indicates the *E. coli* origin of replication (Ori) and a multiple cloning site (MCS). Note that cloning of this plasmid resulted in insertion of (incomplete) TTC1921 and TTHV050 genes. **b**, Relative plasmid content of *T. thermophilus* strains HB27 and HB27Δ*ago* transformed with pMKPnqosGFP. Plasmid content was calculated from the complete DNA isolated from biological triplicates at an $OD_{600\ nm}$ of 0.5. **c**, Gene expression of plasmid encoded genes. Expression values are given in Fragments Per Kilobase of exon per Million fragments mapped (FPKM).

Bacterial Ago from *Rhodobacter sphaeroides* (*Rs*Ago) associates with small RNAs and DNAs derived from extracellular sources such as plasmids, transposons and phages [177]. We therefore analyzed the effect of *Tt*Ago on RNAs encoded by transposases (**Table S6**). We observed no higher expression of transposase genes in HB27Δ*ago* strains. In contrast, compared to HB27, in HB27+P levels of RNA mapped against predicted transposases TTC1169, TT_P0099 and TT_P0211 are significantly (P<0.02) increased (12.6-fold increase, 4.3-fold increase, and 6.6-fold increase, respectively). The fact that these differences were only observed if both *Tt*Ago and the plasmid are present, suggests that transposase expression is induced under these specific conditions. TTC1169 and TT_P0099 are located near CRISPR loci on the genome, suggesting their increased expression is a

result of up-regulation of these CRISPR loci. We only analyzed RNA levels, and although *Tt*Ago appears not to be interfering with RNA directly (see above), we cannot rule out that *Tt*Ago interferes with transposons at the DNA level (for example during the extrachromosomal step of their life-cycle).

Discussion

Only small stochastic changes in gene expression are observed when comparing wild type *Thermus thermophilus* and the derived Δ*ago* mutant. This implies that *Tt*Ago, in contrast to eAgos, is not involved in regulation of host gene expression. In agreement with previous observations [27,267], *Tt*Ago lowers intracellular plasmid DNA levels, even under selective conditions for plasmid propagation. This results in decreased plasmid DNA levels that still allows for survival in the presence of kanamycin while lowering the metabolic burden of high copy number plasmids. Earlier work showed that *Tt*Ago preferentially acquires DNA guides complementary to plasmid DNA and/or plasmid-encoded RNA [27], and demonstrated that *Tt*Ago can cleave both DNA and RNA *in vitro* [27,169]. To investigate the effect of *Tt*Ago on plasmid DNA and on plasmid-encoded RNA, we analyzed new RNA-seq data derived from *T. thermophilus* strains harboring a plasmid. Compared to the Δ*ago* strain, we observed lowered plasmid DNA levels an accordingly lowered levels of plasmid-encoded RNA in the wild type strain. Strikingly, we observed no further reduction of plasmid encoded RNA. This suggests that *Tt*Ago does not directly target RNA *in vivo*, making it a strict DNA-guided DNA-interfering host defense system.

Furthermore, the presence of plasmid DNA itself does not result in up-regulation of host defense genes. This suggests that, unlike phage infection, the presence of plasmid DNA is not registered as a threat. *T. thermophilus* requires a host defense system that is able to distinguish invader DNA from genomic DNA. While CRISPR-Cas systems require incorporation of spacers before being able to target invaders, *Tt*Ago specifically interferes with plasmid DNA without being dependent on genomic-encoded information about the invader. The observation that the combined presence of *Tt*Ago and plasmid DNA correlates with up-regulation of various CRISPR loci and at least part of the CRISPR adaptation machinery suggests that *Tt*Ago-mediated plasmid interference stimulates CRISPR adaptation. Although pAgos and CRISPR-Cas systems sometimes co-occur, often only one of these defense systems is encoded by a genome [4]. This suggests that these systems

function independently. Nevertheless, there are rare examples where the gene encoding pAgo co-localizes with Cas1 and Cas2 (for example in *Methanopyrus kandleri*) or Cas4 (multiple pAgos) [4,27]. Cas1 and Cas2 are known to be essential for CRISPR adaptation [295,297,298]. Also Cas4 has been predicted to be involved in CRISPR adaptation as its forms complexes with Cas1 and Cas2 [299] and additionally Cas4 is fused to Cas1 in several Type I CRISPR-Cas systems [294]. As *cas* genes and pAgo do not strictly co-occur, we hypothesize that pAgo itself is not directly involved in spacer adaptation, but that pAgo-mediated plasmid interference indirectly stimulates CRISPR adaptation. For example, pAgos might generate plasmid DNA degradation products that somehow stimulate expression of genes involved in CRISPR adaptation. Acquisition of new spacers, stimulated by *Tt*Ago, would make future generations resistant to the invader by CRISPR-Cas-mediated defense. An additive effect of two host defense systems (a restriction modification and a CRISPR-Cas system) on total resistance levels has recently been reported [300]. Combined with observation that *Tt*Ago lowers plasmid concentrations even under conditions selecting for plasmid maintenance, this makes *Tt*Ago a valuable addition to the current arsenal of host defense systems.

Acknowledgements

We thank Stan J.J. Brouns for critical reading of the manuscript. This work was financially supported by grants from the Netherlands Organization for Scientific Research (NWO) to JvdO (NWO-TOP, 854.10.003).

Author contributions

DCS, ERW and JvdO conceived and designed the experiment. DCS and ERW performed the experiments. DCS, JJK, PJS and JvdO analyzed the data. DCS wrote the paper. All authors have read and approved the manuscript.

Author information

Correspondence should be addressed to john.vanderoost@wur.nl.

Data deposition

Tables S1-S7 are available in the online version of the published manuscript [278]. The RNA-seq data discussed in this chapter have been deposited in the ENA (European Nucleotide Archive) under the primary accession number PRJEB8709.

Experimental procedures

Strains

T. thermophilus HB27 (ATCC BAA-163, DSM7039 and NBRC101085), which is referred to in this chapter as HB27 or wild type, and the *Tt*Ago-encoding gene knockout strain HB27Δ*ago* [27] were used for the studies described in this chapter (**Fig. 1**).

Transformations

T. thermophilus strains were transformed with plasmid pMKPnqosGFP [285] as described previously [27]. Colonies were selected and cultivated overnight at 65 °C in 20 ml TTH medium [27] in a shaker incubator. 1 ml aliquots were prepared from the overnight cultures in 1.5 ml Eppendorff tubes which were centrifuged in a table top centrifuge at 6,000 rpm for 10 min. Supernatant was removed and cell pellets were stored at -20 °C.

RNA sequencing and analysis

T. thermophilus strains with and without plasmid pMKPnqosGFP were cultivated in triplicates as described previously [27]. Growth medium was supplemented with 30 μg ml^{-1} kanamycin for cultures harboring pMKPnqosGFP. When cultures reached an $OD_{600\ nm}$ of 0.5, RNA was purified using the mirVana RNA isolation kit (Ambion) as described previously [27]. Purified RNA from these biological triplicates was sequenced by Illumina sequencing by BaseClear BV (Leiden, The Netherlands). Total RNA was first assessed for quality on a Bioanalyzer 2100 (Agilent) and the rRNA fraction was depleted using the Ribo-Zero bacteria kit. The rRNA-depleted RNA fraction was further prepared using the Illumina TruSeq RNA library preparation kit (Illumina). The resultant sequencing libraries were checked on a Bioanalyzer (Agilent) and quantified. The libraries were multiplexed, clustered, and sequenced on an Illumina HiSeq 2000 with single-read protocol for 50 cycles. The sequencing run was analyzed with the Illumina CASAVA pipeline with demultiplexing based on sample-specific barcodes. The raw sequencing data produced was processed removing the sequence reads which were of too low quality (only "passing filter" reads were selected) and discarding reads containing adaptor sequences or PhiX control.

T. thermophilus genome was re-annotated using an in-house annotation pipeline SAPP platform (Koehorst *et al.*, manuscript in preparation). Reads of different experiments were all mapped against the *T. thermophilus* genome (consisting of the HB27 chromosome and pTT27 mega-plasmid) and plasmid pMKPnqosGFP plasmid. For the identification of noise,

reads of all experiments also excluding the pMKPnqosGFP plasmid were mapped against the entire *T. thermophilus* genome and corresponding plasmids. Differential expression analysis was performed using the trinity package in combination with RSEM [301].

crRNA analysis

For the analysis of crRNAs the CRISPR cassettes were predicted using the CRT prediction module in SAPP (Koehorst *et al.*, manuscript in preparation). The corresponding regions of the CRISPR cassettes were extracted and analyzed in combination with the gene sequences using the trinity package. To improve mapping, repeat regions were trimmed.

Analysis of CRISPR loci

Triplicate HB27 and HB27Δ*ago* cultures with or without plasmid pMHPnqosGFP were cultivated in medium with and without antibiotics to an $OD_{600\ nm}$ of 0.5, after which genomic DNA was purified using the JGI 'bacterial genomic DNA isolation using CTAB' protocol [302]. Short stretches of each CRISPR locus, encompassing at least a part of the leader sequence and the first spacer-repeat unit, were PCR amplified (for primers see **Table S7**), and resolved on 2% agarose gels. Gels were stained with SYBR Safe Nucleic Acid Stain (Invitrogen) and nucleic acids were visualized using a G:BOX Chemi imager. A comparable method has previously been demonstrated to detect CRISPR adaptation if at least 0.4% of the culture obtained new spacers [297].

Plasmid content analysis

For complete DNA (containing both genomic and plasmid DNA) purification, *T. thermophilus* HB27 and HB27Δ*ago* transformed with pMKPnqosGFP were cultivated in triplicates to an $OD_{600\ nm}$ of 0.5. One $OD_{600\ nm}$ unit was harvested and complete DNA was isolated using the JGI 'bacterial genomic DNA isolation using CTAB' protocol [302]. 1 mg DNA of each purification was resolved on 0.8% agarose gels and stained with SYBR Safe Nucleic Acid Stain (Invitrogen), visualized using a G:BOX Chemi imager and analyzed using GeneTools analysis software (Syngene).

Statistical analysis

For the calculation of *P* values of differences in expression levels of specific genes, FPKM of biological triplicates of each strain were used as the input. *P* values stated in this chapter are calculated by a two-tailed distributed two-sample t-test assuming equal variances.

Supplementary Figure

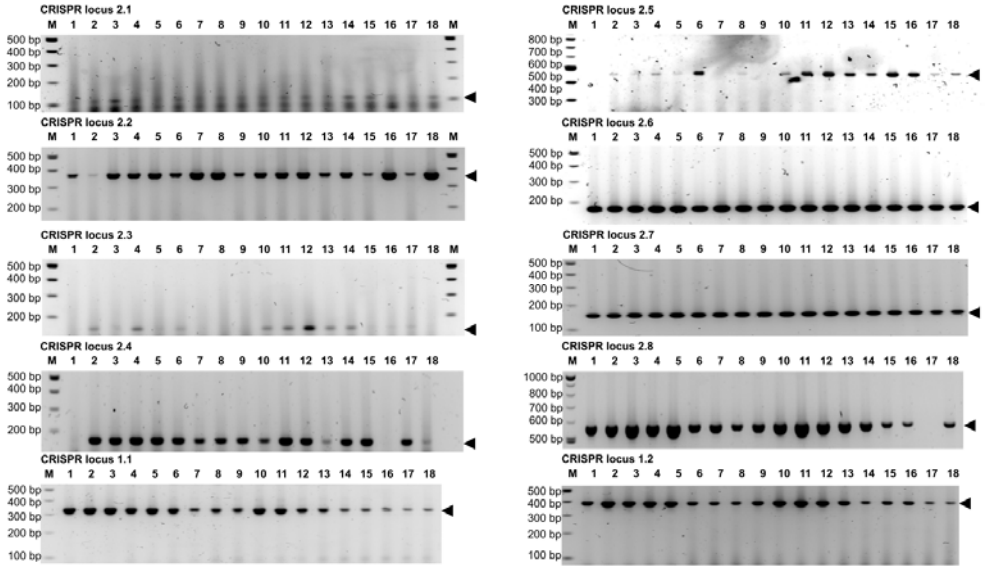


Figure S1 | PCR of CRISPR loci. 1-3: HB27 grown in absence of antibiotics. 4-6: HB27+P grown in absence of antibiotics. 7-9: HB27+P grown in presence of antibiotics. 10-12: HB27Δ*ago* grown in absence of antibiotics. 13-15: HB27Δ*ago*+P grown in absence of antibiotics. 16-18: HB27Δ*ago* grown in presence of antibiotics. M: GeneRuler 100 bp plus DNA ladder (Thermo Scientific). Black triangles indicate expected sizes of PCR products if no new spacers are acquired. If new spacers are acquired a new band ~75 bp larger than the original band is expected. No spacer acquisition was observed.

Chapter 7

CRISPR adaptation triggers plasmid curing

Daan C. Swarts, Cas Mosterd, Mark W.J. van Passel, Stan J.J. Brouns

Adapted from:
'CRISPR interference directs strand specific spacer acquisition'
PLoS ONE. April 2012. Volume 7, Issue 4, e35888

Abstract

CRISPR-Cas is a widespread adaptive immune system in prokaryotes. This system integrates short stretches of DNA derived from invading nucleic acids into genomic CRISPR loci, which function as memory of previously encountered invaders. In *Escherichia coli*, transcripts of these loci are cleaved into small RNAs and utilized by the Cascade complex to bind invader DNA, which is then degraded by Cas3 during CRISPR interference.

We describe how a CRISPR-activated *E. coli* K12 is cured from a high copy number plasmid under non-selective conditions in a CRISPR-mediated way. Cured clones integrated at least one up to five anti-plasmid spacers in genomic CRISPR loci. New spacers are integrated directly downstream of the leader sequence. The spacers are non-randomly selected to target protospacers with an AAG protospacer adjacent motif (PAM), which is located directly upstream of the protospacer. A co-occurrence of PAM deviations and CRISPR repeat mutations was observed, indicating that one nucleotide from the PAM is incorporated as the last nucleotide of the repeat during integration of a new spacer. When multiple spacers were integrated in a single clone, all spacer targeted the same strand of the plasmid, implying that CRISPR interference caused by the first integrated spacer directs subsequent spacer acquisition events in a strand specific manner.

The *E. coli* Type I-E CRISPR-Cas system provides resistance against bacteriophage infection, but also enables removal of residing plasmids. We established that there is a positive feedback loop between active spacers in a cluster – in our case the first acquired spacer - and spacers acquired thereafter, possibly through the use of specific DNA degradation products of the CRISPR interference machinery by the CRISPR adaptation machinery. This loop enables a rapid expansion of the spacer repertoire against an actively present DNA element that is already targeted, amplifying the CRISPR interference effect.

Introduction

Prokaryotes have evolved an adaptive immune system called CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR associated protein) that enables them to counter invasions from viruses and plasmids (reviewed in [303,304,305,306]). This immune system contains genomic CRISPR loci in which genetic material from invaders is incorporated. Memorized invaders can be recognized by expressing incorporated genetic material as RNA, which can guide Cas protein complexes to invader nucleic acid sequences.

The *E. coli* K12 genome encodes only a Type I-E CRISPR-Cas system [307, Makarova, 2011 #979]. This system is capable of providing resistance to bacteriophage infection, prophage induction and plasmid transformation [308,309,310]. Comparative genomics has shown that the *E. coli* K12 genome contains two CRISPR loci with type 2 repeats and a variable spacer content (CRISPR locus 2.1 (12 spacers) and 2.3 (6 spacers)), suggesting that both loci are active [307,311]. CRISPR locus 2.1 is located directly downstream of a *cas* gene operon, while locus 2.3 does not have any *cas* genes encoded in its proximity. Both CRISPR loci have a conserved AT-rich leader sequence that acts as a promoter for expression [312] and consist of 29 nucleotide palindromic repeats that are separated from each other by 32 or 33 nucleotide guide sequences called spacers. CRISPR transcripts are cleaved into mature CRISPR RNAs (crRNA) and these remain bound by the ribonucleoprotein complex Cascade (Cas-complex for antiviral defense, in Type I-E consisting of proteins encoded by *cas* genes *cse1*, *cse2*, *cas7*, *cas5* and *cas6e*) to guide the interference machinery to target DNA sequences (*i.e.* protospacers) [313]. In addition to Cascade, resistance requires the nuclease and helicase Cas3 [308,314,315]. Cas3 is recruited to the target DNA by the Cascade protein Cse1, after which Cas3 nicks the target DNA and further degrades the target DNA by ATP-dependent helicase and ssDNA nuclease activities [240].

In *E. coli* K12, transcription of the Type I-E Cascade-*cas1-cas2* operon, and to some extent the CRISPR array, is repressed by the global transcriptional repressor H-NS (heat-stable nucleoid-structuring protein [312,316,317]. In the *hns* knockout strain of *E. coli* K12 repression of the Cas genes is at least partially relieved [317], resulting in an activated CRISPR-Cas phenotype. Although the expression and interference stages of CRISPR immunity have been studied in *E. coli*, the process of acquiring spacers to modify the viral and plasmid specificity of the immune system has not yet been described. The *Streptococcus*

thermophilus Type II system integrates new spacers against bacteriophages [303,318] and plasmids [319], and thereby acquires resistance to these bacteriophages (BIM: bacteriophage insensitive mutant) or cures itself from the corresponding plasmids (PIM: plasmid interfering mutant). The Type II specific Cas protein Csn2 [294], a calcium-dependent dsDNA binding protein [320], was reported to be essential during the spacer integration process in *S. thermophilus* [303]. In *E. coli*, Cas1 and Cas2 are not required during CRISPR expression or interference [308]. Their strict conservation with CRISPR loci suggests involvement in CRISPR adaptation [321].

Here we describe that *E. coli* K12 Δ*hns* is cured from a high copy number plasmid by integrating new spacers into two CRISPR loci. Based on our observations we propose that active spacers in a cluster are used to expand the range of new spacers against the same target in a strand specific manner.

Results

Spacer integration results in plasmid curing and plasmid interference

Upon prolonged cultivation (~1-2 weeks) at 37 °C under non-selective conditions, *E. coli* Δ*hns* is cured from the 3.7 kb high copy number plasmid pRSF-1b. Out of 75 individual non-selectively propagated clones tested, 59 (79%) were kanamycin sensitive (Kan^S) and 16 (21%) kanamycin resistant (Kan^R). Sequencing of PCR amplicons of CRISPR loci 2.1 and 2.3 showed that between one and five anti-plasmid spacers were integrated in all Kan^S clones (**Fig. 1**, **Table S1**), while Kan^R clones did not contain any new spacers. No plasmid DNA could be isolated from eight out of eight tested Kan^S clones (clone nr. 1, 2, 3, 4, 6, 7, 19, and 27; **Fig. 1**, **Table S1**), confirming that the Kan^S clones were indeed cured from pRSF-1b.

When these clones were retransformed with pRSF-1b a 100- to 1000-fold drop in transformation efficiency was observed for clones with one or two integrated spacers, respectively (**Fig. 2**). These combined results indicate that the Kan^S clones are indeed PIMs. When retransformation efficiencies of PIMs with spacers integrated in either CRISPR 2.1 or 2.3 were compared, no significant differences in efficiencies could be observed, indicating that spacers from both loci are actively transcribed and utilized. Transformation of the PIMs with the target plasmid is not completely inhibited because point mutations in the protospacer at critical positions (seed region or protospacer adjacent motif (PAM)) [322],

or deletions, allow pRSF-1b to 'escape' the CRISPR interference [310]. This explains why PIMs containing multiple anti-plasmid spacers exhibited lower transformation efficiencies as mutation of multiple protospacers or their PAMs simultaneously occurs at lower frequencies (**Fig. 2**). We cannot explain why PIM25, which acquired 5 spacers, does not exhibit lower transformation efficiencies than PIMs with only 2 spacers.

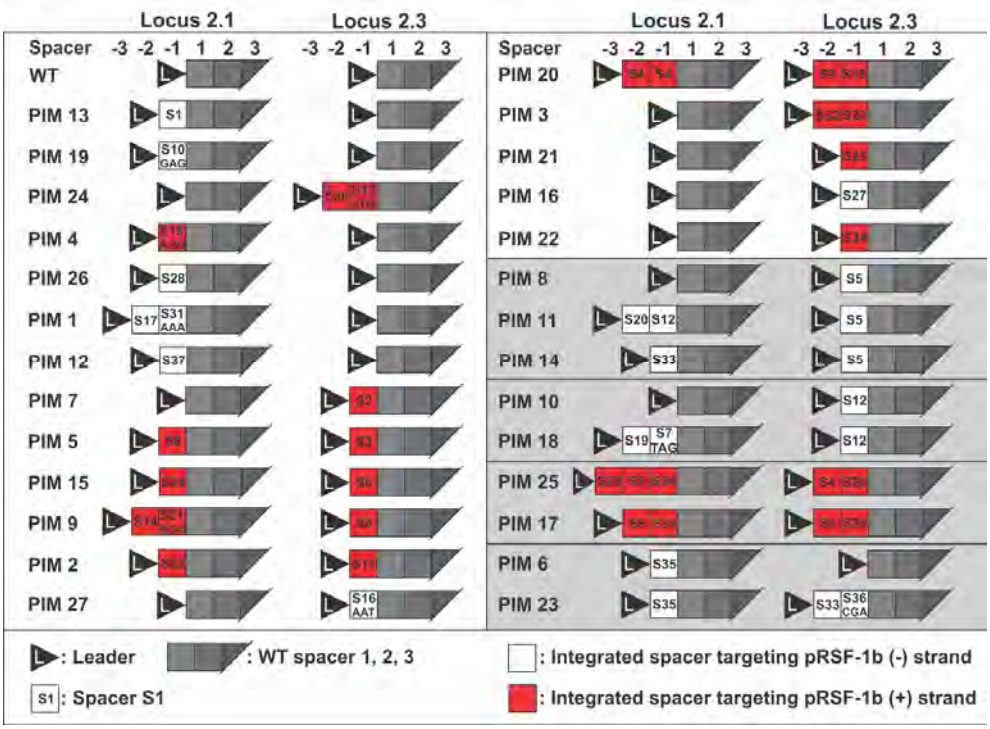


Figure 1 | Graphical representation of spacers integrated in the various PIMs. Both CRISPR locus 2.1 and 2.3 of each PIM are displayed. The newly acquired spacer positions (-3, -2, -1) and original spacer positions (1, 2, 3) correspond to the order of spacers downstream from the leader sequence (displayed as black triangle). White and red spacer boxes indicate that the corresponding protospacer is located on the – or + strand of the plasmid, respectively. PIMs clustered in grey boxes possibly share a common ancestor. Spacers have an AAG PAM unless indicated otherwise. Additional information on spacers is given in Table S1.

Sequencing of the leader-flanking end of CRISPR loci 2.1 and 2.3 of a random selection of 46 Kan^S clones revealed a total of 27 unique PIMs carrying a total of 37 different anti-plasmid spacers (**Fig. 1**, **Table S1**). While 13 PIMs had integrated a single new spacer, 7, 4, 2 and 1 PIMs integrated two, three, four and five new spacers, respectively. Of all different PIMs, 67% had integrated between one and three spacers in the CRISPR 2.1 locus, while

74% had integrated one or two spacers into the CRISPR 2.3 locus, indicating that both clusters are active. New spacers were always integrated directly downstream from the leader-flanking repeat. This suggests that there is a specific signal in the leader sequence to integrate new repeat-spacer units at this position in the CRISPR array. No spacer deletion was observed, indicating that the acquisition of new spacers occurs via addition rather than substitution. This is in agreement with findings in *S. thermophilus,* where repeat-spacer units were also mainly added directly downstream of the leader sequence [303,318,319]. In agreement with our findings, bioinformatic analyses have shown that spacer turnover and internal spacer integration is a rare event in *E. coli* [307,323].

AAG is the dominant protospacer adjacent motif

The protospacer adjacent motif (PAM) is a short conserved nucleotide sequence located in a protospacer flanking region [322]. The analysis of spacer-protospacer pairs from over 150 species has revealed the existence of several PAM consensus sequences which co-occur with specific repeat types [322]. The PAM consensus sequence 5'-AWG-protospacer-3' was identified for *E. coli* [322]. When present, PAMs are essential for CRISPR-interference as a point mutation in the PAM allows bacteriophages to escape the immune system [310,324]. For *E. coli* it was shown that mutations in the PAM result in dramatically lower target DNA binding affinity of the crRNA guided complex Cascade [310], explaining how the bacteriophage genome can avoid being detected.

Of all integrated spacers, 29 (78%) corresponded to protospacers with an AAG PAM, one (3%) with an ATG PAM, and seven (19%) with non-consensus PAM sequences (AAA, AGG (2x), GAG, TAG, CGA, AAT; **Table S1**). Although the functionality of only the ATG PAM has been verified in *E. coli* [310], the majority of integrated spacers in our experiments correspond to protospacers flanking an AAG PAM. It could be argued that spacers are selected randomly followed by natural selection. Clones that have integrated spacers with a consensus PAM (AWG) are cured from the high copy number plasmid pRSF-1b and generally gain an energetic growth advantage [325], which allows them to outgrow clones that have incorporated spacers with non-functional PAMs. However, this would have resulted in a more equal distribution of AAG and ATG PAMs, making the random spacer selection process unlikely. Furthermore, since an AAG triplets are found less frequently on pRSF-1b than ATG triplets (94 times AAG versus 129 times ATG), limited availability ATG is not the reason for AAG PAM selection. Moreover, five spacers were integrated multiple

times in unrelated PIMs and in different CRISPR loci (S4 in PIM 20 (2x) and 25; S8 in PIM 5, 9, 17 and 25; S12 in PIM 10, 11 and 18; S33 in PIM 14 and 23; S34 in PIM 17, 22 and 25) which also argues against random spacer selection. These findings indicate that there is a selection for AAG PAM sequences during spacer acquisition.

It is worth noting that three PIMs (4, 19, and 27) integrated a single anti-plasmid spacer corresponding to the non-PAM consensus sequences AGG, GAG and AAT. Sequencing of pRSF-1b in the corresponding regions excluded the possibility that the plasmid contained mutations at these positions, confirming that these PAMs were indeed non-consensus PAM sequences. The fact that these PIMs were cured from the plasmid, and were less susceptible to retransformation of the target plasmid (**Fig. 2**) indicates that at least non-consensus PAMs AGG and GAG are additionally allowed during CRISPR interference. PIM 1, which integrated a spacer with a non-consensus AAA PAM and one other spacer, shows resistance typical for PIMs with two functional spacers. This indicates that also this PAM is likely to be allowed during CRISPR interference.

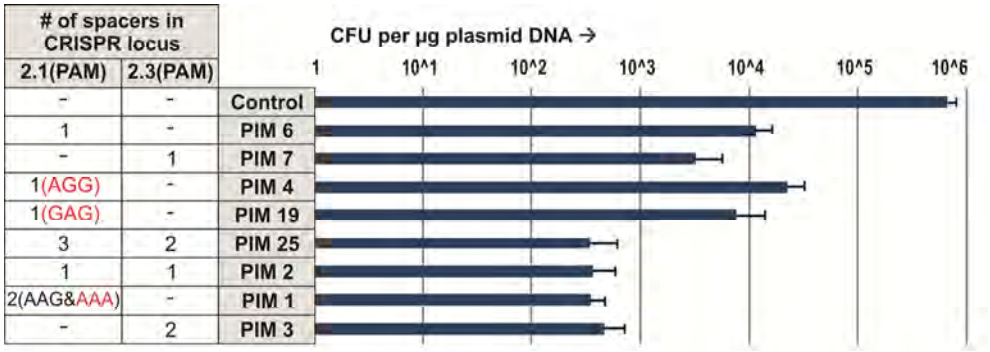


Figure 2 | Effect of integrated spacers on retransformation efficiency. Transformation efficiencies of various PIMs and the control (Wild type *E. coli* K12 W3110) are given in a logarithmic scale as colony forming units (CFU) per µg of pRSF-1b plasmid DNA. For each PIM, the number of spacers integrated in either CRISPR locus 2.1 or 2.3 is given. All spacers have an AAG PAM, unless indicated otherwise. The exact spacer composition of each PIM is given in Table S1.

Counterselection for self-targeting spacers

The locations of the protospacers were mapped on both strands of the plasmid (43% and 57% on the (+) and (-) strand, respectively) and covered regions of the backbone and multiple cloning site (32%), origin of replication (40%) and the kanamycin resistance gene (24%) (**Fig. 3**). This indicates that protospacer acquisition occurs independently of

transcription or direction of replication of the plasmid. Interestingly, only a single spacer (2%) was integrated against the plasmid-encoded *lacI* gene (S36; **Fig. 3**). This observation can be explained by the presence of a nearly identical copy (one nucleotide difference) of the *lacI* gene in the *E. coli* K12 genome. Spacers targeting the plasmid encoded *lacI* gene would therefore also target the *E. coli* genome, leading to lethal DNA damage, and resulting in a counterselection for these variants. This result fits very well with the observation that spacers against a prophage are lethal to *E. coli* [309]. The identified anti-*lacI* spacer in PIM 23 has a non-consensus PAM CGA that possibly prevents self-targeting. The plasmid interfering phenotype of this PIM is likely to be caused by the two additional spacers corresponding to protospacers with AAG PAMs (**Fig. 1**, **Table S1**).

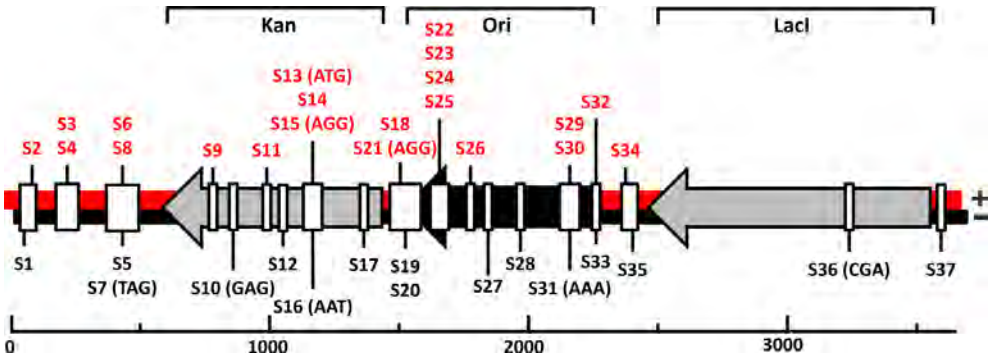


Figure 3 | Linear display of pRSF-1b and locations of protospacers. The (+) and (-) strands and corresponding protospacers are colored red and black, respectively. Kanamycin marker (Kan), Origin of replication (Ori) and *lacI* (LacI) are shown as arrows. Protospacers have an AAG PAM unless indicated otherwise. Sequences of spacers corresponding with the protospacers are given in Table S1.

Nucleotide composition of spacers

The nucleotide content of the 37 unique anti-plasmid spacers was compared with the composition of all possible AAG-flanking protospacers on pRSF-1b (**Fig. 4**). The analysis showed that the integrated spacers displayed no selection bias for GC content. This suggests that GC content of the protospacers, and therefore the local stability of the DNA duplex, plays no major role during spacer selection. In addition to GC content, we also analyzed purine (AG) content of the new spacers (**Fig. 4**), as purine-rich RNA is known to basepair energetically more favorable with DNA than the corresponding DNA:DNA duplex [326,327]. This may be of importance during the hybridization of the crRNA to double stranded target DNA molecules. Again, no apparent bias could be observed compared to the semi-randomly generated spacer set, suggesting that the energetic gain of pairing purine

rich crRNA with DNA by Cascade is not taken into account by the CRISPR adaptation machinery during spacer integration. Also no bias was found for GC or AG content in the seed sequence, which plays an important role in during target DNA binding of Cascade [310].

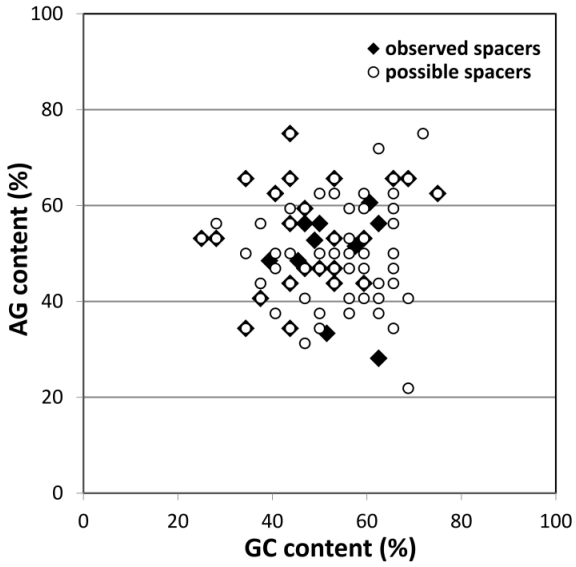


Figure 4 | Graphical representation of AG and GC contents of each observed and possible spacer. Observed spacers (◆) are spacers integrated in CRISPR loci 2.1 and 2.3 (Table S1). These spacers are 32 or 33-mers with various PAMs. Possible spacers (○) are all 32-mers found on pRSF-1b directly downstream of an AAG PAM. Note that some observed spacers did not meet these criteria.

The last nucleotide of the repeat is PAM derived

It has previously been described that repeats of CRISPR 2.1 and 2.3 (consensus: 5'-GWGTTCCCCGCGCCAGCGGGGATAAACCG-3') contain polymorphisms [307]. Some polymorphisms in the repeats have been associated with preventing self-targeting, as self-targeting spacers are often accompanied by degraded repeats [328]. Especially the last 8 nucleotides of the repeat, which determine the first 8 nucleotides of mature crRNAs, appear to be important for the functioning of CRISPR-Cas systems [306]. The Type III-a system of *Staphylococcus epidermidis* uses differential complementarity of these first 8 nucleotides of the crRNA with one protospacer flank to discriminate between self DNA (the CRISPR) and non-self DNA (the target), preventing autoimmunity [329]. Other CRISPR-Cas systems may use PAMs to determine if a sequence will be targeted [310,318,319,322].

Our dataset shows that the last three nucleotides of the repeat (CCG) occasionally carry mutations. Repeat 2 of CRISPR locus 2.3 in the parental strain contains a polymorphism at the last nucleotide, changing the trinucleotide sequence from CCG to CCT. Almost all

PIMs with new spacers in CRISPR 2.3, however, did not carry this polymorphism in their new repeats, indicating that the second repeat in a CRISPR is not duplicated during the spacer integration process.

S16 is preceded by a CCT trinucleotide repeat sequence, and strikingly this spacer corresponds to a protospacer with non-consensus AAT PAM. This combination is apparently functional, as this PIM is cured from the plasmid and is less susceptible for retransformation with pRSF-1b (**Fig. 2**), while carrying only one anti-plasmid spacer. This indicates that S16 facilitates interference although it has a non-consensus PAM and a mutated repeat. S31 in PIM 1 is preceded by a CCA trinucleotide repeat sequence and it has the non-consensus AAA PAM, while spacer S36 in PIM 23 is preceded by repeat sequence CCA and targets a plasmid sequence flanking a non-consensus CGA PAM. Because PIM 1 and 23 each contain additional typical anti-plasmid spacers, it cannot be concluded whether S31 and S23 are functional. However PIM 1 (carrying S31 and typical spacer S17) shows a decrease in transformation efficiency similar to PIMs with two typical anti-plasmid spacers (**Fig. 2**), suggesting that S17 is indeed functional.

Interestingly, the last nucleotide of the repeat preceding the new spacer always matched the third nucleotide of the PAM, both in normal situations (repeat CC<u>G</u>, and AA<u>G</u>, match underlined) and in deviations from normal (CC<u>T</u>^{R} – AA<u>T</u>^{P}; CC<u>A</u>^{R} – CG<u>A</u>^{P}; CC<u>A</u>^{R} – AA<u>A</u>^{P}; **Fig. 5a**). The single nucleotide polymorphism (SNP) at the last position of the repeat and corresponding deviations from the PAM consensus sequence suggests that the last nucleotide of the repeat is derived from the PAM in the target DNA (**Fig. 5b**). Evidence supporting this hypothesis is provided in PIMs 1 and 23 which contain the deviated repeat-spacer unit at the second position in the locus and have a consensus repeat-spacer unit at the first position. Apparently, the repeat SNP is not propagated in the new repeat-spacer unit at the first position in the locus (**Table S1**, PIM1 and PIM23), but reverted to the repeat-consensus by the selection of a normal AAG PAM-containing protospacer. We hypothesize that the protospacer-flanking nucleotide of the PAM is still attached to the selected, to-be-integrated spacer (pre-spacer) [330], and forms the last nucleotide of the proximal repeat after integration is complete (**Fig. 5b**). As a consequence, this nucleotide in the crRNA is always complementary to the protospacer-flanking nucleotide of the PAM (**Fig. 5c**), even when a non-consensus PAM is selected during spacer acquisition.

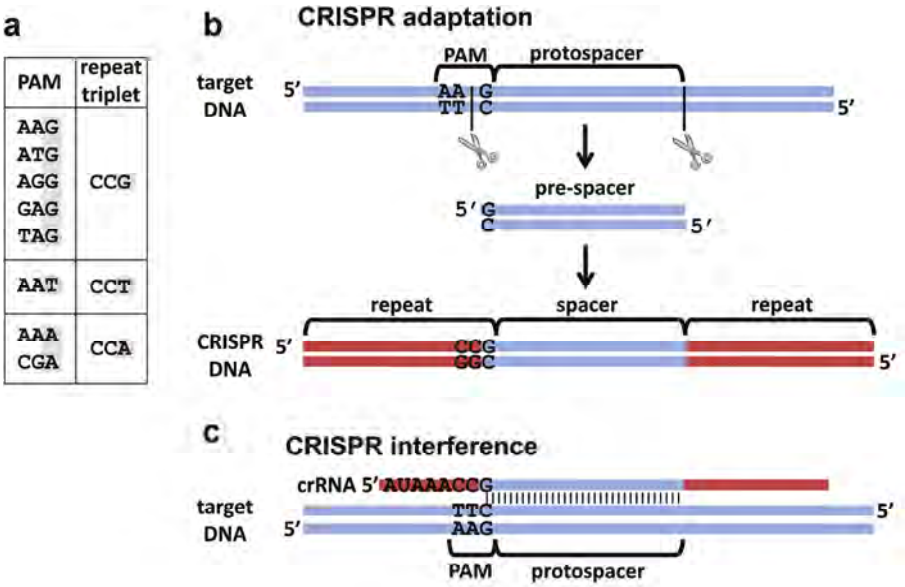


Figure 5 | PAM and repeat-end correlation. a, PAMs of observed spacers and the co-occurring trinucleotide repeat-ends associated with these spacers. Notice that the spacer-proximal nucleotide of the repeat end is identical to the protospacer-proximal nucleotide of the PAM. **b**, Schematic representation of the proposed mechanism for spacer acquisition during CRISPR adaptation. A protospacer with specific PAM is selected after which it is processed into the pre-spacer (at least 33-34 bp), which contains the last nucleotide of the PAM (the pre-spacer could be single-stranded or double-stranded). The pre-spacer is than integrated at the leader proximal end of the CRISPR locus. The nucleotide derived from the PAM forms the last nucleotide of the repeat. **c**, R-loop formation by mature crRNA (61 nucleotides) during CRISPR interference. Notice that the last nucleotide of the repeat (the nucleotide derived from the PAM) is complementary to the target DNA sequence. It remains unknown whether base-pairing between these nucleotides is important for interference.

Spacer integration patterns suggest a positive feedback loop of active spacers

In 14 different PIMs, two or more spacers were integrated (**Fig. 1**, **Table S1**). No preference for a specific target location of subsequently integrated spacers could be detected, such as a location near the target site of the primary integrated spacer. However, all spacers of an individual PIM always targeted the same strand of the plasmid, implying that the primary integrated spacer determines which strand subsequently integrated spacers will target. This suggests a positive feedback loop that may result from interplay between the CRISPR interference machinery (Cascade and Cas3) and the spacer integration machinery. We hypothesize that CRISPR-mediated plasmid degradation by Cas3 [240], recruited by Cascade guided by a crRNA from an active spacer – the first new spacer in this case - generates specific DNA degradation products that are used as precursors for subsequent

new spacers (**Fig. 6**). These findings are in contrast with new spacer integration patterns in *S. thermophilus*, where secondary spacers show no strand selection bias [303,318,331] suggesting that CRISPR acquisition and CRISPR interference by Cas9 [332] are independent processes in *S. thermophilus*.

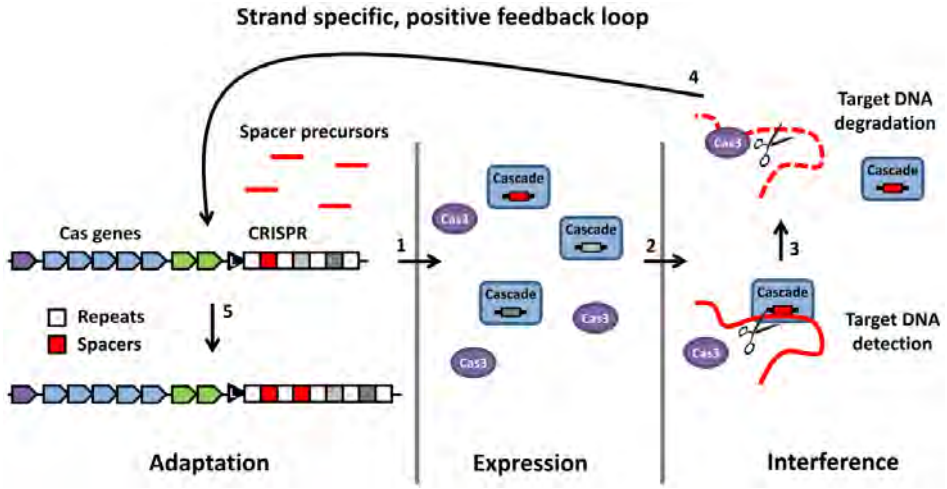


Figure 6 | Model of the strand specific positive feedback loop. Cells with a spacer against a known and actively present invader DNA produce targeting Cascade complexes in the expression stage. In the interference stage, Cascade binds the target dsDNA after which the target is cleaved and degraded by Cas3 [240]. DNA degradation products generated by Cascade and Cas3 (which could be ssDNA or dsDNA) act as precursors for new spacers in the adaptation phase in a strand-specific manner. By integration of these strand-specific precursors, the spacer repertoire against an actively present invader is expanded, completing the positive feedback loop.

Discussion

E. coli K12 is cured from a high copy number plasmid by integrating anti-plasmid spacers in two of its CRISPR loci. New spacers are selected in a non-random process that takes into account the presence of a PAM on the target DNA. We hypothesize that the mechanism of CRISPR adaptation in Type I-E systems involves selection of protospacers including one nucleotide from the PAM, which determines the last nucleotide of the preceding repeat. Spacer analysis further suggests a positive feedback loop between active spacers in a cluster and newly acquired spacers, through interplay of the CRISPR interference and adaptation machinery. Possibly the target DNA degradation products generated by Cascade and Cas3 serve as precursors for the integration of new spacers against the same target (**Fig. 6**).

Increasing the number of spacers targeting an invading DNA element may represent an efficient strategy to expand the repertoire of spacers targeting a specific invader to amplify the CRISPR interference effect. Having multiple active spacers against the same target reduces the chance that invaders evade immunity by point mutation in the seed region of the protospacer or PAM, since point mutations at multiple target sites simultaneously occur at lower frequencies.

Shortly after the publication of the manuscript described in this chapter [333], observations similar to the 'positive feedback loop' have been described in multiple publications. Datsenko *et al.* [334] described that mutated protospacers that lead to escape from interference in *E. coli* K12 lead to stimulated acquisition of additional spacers, and termed the process 'priming'. Like observed for the secondary spacers acquired during the positive feedback loop, priming results in generation of secondary spacers targeting the same strand as the primary spacer. Fineran *et al.* [335] have demonstrated that priming occurs for degenerate target regions with up to 13 mutations, and showed that the number of mismatches, their position and nucleotide type influence the priming strength. Priming has also been observed for Type I-B and I-F CRISPR-Cas systems [336,337], although the strand bias is not observed in these systems. This demonstrates that priming is a conserved mechanism that allows 'outdated' spacers to induce CRISPR adaptation in order to restore CRISPR interference.

Author Contributions

DCS and SJJB conceived and designed the experiment. DCS and CM performed the experiments. DCS, CM, MWJvP and SJJB analyzed the data. DCS and SJJB wrote the paper.

Acknowledgements

We thank Edze R. Westra and John van der Oost for critical reading of the manuscript. This work was financially supported by Veni grants from the Netherlands Organization for Scientific Research (NWO) to SJJB (863.08.014) and MWJvP (863.08.016).

Author information

Correspondence and requests for materials should be addressed to stan.brouns@wur.nl.

Data deposition

Table S1 is available in the online version of the published manuscript [333].

Experimental procedures

Plasmid curing

Escherichia coli K12 W3110 derivate Δ*hns* (JW1225) from the KEIO collection [338] was supplied by the American Type Culture Collection (ATCC). Its kanamycin resistance marker was removed according to protocol described by Datsenko *et al.* [339]. This strain was transformed with high copy number plasmid pRSF-1b (Novagen) (RSF1030 origin of replication, >100 copies/cell, 3.7 kb) [340] as described below. Colonies were picked from an LB-agar plate containing 100 μg ml^{-1} kanamycin and used to inoculate 2YTL medium [308] containing no antibiotics. The culture was transferred daily to fresh 2YTL medium in a shaking incubator for prolonged periods of time (~1-2 weeks). The culture was regularly checked for plasmid loss by plating on non-selective LB-agar plates, followed by replica streaking on selective (containing 100 μg ml^{-1} kanamycin) and non-selective LB-agar plates.

Plasmid loss and transformations

PIM 1, 2, 3, 4, 6, 7, 19 and 25 were cultured in 5 ml LB medium without antibiotics, and were incubated overnight in a shaker incubator at 37 °C. The overnight cultures were miniprepped (GeneJET, Fermentas) and the absence of plasmid DNA in the eluate was verified by nanodrop and agarose gel electrophoresis.

Cells for the plasmid curing experiments and retransformation experiments were made chemically competent using the RuCl method and transformed by applying a heat-shock as described in the QIAexpressionist handbook (QIAGEN). After transformation, cells were plated on an LB-agar plate containing 50 μg ml^{-1} kanamycin. PIM 1, 2, 3, 4, 6, 7, 19 and 25 and the wild-type control strain were retransformed with pRSF-1b and plated on LB-agar plates containing 50 μg ml^{-1} kanamycin. Transformations efficiency was determined as the number of colony forming units per μg plasmid DNA.

Colony PCR

Clones were screened for spacer integration by colony PCR using DreamTaq Green DNA polymerase (Fermentas). New spacers in the CRISPR 2.1 locus were PCR amplified using forward primer BG3474 (5'-AAATGTTACATTAAGGTTGGTG-3') annealing 72 bases upstream of the first repeat and reverse primer BG3475 (5'-GAAATTCCAGACCCGATCC-3') annealing in spacer 4 of this locus. New spacers in the CRISPR 2.3 locus were PCR amplified using forward primer BG3414 (5'-

GGTAGATTTTAGTTTGTATAGAG-3') annealing 164 bases upstream of the first repeat and BG3415 (5'-CAACAGCAGCACCCATGAC-3') annealing in spacer 3 of this locus. PCR product sizes were estimated using agarose gels and SYBR-safe DNA gel stain (Invitrogen). The CRISPR 2.1 and 2.3 loci of 46 Kan^S clones were sequenced by GATC-Biotech (Konstanz, Germany) with BG3474 and BG3414, respectively.

Spacer composition analysis

Nucleotide analyses were carried out using in-house perl scripts. In brief, all 32-mers from plasmid pRSF-1b preceded by the PAM AAG were tested for their nucleotide composition, and compared to the nucleotide composition of all experimentally retrieved spacers.

Chapter 8

Summary and General discussion

Summary

Argonaute proteins (Agos) are proteins that are present in all domains of life. They generally play important roles as nucleases that use single stranded nucleic acid guides to specifically bind and cleave complementary nucleic acid targets. This thesis describes the evolution, the role and the mechanism of prokaryotic Argonaute proteins (pAgos), focusing on the bacterial pAgo from *Thermus thermophilus* (*Tt*Ago) and archaeal pAgo from *Pyrococcus furiosus* (*Pf*Ago).

Chapter 1 gives an overview of evolution, structure and function of both eukaryotic Argonautes (eAgos) and pAgos. Although sequence similarity between eAgos and pAgos and even amongst pAgos is very low, the overall architecture of Agos is conserved throughout the three domains of life. Both eAgos and long pAgos consist of four key domains, of which the core function is conserved: The MID domain (responsible for guide 5'- end binding and guide pre-ordering), the PIWI domain (involved in guide pre-ordering and in cleavage, as it forms the RNase H-like DEDX (X is D or H) catalytic site), the PAZ domain (binds the 3' end of the guide), and the N-domain (involved in unwinding guide-target duplexes and indirectly in target cleavage). Despite this structural homology, small differences in structure can have a major impact on the functionality of different Agos. Whereas pAgos appear to function as stand-alone proteins in host defense pathways, insertion segments in eAgos have enabled them to become the key player in a plethora of complex multi-protein RNA-regulating pathways.

T. thermophilus is a gram-negative thermophilic bacterium with a natural competence system [234], which allows it to import DNA from its environment into the cell. **Chapter 2** describes the role of *Tt*Ago in *T. thermophilus* HB27. A genomic *ago* knockout (HB27Δ*ago*) was constructed and its competence was compared to the competence of wild type HB27 by transforming both strains with plasmid DNA. Transformation of wild type HB27 yielded ~10-fold less colony forming units compared to HB27Δ*ago*. Also after transformation, *Tt*Ago lowered intracellular plasmid concentrations 3 to 5-fold, even when *T. thermophilus* was grown under antibiotic pressure selecting for maintaining the antibiotic resistance-encoding plasmid. This demonstrated that *Tt*Ago interferes both with plasmid transformation and with plasmid propagation.

After heterologous expression of *Tt*Ago in *Escherichia coli*, nucleic acids that co-purified with *Tt*Ago were analyzed [27]. We discovered that catalytically active *Tt*Ago becomes loaded with 5'-phosphorylated ssDNA guides, 13 to 25 nucleotides in length, which were termed small interfering DNAs (siDNAs). We developed a novel DNA sequencing protocol, by which we were able to demonstrate that guides are preferentially acquired from plasmid DNA. Heterologously expressed *Tt*Ago was found to utilize the *in vivo* acquired siDNA guides to cleave double stranded DNA (dsDNA) plasmids *in vitro*.

Using synthetic DNA guides, it was demonstrated that *Tt*Ago cleaves both single stranded (ss)DNA and ssRNA targets complementary to the guide, while negatively supercoiled plasmid dsDNA was nicked on the strand base pairing to the DNA guide. A dsDNA break was obtained by incubating the plasmid with two different *Tt*Ago-guide complexes, targeting either strand of the plasmid. Plasmid DNA was not cleaved when it was linearized, suggesting that energy stored in the supercoils contributes to unwinding of the plasmid DNA, after which *Tt*Ago-siDNA complexes can bind and cleave individual DNA strands. This assumption was supported by the observation that AT-rich dsDNA is cleaved more efficiently than GC-rich. Combined, these results demonstrate that *Tt*Ago functions in host defense by DNA-guided DNA interference [27].

In **Chapter 3** structural insights in the cleavage mechanism of *Tt*Ago are described. The structure of DNA-guide loaded *Tt*Ago bound to 12 and 15 nucleotide target DNAs revealed that the complex remained in a cleavage-incompatible conformation. Upon binding of longer targets of 16 or 19 nucleotides, the complex adapted a cleavage-competent conformation: the PAZ domain rotates, while the 3' end of the guide is released from this domain. Furthermore, movements of three loops in the PIWI domain can be observed. Movement of PIWI loop 2, also termed the 'glutamate finger', results in insertion of a glutamic acid (E) residue into a pocket that contains three aspartic acids (D), thereby completing the DEDD catalytic tetrad of *Tt*Ago. In contrast to the cleavage-incompatible conformation, two divalent cations (Mg^{2+} or Mn^{2+}) are bound by this tetrad in the cleavage-compatible conformation. The cations coordinate a water molecule that allows for in-line attack on the cleavable phosphate group between nucleotide 10 and 11 of the target DNA. Finally, a model for the step-wise target-induced activation of the pAgo-guide complex and divalent metal cation-dependent target DNA cleavage is provided [176].

The characterization of archaeal *Pf*Ago is described in **Chapter 4**. In this chapter it is demonstrated that *Pf*Ago, like *Tt*Ago, interferes with plasmid transformation. *Pf*Ago belongs to the phylogenetic clade of pAgos most closely related to eAgos. Therefore, it was anticipated to utilize RNA guides for its activity. In contrast to our expectations, *Pf*Ago utilizes DNA guides rather than RNA guides, and cleaves DNA targets but not RNA targets. The activity of *Pf*Ago is mediated by a DEDH active site that binds Mn^{2+} or Co^{2+} for target DNA cleavage. Strikingly, the divalent cation Mg^{2+}, which is commonly utilized by eAgos and *Tt*Ago, does not support catalytic activity by *Pf*Ago. Like *Tt*Ago, *Pf*Ago can utilize siDNAs to cleave dsDNA plasmids. Additionally, guide-free *Pf*Ago cleaves plasmid DNA, indicating it does not require siDNAs for activity. Concluding, like *Tt*Ago, *Pf*Ago provides host defense by DNA-guided DNA interference. This suggests a broad conservation of pAgo role and mechanism in bacteria and archaea. As *Pf*Ago is closely related to eAgos, these findings suggest that the ancestor of eAgos might also have mediated DNA-guided DNA interfering protein to protect its host against invading nucleic acids [277].

In **Chapter 5** a mechanism is described by which *Tt*Ago generates siDNAs. While eAgos rely on proteins such as Dicer or Drosha for guide generation, no homologs of these proteins are present in prokaryotes. As *Tt*Ago, but not a catalytic mutant of *Tt*Ago, co-purifies with guides after expression in *E. coli*, it was hypothesized that *Tt*Ago is able to generate its guides independently using its catalytic activity. Using *in vitro* studies it is revealed that guide-free *Tt*Ago, like *Pf*Ago, can cleave dsDNA plasmids under certain conditions. *Tt*Ago degrades (partially) unwound dsDNA, a process which we termed 'DNA chopping'. DNA chopping generates 13 to 25 nucleotide long 5'-phosphorylated DNAs, which can be utilized by *Tt*Ago to cleave complementary target DNAs. Thus, under specific conditions, guide-free *Tt*Ago is able to generate its own guides by DNA chopping [274].

To investigate whether *Tt*Ago, besides interfering with plasmid DNA, also interferes with plasmid-encoded RNA, we performed RNA-seq analysis on RNA isolated from four different *T. thermophilus* strains (HB27, HB27Δ*ago* and both strains harboring a plasmid). This analysis, described in **Chapter 6**, confirmed that *Tt*Ago does not regulate competence or host defense by RNA interference. Additionally, it was demonstrated that the presence of plasmid DNA itself does result in global changes in gene expression levels in *T. thermophilus*. When both *Tt*Ago and plasmid DNA are present, no genes are down-regulated by *Tt*Ago. Interestingly however, when both *Tt*Ago and plasmid DNA are

present, several CRISPR loci and a subset of CRISPR loci-associated genes are up-regulated. Some of these up-regulated genes encode proteins that are involved in CRISPR adaptation. This suggests that *Tt*Ago-mediated plasmid DNA interference activates CRISPR adaptation [278].

CRISPR-Cas is a host defense mechanism that protects its host against invading nucleic acids in three stages (reviewed in [295,298]): the adaptation stage, the expression stage and the interference stage. In the adaptation stage, small pieces of DNA (termed spacers) are acquired from specific invader DNA loci (protospacers) and are integrated in genomic CRISPR loci. During the expression stage, RNA expressed from the CRISPR loci is processed and loaded onto Cas-protein complexes. In the interference stage, these complexes utilize the CRISPR RNA as guide to bind and degrade invader DNA complementary to the guide.

In **Chapter 7** [333] a study of CRISPR adaptation in *E. coli* K12 is presented. By sequencing CRISPR loci of clones that lost their plasmids upon prolonged cultivation, we demonstrate that spacers are acquired from invading plasmid DNA in a non-random fashion, taking into account protospacer adjacent motifs (PAMs) located in the invader DNA. New spacers are integrated at the promoter-proximal end of the CRISPR locus. We revealed that the last nucleotide of the repeat unit upstream the spacer is not derived from the repeat itself, but rather from the PAM flanking the protospacer. Additionally, we showed that integration of a single spacer primes the subsequent rapid integration of additional spacers, creating a positive feedback loop in the adaptive phase of the CRISPR-Cas interference mechanism.

In summary, the work performed during this PhD project has elucidated the role and molecular mechanism of *Tt*Ago and *Pf*Ago proteins, revealing that these prokaryotic Agos mediate a novel DNA-guided host defense mechanism that directly targets DNA. This has contributed to the fundamental understanding of the biological function of pAgos, which are the evolutionary precursors of eAgos. Additionally, the discovery that some pAgos are able to target DNA has raised the exciting possibility of developing pAgos as molecular tools for genome editing.

General discussion

Since the discovery of the existence of Argonaute proteins in prokaryotes in 2000 [3], pAgos from thermophilic bacteria and archaea have initially been used for structural and biochemical studies. Bioinformatics, structures and biochemical characterization studies have revealed that pAgos share many similarities with each other and with eAgos. Best conserved are their domain architecture (N-PAZ-MID-PIWI) and their mechanism to utilize a nucleic acid guide for binding of complementary nucleic acid targets. However, also differences between pAgos and eAgos exist. In this chapter the (potential) roles of pAgos are discussed and dissimilarities that determine mechanistic differences between pAgos and eAgos are described. In addition, the role of the proteins that are encoded by the same operons as specific long and short pAgos is hypothesized upon. Finally, the potential of pAgos as the new generation of genome editing tools is discussed.

The role of pAgos

pAgo-mediated plasmid interference

Bacterial *Tt*Ago and archaeal *Pf*Ago both play a role in host defense by interfering with plasmid DNA in an siDNA dependent manner [27,267,277]. Also *Rhodobacter sphaeroides* Argonaute (*Rs*Ago) interferes with plasmid DNA, but it utilizes RNA guides for this purpose [177]. This suggests that a plasmid-interfering function is conserved in pAgos. Although plasmids are considered selfish genetic elements [341], they often carry genes beneficial to its host. Here we discuss why plasmid-interfering pAgos are valuable assets to their hosts.

In laboratory settings, plasmids generally provide a fitness gain for their host. Cells harboring plasmids are often cultivated under conditions selecting for plasmid maintenance, for example by addition of antibiotics to the medium (selecting for plasmid-encoded antibiotic resistance) or by cultivating in growth medium that lacks essential growth components (selecting for plasmids complementing auxotrophy). This maximizes the benefits of plasmid retention for the host. A major drawback of maintaining plasmid DNA concerns the metabolic costs involved in plasmid propagation (reviewed in [342,343]). Yet, under most laboratory conditions, prokaryotes are grown in energy-rich media, which minimizes drawbacks of plasmid propagation. In nature, however, energy

sources are not as abundant, which increases the drawbacks of plasmid maintenance. Additionally, as prokaryotes often live in ecosystems that allow growth, they are less dependent on plasmid-encoded genes. This makes the plasmids a metabolic burden rather than a fitness benefit for the host. Nevertheless, ecosystems are subjective to change and there might be situations in which plasmids provide the genes required for survival. In fact, some prokaryotes, amongst which *T. thermophilus* and *P. furiosus*, are naturally competent, implying they actively import DNA from their environment [234,270]. As both beneficial and harmful DNA could be imported with these systems, a fine balance between DNA uptake and host defense should be established to maximize fitness.

Previously described prokaryotic host defense systems include restriction-modification systems (reviewed in [344]) and CRISPR-Cas systems (reviewed in [240,298]). Canonical restriction-modification systems methylate the host genome, while cleaving non-methylated invader DNA. CRISPR-Cas protein complexes interfere with invader DNA via small RNA guides which are encoded on the genome in CRISPR loci. Targeted plasmids are eliminated from cells with active CRISPR-Cas systems [319,333]. While both systems might be excellent host defense mechanisms against invading nucleic acids that are not beneficial for the host, they drastically lower the possibility of potentially beneficial plasmids to enter the cell. The effect of pAgos on plasmid transformation is relatively low, with ~10-fold lower transformation efficiency in strains with *Tt*Ago [27,267], and only ~1.5 to 2-fold lower transformation efficiency in strains with *Pf*Ago [277]. In comparison, CRISPR-Cas in *E. coli* can lower plasmid transformation efficiency ~100-fold with a single spacer targeting the transformed plasmid, and ~1000-fold with two spacers targeting the transformed plasmid [27]. Thus, although pAgo provides a certain degree of host defense against plasmid transformation, they appear relatively permissive compared to CRISPR-Cas systems. This would allow at least a part of the population to acquire beneficial genetic traits from their environment. Additionally, the action of pAgos does not stop at interfering at the transformation step. Both *Rs*Ago and *Tt*Ago lower intracellular plasmid concentrations after transformation, even when cultivation takes place under conditions selecting for plasmid maintenance [27,177,278]. We hypothesize that by doing this, pAgos can minimize drawbacks of invading nucleic acids (lowering the metabolic burden), while leaving some room for beneficial genetic treats to enter the cell, allowing a potential gain of fitness for their hosts.

pAgo-mediated transposon and virus interference

Prokaryotes additionally require defense against other invading nucleic acids, such as transposons and viruses, which are less likely to provide a fitness gain. Besides preferentially associating with guides complementary to plasmid DNA, *Rs*Ago preferentially associates with guides complementary to transposon and virus genes [177]. These targeted transposon and virus genes appear to have an extrachromosomal stage during their lifecycle. In addition, it has been described that *Tt*Ago expression is up-regulated upon virus infection [289]. This suggests that pAgos not only interfere with invading plasmid DNA, but possibly also with other invading nucleic acids. Although *Tt*Ago-mediated down-regulation of RNAs encoded by transposase-genes has never been observed [278], it cannot be ruled out that pAgos interferes with transposon and/or virus DNA during their extrachromosomal life stages.

Lytic viruses highjack the metabolism of their host to replicate and lyse the cell before spreading to find new hosts [345]. The strong interfering mechanism of restriction-modification systems [344] and CRISPR-Cas systems [240,298] provide a welcome barrier to this kind of invaders. We tested if *Tt*Ago also provides defense against invading lytic viruses (**Fig. 1**). Unfortunately, only three *T. thermophilus* viruses are characterized in detail: PH75 [346], ϕYS40, [347] and ϕTMA [348]. As we could not acquire infectious PH75 stocks, and as ϕYS40 infects *T. thermophilus* HB8 but not HB27 (data not shown), our single option was ϕTMA. ϕTMA is a lytic virus with a dsDNA genome that is 151,483 bp in length [348]. When we compare the plaquing efficiency of ϕTMA in wild type *T. thermophilus* HB27 with HB27Δ*ago* and HB27Δ*ago::*^{s}*ago*, we observe a small but significant 1.5-fold increase of plaque forming units in HB27Δ*ago* compared to HB27 (**Fig. 1b**, $P<0.01$).

In culture collapse studies, we observe a slightly earlier collapse of the HB27Δ*ago* culture compared to the wild type (**Fig. 1c**). Thus, it appears that *Tt*Ago provides a small fitness benefit during ϕTMA infection, although it does not appear to play a major role in defense against this virus. Possibly, ϕTMA replicates too fast and lyses the cell too early to allow *Tt*Ago to provide defense. Alternatively, the highly specialized ϕTMA virus encodes *Tt*Ago inhibitors or modifies its DNA, making it is resistant to *Tt*Ago mediated host defense. Although *Tt*Ago-mediated defense against ϕTMA is not very strong, it is not unlikely that *Tt*Ago and/or other pAgos provide a strong barrier against other invading viruses.

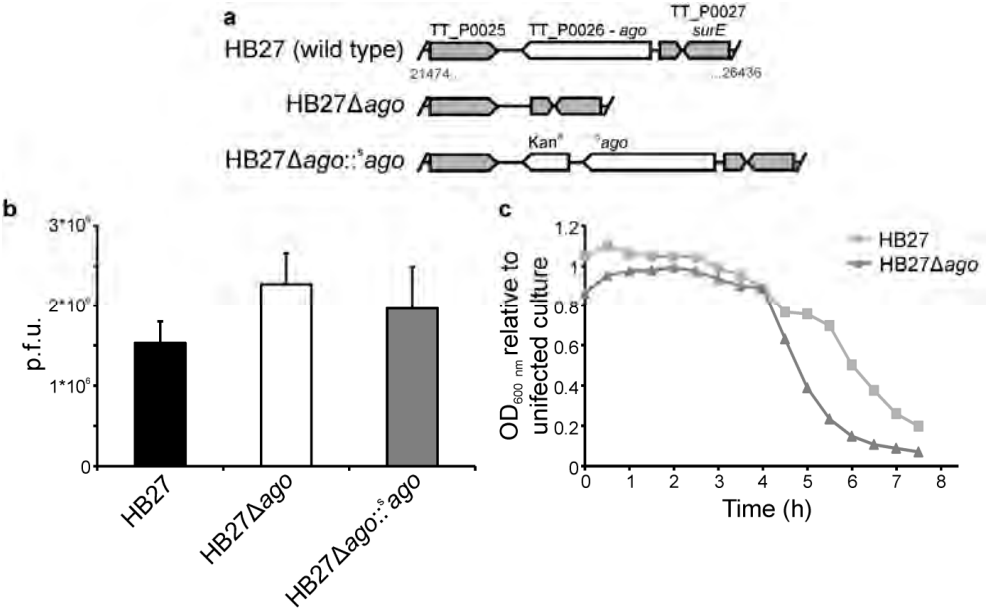


Figure 1 | Effect of *Tt*Ago on ϕTMA infection. a, Overview of ago gene loci of *T. thermophilus* strains HB27 (wild type), HB27Δ*ago* (knockout), and HB27Δ*ago::*^{s}*ago* (HB27Δ*ago* complemented with at *strep(II)-tag-ago* gene fusion insert). Kan^{R}: kanamycin resistance marker. **b**, Plaque forming units (p.f.u.) of ϕTMA infecting different *T. thermophilus* strains. Error bars indicate standard deviations of biological triplicates. **c**, *T. thermophilus* culture collapse after addition of ϕTMA at time=0. Percentage of remaining $OD_{600\ nm}$ units was calculated by comparing $OD_{600\ nm}$ of infected cultures (biological duplicates) with $OD_{600\ nm}$ uninfected controls (biological duplicates).

The role of protecting the host against invading nucleic acids is not restricted to the prokaryotic variants of Agos. Multiple eAgos are involved in transposon and virus silencing (reviewed in [121,122,123,124]). In contrast to targeting the genome of DNA invaders, certain eAgos target the RNA encoded by transposons viruses. Additionally, genomes of both ssRNA viruses and dsRNA viruses are targeted directly. Thus, although the eukaryotic mechanisms for host defense appear to take place at the RNA level rather than at the DNA level, it is concluded that the role of pAgos in host defense is conserved in all domains of life.

Mechanism of pAgos

Although the physiological role of some eAgo and pAgo variants appears to be conserved, some of their mechanistic details are clearly different. eAgos function in RNA-guided RNA interference pathways, although DNA-guided activity should not be ruled out [275]. While

it previously has been established that some pAgos utilize 5'-phosphorylated DNA guides and can cleave ssDNA targets *in vitro* [169], this thesis provides proof for the occurrence of this phenomenon *in vivo*. In addition, *Rs*Ago appears to mediate RNA-guided DNA interference [177], indicating that eAgos and pAgos have varied preferences for guides and targets. Here we discuss possible determinants for guide and/or target preference, as well as the conserved features of seed-pre-ordering and target cleavage.

Distinguishing between DNA and RNA

Currently, it remains unknown how exactly pAgos distinguish between DNA and RNA guides and targets, although several hypotheses exist. Possibly guide-type preference is determined by differential guide generation. In other words, only RNA guides or only DNA guides are generated. We consider it very unlikely that this is the only determinant, as we have demonstrated that *Tt*Ago and *Pf*Ago cannot utilize RNA guides for target cleavage [27,277]. It appears more likely that structural features determine if DNA or RNA guides are bound (reviewed in [230]). For example, the MID domain-located pocket binding the 5' end of the guide is more hydrophobic in *Tt*Ago compared to hAGO2 [178], and even more hydrophobic in and *Pf*Ago [273]. Possibly the 2'-OH groups on the ribose ring that distinguishes RNA from DNA nucleotides is repelled by this hydrophobicity, thereby determining that only DNA guides and/or targets can be bound. Additionally, the 2'-OH groups which are present over the entire RNA backbone may allow to distinguish them from DNA by specific binding or exclusions of these groups. Indeed, it has been observed that several 2'-OH groups located on the RNA guides are specifically bound by eAgos [28,178,179,193,194]. Also RNase H, an evolutionary precursor of the PIWI domain, is able to discriminate between the DNA and RNA strands from its DNA/RNA duplex target by specifically binding or excluding 2'-OH groups [281,282].

To investigate if guide and/or target preference can be deduced from specific amino acid residues, various characterized guide and/or target binding residues of *Tt*Ago [176], hAGO2 [28] and AtAGO1 [349] were annotated in a multiprotein sequence alignment (**Fig. 2**). This alignment demonstrates that multiple residues interacting with guide backbone phosphates are conserved in both RNA-guided and DNA-guided Agos (**Fig. 2**, residues colored red). However, several polar residues in RNA-guided Agos (**Fig. 2**, residues highlighted green) that specifically interact with 2'-OH groups are only found in some RNA-guided Agos. It should be noted however, that these residues are conserved only in a

narrow group of fungal and metazoan eAgos [27]. At the same position, DNA guided Agos sometimes have hydrophobic or negatively charged residues which possibly exclude 2'-OH groups.

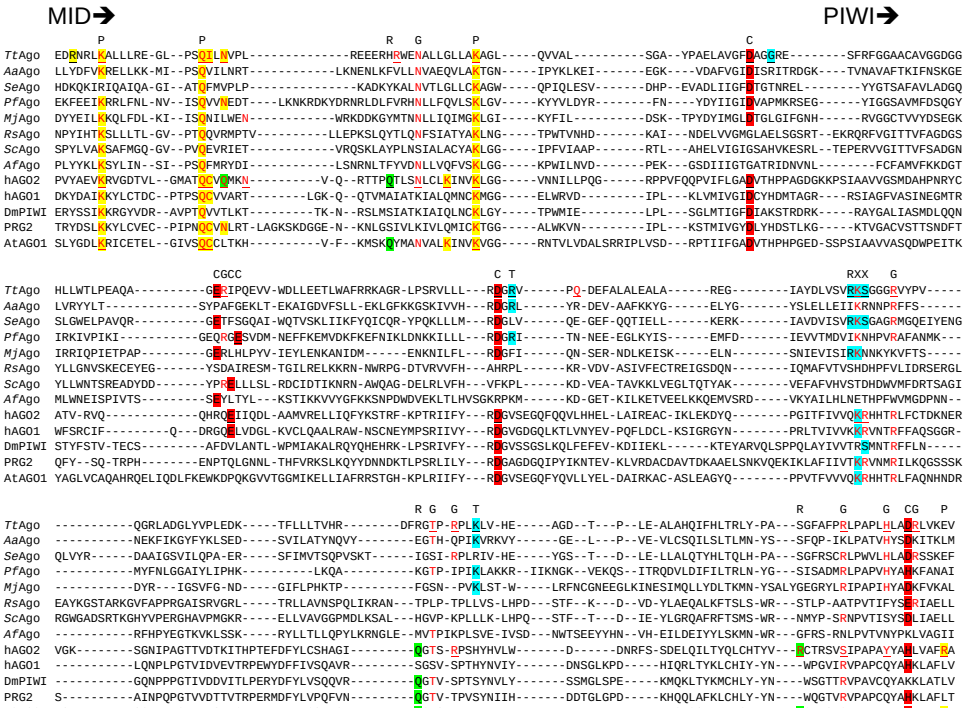


Figure 2 | Alignment of MID-domain 5' end-binding pocket and PIWI domain of various Agos. The alignment is adapted from [230]. Ago sequences are from *Tt*Ago (*Thermus thermophilus* HB8), *Aa*Ago (*Aquifex aeolicus* VF5), *Se*Ago (*Synechococcus elongatus* PCC 7942), *Pf*Ago (*Pyrococcus furiosus* DSM 3638), *Mj*Ago (*Methanocaldococcus jannaschii* DSM 2661), *Rs*Ago (*Rhodobacter sphaeroides* ATCC 17025), *Sc*Ago (*Streptomyces coelicolor* A3 2), *Af*Ago (*Archaeoglobus fulgidus* DSM 4304), hAGO2 (*Homo sapiens*), hAGO1 (*Homo sapiens*), DmPIWI (*Drosophila melanogaster*), PRG2 (*Caenorhabditis elegans*), and AtAGO1 (*Arabidopsis thaliana*). Residue annotation was deduced from structures of *Tt*Ago [176], hAGO2 [28] and the AtAGO1 MID domain [349]. Residues of which the function was deduced from structures are underscored. Note that for all residues that are not underscored, the function is predicted. P: conserved residue for 5'-end phosphate binding (highlighted yellow, colored red). C: DEDX catalytic tetrad residue (highlighted red). G: residue involved in guide backbone phosphate binding (colored red). T: residue involved in target backbone phosphate binding (highlighted blue). X: residue involved in target and/or guide backbone phosphate binding. R: residue involved in RNA guide 2'-OH group binding (highlighted green).

At this point, it is not possible to determine whether a specific pAgo utilizes RNA or DNA guides based on the amino acid sequence and available structures alone. Nevertheless, there

are trends that, alone or combined, could determine which type of guides and/or targets are acquired. It should be noted that these residues are not conserved and that these observations are based on limited structural data. More structural studies should be performed before strong conclusions can be drawn.

The nucleotide specificity loop

For various eAgos, preferential binding of RNA guides with specific 5' ends has been described [86,182,199]. This preference is caused by a structural feature termed the 'nucleotide specificity loop', of which amino acids interact with the base of the first nucleotide of the guide [182,199]. As eukaryotes often encode multiple eAgos, the nucleotide specificity loop plays an important role in guide sorting [199]. Little is known about the nucleotide specificity loop in pAgos. Although specific interactions of the nucleotide specificity loop of *Tt*Ago with a 5' thymine have been reported [169], it is unknown if these interactions are the preferred above interactions with other 5'-end nucleotides. In fact, *Tt*Ago appears to specifically interact with guides with a 5'-end cytosine *in vivo* [27]. Like many eAgos, *Rs*Ago appears to have a preference for guides with a 5'-end uracil [177]. This suggests that binding of guides with a specific 5'-end nucleotide also is important in pAgos.

The seed segment of the guide

The 'seed' is used to describe a specific segments in RNA guides of Agos, prokaryotic small regulatory RNAs and prokaryotic CRISPR RNAs (reviewed in [175]). This segment plays a critical role in target binding. The term 'seed' was first used to describe miRNA nucleotides 2 to 7-8, as perfect base pairing of only these guide nucleotides with complementary target nucleotides appeared to be sufficient for causing translational repression [350,351], which is true for eAgo-mediated translational repression in general (**Fig. 3**) [350,352,353,354]. Although the nucleotide composition of the seed affects binding affinity and the degree of translational repression [355], seed–target mismatches, bulges and G:U wobbles typically decrease target binding affinities much more than can be explained by thermodynamics of the seed-target match alone [225,350,353,356]. This indicates that the structure of the seed plays an important role during target binding. The seed segment is located in a narrow groove of the PIWI lobe in which specific amino acids of the protein interact with backbone phosphates of the DNA or RNA guides, and with the 2'-OH groups of RNA guides [28,168,176,178,179]. The guide is bound such that nucleotides 2 to 6-8 are pre-ordered in

an A-form helix with bases pointed towards the solvent to promote target binding. The pre-ordered helix lowers the entropic cost when the seed of the guide forms a stable duplex with the target [188,357].

During perfect seed–target pairing, additional pairing of the 3' region of the guide to the target, called supplementary pairing, can enhance translational repression (**Fig. 3**) [225,350,352,353,354]. Supplementary pairing requires four contiguous Watson-Crick base pairs in the region of guide nucleotides 13 to 16. If seed pairing is imperfect, 3' compensatory pairing at a minimum of nine nucleotide positions may compensate for a single bulge or mismatches in the seed region [76,77,225,350,353,354,355,358]. Compensatory pairing can decrease the seed pairing requirement to as little as four base pairs [350].

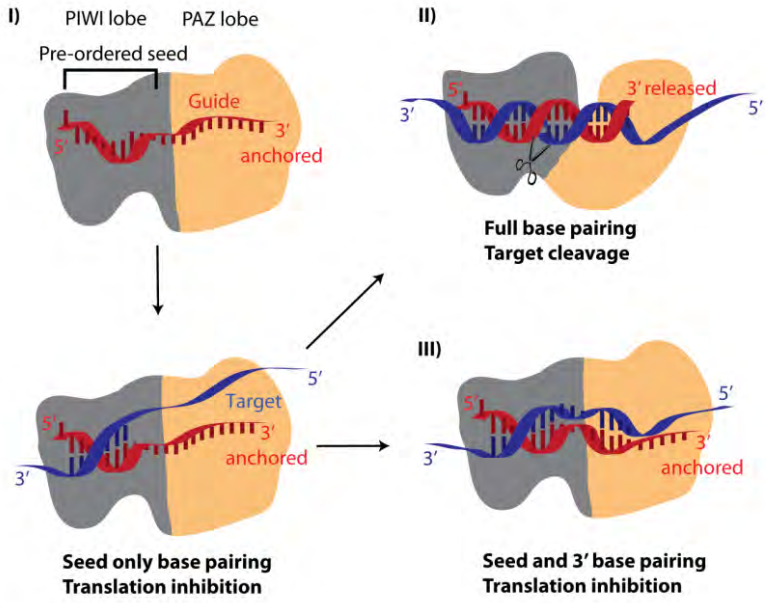


Figure 3 | Schematic representation of guide-target binding. a, The guide (red) contains a seed segment that is pre-ordered in the PIWI lobe. Binding of the target (blue) initiates at this segment. In eAgo pathways, mRNA target binding by the seed alone is sometimes sufficient for translational inhibition. **b**, Upon binding of complementary targets, the 3' end of the guide is released and conformational changes are induced. In Agos with a catalytic DEDX tetrad, this results in target cleavage. **c**, In the case of imperfect target binding, or in pathways with eAgos without catalytic residues, supplementary pairing of the target with the 3' end of the guides enhances translational inhibition. Figure adapted from [175].

Target cleavage

Perfect pairing of the guide to the target results in ATP-independent target cleavage by Ago when it has a complete catalytic DEDX tetrad [68,75,76,176,225,242,258,359,360]. Single and double guide–target mismatches lower the activity of the slicer complex depending on their position [258,356,359,360,361]. Mismatches in the seed and 3' complementary region lower the target cleavage efficiency, whereas mismatches in nucleotides 10 and 11 results in a complete loss of cleavage [242,356,361]. Mismatches are tolerated at position 1, as this nucleotide does not contribute to target binding [225] but instead is required for stable binding of the guide to Ago; binding of the 5'-end nucleotide of the guide in the MID domain binding pocket prevents base pairing of this nucleotide with the target [28,167,176,178,179]. The study of the structures of *Tt*Ago-siDNA complexes bound to targets of varying lengths allowed to propose a detailed mechanistic model of target cleavage [176]. Upon perfect guide-target pairing, conformational changes are induced in the structure of Ago. These changes result in the repositioning of a so-called glutamate finger such that the glutamic acid residue located on the finger can complete the DEDX (where X can be either D or H) catalytic tetrad. This allows two divalent cations to accurately position an activated water molecule that is responsible for a nucleophilic attack of the phosphate di-ester bond, resulting in nicking of the target nucleic acid strand. The DEDX tetrad is conserved in all catalytically active pAgos and eAgos [4,230].

pAgo-interacting proteins

Besides differences in guide and target preferences, the role of eAgos is determined by the proteins with which they interact. Although some pAgo genes co-localize with specific genes, many catalytically active long pAgos (amongst which *Tt*Ago and *Pf*Ago) appear to occur as stand-alone genes. *In vitro* experiments with *Tt*Ago and *Pf*Ago [27,274,277] confirm that these pAgos function as stand-alone proteins, as they do not require assistance of other proteins for guide generation, loading and target cleavage. Nevertheless, we cannot rule out that *in vivo* activity of these pAgos is determined or at least affected by interacting proteins.

To identify proteins that interact with *Tt*Ago and *Pf*Ago *in vivo*, we performed Strep(II)-tag affinity pull-down experiments and immunoprecipitations, respectively. After purification or immunoprecipitation of *Tt*Ago and *Pf*Ago and possible pAgo-interacting proteins,

protein samples were analyzed by Mass spectrometry. Three independently performed *Tt*Ago co-purifications and a single *Pf*Ago co-immunoprecipitation yielded inconsistent results. All purifications appeared to be enriched in various nucleic acid-binding proteins (data not shown), but we could not functionally link *Tt*Ago and *Pf*Ago with these proteins. As the interactions appear to be random, we hypothesize that the co-purified proteins interact with nucleic acids bound by pAgo rather than directly with the pAgo. Thus, no specific interaction partners of these catalytically active long pAgos could be identified. In contrast to *Tt*Ago and *Pf*Ago, the genes of many other catalytically active and inactive long pAgos cluster in (predicted) operons with other genes (**Fig. 4**) [230]. Short pAgos, which consist of the PIWI lobe alone (containing the MID and PIWI domains), also co-occur with specific genes (**Fig. 4**). Here, we describe the proteins encoded by these genes and try to link their (predicted) activity with that of pAgo.

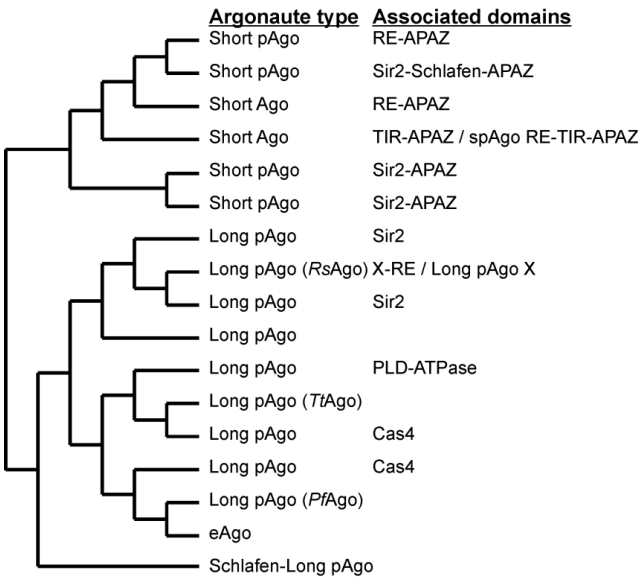


Figure 4 | Schematic phylogenetic tree of Argonaute proteins and associated domains. Simplified phylogenetic tree adapted from [230]. Short pAgo: (MID-PIWI only). RE: Restriction Endonuclease superfamily. APAZ: Analog of PAZ domain. Sir2: Predicted Sir2-like nuclease. TIR: Predicted nuclease of TIP family. Schlafen: Predicted ATPase. Cas4: Cas4 subfamily of restriction endonuclease superfamily. PLD: predicted nuclease of phospholipase D superfamily. X: distinct families of uncharacterized proteins.

Proteins associated with catalytically active long pAgos

Long pAgos that associate with certain genes form specific clades in the phylogenetic tree (**Fig. 4**). There are two types of catalytically active long pAgos which co-occur with other specific genes. These pAgos associate either with Cas4 or with a Phospholipase D (PLD)-superfamily protein.

Cas4 is a CRISPR associated protein that is predicted to be involved in CRISPR adaptation, as it associates with Cas1 and Cas2 [299], the core machinery of CRISPR adaptation (reviewed in [294,295,298]). Additionally, *cas4* is often found in the genetic neighborhood of *cas1* and *cas2* [362], and sometimes *cas4* is fused to *cas1* [363]. More recently, two independent studies have shown that Cas4 has a 5'-3' exonuclease activity [210,364]. pAgo-associated Cas4 could be involved in guide generation, or possibly it uses its (predicted) helicase activity [362] for making dsDNA target available for pAgo binding. Another possibility is that Cas4 is recruited by pAgos to further degrade pAgo targets. In this process, Cas4 possibly generates 3' overhangs on the target DNA, which might stimulate CRISPR adaptation. This would imply interplay between pAgos and CRISPR adaptation, which previously has been suggested based on activation of CRISPR-Cas genes after *Tt*Ago mediated plasmid interference [278].

Members of the PLD superfamily are found in all domains of life and often hydrolyze substrates to generate signaling molecules [365]. Other members of this protein family have been shown to hydrolyze the phosphodiester bond found in the backbone of DNA (reviewed in [365,366]). The PLD domain associated with pAgos is fused to a helicase domain, a combination also found in bacterial restriction-modification systems [212], making it likely that the pAgo-associated PLD domains function as nucleases. Like pAgo-associated Cas4, this protein might play a role in target binding or degradation, or alternatively in guide generation. The latter is not unlikely, as the eukaryotic protein Zucchini (the primary piRNA biogenesis factor) also belongs to the PLD-superfamily [90,366,367]. Although Zucchini itself does not have a helicase domain, piRNA loading is dependent on helicases. As such, it makes sense that pAgo-associated PLD proteins are fused to helicases. It has previously been stated that important proteins for eukaryotic RNAi pathways, such as Dicer and RdRP, are not encoded by prokaryotes [4,170,230], which makes it unlikely that miRNA- or siRNA-like pathways are present in prokaryotes. However, the co-occurrence of genes encoding PLD-superfamily proteins and Agos suggests an ancient piRNA-like pathway could be encoded in prokaryotes.

Proteins associated with catalytically inactive long pAgos

Genes encoding long pAgos that have lost their catalytic DEDX residues primarily associate with genes encoding (predicted) nucleases such as the Sirtuin (Sir2) protein, or with the Methylated adenine Recognition and Restriction (Mrr) protein (**Fig 4**). For these examples,

it is likely that the pAgo plays a role in guide-mediated target binding while the other protein is recruited by pAgo to cleave the nucleic acid target. Sir2 proteins function as protein deacetylases both in prokaryotes and eukaryotes [368]. Some eukaryotic Sir2 proteins deacetylate histones, resulting in chromatin formation. Other eukaryotic and bacterial Sir2 homologs deacetylate proteins to regulate their activity. However, the Sir2 proteins that associate with pAgos lack specific features which are conserved in Sir2 proteins involved in protein deacetylation. Based on the presence of a DxH motif in some of these bacterial Sir2 proteins, it was predicted that they function as nucleases [207]. Although there is no experimental proof for this, it can be hypothesized that pAgos recruit Sir2 for target cleavage. Mrr proteins are type IV restriction endonucleases that specifically cleave N6-adenine-methylated DNA [369]. Additionally, Mrr sometimes plays a role in SOS-response induction [369,370]. Mrr proteins that are encoded by genes that associate with pAgo encoding genes have the conserved (D/E)-(D/E)XK nuclease motif [4]. Possibly pAgo-associated Mrr proteins cleave methylated invader DNA, generating guides for pAgo in the process. Alternatively, the Mrr proteins are recruited by pAgos to cleave their targets, but it is not obvious why a methylation-dependent nuclease would be required for this. Maybe the requirement for methylation lowers the chance on auto-immunity. Alternatively, the requirement for methylation is lost in these proteins.

In addition, some catalytically inactive long pAgos co-occur with Schlafen-superfamily (SLFN) proteins (**Fig. 4**). Upon closer inspection of their genes it becomes clear that these are not co-occurring genes, but gene fusions. These proteins encompass a slfn box, of which the function is unknown, and an AAA domain, suggesting it requires ATP for activity [371]. Some SLFN proteins have a motif that is homologous to superfamily I of RNA helicases [213]. Eukaryotic SLFNs are involved in cell proliferation, induction of immune responses and regulation of viral replication (reviewed in [372]). Possibly the pAgos fused to SLFN guide SLFN activity. Alternatively, if the bacterial SLFN contains helicase activity, it might make double stranded nucleic acids targets available for pAgo binding, or it (further) unwinds targets bound by the pAgo domain.

Proteins associated with short pAgos

Short pAgos lack the N and PAZ domains, which in long pAgos and eAgos play an important role in passenger/target strand removal, and in binding of the 3' end of the guide, respectively. Most likely, the MID and PIWI domains of short pAgos facilitate binding of the 5' end of the guide as well as binding and pre-ordering of the seed segment, like they do

in long pAgos and eAgos (reviewed in [230]). As catalytic residues are lost in all short pAgos [4,230], it is anticipated that they form short pAgo-guide complexes whose only function is to bind target nucleic acids. For catalytic activity, short pAgos are predicted to rely on co-occurring proteins. All proteins encoded by the same operon as short pAgos at least encompass a predicted Analog of PAZ (APAZ) domain (**Fig. 4**) [4,230]. This domain has no detectable sequence homology with the PAZ domain, but is not found in another context than with short pAgos. The domains fused to APAZ domains vary (**Fig. 4**), but many of these domains are putative nucleases. Mrr and Sir2 domains fused to APAZ domains are from different subfamilies than the ones associated with long pAgos but likely fulfill similar roles. Also Toll-Interleukin Receptor (TIR)-like domains fused to APAZ have predicted nuclease activity [230]. In addition, APAZ is sometimes fused to a combination of multiple domains. Examples include Sir2-APAZ-short pAgo, Sir2-Schlafen-APAZ and Mrr-TIR-APAZ. These domain fusions might improve the functionality of the associated short pAgo.

No short eAgos exist

Like some long pAgos and all short pAgos, ~10% of the eAgos lost their catalytic residues. These eAgos function by target binding and sometimes recruit additional proteins to the targeted RNA (reviewed in [230]). It is striking that catalytically inactive eAgos, unlike catalytically inactive pAgos, always encompass the PAZ and N domains. The fact that no short eAgos exist agrees with the proposed phylogenetic scenario that eAgos evolved from catalytically active long pAgos [4,230]. Later in the evolution, eAgos might have lost catalytic residues without losing the N and PAZ domains. Although these catalytically inactive eAgos do not require the N and PAZ domains for cleavage, the N domain might still play a role in guide loading, as removal of the passenger strand of a duplex RNA sometimes is N-domain dependent (reviewed in [230]). The function of the PAZ domain possibly also remained essential by contributing to overall guide binding. Additionally, guides of which the 3' end is bound by PAZ binding pockets are shielded to degradation by nucleases, which increases the lifespan of eAgo-bound guides. Thus, although short pAgos are predicted to be functional without the N and PAZ domains, the same domains seem essential for eAgo activity, even if catalytic residues are lost. Possibly some of the functions of these domains are also important for the few long pAgos that lost their catalytic residues. The observation that short pAgos without exception co-occur with proteins containing at

least the APAZ domain, suggests that this domain also could play a role in guide loading and/or protecting the guide against nucleases.

pAgo as genome editing tool

Genome editing is a type of genetic engineering in which site-specific nucleases are used to generate double-stranded breaks (DSBs) in DNA genomes (reviewed in [373,374]). In eukaryotes, DSBs induce endogenous DNA repair mechanisms such as non-homologous end-joining (NHEJ) and homologous recombination (HR). Error-prone NHEJ often generates indels, which can result in frameshifts. If template DNA is provided, HR can result in integration of the template, allowing gene insertion or replacement. Genome editing is a powerful strategy in biological research, biotechnology, and possibly in future human gene therapy [375]. Currently genome editing relies on a small selection of programmable nucleases.

Both Zinc Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) consist of multiple domains which each recognize specific bases or base patterns, which are fused to a non-specific nuclease domain. By varying the base recognition domains, the DNA binding specificity of the protein can be engineered. The nuclease domain allows cleavage of the targeted DNA sequence. A major drawback of the use of these tools for genome editing is that they require new protein design and generation for every desired target site. In contrast, the CRISPR-Cas9 tool (reviewed in [376]) is an RNA guided protein. Cas9 guides can easily be synthesized, transcribed *in vitro*, or expressed *in vivo*, which makes it easy to alter the specificity of Cas9. Upon binding of DNA complementary to a 20-nucleotide segment of the guide RNA, Cas9 generates DSBs. Due to its high specificity and simplicity of programming, Cas9 has emerged as an extremely powerful molecular tool for genome editing [377,378]. The potential of Cas9 is limited by the need for a protospacer adjacent motif (PAM; usually 5'-NGG-3') next to the targeted sequence, which is required by Cas9 to bind and cleave the target [379].

The limitations of current tools generate the need for alternative technologies for both research and ultimately clinical use of genome editing. pAgos are interesting candidates to fulfill these needs. Like Cas9, pAgos utilize nucleic acid guides to target DNA. Furthermore, pAgos do not require a PAM and are much smaller than Cas9, which makes delivery to cells

easier. It has been demonstrated that both *Tt*Ago and *Pf*Ago can be loaded with synthetic DNA guides to cleave dsDNA at the sequence of choice *in vitro*. However, their high optimum activity temperature limits their use for *in vivo* genome editing in mesophiles. Characterization of mesophilic pAgos might uncover more suitable candidates for genome editing, and have the potential to address the limitations of current genome editing tools.

Future perspectives

Currently, the physiological role of only three pAgos has been characterized [27,177,277], and only for two of these pAgos the biochemistry has been described [27,277]. As the variety of pAgos is much larger than that, additional research needs to be performed to get a better understanding of the functionality and biochemistry of uncharacterized pAgo variants and their co-occurring proteins. Furthermore, future structural studies might give additional insights in Ago mechanisms. Such studies possibly allow identification of residues involved in guide and target binding, and identification of structural domains involved in binding of pAgo-associated domains. These insights are very interesting from a fundamental point of view, and in addition they might unlock the potential of pAgos to form a new class of genome editing tools.

Appendices

Nederlandse samenvatting

Argonaute eiwitten (Agos) komen voor in alle domeinen van het leven, en spelen een belangrijke rol als nuclease die enkelstrengs nucleïnezuren (de 'guide') binden. Deze guide wordt gebruikt door Ago eiwitten om complementaire nucleïnezuren (het 'target') op te sporen en te binden, waarna Ago het target knipt. Deze thesis beschrijft de evolutie, de rol en het mechanisme van prokaryote Argonaute eiwitten (pAgos), en focust op de pAgos uit de bacterie *Thermus thermophilus* (*Tt*Ago) en uit het archaeon *Pyrococcus furiosus* (*Pf*Ago).

Hoofdstuk 1 geeft een overzicht van de evolutie, structuur en functie van eukaryote Argonauten (eAgos) en pAgos. Ondanks dat de aminozuursequentie tussen eAgos en pAgos en zelfs tussen verschillende pAgos erg veel kan afwijken, is de algemene architectuur van Ago eiwitten bewaard gebleven: eAgos en pAgos bestaan uit vier domeinen: Het MID domein (bindt het 5'-uiteinde van de guide), het PIWI domein (knipt het target), het PAZ domein (bindt het 3'-uiteinde van de guide), en het N domein (speelt een rol bij het losmaken van guide-target duplexen, en indirect bij het knippen van targets). Ondanks deze structurele homologie zijn kleine verschillen in structuur aanwezig, welke een grote invloed hebben op de functionaliteit van de verschillende Agos. Waar pAgos lijken te functioneren als op zichzelf staande eiwitten, hebben eAgos insertiesegmenten die het mogelijk maken andere eiwitten te binden. Deze insertiesegmenten hebben het voor eAgos mogelijk gemaakt om onderdeel te worden van complexe RNA-regulerende processen.

T. thermophilus is een gramnegatieve thermofiele bacterie met een natuurlijk competentie systeem [234], waarmee het DNA uit zijn omgeving kan opnemen. **Hoofdstuk 2** beschrijft de rol van *Tt*Ago in *T. thermophilus* HB27. Een genomische *ago* knock-out (HB27Δ*ago*) is gemaakt en de competentie van deze stam is vergeleken met wild-type HB27, door ze beide te transformeren met plasmide DNA. Transformatie van wild-type HB27 resulteerde in tien keer minder kolonies vergeleken met transformatie van HB27Δ*ago*. Ook na transformatie verlaagt *Tt*Ago de intracellulaire plasmideconcentratie drie tot vijf keer, zelfs wanneer deze stammen gecultiveerd werden in aanwezigheid van antibiotica waardoor het plasmide normaalgesproken behouden blijft. Dit laat zien dat *Tt*Ago interfereert met plasmide transformatie en propagatie. Na heterologe expressie van *Tt*Ago in *Escherichia coli* zijn

nucleïnezuren die met *Tt*Ago opzuiverden geanalyseerd. Katalytisch actief *Tt*Ago wordt geladen met enkelstrengs DNA guides, welke 13 tot 25 nucleotiden lang zijn en een fosfaat groep hebben aan hun aan hun 5'-uiteinden. Deze guides zijn 'small interfering DNAs' (siDNAs) genoemd. We hebben een techniek ontwikkeld om de sequentie van siDNAs te bepalen en hebben laten zien dat *Tt*Ago voornamelijk is geladen met siDNAs die verkregen zijn van het plasmide waarvan *Tt*Ago tot expressie werd gebracht. Deze siDNAs maken het mogelijk voor *Tt*Ago om dubbelstrengs plasmide DNA te knippen. Met synthetische siDNAs hebben we aangetoond dat *Tt*Ago enkelstrengs DNA en enkelstrengs RNA kan knippen, en dat het een van de twee strengen van dubbelstrengs DNA kan knippen. Twee verschillende *Tt*Agos geladen met complementaire siDNAs, welke ieder een andere streng van het dubbelstrengs DNA bindt, kunnen dubbelstrengs DNA knippen. Deze resultaten hebben aangetoond dat *Tt*Ago als afweersysteem werkt door met behulp van siDNAs met DNA te interfereren.

Hoofdstuk 3 geeft inzicht in het mechanisme waarmee *Tt*Ago zijn targets knipt. De kristalstructuur van *Tt*Ago geladen met een DNA guide, welke gebonden is aan target DNA van 12 of 15 nucleotiden lang, laat zien dat het complex in een conformatie blijft waarin het geen target kan knippen. Wanneer langere targets van 16 of 19 nucleotiden worden gebonden, vinden er conformationele veranderingen plaats: het PAZ domein roteert terwijl het 3'-uiteinde van de guide los wordt gelaten. Verder bewegen drie 'loops' in het PIWI domein, waardoor loop 2, ook wel de 'glutamaat vinger' genoemd, in nabijheid van de andere katalytische residuen wordt gebracht. Hierdoor wordt het katalytisch DEDD quadruplet gevormd. Hierna worden twee divalente kationen gebonden, welke een geactiveerd watermolecuul coördineren, waarna dit molecuul de fosfaatgroep tussen nucleotide 10 en 11 van het target DNA aanvalt. In dit hoofdstuk wordt ook een model gegeven voor het stap-voor-stap target-geïnduceerde activatie van het pAgo-guide complex en voor het knippen van het target DNA[176].

De karakterisatie van archaeaal *Pf*Ago is beschreven in **Hoofdstuk 4**. *Pf*Ago interfereert net als *Tt*Ago met plasmide transformatie. Hoewel *Pf*Ago tot een fylogenetische aftakking behoort die van alle pAgos het dichtst gerelateerd is aan eAgos, gebruikt *Pf*Ago uitsluitend DNA guides en kan het alleen DNA targets knippen. De activiteit van *Pf*Ago wordt gefaciliteerd door een katalytisch DEDH quadruplet, waarmee Mn^{2+} of Co^{2+} wordt gebonden om target DNA te knippen. Net als *Tt*Ago kan *Pf*Ago siDNAs gebruiken om dubbelstrengs DNA te knippen. Deze bevindingen suggereren dat de rol van pAgo als

afweersysteem tegen invasief DNA aanwezig is in bacteriën en in archaea. Omdat *Pf*Ago gerelateerd is aan eAgos, zou deze ontdekking kunnen betekenen dat ook de evolutionaire voorlopers van eAgos DNA guides gebruikten om DNA targets te knippen.

In **Hoofdstuk 5** is een mechanisme beschreven waarmee *Tt*Ago zijn eigen siDNAs kan maken. Hoewel eAgos afhankelijk zijn van andere eiwitten voor het maken van guides, zijn deze eiwitten niet aanwezig in prokaryoten. Met *in vitro* experimenten wordt aangetoond dat guide-vrij *Tt*Ago onder de juiste condities dubbelstrengs DNA kan afbreken, een proces wat 'DNA chopping' is genoemd. Gedurende DNA chopping worden stukjes DNA van 13 tot 25 nucleotiden lang gegenereerd, welke een fosfaat groep hebben aan hun 5'-uiteinde. Deze stukjes DNA kunnen gebruikt worden als guide door *Tt*Ago om complementaire targets te knippen.

We hebben uitgezocht of *Tt*Ago naast plasmide DNA ook RNA knipt *in vivo*. We hebben RNA-seq analyses uitgevoerd op RNA geïsoleerd uit vier verschillende *T. thermophilus* stammen (HB27, HB27Δ*ago*, en beide stammen met een plasmide). Deze analyse is beschreven in **Hoofdstuk 6** en laat zien dat *Tt*Ago geen invloed heeft op expressie van genen betrokken bij competentie of afweer tegen invasief DNA. Verder laten we zien dat de aanwezigheid van plasmide DNA geen veranderingen aanbrengt in het RNA-expressie patroon van *T. thermophilus*, en dat de veranderingen in plasmide RNA-expressie komen door verminderde aanwezigheid van plasmide DNA. *Tt*Ago heeft dus geen directe invloed op plasmide-gecodeerd RNA. Echter, wanneer *Tt*Ago en plasmide DNA beide aanwezig zijn, wordt de expressie van verschillende genomische CRISPR loci en CRISPR-geassocieerde eiwitten omhoog gereguleerd. Sommige van deze genen zijn betrokken bij CRISPR-adaptatie, wat suggereert dat afbraak van plasmide DNA door *Tt*Ago CRISPR adaptatie stimuleert.

CRISPR-Cas is een adaptief immuunsysteem dat prokaryoten in drie stappen beschermt tegen invasief DNA [295,298]: de adaptatie stap, de expressie stap en de interferentie stap. In de adaptatie stap worden kleine stukjes DNA, verkregen van het invasief DNA, ingebouwd in genomische CRISPR loci. Gedurende de expressie stap wordt er CRISPR RNA afgelezen van deze loci, en wordt dit RNA bewerkt en geladen in Cas-eiwit complexen. In de interferentie-stap gebruiken de Cas-eiwit complexen het stukje CRISPR RNA als guide om invasief DNA te binden en te degraderen.

In **Hoofdstuk 7** is een studie naar CRISPR adaptatie in *E. coli* K12 beschreven. Door middel van het bepalen van de sequentie van CRISPR loci van bacteriën die hun plasmide zijn verloren, laten we zien dat stukjes DNA (spacers) zijn verkregen van het plasmide en zijn ingebouwd in het CRISPR locus. Deze spacers komen van specifieke locaties op het plasmide, waarbij rekening is gehouden met een motief van drie nucleotiden naast de locatie waarvan de spacer is verkregen. Nieuwe spacers worden ingebouwd aan een specifieke kant van de CRISPR locus. Er wordt aangetoond dat een specifieke nucleotide naast het ingebouwde DNA ook verkregen is van het plasmide. Verder laten we zien dat integratie van een enkele spacer kan leiden tot versnelde integratie van extra spacers, wat een positieve feedback loop veroorzaakt in CRISPR-Cas adaptatie.

Het werk dat is beschreven in deze thesis draagt bij aan de fundamentele kennis van de biologische functies van pAgos, welke de evolutionaire voorlopers van eAgos zijn. De ontdekking dat sommige pAgos direct en DNA kunnen knippen op specifieke locaties heeft de spannende mogelijkheid aan het licht gebracht om pAgos te ontwikkelen als moleculair gereedschap waarmee DNA geknipt kan worden op locaties naar keuze.

References

1. Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J 17: 170-180.
2. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 293: 1146-1150.
3. Cerutti L, Mian N, Bateman A (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. Trends in biochemical sciences 25: 481-482.
4. Makarova KS, Wolf YI, van der Oost J, Koonin EV (2009) Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biology direct 4.
5. Green PJ, Pines O, Inouye M (1986) The Role of Antisense Rna in Gene-Regulation. Annual Review of Biochemistry 55: 569-597.
6. Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in Trans. The Plant cell 2: 279-289.
7. Romano N, Macino G (1992) Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol 6: 3343-3353.
8. Vanblokland R, Vandergeest N, Mol JNM, Kooter JM (1994) Transgene-Mediated Suppression of Chalcone Synthase Expression in Petunia-Hybrida Results from an Increase in Rna Turnover. Plant Journal 6: 861-877.
9. Montgomery MK, Fire A (1998) Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. Trends in genetics : TIG 14: 255-258.
10. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391: 806-811.
11. Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286: 950-952.
12. Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101: 25-33.
13. Hammond SM, Bernstein E, Beach D, Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404: 293-296.
14. Elbashir SM, Lendeckel W, Tuschl T (2001) RNA interference is mediated by 21-and 22-nucleotide RNAs. Genes & Development 15: 188-200.
15. Lin H, Spradling AC (1997) A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development 124: 2463-2476.
16. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409: 363-366.
17. Martinez J, Patkaniowska A, Urlaub H, Luhrmann R, Tuschl T (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110: 563-574.
18. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science 305: 1437-1441.
19. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell 123: 607-620.
20. Rand TA, Petersen S, Du F, Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell 123: 621-629.
21. Apollonius, Seaton RC (1912) Apollonius Rhodius, the Argonautica. London,
New York,: W. Heinemann; The Macmillan co. xiv p., 1 l., 431, 431 p. p.

22. Roper CFE, Sweeney MJ, Nauen CE (1984) Cephalopods of the world : an annotated and illustrated catalogue of species of interest to fisheries. Rome: United Nations Development Programme : Food and Agriculture Organization of the United Nations. viii, 277 p. p.
23. Aristotle, Ross WD, Smith JA (1910) The works of Aristotle translated into English under the editorship of W. D. Ross. Oxford,: Clarendon press.
24. Finn JK, Norman MD (2010) The argonaut shell: gas-mediated buoyancy control in a pelagic octopus. Proc Biol Sci 277: 2967-2971.
25. Hock J, Meister G (2008) The Argonaute protein family. Genome biology 9: 210.
26. Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, Esteban L, Serra E, Rovira C, Robello C, Cayota A (2010) Cloning, characterization and subcellular localization of a Trypanosoma cruzi argonaute protein defining a new subfamily distinctive of trypanosomatids. Gene 466: 26-35.
27. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, Wang Y, Patel DJ, Berenguer J, Brouns SJ, van der Oost J (2014) DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507: 258-261.
28. Elkayam E, Kuhn CD, Tocilj A, Haase AD, Greene EM, Hannon GJ, Joshua-Tor L (2012) The structure of human argonaute-2 in complex with miR-20a. Cell 150: 100-110.
29. Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev 12: 3715-3727.
30. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305: 1434-1437.
31. Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, 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.
32. Faehnle CR, Joshua-Tor L (2010) Argonaute MID domain takes centre stage. EMBO Rep 11: 564-565.
33. Kwak PB, Tomari Y (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol 19: 145-151.
34. Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature reviews Genetics 14: 100-112.
35. Dueck A, Meister G (2014) Assembly and function of small RNA - argonaute protein complexes. Biological chemistry 395: 611-629.
36. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev 22: 2773-2785.
37. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC (2008) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803-806.
38. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H (2008) Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature 453: 793-797.
39. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J (2008) An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798-802.
40. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science 320: 1077-1081.
41. Chung WJ, Okamura K, Martin R, Lai EC (2008) Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr Biol 18: 795-802.
42. Okamura K, Balla S, Martin R, Liu N, Lai EC (2008) Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol 15: 998.
43. Knight SW, Bass BL (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science 293: 2269-2271.
44. Provost P, Dishart D, Doucet J, Frendewey D, Samuelsson B, Radmark O (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. Embo Journal 21: 5864-5874.

45. Zhang HD, Kolb FA, Brondani V, Billy E, Filipowicz W (2002) Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. Embo Journal 21: 5875-5885.
46. Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science 315: 241-244.
47. Gredell JA, Dittmer MJ, Wu M, Chan C, Walton SP (2010) Recognition of siRNA asymmetry by TAR RNA binding protein. Biochemistry 49: 3148-3155.
48. Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nature reviews Genetics 10: 94-108.
49. Tomari Y, Du T, Zamore PD (2007) Sorting of Drosophila small silencing RNAs. Cell 130: 299-308.
50. Tahbaz N, Kolb FA, Zhang H, Jaronczyk K, Filipowicz W, Hobman TC (2004) Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. EMBO Rep 5: 189-194.
51. Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M (2010) In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. Molecular cell 39: 282-291.
52. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Molecular cell 39: 292-299.
53. Miyoshi T, Takeuchi A, Siomi H, Siomi MC (2010) A direct role for Hsp90 in pre-RISC formation in Drosophila. Nat Struct Mol Biol 17: 1024-1026.
54. Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, Zhou X, Qi Y (2012) Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes. Molecular cell 46: 859-870.
55. Colmenares SU, Buker SM, Buhler M, Dlakic M, Moazed D (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. Molecular cell 27: 449-461.
56. Sijen T, Steiner FA, Thijssen KL, Plasterk RH (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science 315: 244-247.
57. Dunoyer P, Himber C, Ruiz-Ferrer V, Alioua A, Voinnet O (2007) Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. Nat Genet 39: 848-856.
58. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. Cell 127: 565-577.
59. Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. The Plant cell 25: 2383-2399.
60. Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. EMBO J 21: 4663-4670.
61. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23: 4051-4060.
62. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. Nature 425: 415-419.
63. Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev 17: 3011-3016.
64. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. Science 303: 95-98.
65. Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. Nature reviews Genetics 11: 597-610.
66. Kurihara Y, Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A 101: 12753-12758.
67. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A 102: 3691-3696.
68. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. Cell 115: 199-208.

69. Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. Cell 115: 209-216.
70. Kawamata T, Seitz H, Tomari Y (2009) Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. Nat Struct Mol Biol 16: 953-960.
71. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature 465: 584-589.
72. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND, Wolfe SA, Giraldez AJ (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science 328: 1694-1698.
73. Ruby JG, Jan CH, Bartel DP (2007) Intronic microRNA precursors that bypass Drosha processing. Nature 448: 83-86.
74. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 130: 89-100.
75. Hutvagner G, Zamore PD (2002) A microRNA in a multiple-turnover RNAi enzyme complex. Science 297: 2056-2060.
76. Yekta S, Shih IH, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. Science 304: 594-596.
77. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136: 215-233.
78. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature reviews Genetics 9: 102-114.
79. Fabian MR, Sonenberg N (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol 19: 586-593.
80. Pfaff J, Meister G (2013) Argonaute and GW182 proteins: an effective alliance in gene silencing. Biochem Soc Trans 41: 855-860.
81. Eulalio A, Huntzinger E, Izaurralde E (2008) Getting to the root of miRNA-mediated gene silencing. Cell 132: 9-14.
82. Nottrott S, Simard MJ, Richter JD (2006) Human let-7a miRNA blocks protein production on actively translating polyribosomes. Nat Struct Mol Biol 13: 1108-1114.
83. Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z (2007) An mRNA m7G cap binding-like motif within human Ago2 represses translation. Cell 129: 1141-1151.
84. Chendrimada TP, Finn KJ, Ji X, Baillat D, Gregory RI, Liebhaber SA, Pasquinelli AE, Shiekhattar R (2007) MicroRNA silencing through RISC recruitment of eIF6. Nature 447: 823-828.
85. Ding L, Han M (2007) GW182 family proteins are crucial for microRNA-mediated gene silencing. Trends Cell Biol 17: 411-416.
86. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T (2006) A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442: 203-207.
87. Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3' end formation of PIWI-interacting RNAs in vitro. Molecular cell 43: 1015-1022.
88. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313: 320-324.
89. Fukuhara S, Nishimasu H, Bonnefond L, Matsumoto N, Ishitani R, Nureki O (2012) Expression, purification, crystallization and preliminary X-ray crystallographic analysis of Zucchini from Drosophila melanogaster. Acta Crystallogr Sect F Struct Biol Cryst Commun 68: 1346-1350.
90. Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, Ishitani R, Siomi H, Siomi MC, Nureki O (2012) Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature 491: 284-287.
91. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128: 1089-1103.

92. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science 315: 1587-1590.
93. Luteijn MJ, Ketting RF (2013) PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nature reviews Genetics 14: 523-534.
94. Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, Zamore PD (2010) Target RNA-directed trimming and tailing of small silencing RNAs. Science 328: 1534-1539.
95. Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. Cell 136: 656-668.
96. Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology 12: 246-258.
97. Aravin AA, Bourc'his D (2008) Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes Dev 22: 970-975.
98. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 22: 908-917.
99. Darricarrere N, Liu N, Watanabe T, Lin H (2013) Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proc Natl Acad Sci U S A 110: 1297-1302.
100. Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H (2013) A major epigenetic programming mechanism guided by piRNAs. Developmental cell 24: 502-516.
101. Rozhkov NV, Hammell M, Hannon GJ (2013) Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev 27: 400-412.
102. Sienski G, Donertas D, Brennecke J (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151: 964-980.
103. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Toth KF (2013) Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev 27: 390-399.
104. Wang SH, Elgin SC (2011) Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci U S A 108: 21164-21169.
105. Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC (2006) Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell 127: 747-757.
106. Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. EMBO J 26: 5007-5019.
107. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell 99: 123-132.
108. Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, Conte D, Jr., Mello CC (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell 139: 123-134.
109. Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen CC, Chaves DA, Duan S, Kasschau KD, Fahlgren N, Yates JR, 3rd, Mitani S, Carrington JC, Mello CC (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Molecular cell 36: 231-244.
110. Seth M, Shirayama M, Gu W, Ishidate T, Conte D, Jr., Mello CC (2013) The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Developmental cell 27: 656-663.
111. Wedeles CJ, Wu MZ, Claycomb JM (2013) Protection of germline gene expression by the C. elegans Argonaute CSR-1. Developmental cell 27: 664-671.

112. Djikeng A, Shi H, Tschudi C, Ullu E (2001) RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA 7: 1522-1530.
113. Tschudi C, Shi H, Franklin JB, Ullu E (2012) Small interfering RNA-producing loci in the ancient parasitic eukaryote Trypanosoma brucei. BMC genomics 13: 427.
114. Patrick KL, Shi H, Kolev NG, Ersfeld K, Tschudi C, Ullu E (2009) Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote Trypanosoma brucei. Proc Natl Acad Sci U S A 106: 17933-17938.
115. Shi H, Tschudi C, Ullu E (2006) An unusual Dicer-like1 protein fuels the RNA interference pathway in Trypanosoma brucei. RNA 12: 2063-2072.
116. Barnes RL, Shi H, Kolev NG, Tschudi C, Ullu E (2012) Comparative genomics reveals two novel RNAi factors in Trypanosoma brucei and provides insight into the core machinery. PLoS pathogens 8: e1002678.
117. Shi H, Chamond N, Djikeng A, Tschudi C, Ullu E (2009) RNA interference in Trypanosoma brucei: role of the n-terminal RGG domain and the polyribosome association of argonaute. J Biol Chem 284: 36511-36520.
118. Katiyar-Agarwal S, Jin H (2010) Role of small RNAs in host-microbe interactions. Annu Rev Phytopathol 48: 225-246.
119. Baulcombe D (2004) RNA silencing in plants. Nature 431: 356-363.
120. Schramke V, Allshire R (2004) Those interfering little RNAs! Silencing and eliminating chromatin. Current opinion in genetics & development 14: 174-180.
121. Ding SW (2010) RNA-based antiviral immunity. Nat Rev Immunol 10: 632-644.
122. Sabin LR, Hanna SL, Cherry S (2010) Innate antiviral immunity in Drosophila. Curr Opin Immunol 22: 4-9.
123. Qu F (2010) Antiviral role of plant-encoded RNA-dependent RNA polymerases revisited with deep sequencing of small interfering RNAs of virus origin. Mol Plant Microbe Interact 23: 1248-1252.
124. Bronkhorst AW, van Rij RP (2014) The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol 7: 19-28.
125. Ding SW, Voinnet O (2014) Antiviral RNA silencing in mammals: no news is not good news. Cell Rep 9: 795-797.
126. Bronkhorst AW, van Cleef KW, Vodovar N, Ince IA, Blanc H, Vlak JM, Saleh MC, van Rij RP (2012) The DNA virus Invertebrate iridescent virus 6 is a target of the Drosophila RNAi machinery. Proc Natl Acad Sci U S A 109: E3604-3613.
127. Kemp C, Mueller S, Goto A, Barbier V, Paro S, Bonnay F, Dostert C, Troxler L, Hetru C, Meignin C, Pfeffer S, Hoffmann JA, Imler JL (2013) Broad RNA interference-mediated antiviral immunity and virus-specific inducible responses in Drosophila. J Immunol 190: 650-658.
128. Bronkhorst AW, Miesen P, van Rij RP (2013) Small RNAs tackle large viruses RNA interference-based antiviral defense against DNA viruses in insects. Fly 7: 216-223.
129. Vodovar N, Goic B, Blanc H, Saleh MC (2011) In Silico Reconstruction of Viral Genomes from Small RNAs Improves Virus-Derived Small Interfering RNA Profiling. Journal of Virology 85: 11016-11021.
130. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW (2010) Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. Proc Natl Acad Sci U S A 107: 1606-1611.
131. Kopek BG, Perkins G, Miller DJ, Ellisman MH, Ahlquist P (2007) Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. PLoS Biol 5: e220.
132. Banerjee AK (1987) Transcription and replication of rhabdoviruses. Microbiol Rev 51: 66-87.
133. Ho T, Pallett D, Rusholme R, Dalmay T, Wang H (2006) A simplified method for cloning of short interfering RNAs from Brassica juncea infected with Turnip mosaic potyvirus and Turnip crinkle carmovirus. J Virol Methods 136: 217-223.
134. Yoo BC, Kragler F, Varkonyi-Gasic E, Haywood V, Archer-Evans S, Lee YM, Lough TJ, Lucas WJ (2004) A systemic small RNA signaling system in plants. The Plant cell 16: 1979-2000.
135. Molnar A, Csorba T, Lakatos L, Varallyay E, Lacomme C, Burgyan J (2005) Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. J Virol 79: 7812-7818.

136. Zambon RA, Vakharia VN, Wu LP (2006) RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster. Cellular Microbiology 8: 880-889.
137. Li H, Li WX, Ding SW (2002) Induction and suppression of RNA silencing by an animal virus. Science 296: 1319-1321.
138. Mueller S, Gausson V, Vodovar N, Deddouche S, Troxler L, Perot J, Pfeffer S, Hoffmann JA, Saleh MC, Imler JL (2010) RNAi-mediated immunity provides strong protection against the negative-strand RNA vesicular stomatitis virus in Drosophila. Proc Natl Acad Sci U S A 107: 19390-19395.
139. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, Andino R (2006) The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev 20: 2985-2995.
140. Wang H, Buckley KJ, Yang XJ, Buchmann RC, Bisaro DM (2005) Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. Journal of Virology 79: 7410-7418.
141. Morel JB, Godon C, Mourrain P, Beclin C, Boutet S, Feuerbach F, Proux F, Vaucheret H (2002) Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. The Plant cell 14: 629-639.
142. Lu R, Folimonov A, Shintaku M, Li WX, Falk BW, Dawson WO, Ding SW (2004) Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. Proc Natl Acad Sci U S A 101: 15742-15747.
143. Schott DH, Cureton DK, Whelan SP, Hunter CP (2005) An antiviral role for the RNA interference machinery in Caenorhabditis elegans. Proc Natl Acad Sci U S A 102: 18420-18424.
144. Wilkins C, Dishongh R, Moore SC, Whitt MA, Chow M, Machaca K (2005) RNA interference is an antiviral defence mechanism in Caenorhabditis elegans. Nature 436: 1044-1047.
145. Voinnet O (2005) Non-cell autonomous RNA silencing. FEBS Lett 579: 5858-5871.
146. Grishok A (2005) RNAi mechanisms in Caenorhabditis elegans. FEBS Lett 579: 5932-5939.
147. Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F, Jr., Hohn T, Pooggin MM (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233-6246.
148. Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, Mudge J, Wilusz J, Olson KE, Blair CD, Ebel GD (2010) C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. PLoS Negl Trop Dis 4: e856.
149. van Mierlo JT, van Cleef KW, van Rij RP (2010) Small Silencing RNAs: Piecing Together a Viral Genome. Cell Host Microbe 7: 87-89.
150. Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, Olson KE, Blair CD (2010) Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells. PLoS Negl Trop Dis 4: e848.
151. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, Monighetti C, Campbell CL (2011) Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol 11: 45.
152. Keene KM, Foy BD, Sanchez-Vargas I, Beaty BJ, Blair CD, Olson KE (2004) RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of Anopheles gambiae. Proc Natl Acad Sci U S A 101: 17240-17245.
153. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, Fazakerley JK, Kohl A, Fragkoudis R (2013) Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol 94: 1680-1689.
154. Burgyan J, Havelda Z (2011) Viral suppressors of RNA silencing. Trends Plant Sci 16: 265-272.
155. Shimura H, Pantaleo V (2011) Viral induction and suppression of RNA silencing in plants. Biochim Biophys Acta 1809: 601-612.
156. Wu Q, Wang X, Ding SW (2010) Viral suppressors of RNA-based viral immunity: host targets. Cell Host Microbe 8: 12-15.

157. Schnettler E, Sterken MG, Leung JY, Metz SW, Geertsema C, Goldbach RW, Vlak JM, Kohl A, Khromykh AA, Pijlman GP (2012) Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells. J Virol 86: 13486-13500.
158. Mochizuki K, Gorovsky MA (2005) A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes Dev 19: 77-89.
159. Sandoval PY, Swart EC, Arambasic M, Nowacki M (2014) Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting. Developmental cell 28: 174-188.
160. Shi H, Barnes RL, Carriero N, Atayde VD, Tschudi C, Ullu E (2014) Role of the Trypanosoma brucei HEN1 family methyltransferase in small interfering RNA modification. Eukaryotic Cell 13: 77-86.
161. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307: 932-935.
162. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD (2007) The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol 17: 1265-1272.
163. Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5 '-end-specific recognition of guide RNA by the A-fulgidus Piwi protein. Nature 434: 666-670.
164. Yuan YR, Pei Y, Ma JB, Kuryavyi V, Zhadina M, Meister G, Chen HY, Dauter Z, Tuschl T, Patel DJ (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Molecular cell 19: 405-419.
165. Parker JS, Roe SM, Barford D (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature 434: 663-666.
166. Rashid UJ, Paterok D, Koglin A, Gohlke H, Piehler J, Chen JC (2007) Structure of Aquifex aeolicus argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. The Journal of biological chemistry 282: 13824-13832.
167. Wang YL, Juranek S, Li HT, Sheng G, Tuschl T, Patel DJ (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature 456: 921-U972.
168. Wang YL, Sheng G, Juranek S, Tuschl T, Patel DJ (2008) Structure of the guide-strand-containing argonaute silencing complex. Nature 456: 209-U234.
169. Wang YL, Juranek S, Li HT, Sheng G, Wardle GS, Tuschl T, Patel DJ (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature 461: 754-U753.
170. Shabalina SA, Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol 23: 578-587.
171. Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nature reviews Molecular cell biology 9: 22-32.
172. Jinek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. Nature 457: 405-412.
173. Ketting RF (2011) The many faces of RNAi. Developmental cell 20: 148-161.
174. Meister G (2013) Argonaute proteins: functional insights and emerging roles. Nature reviews Genetics 14: 447-459.
175. Kunne T, Swarts DC, Brouns SJ (2014) Planting the seed: target recognition of short guide RNAs. Trends in microbiology 22: 74-83.
176. Sheng G, Zhao H, Wang J, Rao Y, Tian W, Swarts DC, van der Oost J, Patel DJ, 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.
177. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA (2013) Bacterial argonaute samples the transcriptome to identify foreign DNA. Molecular cell 51: 594-605.
178. Schirle NT, MacRae IJ (2012) The crystal structure of human Argonaute2. Science 336: 1037-1040.
179. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ (2012) Structure of yeast Argonaute with guide RNA. Nature 486: 368-374

180. Parker JS, Roe SM, Barford D (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J 23: 4727-4737.
181. Boland A, Tritschler F, Heimstadt S, Izaurralde E, Weichenrieder O (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep 11: 522-527.
182. Frank F, Sonenberg N, Nagar B (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature 465: 818-822.
183. Parker JS (2010) How to slice: snapshots of Argonaute in action. Silence 1: 3.
184. Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature 426: 465-469.
185. Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou MM (2003) Structure and conserved RNA binding of the PAZ domain. Nature 426: 468-474.
186. Ma JB, Ye K, Patel DJ (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature 429: 318-322.
187. Lingel A, Simon B, Izaurralde E, Sattler M (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol 11: 576-577.
188. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Molecular cell 33: 204-214.
189. Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, Chowdhury D, Dykxhoorn DM, Tsai P, Hofmann O, Becker KG, Gorospe M, Hide W, Lieberman J (2009) miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. Molecular cell 35: 610-625.
190. Zander A, Holzmeister P, Klose D, Tinnefeld P, Grohmann D (2014) Single-molecule FRET supports the two-state model of Argonaute action. RNA biology 11: 45-56.
191. Nowotny M, Yang W (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. EMBO J 25: 1924-1933.
192. Nowotny M (2009) Retroviral integrase superfamily: the structural perspective. EMBO Rep 10: 144-151.
193. Faehnle CR, Elkayam E, Haase AD, Hannon GJ, Joshua-Tor L (2013) The making of a slicer: activation of human Argonaute-1. Cell Rep 3: 1901-1909.
194. Nakanishi K, Ascano M, Gogakos T, Ishibe-Murakami S, Serganov AA, Briskin D, Morozov P, Tuschl T, Patel DJ (2013) Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. Cell Rep 3: 1893-1900.
195. Kuhn CD, Joshua-Tor L (2013) Eukaryotic Argonautes come into focus. Trends in biochemical sciences 38: 263-271.
196. Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O (2011) Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A 108: 10466-10471.
197. Huntzinger E, Kuzuoglu-Ozturk D, Braun JE, Eulalio A, Wohlbold L, Izaurralde E (2013) The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res 41: 978-994.
198. Pfaff J, Hennig J, Herzog F, Aebersold R, Sattler M, Niessing D, Meister G (2013) Structural features of Argonaute-GW182 protein interactions. Proc Natl Acad Sci U S A 110: E3770-3779.
199. Frank F, Hauver J, Sonenberg N, Nagar B (2012) Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. EMBO J 31: 3588-3595.
200. Hur JK, Zinchenko MK, Djuranovic S, 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.
201. Hauptmann J, Dueck A, Harlander S, Pfaff J, Merkl R, Meister G (2013) Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nat Struct Mol Biol 20: 814-817.
202. Yoda M, Kawamata T, Paroo Z, Ye X, Iwasaki S, Liu Q, Tomari Y (2010) ATP-dependent human RISC assembly pathways. Nat Struct Mol Biol 17: 17-23.

203. Burroughs AM, Iyer LM, Aravind L (2013) Two novel PIWI families: roles in inter-genomic conflicts in bacteria and Mediator-dependent modulation of transcription in eukaryotes. Biology direct 8: 13.
204. Makarova KS, Wolf YI, Koonin EV (2013) Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41: 4360-4377.
205. Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5: e9490.
206. Makarova KS, Aravind L, Wolf YI, Koonin EV (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biology direct 6: 38.
207. Iyer LM, Makarova KS, Koonin EV, Aravind L (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 32: 5260-5279.
208. Kinch LN, Ginalski K, Rychlewski L, Grishin NV (2005) Identification of novel restriction endonuclease-like fold families among hypothetical proteins. Nucleic Acids Res 33: 3598-3605.
209. Knizewski L, Kinch LN, Grishin NV, Rychlewski L, Ginalski K (2007) Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. BMC Struct Biol 7: 40.
210. Zhang J, Kasciukovic T, White MF (2012) The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. PLoS One 7: e47232.
211. Lemak S, Beloglazova N, Nocek B, Skarina T, Flick R, Brown G, Popovic A, Joachimiak A, Savchenko A, Yakunin AF (2013) Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from Sulfolobus solfataricus. J Am Chem Soc 135: 17476-17487.
212. Grazulis S, Manakova E, Roessle M, Bochtler M, Tamulaitiene G, Huber R, Siksnys V (2005) Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. Proc Natl Acad Sci U S A 102: 15797-15802.
213. Geserick P, Kaiser F, Klemm U, Kaufmann SH, Zerrahn J (2004) Modulation of T cell development and activation by novel members of the Schlafen (slfn) gene family harbouring an RNA helicase-like motif. Int Immunol 16: 1535-1548.
214. Aravind L, Koonin EV (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res 27: 4658-4670.
215. Rana RR, Zhang M, Spear AM, Atkins HS, Byrne B (2013) Bacterial TIR-containing proteins and host innate immune system evasion. Med Microbiol Immunol 202: 1-10.
216. Brikos C, O'Neill LA (2008) Signalling of toll-like receptors. Handb Exp Pharmacol: 21-50.
217. Palsson-McDermott EM, O'Neill LA (2007) Building an immune system from nine domains. Biochem Soc Trans 35: 1437-1444.
218. Burch-Smith TM, Dinesh-Kumar SP (2007) The functions of plant TIR domains. Sci STKE 2007: pe46.
219. Boubakri H, de Septenville AL, Viguera E, Michel B (2010) The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. EMBO J 29: 145-157.
220. Conaway RC, Sato S, Tomomori-Sato C, Yao T, Conaway JW (2005) The mammalian Mediator complex and its role in transcriptional regulation. Trends in biochemical sciences 30: 250-255.
221. Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, Kennedy S (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature 465: 1097-1101.
222. Ender C, Meister G (2010) Argonaute proteins at a glance. J Cell Sci 123: 1819-1823.
223. Li Y, Lu J, Han Y, Fan X, Ding SW (2013) RNA interference functions as an antiviral immunity mechanism in mammals. Science 342: 231-234.
224. Maillard PV, Ciaudo C, Marchais A, Li Y, Jay F, Ding SW, Voinnet O (2013) Antiviral RNA interference in mammalian cells. Science 342: 235-238.
225. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD (2012) Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell 151: 1055-1067.

226. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell 133: 116-127.
227. Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, Sehr P, Pillai RS, Carlomagno T (2011) Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. Structure 19: 172-180.
228. Tian Y, Simanshu DK, Ma JB, Patel DJ (2011) Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. Proc Natl Acad Sci U S A 108: 903-910.
229. van Wolfswinkel JC, Claycomb JM, Batista PJ, Mello CC, Berezikov E, Ketting RF (2009) CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell 139: 135-148.
230. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J (2014) The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol 21: 743-753.
231. Ketting RF (2011) microRNA Biogenesis and Function : An overview. Advances in experimental medicine and biology 700: 1-14.
232. Joshua-Tor L, Hannon GJ (2011) Ancestral roles of small RNAs: an Ago-centric perspective. Cold Spring Harbor perspectives in biology 3: a003772.
233. Koyama Y, Hoshino T, Tomizuka N, Furukawa K (1986) Genetic-Transformation of the Extreme Thermophile Thermus-Thermophilus and of Other Thermus Spp. Journal of Bacteriology 166: 338-340.
234. Averhoff B (2009) Shuffling genes around in hot environments: the unique DNA transporter of Thermus thermophilus. FEMS Microbiol Rev 33: 611-626.
235. Gregory ST, Dahlberg AE (2008) Transposition of an insertion sequence, ISTth7, in the genome of the extreme thermophile Thermus thermophilus HB8. FEMS Microbiol Lett 289: 187-192.
236. Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature 434: 666-670.
237. Collin RG, Morgan HW, Musgrave DR, Daniel RM (1988) Distribution of Reverse Gyrase in Representative Species of Eubacteria and Archaebacteria. Fems Microbiology Letters 55: 235-239.
238. Charbonnier F, Forterre P (1994) Comparison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaebacteria. J Bacteriol 176: 1251-1259.
239. Duguet M (1993) The helical repeat of DNA at high temperature. Nucleic Acids Res 21: 463-468.
240. Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJ, van der Oost J (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular cell 46: 595-605.
241. Bates AD, Maxwell A (2005) DNA topology. Oxford ; New York: Oxford University Press. xviii, 198 p. p.
242. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J 20: 6877-6888.
243. Travers A, Muskhelishvili G (2005) DNA supercoiling - A global transcriptional regulator for enterobacterial growth? Nature Reviews Microbiology 3: 157-169.
244. Rocha EPC, Danchin A (2002) Base composition bias might result from competition for metabolic resources. Trends in Genetics 18: 291-294.
245. Laptenko O, Kim SS, Lee J, Starodubtseva M, Cava F, Berenguer J, Kong XP, Borukhov S (2006) pH-dependent conformational switch activates the inhibitor of transcription elongation. Embo Journal 25: 2131-2141.
246. Cava F, Laptenko O, Borukhov S, Chahlafi Z, Blas-Galindo E, Gomez-Puertas P, Berenguer J (2007) Control of the respiratory metabolism of Thermus thermophilus by the nitrate respiration conjugative element NCE. Molecular Microbiology 64: 630-646.
247. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, Vanderpool CK, Tjaden B (2013) Computational analysis of bacterial RNA-Seq data. Nucleic Acids Research 41.

248. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
249. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D (2013) Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics 14: 193-202.
250. Uliczka F, Pisano F, Kochut A, Opitz W, Herbst K, Stolz T, Dersch P (2011) Monitoring of Gene Expression in Bacteria during Infections Using an Adaptable Set of Bioluminescent, Fluorescent and Colorigenic Fusion Vectors. PLoS One 6.
251. Peters L, Meister G (2007) Argonaute proteins: mediators of RNA silencing. Molecular cell 26: 611-623.
252. Kawamata T, Tomari Y (2010) Making RISC. Trends in biochemical sciences 35: 368-376.
253. Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12: 340-349.
254. Lingel A, Sattler M (2005) Novel modes of protein-RNA recognition in the RNAi pathway. Current opinion in structural biology 15: 107-115.
255. Martinez J, Tuschl T (2004) RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev 18: 975-980.
256. Schwarz DS, Tomari Y, Zamore PD (2004) The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol 14: 787-791.
257. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411: 494-498.
258. Ameres SL, Martinez J, Schroeder R (2007) Molecular basis for target RNA recognition and cleavage by human RISC. Cell 130: 101-112.
259. Yang W, Lee JY, Nowotny M (2006) Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Molecular cell 22: 5-13.
260. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell 121: 1005-1016.
261. Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A 90: 6498-6502.
262. Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. Journal of Applied Crystallography 40: 658-674.
263. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60: 2126-2132.
264. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography 66: 213-221.
265. Pratt AJ, MacRae IJ (2009) The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. Journal of Biological Chemistry 284: 17897-17901.
266. Yuan YR, Pei Y, Ma JB, Kuryavyi V, Zhadina M, Meister G, Chen HY, Dauter Z, Tuschl T, Patel DJ (2005) Crystal structure of A-aeolicus Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Molecular cell 19: 405-419.
267. Blesa A, Cesar CE, Averhoff B, Berenguer J (2015) Noncanonical Cell-to-Cell DNA Transfer in Thermus spp. Is Insensitive to Argonaute-Mediated Interference. J Bacteriol 197: 138-146.
268. Fiala G, Stetter KO (1986) Pyrococcus-Furiosus Sp-Nov Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100-Degrees C. Arch Microbiol 145: 56-61.
269. Farkas J, Stirrett K, Lipscomb GL, Nixon W, Scott RA, Adams MW, Westpheling J (2012) Recombinogenic properties of Pyrococcus furiosus strain COM1 enable rapid selection of targeted mutants. Appl Environ Microbiol 78: 4669-4676.
270. Lipscomb GL, Stirrett K, Schut GJ, Yang F, Jenney FE, Jr., Scott RA, Adams MW, Westpheling J (2011) Natural competence in the hyperthermophilic archaeon Pyrococcus furiosus facilitates genetic

manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol 77: 2232-2238.
271. Farkas J, Chung D, DeBarry M, Adams MW, Westpheling J (2011) Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon Pyrococcus furiosus needed for construction of a stable replicating shuttle vector. Appl Environ Microbiol 77: 6343-6349.
272. Waege I, Schmid G, Thumann S, Thomm M, Hausner W (2010) Shuttle vector-based transformation system for Pyrococcus furiosus. Appl Environ Microbiol 76: 3308-3313.
273. Willkomm S, Zander A, Gust A, Grohmann D (2015) A prokaryotic twist on argonaute function. Life (Basel) 5: 538-553.
274. Swarts DC, Zhu Y, Timmers L, Van der Oost J (2015) Acquisition of DNA guides by bacterial Argonaute. Manuscript in preparation.
275. Smalheiser NR, Gomes OL (2014) Mammalian Argonaute-DNA binding? Biology direct 10: 27.
276. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP (2010) Binding and cleavage of CRISPR RNA by Cas6. Rna-a Publication of the Rna Society 16: 2181-2188.
277. Swarts DC, Hegge JW, Hinojo I, Shiimori M, Ellis M, Dumrongkulraksa J, Terns RM, Terns MP, Van der Oost J (2015) Molecular characterization of Argonaute from the hyperthermophilic archaeon *Pyrococcus furiosus*. Nucleic Acids Res.
278. Swarts DC, Koehorst JJ, Westra ER, Schaap PJ, Van der Oost J (2015) Effect of Argonaute on gene expression in *Thermus thermophilus*. PLoS One, 10(4): e0124880.
279. Owczarzy R, You Y, Moreira BG, Manthey JA, Huang L, Behlke MA, Walder JA (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. Biochemistry 43: 3537-3554.
280. Record MT, Jr., Ha JH, Fisher MA (1991) Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. Methods in enzymology 208: 291-343.
281. Tadokoro T, Kanaya S (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. FEBS J 276: 1482-1493.
282. Cerritelli SM, Crouch RJ (2009) Ribonuclease H: the enzymes in eukaryotes. FEBS J 276: 1494-1505.
283. Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35: W43-46.
284. Westra ER, Buckling A, Fineran PC (2014) CRISPR-Cas systems: beyond adaptive immunity. Nat Rev Microbiol 12: 317-326.
285. Cava F, de Pedro MA, Blas-Galindo E, Waldo GS, Westblade LF, Berenguer J (2008) Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. Environmental Microbiology 10: 605-613.
286. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics 11: 119.
287. Ohyama H, Sakai T, Agari Y, Fukui K, Nakagawa N, Shinkai A, Masui R, Kuramitsu S (2014) The role of ribonucleases in regulating global mRNA levels in the model organism Thermus thermophilus HB8. BMC genomics 15: 386.
288. Bruggemann H, Chen C (2006) Comparative genomics of Thermus thermophilus: Plasticity of the megaplasmid and its contribution to a thermophilic lifestyle. J Biotechnol 124: 654-661.
289. Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, Shinkai A (2010) Transcription Profile of Thermus thermophilus CRISPR Systems after Phage Infection. Journal of Molecular Biology 395: 270-281.
290. Westra ER, Van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski JM, Best A, Bondy-Denomy J, Davidson A, Boots M, Buckling A (2015) Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. Current Biology In Press.
291. Best A, Hoyle A (2013) The evolution of costly acquired immune memory. Ecol Evol 3: 2223-2232.
292. Harrison E, Brockhurst MA (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends Microbiol 20: 262-267.

293. Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome biology 8: R61.
294. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. Nature reviews Microbiology 9: 467-477.
295. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, van der Oost J (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annual review of genetics 46: 311-339.
296. Heler R, Marraffini LA, Bikard D (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol Microbiol 93: 1-9.
297. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res 40: 5569-5576.
298. van der Oost J, Westra ER, Jackson RN, Wiedenheft B (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nature reviews Microbiology 12: 479-492.
299. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R (2012) Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. J Bacteriol 194: 2491-2500.
300. Dupuis ME, Villion M, Magadan AH, Moineau S (2013) CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. Nat Commun 4: 2087.
301. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644-652.
302. William S, Helene Feil A, Copeland A (2012) Bacterial genomic DNA isolation using CTAB. http://jgidoegov/collaborate-with-jgi/pmo-overview/protocols-sample-preparation-information/.
303. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709-1712.
304. Karginov FV, Hannon GJ (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. Molecular cell 37: 7-19.
305. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167-170.
306. Jore MM, Brouns SJ, van der Oost J (2011) RNA in Defense: CRISPRs Protect Prokaryotes against Mobile Genetic Elements. Cold Spring Harbor perspectives in biology.
307. Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJ (2010) Diversity of CRISPR loci in Escherichia coli. Microbiology 156: 1351-1361.
308. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321: 960-964.
309. Edgar R, Qimron U (2010) The Escherichia coli CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. Journal of bacteriology 192: 6291-6294.
310. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proceedings of the National Academy of Sciences of the United States of America 108: 10098-10103.
311. Touchon M, Rocha EP (2010) The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. PLoS One 5: e11126.
312. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Molecular microbiology 75: 1495-1512.
313. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders AP, Dickman MJ, Doudna JA, Boekema EJ, Heck AJ, van der Oost J, Brouns SJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature structural & molecular biology 18: 529-536.

314. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. The EMBO journal 30: 1335-1342.
315. Howard JA, Delmas S, Ivancic-Bace I, Bolt EL (2011) Helicase dissociation and annealing of RNA-DNA hybrids by Escherichia coli Cas3 protein. The Biochemical journal 439: 85-95.
316. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. Molecular microbiology 77: 1367-1379.
317. Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, Wurm R, Raine A, Mescher M, Van Heereveld L, Mastop M, Wagner EG, Schnetz K, Van Der Oost J, Wagner R, Brouns SJ (2010) H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Molecular microbiology 77: 1380-1393.
318. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R (2008) Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. Journal of bacteriology 190: 1401-1412.
319. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468: 67-71.
320. Nam KH, Kurinov I, Ke A (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. The Journal of biological chemistry 286: 30759-30768.
321. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. Trends in biochemical sciences 34: 401-407.
322. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155: 733-740.
323. Touchon M, Charpentier S, Clermont O, Rocha EP, Denamur E, Branger C (2011) CRISPR distribution within the Escherichia coli species is not suggestive of immunity-associated diversifying selection. Journal of bacteriology 193: 2460-2467.
324. Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of bacteriology 190: 1390-1400.
325. Dykhuizen DE, Hartl DL (1983) Selection in chemostats. Microbiol Rev 47: 150-168.
326. Hall KB, McLaughlin LW (1991) Thermodynamic and structural properties of pentamer DNA.DNA, RNA.RNA, and DNA.RNA duplexes of identical sequence. Biochemistry 30: 10606-10613.
327. Roberts RW, Crothers DM (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science 258: 1463-1466.
328. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? Trends in genetics : TIG 26: 335-340.
329. Marraffini LA, Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463: 568-571.
330. Al-Attar S, Westra ER, van der Oost J, Brouns SJ (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biological chemistry 392: 277-289.
331. van der Ploeg JR (2009) Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 155: 1966-1976.
332. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic acids research.
333. Swarts DC, Mosterd C, van Passel MW, Brouns SJ (2012) CRISPR interference directs strand specific spacer acquisition. PLoS One 7: e35888.

334. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 3: 945.
335. Fineran PC, Gerritzen MJ, Suarez-Diez M, Kunne T, Boekhorst J, van Hijum SA, Staals RH, Brouns SJ (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci U S A 111: E1629-1638.
336. Richter C, Dy RL, McKenzie RE, Watson BN, Taylor C, Chang JT, McNeil MB, Staals RH, Fineran PC (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res 42: 8516-8526.
337. Li M, Wang R, Zhao D, Xiang H (2014) Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res 42: 2483-2492.
338. Baba T, Huan HC, Datsenko K, Wanner BL, Mori H (2008) The applications of systematic in-frame, single-gene knockout mutant collection of Escherichia coli K-12. Methods in molecular biology 416: 183-194.
339. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97: 6640-6645.
340. Conrad SE, Wold M, Campbell JL (1979) Origin and direction of DNA replication of plasmid RSF1030. Proceedings of the National Academy of Sciences of the United States of America 76: 736-740.
341. Velmurugan S, Mehta S, Uzri D, Jayaram M (2003) Stable propagation of 'selfish' genetic elements. Journal of biosciences 28: 623-636.
342. Glick BR (1995) Metabolic load and heterologous gene expression. Biotechnol Adv 13: 247-261.
343. Ricci JCD, Hernandez ME (2000) Plasmid effects on Escherichia coli metabolism. Critical Reviews in Biotechnology 20: 79-108.
344. Wilson GG, Murray NE (1991) Restriction and modification systems. Annual review of genetics 25: 585-627.
345. Clokie MR, Millard AD, Letarov AV, Heaphy S (2011) Phages in nature. Bacteriophage 1: 31-45.
346. Pederson DM, Welsh LC, Marvin DA, Sampson M, Perham RN, Yu M, Slater MR (2001) The protein capsid of filamentous bacteriophage PH75 from Thermus thermophilus. J Mol Biol 309: 401-421.
347. Naryshkina T, Liu J, Florens L, Swanson SK, Pavlov AR, Pavlova NV, Inman R, Minakhin L, Kozyavkin SA, Washburn M, Mushegian A, Severinov K (2006) Thermus thermophilus bacteriophage phiYS40 genome and proteomic characterization of virions. J Mol Biol 364: 667-677.
348. Tamakoshi M, Murakami A, Sugisawa M, Tsuneizumi K, Takeda S, Saheki T, Izumi T, Akiba T, Mitsuoka K, Toh H, Yamashita A, Arisaka F, Hattori M, Oshima T, Yamagishi A (2011) Genomic and proteomic characterization of the large Myoviridae bacteriophage varphiTMA of the extreme thermophile Thermus thermophilus. Bacteriophage 1: 152-164.
349. Zha X, Xia Q, Yuan YA (2012) Structural insights into small RNA sorting and mRNA target binding by Arabidopsis Argonaute Mid domains. FEBS Lett 586: 3200-3207.
350. Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. PLoS Biol 3: e85.
351. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115: 787-798.
352. Chorn G, Zhao L, Sachs AB, Flanagan WM, Lim LP (2010) Persistence of seed-based activity following segmentation of a microRNA guide strand. RNA 16: 2336-2340.
353. Doench JG, Sharp PA (2004) Specificity of microRNA target selection in translational repression. Genes Dev 18: 504-511.
354. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell 27: 91-105.
355. Ui-Tei K, Naito Y, Nishi K, Juni A, Saigo K (2008) Thermodynamic stability and Watson-Crick base pairing in the seed duplex are major determinants of the efficiency of the siRNA-based off-target effect. Nucleic Acids Res 36: 7100-7109.

356. Schwarz DS, Ding HL, Kennington L, Moore JT, Schelter J, Burchard J, Linsley PS, Aronin N, Xu ZS, Zamore PD (2006) Designing siRNA that distinguish between genes that differ by a single nucleotide. PLoS genetics 2: 1307-1318.
357. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281-297.
358. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120: 15-20.
359. Dahlgren C, Zhang HY, Du Q, Grahn M, Norstedt G, Wahlestedt C, Liang Z (2008) Analysis of siRNA specificity on targets with double-nucleotide mismatches. Nucleic Acids Res 36: e53.
360. Haley B, Zamore PD (2004) Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol 11: 599-606.
361. Pusch O, Boden D, Silbermann R, Lee F, Tucker L, Ramratnam B (2003) Nucleotide sequence homology requirements of HIV-1-specific short hairpin RNA. Nucleic Acids Res 31: 6444-6449.
362. Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43: 1565-1575.
363. Viswanathan P, Murphy K, Julien B, Garza AG, Kroos L (2007) Regulation of dev, an operon that includes genes essential for Myxococcus xanthus development and CRISPR-associated genes and repeats. J Bacteriol 189: 3738-3750.
364. Lemak S, Nocek B, Beloglazova N, Skarina T, Flick R, Brown G, Joachimiak A, Savchenko A, Yakunin AF (2014) The CRISPR-associated Cas4 protein Pcal_0546 from Pyrobaculum calidifontis contains a [2Fe-2S] cluster: crystal structure and nuclease activity. Nucleic Acids Res 42: 11144-11155.
365. Peng X, Frohman MA (2012) Mammalian phospholipase D physiological and pathological roles. Acta Physiol (Oxf) 204: 219-226.
366. Nureki O (2014) Is zucchini a phosphodiesterase or a ribonuclease? Biomed J 37: 369-374.
367. Voigt F, Reuter M, Kasaruho A, Schulz EC, Pillai RS, Barabas O (2012) Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. RNA 18: 2128-2134.
368. Denu JM (2005) The Sir2 family of protein deacetylases. Current Opinion in Chemical Biology 9: 431-440.
369. Heitman J, Model P (1987) Site-specific methylases induce the SOS DNA repair response in Escherichia coli. J Bacteriol 169: 3243-3250.
370. Aertsen A, Michiels CW (2005) Mrr instigates the SOS response after high pressure stress in Escherichia coli. Mol Microbiol 58: 1381-1391.
371. Neumann B, Zhao L, Murphy K, Gonda TJ (2008) Subcellular localization of the Schlafen protein family. Biochem Biophys Res Commun 370: 62-66.
372. Mavrommatis E, Fish EN, Platanias LC (2013) The schlafen family of proteins and their regulation by interferons. J Interferon Cytokine Res 33: 206-210.
373. Ain QU, Chung JY, Kim Y (2014) Current and future delivery systems for engineered nucleases ZFN, TALEN and RGEN. J Control Release.
374. Kim H, Kim JS (2014) A guide to genome engineering with programmable nucleases. Nature reviews Genetics 15: 321-334.
375. Cox DB, Platt RJ, Zhang F (2015) Therapeutic genome editing: prospects and challenges. Nat Med 21: 121-131.
376. Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346: 1258096.
377. Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nature methods 10: 957-963.
378. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157: 1262-1278.
379. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816-821.

Acknowledgements

This thesis would not be completed without the direct and indirect help of many colleagues, collaborators, friends and family. I would like to thank all of you and some of you in particular.

Ik kan me nog goed herinneren dat **John** het voorstel voor het Argonaute project aan mij gaf en zei: 'Lees het en laat maar weten of je het wil doen'. Hoewel ik graag een PhD project buiten Wageningen wilde doen, ben ik heel erg blij dat ik ben gebleven. John, bedankt voor de inspirerende en erg leuke begeleiding. Hoewel we regelmatig op dead-ends leken te stuiten, wist jij me met je aanstekelijke enthousiasme altijd weer te motiveren om een uitweg te vinden. Je ongelimiteerde hoeveelheid ideeën hielpen altijd om grote stappen in de goede richting te zetten. Een betere promotor had ik me niet kunnen wensen!

Stan, je hebt met je inzicht en kennis voortdurend grote bijgedragen geleverd aan het project, in het begin toen je nog nauw bij het project betrokken was, tot aan het einde, waar je zijdelings alsnog vaak met goede ideeën en oplossingen kwam. Heel erg bedankt hiervoor! Het was erg leuk het kantoor met je te delen en bovendien zeer nuttig omdat ik altijd van je zeer uitgebreide kennis gebruik kon maken in de gevallen dat PubMed en zelfs Google me niet aan een antwoord konden helpen.

Matthijs, ik ben heel erg blij dat ik dit project samen met jou mocht beginnen. Hoewel we in het begin van het project op de meeste dead-ends terecht zijn gekomen (doordat we nog niet zoveel wisten van pAgos en van alles probeerden), vond ik dit misschien wel de leukste periode. Dit komt omdat we in deze fase de meest spannende ontdekkingen hebben gedaan en daarnaast omdat ik het erg gezellig vond om met je te werken. Ik heb extreem veel van je gestructureerde manier van werken en scherpe analyses geleerd. Ik hoop dat ik jou ook nog wat bij heb kunnen brengen, bijvoorbeeld de kunst van het smileys lezen ;-).

Ik vind het zeer leuk **Edze**, dat jij het Argonaute-team tijdelijk hebt versterkt! Je kritische blik en daadkracht zijn van grote waarde geweest voor het project en daarnaast vond het ik het altijd erg gezellig met je, binnen en buiten het lab. Behalve die keer dat je aan het mokken was dat je niet kon schaatsen omdat we een geweldige trip naar Vancouver in het

verschiet hadden. Deze trip was toch zeker een van de hoogtepunten van mijn PhD, daar ben je het achteraf vast met mij over eens! Bedankt voor alle leuke momenten!

Jorrit, ik vind het bijzonder leuk dat je via Research Methods in Microbiology en daarna je MSc thesis uiteindelijk bent geëindigd als PhD-student op het Argonaute project. Behalve dat ik er erg blij mee ben dat iemand door gaat met het project (waar nog veel interessante ontdekkingen in het verschiet liggen!), vond ik het altijd erg leuk om met je samen te werken en om samen een biertje te drinken. Heel veel succes met je project. We houden contact over de ontwikkelingen!

I am very grateful that so many talented BSc and MSc students contributed to the Argonaute project. Without the contribution of **Ruben**, **Bernard**, **Jorijn**, **Giel**, **Remco**, **Indra**, **Yifan**, **Jorrit**, **Lisa** and **Isma**, the current knowledge on prokaryotic Argonaute proteins would be very limited. Working with you truly enriched my PhD, and I hope you learned as much from me as I have learned from you. I am happy to see that many of you chose to continue your careers in (academic) science, and I am sure that our paths will cross again in the future.

I had the honour to work with various experts from different research fields. By combining the knowledge from these fields we together were able to investigate pAgos from different perspectives. It was a pleasure to work with all of you! **Kira** and **Eugene**, you did some very thorough and insightful bioinformatic analyses, which provided the first clues for which direction to take in this project. **Dinshaw** and **Yanli**, your advice in the early stages of the project was very valuable, and I am glad you could provide such valuable structural insights, for which I also have to thank **Kotaro**. **René**, je kennis over eukaryote Argonauten is erg waardevol geweest voor ons Argonaute review! **José**, but also **Aurelio** and **Carolina**, thanks for teaching me to work with *Thermus thermophilus* prior to my PhD, and for the useful tips during my PhD! **Bram** and **Helen**, the Mass Spectrometry experiments you performed generated really useful insights to determine the direction of our research. **Mark D.**, bedankt voor de hulp met mijn eigen RNA-seq analyse, en **Peter** en **Jasper**, bedankt voor de daarna uitgevoerde grootschalige bioinformatische analyse van RNA-seq data. **Cas** en **Mark vd P.**, het was een erg leuke samenwerking op het CRISPR-project. **Mike**, I really enjoyed the collaboration with you and your group. It was great that after all our interesting and fun discussions at conferences, we could work together on the *Pf*Ago project. The first series of

single-molecule experiments look very promising, **Chirlmin** and **Malwina**, I am sure this project will yield some nice results!

Een essentiële pilaar voor het succes van het Laboratorium voor Microbiologie is het uitzonderlijk getalenteerd ondersteunend personeel. **Anja** en **Carolien**, bedankt voor de talloze keren dat jullie me geholpen hebben met administratieve kwesties als declaraties, urenregistratie en vele andere zaken. **Wim**, jij bent de ware held van Microbiologie, zonder jou loopt alles waarschijnlijk in het honderd. Ook bedankt voor het ophalen en thuisbrengen iedere dag, toen ik mijn knieschijf had gebroken. Alle technicians bedankt voor vele bestellingen en het draaiend houden van het lab. **Sjon**, het was fijn dat je zoveel last-minute bestellingen voor me kon doen! **Philippe**, bedankt voor het bewaken van mijn veiligheid door me elke keer aan het dragen van een labjas te herinneren, maar vooral voor het constant verbeteren van de labcondities. **Willem**, je hebt een plezierige en zeer stimulerende werksfeer gecreëerd, daar kun je trots op zijn! Besides Stan, Matthijs, Edze and Jorrit, I shared my office with **Pawel**, **Slav**, **Peter**, **Marnix** and **Tim**. It was great that I could bother all of you with my work-related questions every now and then, but I also highly appreciated our non-work related discussions both in and outside our office. It was inspiring to be part of the **Host Defense-team**, which besides researchers investigating pAgos, consisted of researchers focussing on the CRISPR field (which is almost as cool as the Argonaute field!). I really appreciated that we could share all our ups and downs and were able to help each other with our research. I also have to thank all other members of the Bacterial Genetics group (**Servé**, **Mark L.**, **John R.**, **Vincent**, **Bram**, **Raymond**, **Tessa**, **Nico**, **Bas**, **Franklin**, **Teunke**, **Melvin**, **Sjoerd**, **Elleke**, **Tom**, **Tijn**, **Aleks**, **Ioannis**, **Emiel**, **Shreyans** and **Prarthana**), for that! In addition I have to thank them and all other members of **Laboratory of Microbiology** and **Systems and Synthetic Biology** (especially **Ruud**, **Jasper S.** and **Rob**) for the fun diners, drinks, barbeques, trips, lunches, coffee breaks and parties. I wish you all the best in your scientific careers and hope to see you again in the future!

Ik ben mijn vrienden heel erg dankbaar dat ze mijn gedachten zo nu en dan van mijn onderzoek af konden zetten. **Geert** & **Sandra**, **Jochem** & **Ileana**, **Stijn**, **Joris**, en **Dorien**, het is altijd als vanouds als we samen zijn. Ik ben blij dat we elkaar ondanks de afstand regelmatig zien, en hoop dat dit nog lang zo doorgaat! **Annemiek** en **Bart**, ik heb afgelopen jaren altijd bijzonder uitgekeken naar de BAFD-competitie avondjes. Naast het spel en het

heerlijke eten was het leuk om van alles en nog wat met jullie te bespreken. Ook waardeer ik de contacten met mijn **oud-huisgenoten**, het is altijd top met jullie! Met mijn jaarclub, **(AL)TIET ZAT**, heb ik goede weekendjes weg gehad. Al zie ik niet iedereen van jullie even vaak meer, is het altijd mooi om weer samen te zijn. **Koen** en **Kasper**, jullie zie ik vaak en daar ben ik erg blij mee! Ondanks mijn regelmatige afwezigheid bij activiteiten van Het Dispuut **De Vikingen**, was het altijd mooi om langs te komen. Zeker de weekenden waren altijd een goede mogelijkheid om even aan mijn PhD te ontsnappen, ODIN ODIN! Ik heb via Fieke ook veel van haar **vrienden en oud-huisgenoten** leren kennen, bedankt voor de etentjes, borrels en andere leuke activiteiten! Hockey bij de **WMHC** was de afgelopen vier jaar een van mijn belangrijkste ontspanningsmogelijkheden. Niet alleen op het veld, maar zeker ook daarnaast met mijn **(oud-)teamgenoten**, en met iedereen die ik via hen ken. Ik hoop met jullie in contact te blijven, dat gaat zeker lukken!

Willem, **Lidwien** en **Roel**, ik voel me altijd erg welkom bij jullie thuis. Erg bedankt voor alle leuke bezoekjes, tripjes en gezelligheid. **Ome Luc**, het was altijd erg leuk om samen ergens te gaan eten en het over reizen te hebben! **Willem**, en **Luuk** & **Judith**, het is altijd fijn om jullie te zien en het over van alles en nog wat te hebben. **Willem**, het is goed om te zien dat je je plekje hebt gevonden in Eindhoven en ik vind het erg interessant om over je baan te horen! **Luuk**, ik vind het heel leuk dat je door je studie de details van mijn onderzoek ook begrijpt, al zijn daardoor onze gesprekken niet altijd even interessant voor de rest van de familie. Hopelijk kunnen we het binnenkort ook over de details van jouw eerste baan hebben. **Pap en mam**, bedankt dat ik altijd kan 'thuiskomen' in Boekel. Het is altijd erg ontspannend om thuis te zijn, maar ook om erop uit te gaan met jullie! Jullie hebben altijd een onuitputbare interesse in de details van mijn PhD, nu kunnen jullie alles nog eens rustig nalezen ;-). Bedankt dat jullie het mogelijk hebben gemaakt me te ontwikkelen en scholen tot een onderzoeker!

Het laatste stukje is gereserveerd voor de belangrijkste persoon in mijn leven, **Fieke**. Bij jou kan ik me echt ontspannen. Bedankt voor het zorgen voor de nodige afleiding van mijn PhD, maar ook voor steun in de zwaardere periodes, waarin ik niet altijd even gezellig was. Ik vind het heel fijn om bij je te zijn en om samen nieuwe dingen te ontdekken. Ik kijk uit naar ons volgende avontuur, en hoop dat we er samen nog vele mogen beleven!

Daan

List of publications

DC Swarts, EM Timmers, Y Zhu, J van der Oost. Acquisition of DNA guides by prokaryotic Argonaute. *Manuscript in preparation*.

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

DC Swarts, JJ Koehorst, ER Westra, PJ Schaap, J van der Oost (2015) *Effects of Argonaute on gene expression in Thermus thermophilus*. **PLoS ONE**, 10(4): e0124880.

DC Swarts, K Makarova, Y Wang, K Nakanishi, RF Ketting, EV Koonin, DJ Patel, J van der Oost (2014) *The evolutionary journey of Argonaute proteins*. **Nature Structural & Molecular Biology** Vol. 21, No. 9, 743-753.

J van der Oost, **DC Swarts**, MM Jore (2014) *Prokaryotic Argonautes – variations on the RNA interference theme*. **Microbial Cell**, Vol. 1, Issue 5, 158-159.

T Künne, **DC Swarts**, SJJ Brouns (2014) *Planting the seed: Target recognition of short guide RNAs*. **Trends in Microbiology**, Vol. 22, Issue 2, 74-83.

DC Swarts*, MJ Jore*, ER Westra, Y Zhu, JH Janssen, AP Snijders, Y Wang, DJ Patel, J Berenguer, SJJ Brouns, J van der Oost (2014) *DNA-guided DNA interference by a prokaryotic Argonaute*. **Nature**, Vol. 507, Issue 7491, 258-261.

G Sheng*, H Zhao*, J Wang, Y Rao, W Tian, **DC Swarts**, J van der Oost, DJ Patel, Y Wang (2014) *Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage*. **PNAS**, Vol. 111, Issue 2, 652-657.

ER Westra, **DC Swarts**, RHJ Staals, MM Jore, SJJ Brouns, J van der Oost (2012) *The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity.* **Annual Reviews Genetics**, Vol. 46, 311-339.

DC Swarts, C Mosterd, MWJ van Passel, SJJ Brouns (2012) *CRISPR Interference Directs Strand Specific Spacer Acquisition*. **PLoS ONE**, 7(4): e35888.

CM Plugge, AM Henstra, P Worm, **DC Swarts**, AH Paultisch-Fuchs, JCM Scholten, A Lykidis, AL Lapidus, E Goltsman, E Kim, E McDonald, L Rohlin, BR Crable, RP Gunsalus, AJM Stams, MJ McInerney (2012) *Complete genome sequence of Syntrophobacter fumaroxidans strain* (MPOBT). **Standards in Genomic Sciences**, Vol. 7 Iss. 1 91-106.

F Blombach, H Launay, V Zorraquino, **DC Swarts**, LD Cabrita, D Benelli, J Christodoulou, P Londei , J van der Oost (2011) *An HflX-Type GTPase from Sulfolobus solfataricus Binds to the 50S Ribosomal Subunit in All Nucleotide-Bound States.* **Journal of Bacteriology**, Vol. 193 no. 11 2861-2867.

*These authors contributed equally

Peer-reviewed protocols

DC Swarts, ER Westra, SJJ Brouns, J van der Oost (2014) *Purification and Sequencing of DNA guides from Prokaryotic Argonaute*. **Bio-protocol**, Vol 4, Issue 22

DC Swarts, MM Jore, J van der Oost (2014) *Expression and Purification of the Thermus thermophilus Argonaute Protein*. **Bio-protocol**, Vol 4, Issue 19

Patent applications

J van der Oost, **DC Swarts**, AP May, RE Haurwitz. *DNA-guided DNA interference by a prokaryotic Argonaute.* PCT/US14/33885 (filed February 13, 2014).

J van der Oost, **DC Swarts**. *Prokaryotic Argonaute.* UK1206574.3 (filed April 11, 2013).

Co-author affiliations

Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands.
John van der Oost, Matthijs M. Jore[a], Edze R. Westra[b], Yifan Zhu, Jorijn H. Janssen[c], Stan J.J. Brouns, Jorrit W. Hegge, Ismael Hinojo, Elizabeth M. Timmers, Cas Mosterd[d].

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD20894, USA.
Kira S. Makarova, Eugene V. Koonin.

Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China.
Yanli Wang, Gang Sheng, Hongtu Zhao, Jiuyu Wang, Yu Rao, Wenwen Tian.

Department of Chemistry and Biochemistry, Ohio State University, Columbus, OH, 43210, USA.
Kotaro Nakanishi.

Institute for Molecular Biology, Ackermannweg 4, D-55128 Mainz, Germany.
René F. Ketting.

Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, 10065, USA.
Dinshaw J. Patel.

Cancer Research UK, Clare Hall Laboratories, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3LD, United Kingdom.
Ambrosius P.L. Snijders.

Centro de Biología Molecular Severo Ochoa. UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain.
José Berenguer.

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.

Systems and Synthetic Biology, Department of Agrotechnology and Food Sciences, Wageningen University Dreijenplein 10, 6703 HB Wageningen, The Netherlands.
Jasper J. Koehorst, Peter J. Schaap, Mark W.J. van Passel[e].

Current address:
[a]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom
[b]Biosciences, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9EZ, United Kingdom
[c]Bioprocess Engineering, Department of Agrotechnology and Food Sciences, Wageningen University, Droevenstaalsesteeg 1, 6708PB Wageningen, The Netherlands.
[d]Department de Biochimie, et de Microbiologie, Faculté des Sciences et de Génie,Université Laval, Québec, Canada
[e]National Institute for Public Health and the Environment, RIVM, 3720 BA Bilthoven, The Netherlands.

About the author

Daniël (Daan) Christianus Swarts was born on the 12th of September, 1986, in Boekel, The Netherlands. In 2004, he finished his pre-university education (VWO) at the Commanderij College in Gemert, and started the study Molecular Life Sciences at Wageningen University. He included a minor in Communication Sciences and a minor in Consumer Sciences in the Bachelor program. During the Master program of this study, Daan specialized in Biological Chemistry. His two Master theses were performed at the Laboratory of Microbiology of Wageningen University. Under supervision of Dr Caroline M. Plugge, he studied the possibilities for developing a genetic system for *Syntrophobacter fumaroxidans*. Under supervision of Dr Stan J.J. Brouns, he investigated the CRISPR-adaptation stage of *Escherichia coli*. In addition, Daan performed an internship at the CBMSO in Madrid under supervision of Prof. José Berenguer and Dr Aurelio Hidalgo. This internship focused on the cloning, expression and purification of thermostable CRISPR-Cas and Argonaute proteins. After obtaining his MSc degree in 2011, Daan started a PhD project at the Laboratory of Microbiology in the Bacterial Genetics department, under supervision of promotor Prof. Dr John van der Oost and co-promotor Dr Stan J.J. Brouns. His work on the characterization of prokaryotic Argonaute proteins (pAgo) is presented in this thesis.

Overview of completed training activities

Discipline specific activities

Meetings

- GBM Mosbacher Kolloquium – Mechanisms of RNA mediated regulation. Mosbach, Germany (2011).
- NWO-ALW Platform Molecular Genetics. Lunteren, The Netherlands (2011). *
- Keystone symposia – Gene silencing by small RNAs. Vancouver, Canada (2012). *, #
- Cold Spring Harbor symposia – Regulatory & Non-coding RNAs. Cold Spring Harbor, USA (2012). *, #
- NWO-ALW Platform Molecular Genetics Lunteren, The Netherlands (2012). #
- Regulating with RNA in bacteria. Würzburg, Germany (2013). #
- NWO-ALW Platform Molecular Genetics Lunteren, The Netherlands (2013). *
- ITQB RNA Meeting. Lisbon, Portugal (2014). #
- Spring School 'Host-Microbe interactomics'. Wageningen, The Netherlands (2014). #
- Antiviral defence mechanisms symposium. Wageningen, The Netherlands (2014). #, &

#Oral presentation, *Poster presentation, &Co-organiser

Course

- Safe handling of Radioactive Materials and Sources (level 5b)

General courses

- Techniques for Presenting and Writing Scientific Papers
- Project & Time management
- VLAG PhD Week
- PhD Competence Assessment
- Career perspectives
- Last stretch of the PhD Programme

Optionals

- Literature research
- PhD Study trip to USA and Canada
- Bacterial Genetics Group Meetings (weekly)
- Laboratory of Microbiology PhD Meetings (biweekly)
- Laboratory of Microbiology/Systems and Synthetic Biology Seminars

The research described in this thesis was financially supported by a TOP-grant from the Netherlands Organization of Scientific Research (NWO) to Prof. Dr John van der Oost (grant number 854.10.003).

Cover design by Daan C. Swarts
Front: *Tt*Ago with bound DNA guide and DNA target (PDB: 4NCA and PDB: 3BSE)
Back: Stereoview of *Tt*Ago with bound DNA guide and DNA target (PDB: 4NCA)

Thesis layout by Daan C. Swarts

Printed by Gildeprint Drukkerijen, Enschede