https://doi.org/10.15388/vu.thesis.589
https://orcid.org/0000-0003-3901-8868

VILNIUS UNIVERSITY

Edvardas Golovinas

Structural and Biochemical Studies of an Argonaute and its Associated Protein from *Archaeoglobus fulgidus*

DOCTORAL DISSERTATION

Natural sciences,
Biochemistry N 004

VILNIUS 2024

This dissertation was written between 2017 and 2024 at Vilnius University Life Sciences Centre, Institute of Biotechnology.

The dissertation is defended on an external basis.

Academic consultants:

Dr Elena Manakova (Vilnius University, Natural sciences, Biochemistry – N 004);

Dr Mindaugas Zaremba (Vilnius University, Natural sciences, Biochemistry – N 004).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr Rolandas Meškys (Vilnius University, Natural sciences, Biochemistry – N 004).

Members:

Prof. Dr Vilmantė Borutaitė (Lithuanian University of Health Sciences, Natural sciences, Biochemistry – N 004);

Dr Rūta Gerasimaitė (Max Planck Institute, Germany, Natural sciences, Biochemistry – N 004);

Dr Stephen Knox Jones, Jr. (Vilnius University, Natural sciences, Biochemistry – N 004);

Dr Daan Swarts (Wageningen University, The Netherlands, Natural sciences, Biochemistry – N 004).

The dissertation shall be defended in English at a public meeting of the Dissertation Defence Panel at 3:00 p.m. on 29th February 2024 in Room R105 of the Vilnius University Life Sciences Centre and remotely.

Address: Saulėtekio Ave. 7, R105, Vilnius, Lithuania

Tel. +37068165673; e-mail: edvardas.golovinas@gmc.vu.lt

The text of this dissertation can be accessed at the library of Vilnius University, as well as on the website of Vilnius University: www.vu.lt/lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.589
https://orcid.org/0000-0003-3901-8868

VILNIAUS UNIVERSITETAS

Edvardas Golovinas

Struktūriniai ir biocheminiai Argonaute ir su juo asocijuoto baltymo iš *Archaeoglobus fulgidus* tyrimai

DAKTARO DISERTACIJA

Gamtos mokslai,
Biochemija N 004

VILNIUS 2024

Disertacija rengta 2017 – 2024 metais Vilniaus universiteto Gyvybės mokslų centre, Biotechnologijos institute.
Disertacija ginama eksternu.

Moksliniai konsultantai:
dr. Elena Manakova (Vilniaus universitetas, gamtos mokslai, biochemija – N 004);
dr. Mindaugas Zaremba (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).
Gynimo taryba:
Pirmininkas – **prof. dr. Rolandas Meškys** (Vilnius universitetas, gamtos mokslai, biochemija – N 004).
Nariai:
prof. dr. Vilmantė Borutaitė (Lietuvos sveikatos mokslų universitetas, gamtos mokslai, biochemija – N 004);
dr. Rūta Gerasimaitė (Makso Planko institutas, Vokietija, gamtos mokslai, biochemija – N 004);
dr. Stephen Knox Jones, Jr. (Vilniaus universitetas, gamtos mokslai, biochemija – N 004);
dr. Daan Swarts (Vageningeno universitetas, Nyderlandai, gamtos mokslai, biochemija – N 004).
Disertacija ginama viešame Gynimo tarybos posėdyje anglų kalba 2024 m. vasario mėn. 29 d. 15.00 val. Vilniaus universiteto Gyvybės mokslų centro R105 auditorijoje ir nuotoliniu būdu. Adresas: Saulėtekio al. 7, R105, Vilnius, Lietuva, tel. +37068165673 ; el. paštas edvardas.golovinas@gmc.vu.lt

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: https://www.vu.lt/naujienos/ivykiu-kalendorius

CONTENTS

ABBREVIATIONS

aa	amino acid
AFM	atomic force microscopy
Ago	Argonaute
ALEX	alternating laser excitation
APAZ	analogue of PAZ
APD	avalanche photodiode
APS	1-(3-aminopropyl)-silatrane
AU	arbitrary units
bp	base pair
BSA	bovine serum albumin
Cryo-EM	cryogenic-sample electron microscopy
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
FPGA	field-programmable gate array
FRET	Förster resonance energy transfer
gDNA	guide DNA
gRNA	guide RNA
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMM	hidden Markov model
HPLC	high-performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
L-Ara	L-Arabinose
LB	lysogenic broth
LCR	ligase chain reaction
MID	“Middle” – a structural domain of Argonaute proteins
miRISC	microRNA-containing RISC
miRNA	microRNA
MS	mass spectrometry
nt	nucleotide
PAAG	polyacrylamide gel
PAGE	polyacrylamide gel electrophoresis
PAZ	PIWI-Argonaute-Zwille domain
PDB	Protein data bank
PIWI	P-element induced wimpy testis

piRNA	PIWI-interacting RNA
PMSF	phenylmethylsulphonyl fluoride
PNK	polynucleotide kinase
RISC	RNA-induced silencing complex
RITS	RNA-induced transcription silencing
RNase	ribonuclease
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulphate
SEC	size-exclusion chromatography
Sir2	silent mating-type information regulation 2
siDNA	small interfering DNA
siRNA	small interfering rna
MALS	multi-angle light scattering
ss	single-stranded
tDNA	target DNA
TIR	Toll/interleukin-1 receptor
TIR	total internal reflection
Tris	tris(hydroxymethyl)aminomethane
tRNA	target RNA
TTL	transistor-transistor logic
WT	wild-type

REPRODUCTION LICENCES

RNA Interference in Eukaryotes and Prokaryotes. *Biochem*. 83, 483–497 (2018) DOI: 10.1134/S0006297918050024 under licence No. 5581860231969.

Figure 4, Figure 6, and Figure 7 reproduced from Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The Expanded Universe of Prokaryotic Argonaute Proteins. *MBio* 9, 1–20 (2018) DOI: 10.1128/mBio.01935-18 under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0).

Figure 8 reproduced from Koopal, B., Mutte, S. K. & Swarts, D. C. A long look at short prokaryotic Argonautes. *Trends Cell Biol.* 1–14 (2022) DOI: 10.1016/j.tcb.2022.10.005 under licence No. 5592560901663.

Figure 9 reproduced from Lisitskaya, L., Aravin, A. A. & Kulbachinskiy, A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nat. Commun*. 9, 5165 (2018) DOI: 10.1038/s41467-018-07449-7 under Creative Commons Attribution 4.0 International Licence (CC BY 4.0).

Figure 10 reproduced from Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D. C., van der Oost, J., Patel, D. J. & Wang, Y. Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci.* 111, 652–657 (2014) DOI: 10.1073/pnas.1321032111 under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 11 reproduced from Swarts, D. C. Prokaryotic Argonautes Function beyond Immunity by Unlinking Replicating Chromosomes. *Cell* 182, 1381–1383 (2020) DOI: 10.1016/j.cell.2020.08.037 under licence No. 5621950366768.

Figure 12 reproduced from Wang, X., Li, X., Yu, G., Zhang, L., Zhang, C., Wang, Y., Liao, F., Wen, Y., Yin, H., Liu, X., Wei, Y., Li, Z., Deng, Z. & Zhang, H. Structural insights into mechanisms of Argonaute protein-associated NADase activation in bacterial immunity. *Cell Res.* (2023) DOI: 10.1038/s41422-023-00839-7 under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 13 reproduced from Zeng, Z., Chen, Y., Pinilla-Redondo, R., Shah, S. A., Zhao, F., Wang, C., Hu, Z., Wu, C., Zhang, C., Whitaker, R. J., She, Q.

& Han, W. A short prokaryotic Argonaute activates membrane effector to confer antiviral defense. *Cell Host Microbe* 30, 1–14 (2022) DOI: 10.1016/j.chom.2022.04.015 under licence No. 5631871064531

INTRODUCTION

Argonaute (Ago) proteins can be found in all three domains of life – bacteria, archaea, and eukaryotes. The first-discovered and best-characterized of these are the eukaryotic Argonaute proteins (eAgos)[1–3]. Being the core of the RNA interference (RNAi) machinery, they employ short RNA guides to target RNA for gene expression regulation and host defence[2]. The structural and mechanistic diversity of prokaryotic Agos (pAgos) is much greater, however: phylogenetic analysis showed that pAgos can be separated into three distinct clades – long-A, long-B, and short pAgos[4,5]. The long pAgos resemble eAgos in that they comprise the main structural domains – N-terminal, MID, PIWI, and PAZ, including two linker domains L1 and L2. Short pAgos, on the other hand, lack the N-terminal part and only bear the MID and PIWI domains, usually associating with another protein, containing an effector and APAZ domains [5]. The best-studied is the long-A clade, to which belong all characterized catalytically active pAgo nucleases, with studies on short pAgos gaining interest and momentum. Some long and short pAgos have been demonstrated to have defensive roles, protecting the host from invading phages and plasmids by invader degradation or abortive infection, or performing other functions, like chromosome decatenation[6–12]. The long-B clade, on the other hand, is severely lacking in attention and, therefore, in findings and information, with only one pAgo well-described[13–16]. One of the likely reasons for this is that all known long-B pAgos are catalytically inactive, having substitutions of key residues in the catalytic site of the PIWI nuclease domain, which may limit their potential use as nucleases in various tools, akin to CRISPR-Cas. Indeed, quite a few catalytically-active pAgos have been developed into proof-of-concept tools[17–19], with some of them using catalytically-dead mutants of those pAgos[20,21]. In fact, the applicability of innately catalytically inactive short pAgo has also been demonstrated[10]. Therefore, the usefulness of such catalytically inactive pAgos cannot be dismissed, as has been discussed previously[22]. The possibility of advanced tool development necessitates a profound and comprehensive grasp of the mechanisms at hand, however. Therefore, it is key to characterize such potential candidates and dispel any misconceptions and uncertainties, stemming from previous results or lack thereof, potentially unveiling some unexplored features and paths of research.

The major subject of this doctoral thesis is a truncated long-B prokaryotic Argonaute and its associated protein from *Archaeoglobus fulgidus*. Although

first described some time ago, the information available on it seems incomplete and conflicting. **The goal** was to characterize the Argonaute and its associated protein structurally and biochemically. To that end, the following objectives were undertaken:

1. To test the oligomeric state of AfAgo *in vitro*.
2. To determine AfAgo specificity for the guide and target strands.
3. To probe complex formation between AfAgo and an associated protein.
4. If a complex is formed, investigate structurally and biochemically whether it differs from stand-alone AfAgo.

Scientific novelty and practical value

AfAgo has been used as a model for the characterization of eukaryotic and prokaryotic Argonaute proteins since it was one of the first to be described structurally with structures available in both apo- and nucleic acid-bound forms[23–26]. It is a curious case in pAgo evolution, since it classifies phylogenetically as a long-B type pAgo, albeit comprising only the MID and PIWI domains, a characteristic thought to be exclusive to short pAgos[4,5]. It has a catalytically inactive PIWI domain and has not been shown to possess any activity *in vivo*. This work demonstrates several important findings. First, all known AfAgo structures available in PDB show a dimeric form of AfAgo, with reasonable dimerization interfaces not further investigated by the researchers[23–26]. Using single-molecule Förster resonant energy transfer, atomic force microscopy, small-angle X-ray scattering, size-exclusion chromatography-coupled multi-angle light scattering, and structural analysis, the results presented in this work demonstrate that AfAgo is capable of forming homodimeric complexes and looped dsDNA structures *in vitro*, a feature previously unobserved among pAgos. Secondly, some Agos exhibit specificity for the nature of the guide and target (RNA *vs*. DNA) strands, the 5′-terminal nucleotide, and the 5′-phosphorylation state of the guide strand, and can use the guide strand for target binding[22,27], which has not been previously described for AfAgo. Using deep sequencing of co-purified RNA, X-ray crystallography, and EMSA, it was revealed that AfAgo has a preference for a ssRNA guide bearing a 5′-P-AUU terminus and is capable of using the guide for ssDNA targeting *in vitro*. Lastly, a previously undescribed protein has been identified in the same operon, upstream of AfAgo in *A. fulgidus* DSM 8774 and reconstructed in DSM 4304, where it was obscured by a dinucleotide deletion and a resulting frameshift. Results demonstrate that

this protein, dubbed AfAgo-N, is structurally equivalent to the N-L1-L2 domains of long pAgos and forms a heterodimeric complex with the AfAgo through the same dimerization surface, which AfAgo uses to form homodimers, forming fAfAgo, structurally similar to short PAZ-less pAgos. fAfAgo also exhibits RNA-guided DNA targeting activity, as does a single peptide chain fused scfAfAgo. These findings significantly broaden the knowledge of potential hitherto unknown pAgo features, e.g., homodimerization and its possible functions, a split state – a heterodimeric PAZ-less pAgo with a non-effector N-domain, sequence specificity beyond the first 5′-nucleotide, and pave the way for future research.

The major findings presented for defence in this dissertation

1. AfAgo forms homodimeric complexes and looped dsDNA structures *in vitro*.
2. AfAgo shows specificity for the 5′-terminal AUU sequence of the guide RNA *in vivo*.
3. AfAgo exhibits RNA-guided ssDNA targeting *in vitro*.
4. AfAgo forms a heterodimeric complex with AfAgo-N, forming fAfAgo, reminiscent of a long PAZ-less pAgo.
5. fAfAgo and the fused scfAfAgo exhibit RNA-guided DNA targeting with a higher affinity than stand-alone AfAgo.

1. LITERATURE OVERVIEW

1.1 Eukaryotic Argonautes

Back in 1998, while studying plant development in a model system of *Arabidopsis thaliana*, Bohmert *et al.* derived a plant with a genetic mutation that significantly affected the morphology of *A. thaliana* saplings[1]. The mutation involved a putative 115 kDa protein, the function of which at the time was unknown, though it was noted that the protein sequence had similarities with a novel gene family present in humans as well as nematodes. Bohmert and colleagues, noting some visual similarity of *A. thaliana* to the pelagic octopus *Argonauta argo* (and erroneously calling it a squid), were compelled to name the newly discovered protein Argonaute, thus coining the term for the whole protein family[1]. At that time, no similar proteins in bacteria or yeast were known. Delving deeper into the mechanisms of Argonaute (Ago) proteins, it was discovered that these eukaryotic proteins execute their function *via* a mechanism called RNA interference, or RNAi for short[28–31]. This mechanism was shown to involve short RNA of either endogenous or exogenous origin and the result of this mechanism was regulation of gene expression and has been observed, although not fully explained in terms of additional factors other than RNA, previously[32–36]. This scientific breakthrough was one of the pivotal advances in biology, scoring the worm soakers/injectors[1][33,37–39] Andrew Fire and Craig C. Mello a Nobel prize in 2006[40], and paved the way for many future discoveries, including the elucidation of certain mechanisms of gene regulation and the development of tools and methods involving said mechanisms[3,28,41–43].

Argonaute proteins – the key players in RNAi – are widespread and conserved across the eukaryotic kingdom, with one of the notable exceptions being *Saccharomyces cerevisiae*, which not only lacks Ago proteins but seems to not have any short RNA-centred gene silencing machinery, present, however, in other budding yeast[44,45]. In eukaryotes, Argonautes can be

[1] As per P. Medawar's comment in *Advice to a Young Scienist* on injecting mice, "Few hypodermic needles are large enough for even the smallest mouse to pass through <…>," the opposite problem exists with worms: you'd be hard-pressed to find a hypodermic needle small enough so that the worm doesn't end up like an olive on a spear or explode like an overinflated party balloon, oozing the experimental liquids – and its innards – all over. Well worthy of a Nobel prize.

grouped into three major paralogous clades: the AGO clade – ubiquitously expressed and interacting predominantly with microRNA and short interfering RNA; the PIWI clade, which is involved exclusively with a class of small RNA termed PIWI-interacting RNA; and the WAGO clade, the latter found only in worms (hence, Worm AGO), described as secondary Argonautes since they depend on other Agos for their function[44,46–48]. Many organisms encode multiple genes for Agos, mostly those belonging to different clades. E.g., *Drosophila melanogaster* has 5 (2 AGO and 3 PIWI), humans and mice (*Mus musculus*) have 8 (4 AGO and 4 PIWI), there are 10 (all AGO) in the originally studied *A. thaliana* and a whopping 27 AGO genes in *C. elegans*[2,49,50].

Eukaryotic Argonautes are very conserved in their structure. They are universally composed of four domains (or six, depending on whether you count the linkers as domains): N-terminal, PAZ (PIWI-Argonaute-Zwille), MID (Middle), and PIWI, along with two linkers – L1 and L2 – positioned between N and PAZ, and PAZ and MID domains, respectively (Figure 1)[2,3,49,51]. The four domains form a bilobed structure with a channel between N-PAZ and MID-PIWI lobes, in which nucleic acid binding occurs, with the active site located in the PIWI domain, containing an RNase H fold with a catalytic tetrad DEDX (where X is D, H, or K). This catalytic site coordinates two magnesium ions and facilitates the cleavage of RNA targets complementary to the guide RNA[52,53]. The functions of the four domains of Argonautes have previously been described (or at least speculated upon) in detail. The N-terminal domain participates in the separation of the guide and target strands, the MID domain contains a pocket that binds the 5′-terminus of the guide RNA, mostly through a lysine residue, while the PAZ domain is responsible for the binding of the 3′-terminus[54]. As mentioned above, the catalytic site is located in the PIWI domain[2,55,56].

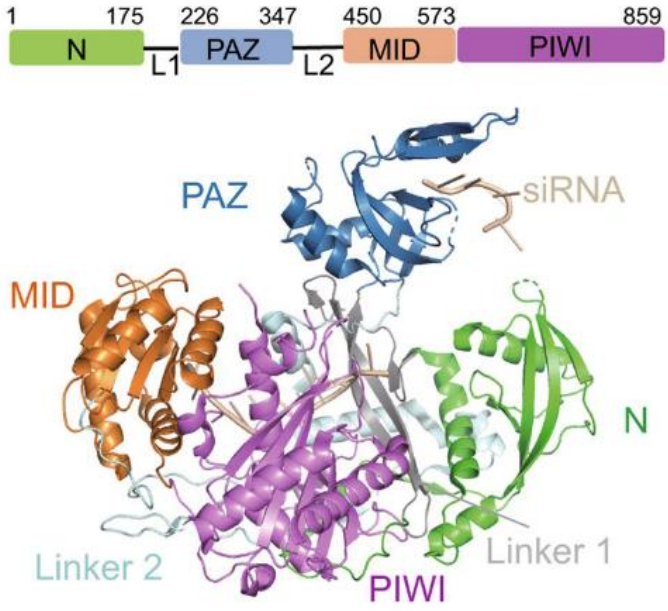


Figure 1. Domain organization of eAgos. Top: A linear representation of hAgo2 domain organization, complete with numbered residues and domains indicated. Bottom: Structure of hAgo2 (PDB ID: 4F3T). Domains are named and coloured according to the linear representation: N-terminal domain, green; L1, grey; PAZ, blue; L2, cyan; MID, orange; PIWI, purple. Reproduced and modified from Jin *et al*., 2021[49].

Not all Argonautes, however, are catalytically active – some have a defective tetrad and therefore lack catalytic activity[2,57]. The catalytically active Agos can cleave the target RNA between the 10th and 11th nucleotide from the 5′-end of the guide[58,59]. Cleavage, however, requires perfect complementarity between the guide and the target RNA, especially in the "seed" region, which is canonically located between the 2nd and 8th nucleotide from the guide 5′-end[24,51,60,61]. This region of complementarity is crucial for the formation of a stable complex between the guide-bearing Argonaute and target RNA.

The other main component of RNAi is small non-coding RNA molecules, which can be divided into several classes and differ not only in the mechanism of their biogenesis but also in their function, although several distinct pathways can operate within a single cell[2,28,44,62–64]. One class of these RNA molecules is small interfering RNA (siRNA, Figure 2). These siRNAs usually have an exogenous origin (exo-siRNA), being processed from double-stranded nucleic acids, like viral or synthetic RNA, although repeats,

transposons, and hairpins found in genomic transcripts can also be a source of endo-siRNAs[28]. In either case, once the RNA enters the cytoplasm, it is cut into short RNA duplexes by the Dicer endonuclease, containing a helicase domain, PAZ domain (required for binding of precursor RNA 3′-end), and two RNase III domains, each of the latter introducing a break in one of the strands of double-stranded RNA, forming 21-25 nt dsRNA, with the exact length depending on the distance between the PAZ and RNase III domains[2,28,44,62–66].

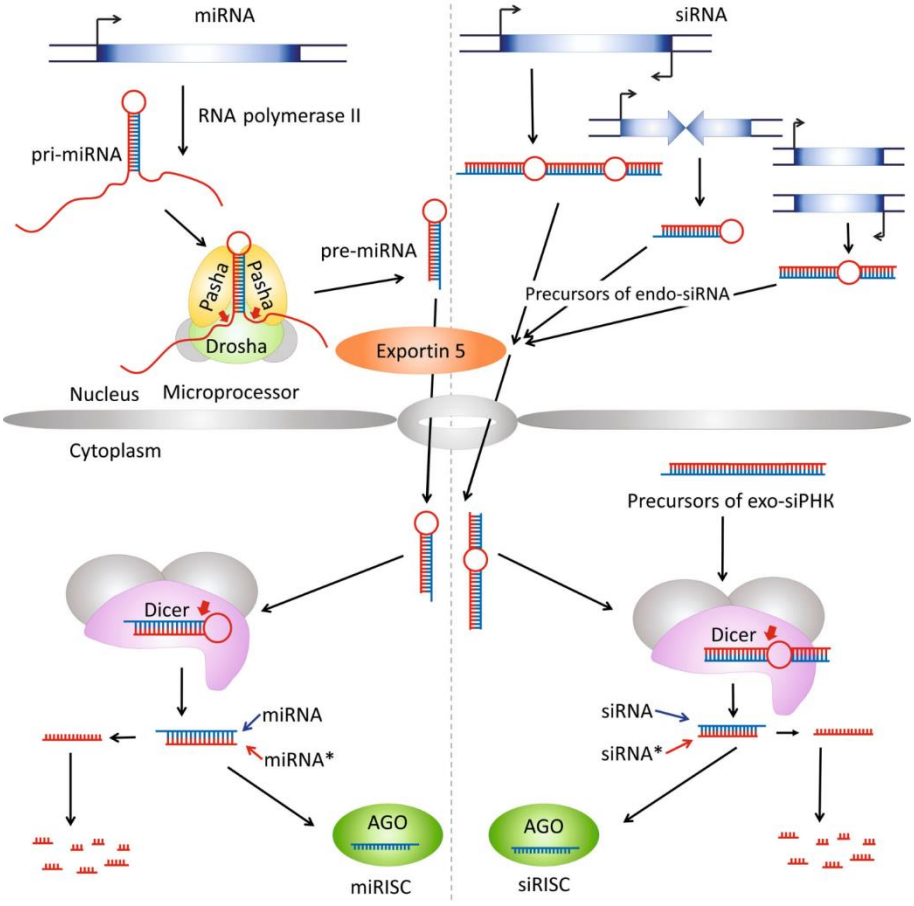


Figure 2. Biogenesis pathways of miRNA and siRNA. Reproduced from Olina *et al*., 2018[2].

Another class of small RNAs, called microRNAs (miRNAs, Figure 2), have their origin in the genome, with as many as 40% of miRNAs believed to stem from introns and even exons of certain genes, and so their biogenesis starts inside the nucleus[2,28,44,63,64,67]. In the canonical biogenesis

pathway, miRNAs are usually synthesized by RNA polymerase II, yielding primary miRNA (pri-miRNA), which can be several hundred to several thousand nucleotides long, forming a hairpin of ~10 nucleotides, although there may be additional hairpins if the transcript encodes several miRNAs[28,44,63,67]. The transcript undergoes processing by a protein complex called the microprocessor, consisting minimally of two proteins: Drosha, an RNase III family protein, which cleaves off the hairpin overhangs, and DGCR8 (DiGeorge syndrome critical region 8) in humans (or Pasha – partner of Drosha – in *D. melanogaster* and *C. elegans*), which contains a dsRNA-binding domain[47,68]. The partially processed product termed the precursor-miRNA (pre-miRNA), is then exported from the nucleus into the cytoplasm by Exportin-5, where another protein of the RNase III family, Dicer, binds the 3′- and 5′-ends of the pre-miRNA and cleaves off the loop, which joins the two arms of the hairpin, which yields a dsRNA molecule about 22 nucleotides long, termed mature miRNA, from which one strand is loaded into the Argonaute.

Lastly, the largest class of small non-coding RNAs in animal cells are PIWI-associated RNAs (piRNAs), so named for their interaction with PIWI proteins[2,28,63,64,69,70]. Unlike the previously mentioned siRNA and miRNA, piRNA is mainly found in germline cells, where their main function, in tandem with PIWI proteins, is to silence mobile genetic elements and retrotransposons in these cells[28,64,71–75]. piRNAs are markedly distinct from miRNAs and siRNAs: firstly, in size and structure – mature piRNAs are longer (26-32 nucleotides) and their precursors have no definite secondary structure motifs; secondly, in biogenesis pathways – piRNAs do not depend on Dicer and Drosha activity[76,77]. Though the complete picture of the biogenesis of piRNAs is yet to be elucidated, possible mechanisms have been described (Figure 3)[76,77]. In this mechanism, the precursors of piRNA are transcribed as single-stranded RNA from genomic loci called piRNA clusters, although active transposons are also a valid source of such transcripts[72,75,78]. In the former case, the long non-coding transcripts give rise to primary piRNA, while the latter gives rise to secondary piRNAs[72,76,77,79]. These transcripts are then transported out of the nucleus and are processed further *via* cleavage by the PIWI protein in the ping-pong amplification mechanism by an endonuclease Zucchini (Zuc) in the Zuc-dependent processing pathway[77]. The Zuc-dependent cleavage products are then bound by a PIWI protein, after which the transcripts are further processed to trim and methylate the 3′-end by a Trimmer nuclease and Hen1

methyltransferase, respectively[80]. In the case of ping-pong amplification, the piRNA-loaded PIWI protein can bind transcripts that are complementary to the bound piRNA and cleave them. This cleavage results in the accumulation of RNA fragments that can also be loaded into PIWI family proteins and participate in a further round of complementary strand cleavage, which results in the accumulation of piRNAs, complementary to both strands of the transposon (Figure 3)[78].

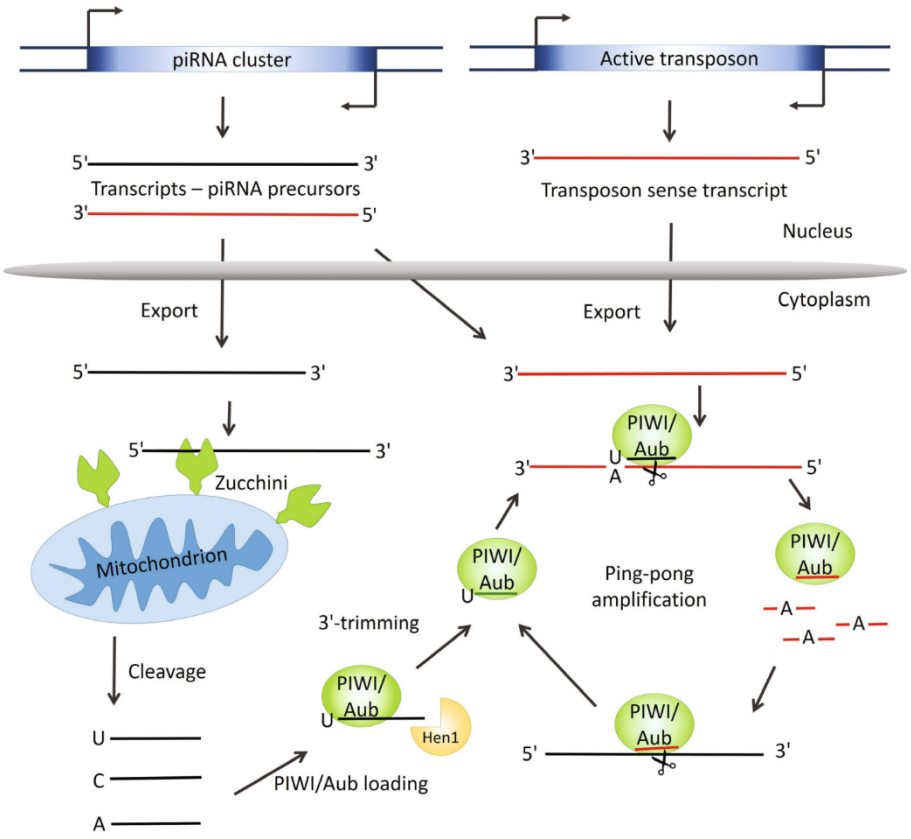


Figure 3. Pathways of piRNA biogenesis. Reproduced from Olina *et al*., 2018[2].

Although the different classes of small RNA molecules are generated by different pathways, all mature small RNAs associate with a protein from the Argonaute family, to form the functioning RNA-induced silencing complex (RISC) or the RNA-induced transcription silencing complex (RITS)[23,52,63,64,68,81]. The formation of RISC (or RITS, for that matter) is a key process in RNA-mediated gene silencing. Since both RISC and RITS act by annealing to their target NAs, the small RNA duplexes must be

unwound to generate ssRNAs, one of which remains in the active complex, while the other is discarded, making a fully functional RISC or RITS complex. Therefore, the active complex formation can be divided into two distinct steps – AGO loading and duplex unwinding.

Small RNAs – siRNAs and miRNAs – are loaded selectively and this selection depends on several features of those RNAs, like the duplex structure and its thermodynamic asymmetry, which dictates that the strand with the less stably paired 5′-end is loaded into the AGO proteins[63,82–84]. In flies, small RNAs are strictly sorted based on their intrinsic structural property: a mismatch at nucleotide positions 9-10 favours miRNA loading into DmAgo1, while siRNA, lacking the mismatch, are loaded into DmAgo2[85]. Another key factor in strand selection is the identity of the 5′-terminal nucleotide of the guide strand – DmAgo1 preferentially associates with miRNAs bearing a 5′-U, while DmAgo2 prefers 5′-C in the guide strand[86]. Meanwhile, humans have no such selectivity for small RNA loading and all four AGO proteins (AGO1-4) can be loaded with both miRNA and siRNA duplexes, though preferably with mismatches at positions 8-11[63]. The AGO proteins are not alone in their task of loading a small RNA. In *D. melanogaster*, e.g., R2D2 and Dicer-2 form a RISC-loading complex (RLC), which senses the thermodynamic asymmetry of the duplex and binds the more stable 5′-end, leaving the less stable end for Dicer-2, thus orienting the duplex for DmAgo2 loading[84,87]. In mammals, the RLC is formed by Dicer and a dsRNA-binding partner protein – the TAR RNA binding protein (TRBP)[78]. The transfer of small RNA duplexes in flies and humans is an ATP-dependent process, which also requires RNA chaperones[86].

Loading is followed by the second step of RISC assembly – unwinding[64,86]. While this process is so far quite understudied, it is accepted that it can follow two distinct mechanisms – slicer-dependent (DmAgo2) and slicer-independent (DmAgo1). In the latter case, duplex mismatches in the seed region and positions 12-16 on the guide strand are detected and facilitate the unwinding of the miRNA[63,64,86]. In flies and humans, however, the AGO2 proteins have slicer activity and can cleave the passenger strand of the siRNA duplex, with cleavage products being released from the AGO protein[88]. This is possible, however, only with perfect strand complementarity, therefore this slicer-dependent pathway is reserved mainly for siRNA and rarely involves miRNA duplexes. Other proteins have also been shown to be involved. E.g., C3PO (Component 3 of Promoter of RISC)

acts as a nuclease, removing the passenger strand after cleavage[89], while helicases unwind the duplex in flies and humans[63,64,86,90]. After duplex unwinding, the AGO protein, now bearing a mature RNA guide, can finally take another step towards executing its function in gene silencing. For that, however, the AGO (with any protein that might use AGO as a partner for its function) must first recognize its target RNA.

Once bound to an Argonaute protein, all small non-coding RNAs function in the same way – as guides for gene silencing by the action of Argonaute or partner proteins, working *via* base-pairing interactions of the guide RNA with target transcripts. The guide sits in the Ago protein with its seed region (bases 2-8) exposed to the solvent and pre-arranged in a quasi-helical geometry, which creates an optimal site for RNA binding[24–26]. The geometry of the guide RNA in the binary duplex does not strictly follow an A- or B-form geometry: most often there is a kink between bases 6 and 7 and destacking of said bases, breaking the A-form structure in this region of the guide[91–93]. The protruding nucleotides 2-8 are, in fact, critical for target recognition and are used by the RISC to probe RNA targets as the complex scours the cell and account for the great efficiency with which RISC can locate its targets. The importance of the seed region is most evident in mammalian Agos, acting mainly in the miRNA-mediated repression without cleavage of target mRNAs that are imperfectly complementary, with regions outside the seed, like the 3′-supplementary region, guide positions 13-16, enhancing seed pairing[64,94]. Agos acting *via* target cleavage require more extensive base pairing between the guide and target strands. DmAgo2, e.g., relies not only on the seed region of the guide siRNA but also uses the extended 3′-supplementary region (guide pos. 12-17) for target binding[94]. The seed region nucleotide positions 4-5 are most sensitive to mismatches and greatly reduce (~600-fold) guide-target affinity, while mismatches at positions 15-16 have a less severe impact on affinity (~250-fold decrease)[94]. The two regions – the seed and the extended 3′-supplementary region – are not functionally equivalent, however, even in cases of full complementarity. As mentioned previously, the RISC first rapidly scans seed-matching sites and once the target is found, base pairing propagates through the central and the extended 3′-supplementary regions to full complementarity[94]. This is accompanied by conformational rearrangement of both the Ago protein and the guide strand[95–97]. In the case of siRNA, this propagation of base pairing precludes any cleavage of partially-matched off-targets[95,97].

Once the small guide RNA-eAgo complex finds an appropriate target, translation repression or target degradation follows[2,63]. Non-cleaving pathway involves additional effector proteins and a cascading pathway. In miRNA targeting, a protein GW182 binds the gRNA-AGO, making contacts with the AGO and PABC (poly(A) binding protein cytoplasmic), bound to the mRNA poly(A) tail. Subsequently, PAN2 and PAN3 poly(A) specific ribonuclease subunits are recruited, followed by another endonuclease CCR4-NOT1[63]. This sequence of events results in the deadenylation of the mRNA poly(A) tail. The now deadenylated mRNA is then decapped by DCP2 and other additional factors and lastly degraded by 5′→3′ exoribonuclease 1 (XRN1)[63].

As we conclude our exploration of eukaryotic Argonautes and their pivotal roles in RNA interference, siRNA, miRNA, and piRNA pathways, it becomes evident that the RNA-guided silencing mechanisms these proteins govern are highly conserved and fundamental to gene regulation in eukaryotic organisms. Their discovery and understanding have revolutionized molecular biology, offering new insights into gene expression control, development, and disease. Now, let's shift our focus to prokaryotic Argonautes, which present a fascinating contrast in structure, function, and evolutionary history. By delving into the world of prokaryotic Argonautes, we will uncover intriguing adaptations and unexpected partnerships that challenge our understanding of these ancient nucleic acid-guided systems.

1.2 Prokaryotic Argonautes – gaining focus and momentum

Eukaryotes are, obviously, not the only ones wielding defence systems based on the recognition of genetic targets by short nucleic acid guides. Along with the ubiquitously elsewhere described CRISPR/Cas systems, prokaryotic Argonaute proteins (pAgos) were discovered, though their functions remained obscure for some time [98] and, indeed, quite a few questions still remain about the exact mechanisms and functions of pAgos. As understudied as they are, however, pAgos were key in elucidating the mechanisms and functions of eAgos. Furthermore, in recent years the interest in pAgos has been growing increasingly faster. Not only are pAgos of purely scientific interest – they may help solve some curious questions about the evolutionary relationships between pAgos and eAgos, prokaryotic and eukaryotic immune systems, like defence against mobile genetic elements and viruses – but also the markedly accelerated rates at which new pAgos are discovered and described is no doubt

spurred by other successful research on prokaryotic defence systems, namely, CRISPR/Cas, in hopes, likely, to rival and maybe even supersede these systems in the field of genome engineering and associated tools. One of the reasons is that although CRISPR/Cas is an invaluable tool, there are drawbacks to it. E.g., one such limitation is the need for a Protospacer Adjacent Motif (PAM) site next to the target sequence by the CRISPR/Cas proteins[99]. This greatly limits potential targeting sites by these systems. In contrast, pAgos do not require a PAM site and rely solely on the guide nucleic acid to find their targets. pAgos are not without their drawbacks, however. Many pAgos come from (hyper)thermophilic organisms, which renders them all but useless in the context of mammalian genome editing. Even more importantly, many discovered pAgos are catalytically inactive and therefore cannot cleave their targets, and those that can, only use guides to cleave single-stranded nucleic acids, not without the help of some factors (elevated temperature, helicases, etc.) to unwind and melt the dsNA before cleavage. With all that considered, the pAgo field of research is, at this time, bustling with new research: pAgos are being investigated not only to probe their functions *in vivo* but also to harness them as tools – with notable success – in diagnostics, microscopy, DNA assembly and fingerprinting[20,100].

Although some structural and mechanistic features of eAgos and pAgos overlap, it is the study of pAgos – their structures, biochemical characteristics and such – that shed light on their eukaryotic counterparts. And while the previous chapter dealt with describing eAgos in detail first, not only for purely historical reasons, but also for the sake of comparison of the different mechanisms between eAgos and pAgos, further in the text we will dive more deeply into pAgos, the main topic of this dissertation.

1.2.1. pAgos: distinct defenders from across the ages

Host-parasite interactions have been around long before there has been life. Ever since there was some genetic information capable of (self-) replication, genetic parasites abounded, using every opportunity to hijack the host's function and machinery for their own benefit. Since it is virtually impossible for any cellular form of life to entirely eliminate genetic parasites, the history of life has always been, and will continue to be, a perpetual arms race between the host and the parasite, where each evolves new and diverse strategies for offence, defence, and defence countering to subvert the strategies and workings of the other. Thus, most cellular life forms combine multiple systems to fend off parasites. The mechanisms described in the chapters above

– the NA-guided target recognition, resulting, in some cases, in the cleavage or suppression of the target, that often happened to be a genetic parasite, likely stem from the primordial RNA world of days yonder, as they are based on the most fundamental principle of life as we know it – the complementary nature of nucleic acids.

As has become evident, eAgos are exactly such a system, yet highly conserved and specific to eukaryotes. Further studies, however, soon discovered homologues of these proteins in prokaryotic genomes[98], with later claims indicating that around 32% of sequenced archaeal and 9% of eubacterial genomes encode Argonaute proteins[4,101], many of those were found in the so-called defence islands, which are clusters of genes in prokaryotic genomes responsible for the defence of the host from parasitic genetic elements (phages, plasmids, mobile genome elements)[98,102]. With time, some 1700 pAgos have been discovered by the year 2020 and the claimed percentage shifted to ~25% in archaea and 10–20% in bacteria[22]. Though in some sense pAgos are quite similar to eAgos, with regards to the general domain organization and the basic functions of the canonical structures and functions of N, PAZ, MID, and PIWI domains could be considered overlapping, curiously, pAgos also exhibit noticeable divergence from their eukaryotic counterparts[46,98] (or, rather, *vice versa*, considering the evolutionary trajectory). Even more significantly, phylogenetic analysis revealed, that eAgos are but a small branch on the tree of all known Agos (Figure 4) [98] and the greater diversity of pAgos compared to their eukaryotic counterparts strongly suggests that these systems first developed and diversified in prokaryotes, and only later were picked up by eukaryotes, which also acquired additional proteins such as Dicer[46,98,101,103]. In fact, the phylogenetic analysis of pAgos based on the sequence similarities of their more conserved domains has shown that pAgos do not follow the phylogenetic trends of their host organisms, suggesting that they may have been acquired and diversified broadly among prokaryotes by horizontal gene transfer[46]. Some more significant examples of the diverse features characteristic (yet some are not exclusive) to pAgos include loss or gain of functional domains, loss of the catalytic DEDX motif in the PIWI domain, gene localization in defence islands, association with other domains or proteins and forming functional complexes, use of Mg^{2+} ion in the MID domain for guide 5′-end binding (in eAgos, a lysine residue performs this function), and use of DNA guides, etc.

1.2.2. Functions and diversity

It was tempting to speculate on the evolutionary relationship of eAgos and pAgos, with the former apparently being a descendant of ancient bacterial, archaeal, and phage proteins, involved in RNA processing and DNA repair[103]. The uniqueness of pAgos became more obvious with an in-depth analysis[98]: it was noticed that some pAgos lack the N and PAZ domains (basically, half the polypeptide chain) or have some putatively analogous domain termed APAZ (Analogue of PAZ) in lieu of them. According to this, they could be classified into short and long pAgos, respectively. Further, some associate with additional domains or effector proteins. Indeed, research has shown that it is even more complicated and elaborate than that. Ryazansky *et al*. [4] demonstrated that, in agreement with Makarova *et al*.[98], pAgos could be classified into short and long pAgos but further divided the long pAgos into two separate phylogenetic clades based on the sequence alignment of the MID-PIWI domains – long-A and long-B (Figure 4)[4,98]. The majority of pAgos that belong to the long pAgo clades have a domain architecture similar to that of eAgos, including, as mentioned previously, N, PAZ, MID, and PIWI domains (Figure 6a) (with some exceptions). There are some key differences, variations, and features in specific structural elements involved in nucleic acid binding and processing, both between the members of the long pAgo clades, and between long pAgos and other Agos[4,27]. The long-A clade pAgos mostly contain both the active PIWI and a normal-sized PAZ domain (Figure 4, Figure 6). The intact catalytic activity is, most likely, the reason why they are the most well-studied clade, with such members as TtAgo, AaAgo, PfAgo, CbAgo from *Thermus thermophilus, Aquifex aeolicus*, *Pyrococcus furiosus*, and *Clostridium butyricum*, respectively, among others (Figure 4)[4,27,104]. All known long-B pAgos, in contrast, lack the canonical catalytic DEDX tetrad of the PIWI domain, with substitutions present for the critical amino acids in the catalytic site, and hence cannot enact their function *via* target cleavage. This is also true for some of the long-A pAgos – not all of them possess an intact catalytic site, hinting at a likelihood of several independent events of the loss of the catalytic activity. Further, many members of both long pAgo clades have substitutions in key residues, interacting with the guide 5′-end. Notice the trend? Exceptions are the rule, it seems.

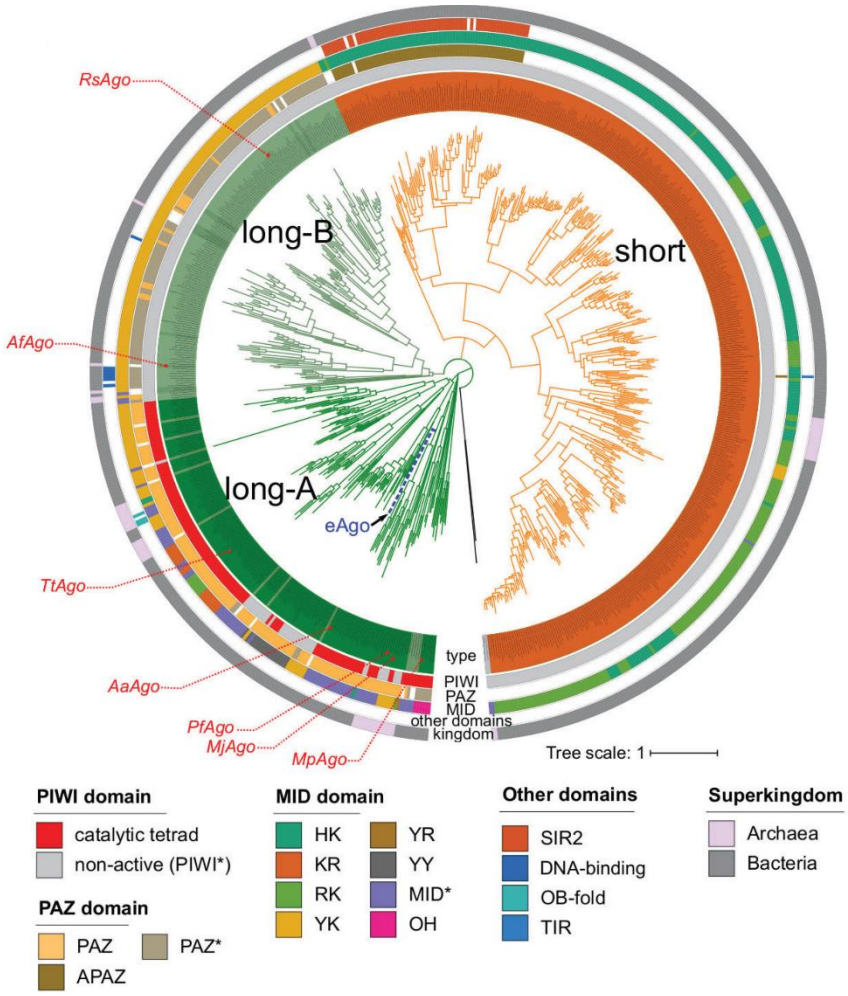


Figure 4. A phylogenetic tree of pAgos, based on the sequence alignment of MID-PIWI domains. Annotation is as follows, from the outer circle in: superkingdom, to where the pAgo belongs; PAZ/APAZ domain presence and type; 5′-end guide binding motif of the MID domain, with the first two conserved residues indicated; PIWI domain type, based on the DEDX catalytic tetrad; type of protein, where long-A pAgos are coloured green and truncated long-A pAgos lacking the PAZ domain are light green, long-B pAgos are light green and their corresponding truncated (PAZ-lacking) variants are green, and orange indicates short pAgos. Reproduced and modified from Ryazansky *et al*., 2018[4].

The insight gleaned from the study by Ryazansky *et al*. on the MID domain revealed the key six amino acid motif that is highly conserved in the MID domains of most pAgos. Four of the highly conserved residues – Y/R, K, Q, and K – form hydrogen bonds with a bound divalent metal ion (Mg^{2+} or Mn^{2+}) and with the 5′-phosphate of the guide and the 3rd phosphate of the guide,

anchoring them in the basic pocket in the MID domain[4,25,26,46,96,97]. Further, Y forms a stacking interaction with the first base of the guide *via* its aromatic ring, which R cannot do as it is non-aromatic[2], as demonstrated in the TtAgo-gDNA binary complex, where R418 is distal from g1T (Figure 5a). Two semiconserved residues T and N interact with the base of the nucleotide in the 2nd position of the guide and its phosphate group. Thus, the conserved motif found in most pAgos can be defined as YKQTNK. These residues, however, are somewhat different between long-A and long-B pAgos: while long-B pAgos seldom have anything other than YK, long-A, on the other hand, feature YY, KR, HK, and KR residues (Figure 4, Figure 6). Altogether, some long pAgos do not contain the "canonical" YKQTNK motif and bear at least one amino acid substitution in their MID* domain: they lack the first Y/H/R, the second K/R/Y and/or the third Q and last K (Figure 6). Interestingly, a separate clade of pAgos exists with a more hydrophobic 5′-end binding pocket (Figure 4)[4,22,46,49]. As exemplified by MpAgo from *Marinitoga piezophila*[13,22,27,105,106], these pAgos likely prefer 5′-OH over 5′-P, substituting the lack of Mg^{2+} ion-mediated contacts with those formed by the second and third phosphate and a lysine residue and the lack of first Y and its stacking with the first base is compensated by a preceding Y or F, which interacts with the same base from another angle[13]. As with usual exceptions in the world of pAgos, there are instances where pAgos that usually bind 5′-phosphorylated guides can also use 5′-OH guides for target cleavage, e.g., SeAgo from *Synechoccoccus elongatus*, KmAgo from *Kurthia massiliensis*, CbAgo, LrAgo from *Limnothrix rosea*, but then the cleavage site gets shifted by one nucleotide, and cleavage efficiency can be affected[107–109].

[2] It is difficult to gauge cation-aromatic (and, more broadly, cation-π) contacts due to the complexity of electrostatic potential surfaces in the aromatics. These interactions can be attractive or repulsive [228] and the proximity of the cation to the aromatic in a crystal structure alone hardly gives any information on the energetic consequences of the interaction, as geometries can be more or less favourable. Further, in the case of arginine-aromatic interactions, not only the charge of the guanidinium group is relevant, but also the low solvation and the van der Waals interactions of the R aliphatic chain with the aromatic ring, as discussed by Dougherty[229] and Gallivan&Dougherty[228], and references therein.

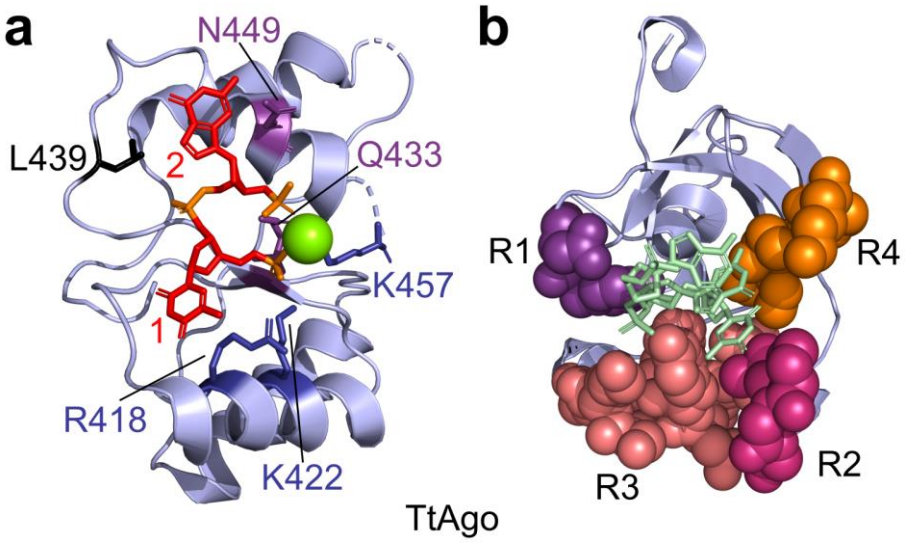


Figure 5. Binding of guide termini by TtAgo (PDB ID: 3DLH). (a) Structure of TtAgo binary complex with 5′-P gDNA with conserved residues of the YKQTNK motif indicated (note the non-conserved R418 and L439). (b) Structural regions R1 through R4 of the guide 3′-end binding pocket of the TtAgo PAZ domain are shown as spheres. Prepared acc. to Ryazansky *et al*., 2018[4].

Another variable feature of pAgos is the PAZ domain, which has the function of guide 3′-end binding. While the majority of short pAgos lack the PAZ domain altogether, some possess a so-called Analogue of PAZ (APAZ) domain. Further, many of the long pAgos also have a truncated PAZ domain, termed PAZ* (Figure 4, Figure 6)[4]. However, the "canonical" PAZ domain, best exemplified by long-A pAgo TtAgo[96] and long-B RsAgo from *Rhodobacter sphaeroides*[13], not only does not have a strictly conserved fold, the amino acid sequences of the PAZ domains from various pAgo proteins are also divergent[4]. The full-length PAZ domain features a hydrophobic 3′-end binding pocket formed by two subdomains, each bearing two nucleic acid binding regions – R1 and R4 in the first subdomain and R2 and R3 in the second (Figure 5b)[52,81,110], though a fully intact PAZ is not really necessary for the binding of the guide 3′-end, as observed in MpAgo and RsAgo, and, indeed, AfAgo [13,16,23,24,26,105] and almost half of the long pAgos – more specifically, most long-B pAgos – lack the R3 nucleic acid binding region and therefore the bilobed structure and the nucleic acid binding pocket of the PAZ domain[4], though this results in a somewhat different orientation of the guide in the case of PAZ*. This raises an interesting point: while most long-B pAgos, which are catalytically inactive, make do with a

PAZ*, the catalytically active long-A pAgos likely need the intact PAZ for a more precise guide-target duplex formation to prevent premature target cleavage. Further, since the 3′-end is released from the PAZ domain upon target recognition, binding of the 3′-end of the guide may shield it from host nucleases while pAgo is searching for a target.

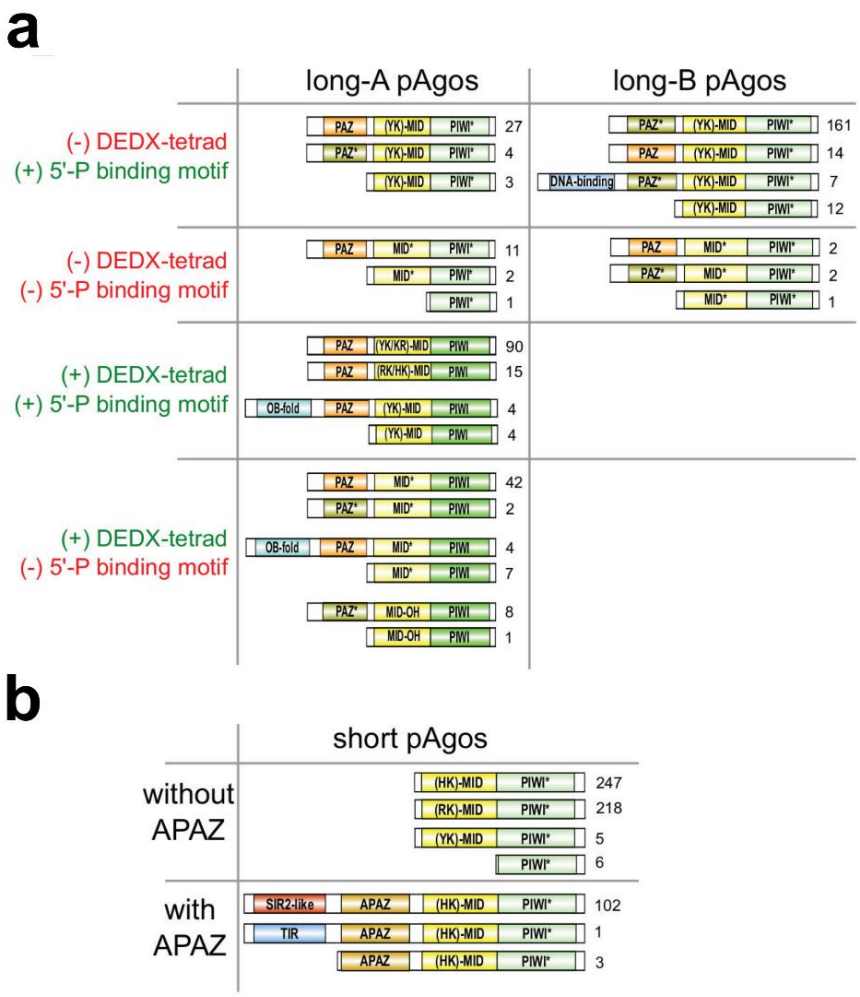


Figure 6. Domain organization of long (a) and short (b) pAgos. The occurrence frequency of each type in a redundant set of proteins is indicated by the numbers on the right. Four types of long pAgos can be distinguished based on the presence of the DEDX tetrad in the PIWI domain and the putative 5′-P guide binding motif in the MID domain. Green PIWI domains contain the active DEDX catalytic tetrad, while turquoise PIWI* domains lack this tetrad. Yellow MID domains have different types of the 5′-end guide binding motif, as shown, while light-yellow MID* domains have substitutions of the critical amino acid residues. Orange

PAZ domains have the full-sized pocket responsible for 3′-end guide binding, while light-brown PAZ* variants lack the second subdomain. The OB-fold is the nucleic acid-binding domain, and the DNA-binding domains are the Schlafen domain. Reproduced and modified from Ryazansky *et al*., 2018[4].

The majority (~59%) of pAgos discovered so far, however, belong to the so-called short pAgo clade[4,5]. The vast majority (94%) of those are found in bacteria and only a small fraction (6%) in archaea. The short pAgos have a distinct feature of lacking the N and PAZ domains, comprising only the MID and PIWI domains, with some being even shorter, lacking the MID domain also. There are, however, exceptions, with some 18% of the short pAgos bearing the APAZ domain, which shows no sequence similarity to the PAZ domain (Figure 4) and was suggested to be the functional analogue but not a homologue of the PAZ or N domain[4,5,98,111]. Short pAgos can also cluster with APAZ-containing proteins rather than bearing the domain in a single peptide chain (Figure 6b)[4,5,10]. Multiple alignment of APAZ domains and domain architecture analysis of APAZ-containing proteins revealed that they fall into five distinct groups, which Ryazansky *et al*. termed Ia, Ib, IIa, IIb, and III, the latter being the smallest of the bunch (Figure 7)[4]. Group Ia and Ib have a common feature of also bearing a Sir2 domain, forming a Sir2-APAZ protein type. Group Ib is different from Ia in that most proteins from Ib lack MID and PIWI domains, yet have Sir2-APAZ. Both groups, however, also feature proteins that comprise only the APAZ domain. Group IIa is characterized by the presence of a TIR domain on the N-terminus of APAZ (TIR-APAZ), however, some proteins that fall into this group lack the TIR domain. The most diverse is group IIb, characterized by proteins featuring Sir2 and DUF4365 domains fused to APAZ, while some examples comprise only the APAZ domain (Figure 7)[4]. It has previously been suggested that the TIR, Sir2, and DUF4365 domains might have some nuclease activity, which would compensate for the lack of a catalytically active PIWI in short pAgos[4,98,112]. And, indeed, while DUF4365 domain belongs to the Mrr PD-(D/E)XK nuclease subfamily and is involved in assisting the short pAgo in guide-mediated target cleavage[98,113], the TIR and Sir2 domains have been shown to mediate abortive infection responses by depleting intracellular $NAD(P)^+$, as exhibited by SPARTA/SPARSA and pAgo-unrelated Thoeris, CBASS (TIR-STING fusion) systems[9–11,114–117], a cell death mechanism also featured in eukaryotes[118–120]. Finally, proteins from group III feature only the APAZ domain and are similar to most group IIb proteins (Figure 7).

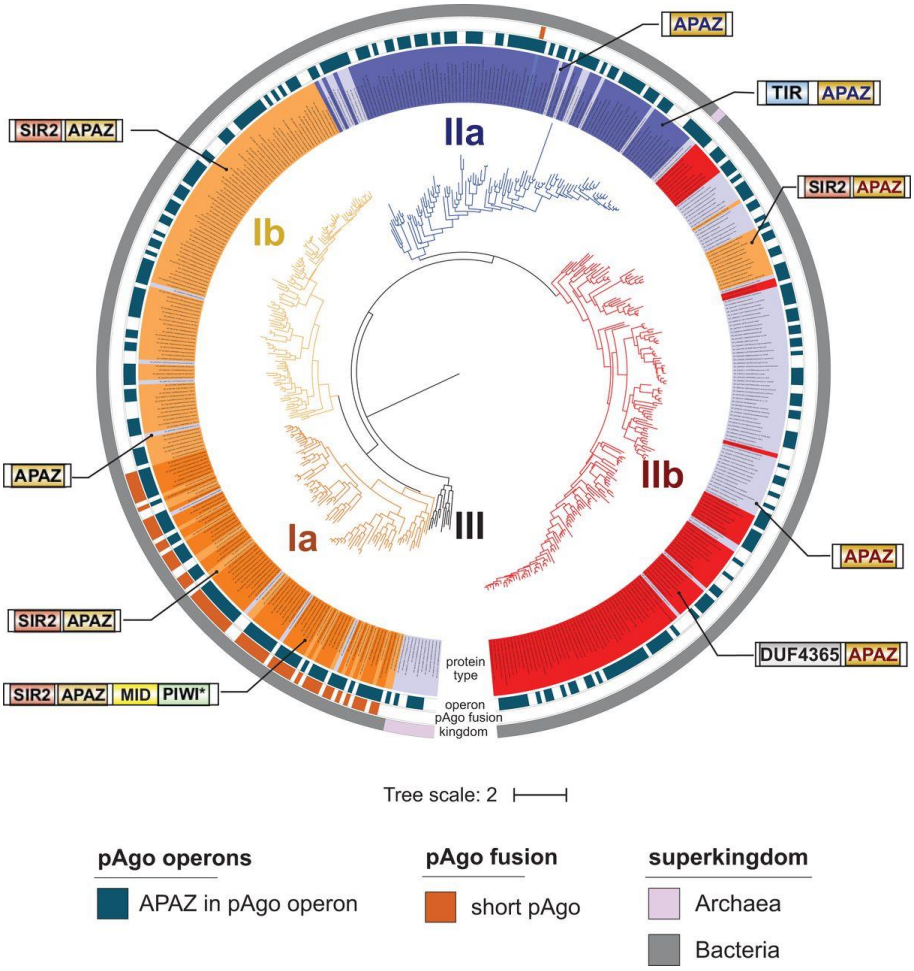


Figure 7. The APAZ domains were subjected to phylogenetic analysis, resulting in the construction of a circular phylogenetic tree illustrating the five distinct groups. The tree was annotated with different information in concentric circles. The innermost circle represents the protein types, with colours indicating the phylogenetic groups and domain compositions. Callouts with corresponding domain schemes exemplify each protein structure type. Isolated APAZ domains in all groups are depicted in light blue. The next circle indicates whether the APAZ-containing protein is found in a pAgo coding operon. The following circle represents APAZ domains fused with short pAgo proteins. Finally, the outermost circle indicates the superkingdom to which the corresponding APAZ-containing protein belongs. Reproduced from Ryazansky *et al*., 2018[4].

Indeed, these results were corroborated and expanded by Koopal *et al*. 2022a[10]. They showed that short pAgos not only cluster with APAZ-containing proteins, but also form four distinct phylogenetic subclades (Figure

8), which they termed S1A, S1B, S2A, and S2B for Sir2-APAZ-pAgo fusions, operon with Sir2-APAZ and pAgo, operons with (Mrr-)TIR-APAZ and pAgo, and operons consisting mostly of DHS-like-APAZ and DUF4365-APAZ, respectively. However, Koopal *et al*. further distinguish two other variants of short pAgos: a small SiAgo-like subclade, characterised by a pAgo system from *Sulfolobus islandicus*, which features a short pAgo forming a complex with SiAga1 and an effector transmembrane domain-bearing effector SiAga2, and other truncated pAgos, which they term pseudo-short pAgos, exemplified by AfAgo. Both of these variants, according to Koopal *et al*., do not associate with APAZ domains. However (spoiler alert), that may not be entirely true.

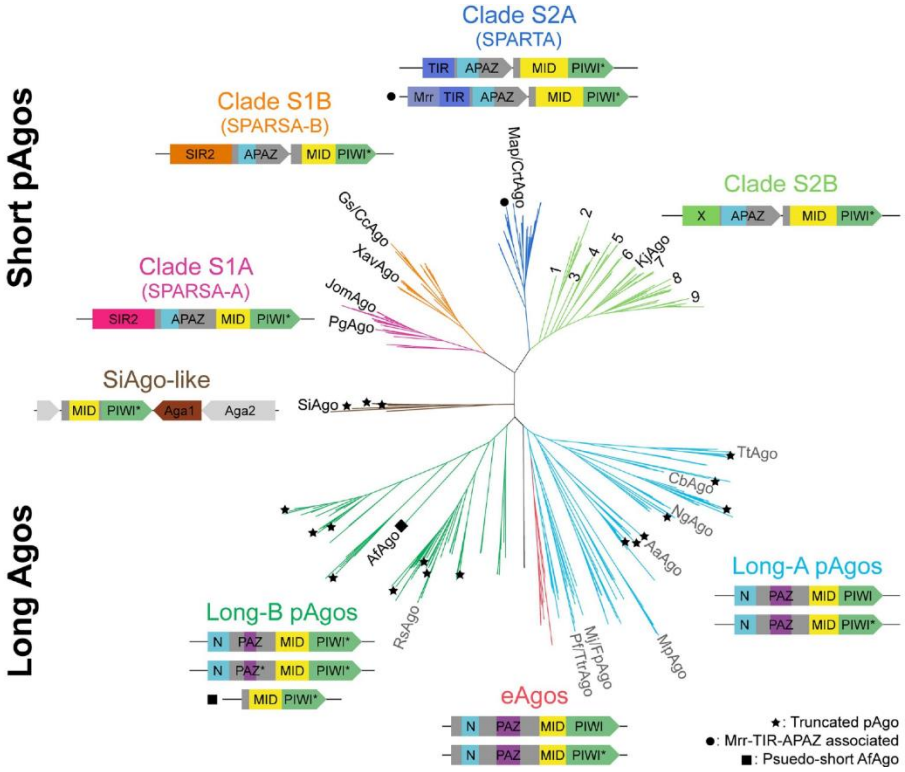


Figure 8. Phylogeny of prokaryotic Argonaute (Ago) proteins by Koopal *et al*., 2022b[5]. The unrooted phylogenetic tree was constructed using maximum likelihood analysis, incorporating all pAgo homologs identified in the RefSeq database with scaffold-level assemblies. Additionally, experimentally characterized pAgos, eAgos, and SiAgo homologs were also included. The phylogenetic tree reveals several distinct clades: Clade S1A, which features the fusion of Sir2 with an APAZ domain in pAgo (SPARSA-A); Clade S1B, consisting of operons with pAgo and Sir2-APAZ (SPARSA-B); Clade S2A, comprising operons with pAgo and TIR-APAZ (SPARTA); and Clade S2B, containing operons with pAgo and APAZ fused to various domains (denoted as "X"), such as Mrr, DUF4365, and DHS-like domains.

Truncated pAgos in long-A, long-B, and SiAgo-like clades are represented by ★, while short pAgos associated with Mrr-TIR-APAZ are denoted by ●. Additionally, pseudo-short AfAgo is indicated by ■. The phylogenetic analysis provides valuable insights into the evolutionary relationships and diversification of prokaryotic Argonaute proteins, shedding light on their functional associations and domain arrangements. Reproduced from Koopal *et al*., 2022b[5].

Whether long, short or truncated, all pAgos eventually need to acquire guide nucleic acids to perform some function of host defence *via* the canonical guide-mediated target recognition pathway, although initial steps can, in some cases, occur with a guide-less pAgo[121,122]. While the guide acquisition and loading steps in pAgos are still poorly understood, there are several characterised or purported mechanisms. In general, what separates pAgos from eAgos is that pAgos do not need (and don't have) any other companion proteins, like eukaryotic Drosha and Dicer, to be loaded with guide nucleic acids. Some pAgos have been demonstrated to generate guides by "chopping" ss- or dsDNA into short fragments, which are then loaded into pAgo and can be used for target recognition. E.g., guide-independent activities have been observed for TtAgo, PfAgo, MjAgo from *Methanocaldococcus jannaschii*, CbAgo, LrAgo, KmAgo, SeAgo, and EmaAgo from *Exiguobacterium marinum*[7,12,104,107,109,121–123]. Of these, only KmAgo has not yet been explicitly demonstrated to directly auto-load guides arising from independently processed DNA, though it has been shown to associate with short ssDNA fragments when purified from *E. coli*[108]. These DNA fragments in KmAgo seem to arise from both the host genome and the plasmid vector and are sampled uniformly[108]. Host genome sampling has also been observed with TtAgo, SeAgo, CbAgo, EmaAgo and several others, as described by Lisitskaya *et al*., 2023[6,7,104,109]. Interestingly, often the most-sampled sites are observed at replication origin and termination sites *ori*, *terA* and *terC*, respectively[6,8,104,108,109,124]. Replication termination site sampling makes sense in light of possible replication fork stalling, where dsDNA ends are formed when RecBCD or AddAB attempt to repair the stall or double-stranded break, generated by the replication forks colliding at the end of replication. The involvement of DSB repair machinery, although not strictly necessary for siDNA (small interfering DNA) generation, has been speculated on and demonstrated previously[6,8,104,124]. Further, pAgos may also sample the transcriptome/RNA degradome for guides or use siDNA-assisted DNA target cleavage products as new guides for a new cycle of DNA cleavage (albeit demonstrated *in vitro* only), similarly to the ping-pong amplification observed with piRNA[14,125]. This may raise the question of self *vs*. non-self discrimination by pAgos. One potential explanation could be

the higher copy numbers of invader DNA, which, in the end, results in a higher abundance of guides and more efficient targeting. Another is if pAgos target DSBs and replication intermediates, these can form in more plentiful numbers in multiple copies of phages and plasmids, compared to the host genome. Also, the relative lack of Chi sites in foreign DNA allows for more efficient processing by the RecBCD/AddAB machinery, potentially producing multiple guides and enabling efficient targeting of replicons with a low copy number.

Regardless of the origin of the guides, pAgos can have some selectivity for the particular nature of the guides they use. As discussed in more detail in the section on the MID domain, the different residues of the guide 5′-end binding pocket can allow the pAgo to discriminate 5′-P and 5′-OH guides, while some can use both. Additionally, akin to some eAgos, some pAgos can be selective for a specific 5′-terminal nitrogenous base of the guide strand. E.g., MjAgo, TtAgo, RsAgo, GsSir2/Ago from *Geobacter sulfurreducens*, MapAgo from *Maribacter polysiphoniae*, and FpAgo from *Ferroglobus placidus* bind guides with 5′-terminal purines, cytosine, uracil, adenosine-uracil dinucleotide, adenosine, and guanosine, respectively, however quite a few more are non-selective[7,9,10,13,14,126,127].

While the first nitrogenous base is flipped out and sequestered in its own binding pocket, rendering it unavailable for guide-target pairing, nucleotides 2-8 of the guide strand – the seed region – are available for target base pairing. Like in eAgos, where the gRNA is pre-arranged in an A-form-like geometry (see p. 32), facilitating target binding, in MpAgo, the gRNA is kinked sharply between the 6th and 7th base by a Y166 residue of the PAZ domain, which disrupts the MpAgo-bound gRNA from the A-form conformation[105]. Destacking also occurs in TtAgo, but between the 7th and 8th base of the guide[110].

These conformational changes likely pre-pay the entropic cost of duplex formation and hence facilitate more efficient target binding. Indeed, as with eAgos, pAgos enhance guide-target duplex formation greatly[20]. While eAgos are sensitive to mismatches between the guide and the target strands in the seed region, one cannot be so strict when defining pAgos as mismatches can affect the pAgo activity in different ways. E.g., TtAgo and RsAgo have been shown to accommodate nucleotide bulges in the seed region of the guide strand and, to a lesser degree, the target strand, which results in local distortions of the double helix and shifting of the cleavage site, lower affinity

to the target and quicker release of the imperfect duplex[16,128]. For TtAgo, cleavage activity was reduced irrespective of the bulge position and in some cases – when closer to the cleavage site – abolished outright[96,128]. MpAgo showed peculiar sensitivity to mismatches: while mismatches at positions 5, 7, and 8 decreased cleavage efficiency, dinucleotide mismatches up to positions 3 and 4 had no such effect, with cleavage being markedly reduced with mismatches between positions 4 – 5 and 15 – 16[105,106]. Such a decrease in efficiency with various guide-target mismatches has also been shown in other pAgos[108,129]. In contrast, target cleavage is not significantly impacted in CbAgo and even stimulated in LrAgo and SeAgo when non-complementary interactions in the seed region are present[107,109]. Furthermore, contrary to eAgos, which are insensitive to mismatches in the 3′-supplementary region, SeAgo, CbAgo, LrAgo and some others exhibit strong inhibition of target cleavage when mismatches at positions 12 – 15 are present, while reports on TtAgo claim unaltered activity with up to six contiguous (pos. 13 – 18) mismatches, decreased activity with eight (pos. 12 – 19) contiguous mismatches [96,107,109,129,130]. KmAgo is also an interesting case. with findings of two studies showing intriguing variations in how mismatches impact the activity of KmAgo and painting distinct pictures. In one study[108], a comprehensive analysis involving different combinations of guide and target nucleic acids revealed that mismatches had varied effects depending on the interacting molecules. For gDNA/tDNA interactions, mismatches in the 3′-supplementary region of guides had the most pronounced impact on DNA cleavage, while central region mismatches and those at positions 4 – 5 in the seed region also reduced cleavage efficiency. Interestingly, when gDNA was paired with tRNA, mismatches in the 3′-supplementary region seemed inconsequential, with the most significant decrease in efficiency occurring at the cleavage site and in the middle of the seed region. However, gRNA/tDNA reactions were highly sensitive to mismatches in both the seed and central regions. Surprisingly, these mismatches not only influenced efficiency but also introduced changes in the pattern of target cleavage, leading to additional cleavage products at noncanonical sites. In stark contrast, the second study[123], which predominantly focused on the seed region (pos. 2 – 8) of guides, revealed a somewhat opposite effect. Here, mismatches in the seed region significantly stimulated KmAgo activity. Moreover, mismatches at the cleavage site (guide pos. 10 – 11) and the 3′-supplementary region of DNA guides (pos. 13 – 15) displayed no substantial effect on DNA cleavage. Instead, certain mismatches in positions 7 and 12 mildly reduced DNA cleavage efficiency. Intriguingly,

mismatches within the 3′-supplementary region had negligible effects on RNA cleavage, whereas a mismatch at the cleavage site (guide pos. 11) decreased RNA cleavage efficiency. The introduction of dinucleotide mismatches revealed a complex scenario where some stimulated DNA cleavage, while others dramatically reduced it. Additionally, single-nucleotide mismatches in specific positions within 5′-P-gRNA sequences had distinct effects on DNA and RNA cleavage. These contrasting results highlight the contextual nature of the response of KmAgo to mismatches, emphasizing that the impact of mismatches on its activity can be highly dependent on the nature of the nucleic acid and the region of the guide-target complex under consideration.

As we further explore the mechanisms governing pAgo response to target binding, it's essential to consider the central player in target cleavage – the PIWI domain. Structural and biochemical studies of both eAgos and pAgos showed that a conserved catalytic amino acid tetrad DEDX (where X is D, H or K) which chelates divalent metal ions Mg^{2+} and Mn^{2+} and is responsible for the target cleavage[46,81,97,98,104,105,110]. And while the majority (~79%) of long-A pAgos have the canonical catalytic tetrad, hinting that they are catalytically active, all known long-B pAgos contain substitutions of the critical amino acid residues[4,27,46,49,104]. In catalytically active pAgos, the catalytic site holds the so-called glutamate finger, in which the catalytic glutamate is positioned. Depending on the presence or absence of a target, this glutamate finger can adopt different conformations: while there is no target, the glutamate finger adopts an "unplugged" conformation when the complete tetrad does not form and sometimes no divalent metal ions are bound, and switches to a "plugged in" conformation when an extended guide-target duplex is formed (Figure 9). This duplex formation causes conformational rearrangements in PIWI and PAZ domains, shifting the glutamate finger into a "plugged-in" conformation, facilitating metal ion binding and target cleavage[97,105,110,126,131,132]. This is not always the case, though. RsAgo, e.g., remains in the "unplugged" conformation, though it does lack the catalytic tetrad anyway[16].

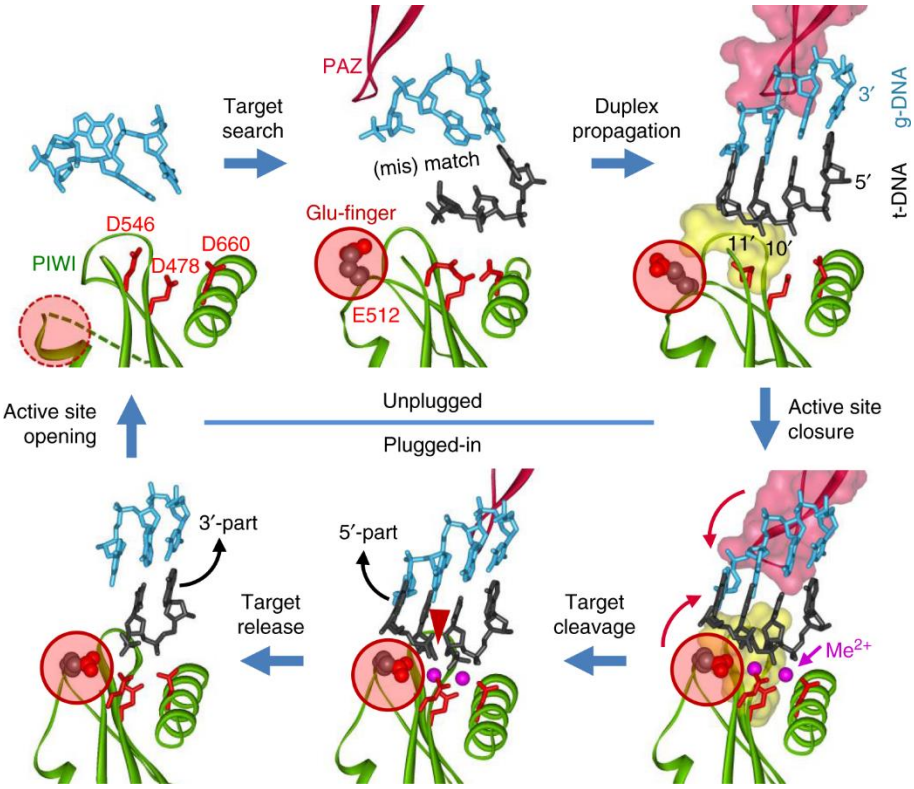


Figure 9. TtAgo exhibits conformational changes in the active site during target recognition and cleavage[97,131]. Starting with the "unplugged" conformation (top left), target recognition and duplex propagation (clockwise) go together with conformational changes in the PIWI and PAZ domains (red arrows) and this in turn "plugs in" the glutamate finger (encircled in red), facilitating metal ion binding (magenta spheres) and lead to target cleavage (bottom left). Afterwards, the target is released, returning TtAgo to the initial "unplugged" state (top left). PDB accession numbers, from the top left, clockwise: 3DLH, 3F73, 4N41, 4NCB, 4NCA, 4N76. The red and yellow surfaces represent the PAZ domain region proximal to the guide and PIWI domain region proximal to the scissile phosphate of the target, respectively. Reproduced from Lisitskaya *et al*., 2018[27].

The catalytic cycle is well described structurally for TtAgo (Figure 9, Figure 10)[97,131]. The Guide-target duplex propagation from 12 – 15 mer to 16 – 19 mer states is followed by the release of the 3′-end of the guide from the PAZ pocket. This is followed by large structural transitions around the catalytic pocket, positioning E512 within proximity of the scissile phosphate and completing the catalytic tetrad, coordinating two Mg^{2+} cations and several water molecules. In contrast to canonical RNase H enzymes, where the catalytic glutamate directly coordinates the divalent cation, here the Mg^{2+} is coordinated *via* bridging waters. One Mg^{2+}-coordinated water molecule is

prepared for the nucleophilic attack on the scissile phosphate (Figure 10b, red arrow). The proposed transition state of an SN2 reaction is substantiated by the position of the nucleophilic water, the scissile phosphate and the O3 phosphate-leaving group (Figure 10c). Following cleavage in TtAgo, the catalytic E512 remains coordinated with one of the two Mg^{2+} cations and two bridging waters. The other Mg^{2+} cation is coordinated by the newly-generated 5′-phosphate and 3′-OH oxygens and two water molecules, in addition to D478 and D660.

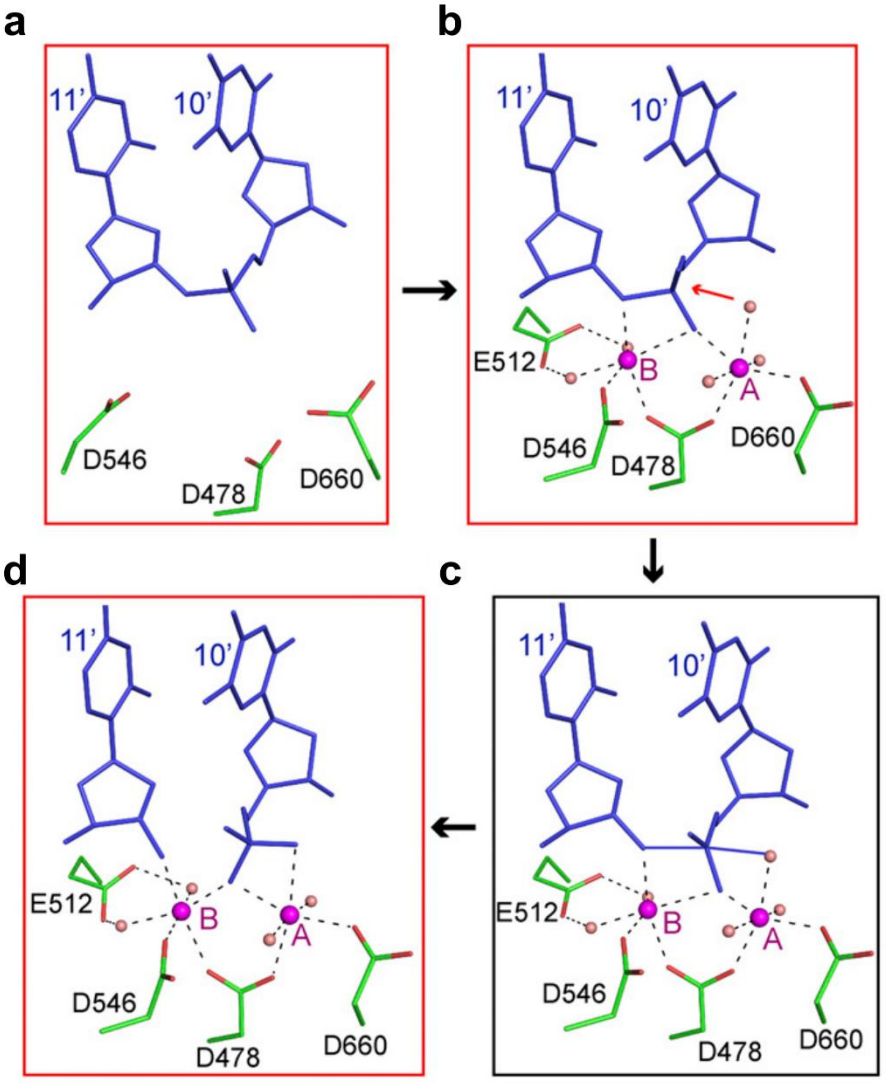


Figure 10. The tDNA cleavage mechanism by TtAgo, proposed by Sheng *et al*.[131], visualized as crystal structure snapshots (a – PDB ID: 4N47, b – PDB ID: 4NCB, and d – PDB ID: 4N76) and a suggested transition state (c). The proposed pathway leads to cleavage of the 10′ – 11′ phosphodiester bond in the TtAgo-gDNA-tDNA complex. (a) The catalytic E512 is outside the catalytic site in the cleavage-incompatible arrangement. (b) E512 shifts closer to the scissile phosphate in the cleavage-compatible conformation. (c) Proposed model of cleavage reaction transition state. (d) Catalytic site arrangement after DNA cleavage. Reproduced and modified from Sheng *et al*., 2014[131].

Target release after cleavage at this time remains quite poorly understood. While in eAgos and certain mesophilic pAgos it is a rate-limiting step[6,53,108,133,134], in thermophilic pAgos it does not seem to be the case.

It is unknown whether any additional protein factors aid pAgos in target dissociation post-cleavage, however, there are a few possible explanations. First, for thermophilic pAgos, the cleaved target may dissociate freely simply due to temperature-induced melting of the guide-cleaved target duplex. In the context of structural insights into TtAgo[131], the process can be explained as follows: as the cleaved 5′-part of the target dissociates, the complementary portion of the guide, specifically its 3′-end, becomes disordered. This disordering of the guide's 3′-end facilitates its binding within the PAZ pocket, which in turn promotes the release of the target. This process is essentially the reverse of target binding. More investigation is needed to elucidate the exact mechanism of target release after cleavage, however. Although information is somewhat scarce beyond guided nucleic acid cleavage for the sake of host defence, pAgos can have more extensive functions *in vivo*. This is true for catalytically active pAgos, as well as inactive, short pAgos. One interesting example is chromosome decatenation by TtAgo after replication (Figure 11)[8]. Prokaryotic chromosomes are circular, bearing a single *ori* site, where replication initiates bi-directionally, and *ter* sites where replication forks collide and replication terminates. This process results in two new circular chromosomes, which are linked, or *catenated* (I will leave the imagining of the topology to you). These linked circles require decatenation – unlinking, a process performed by topoisomerases or gyrases[3] (e.g., DNA topoisomerase IV in *E. coli* and gyrase A in *T. thermophilus*, its sole type II topoisomerase)[8,135]. Jolly *et al*. demonstrated in *T. thermophilus* that TtAgo knockouts subjected to a gyrase inhibitor ciprofloxacin failed to successfully decatenate and segregate chromosomes into cells and septate them after replication, resulting in cell elongation and the formation of cellular filaments. And while TtAgo was able to restore normal cell replication, it did so only up to a certain concentration of ciprofloxacin (Figure 11). This demonstrates that TtAgo participates in chromosome decatenation upon gyrase A inhibition[8]. This has also been demonstrated for SeAgo and LrAgo[124].

[3] There are two types of topoisomerases: type I produces transient ss breaks, while type II – ds breaks. Topoisomerase I in both prokaryotes and eukaryotes is type I, while eukaryotic topoisomerase II, bacterial topoisomerase IV and gyrase (the latter two are homologues) are type II. And while one may think that DNA gyrase is the minimal functional component for decatenation, it is rather topoisomerase IV, as shown experimentally by Zechiedrich et al., 1997[230].

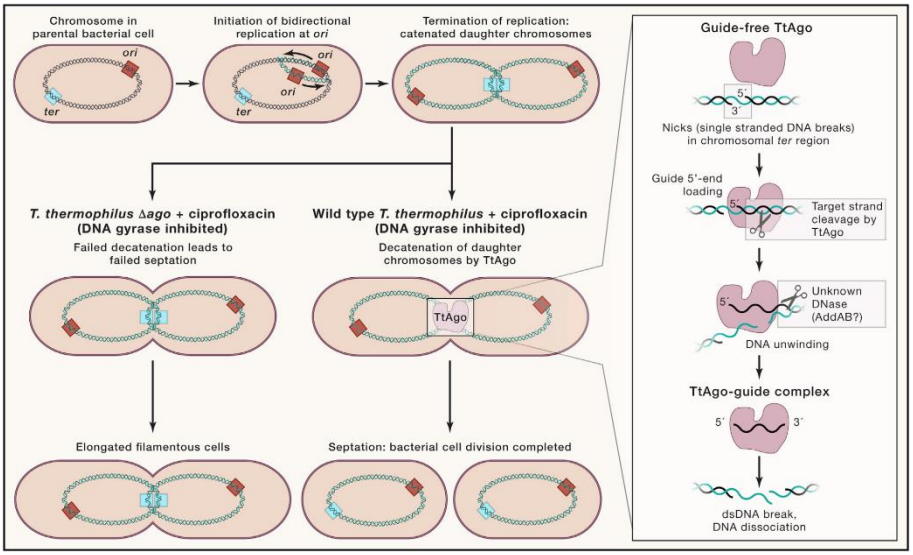


Figure 11. Models for TtAgo DNA guide generation and loading, and TtAgo-mediated chromosome decatenation. Reproduced from Swarts, 2020[136].

Quite a few studies on activities of short pAgos in host defence beyond nucleic acid degradation have popped up recently (some briefly mentioned above, pp. 32 – 34), several even elaborating on mechanistic details with the help of cryo-EM[9–11,116,137–140]. One of these activities is NAD^+ depletion by SPARTA/SPARSA systems, a form of abortive infection – a means of last resort to induce cell dormancy or death to prevent the spread of invaders[141,142]. As mentioned previously, short pAgos lost their function as guided nucleases but associate with various effector proteins and/or domains, e.g. TIR and Sir2. Proteins bearing these domains have been demonstrated to be NADases[11,114,115,143,144]. Therefore, short pAgos have taken the role of sensors in prokaryotic immune systems. Upon the detection of invading nucleic acids, the nucleoprotein complex undergoes conformational changes. These changes subsequently activate the associated TIR and Sir2 effectors, which can operate either through oligomerization-dependent or -independent mechanisms (Figure 12). This activation ultimately leads to the depletion of intracellular $NAD(P)^+$, which, in turn, results in cell death. In the case of SPARSA, no oligomerization is observed, including in structural studies, since the Sir2 domain occupies the dimerization interface, preventing the formation of higher-order complexes and Sir2 domain activation likely occurs thanks to the increased domain flexibility[9,11,116].

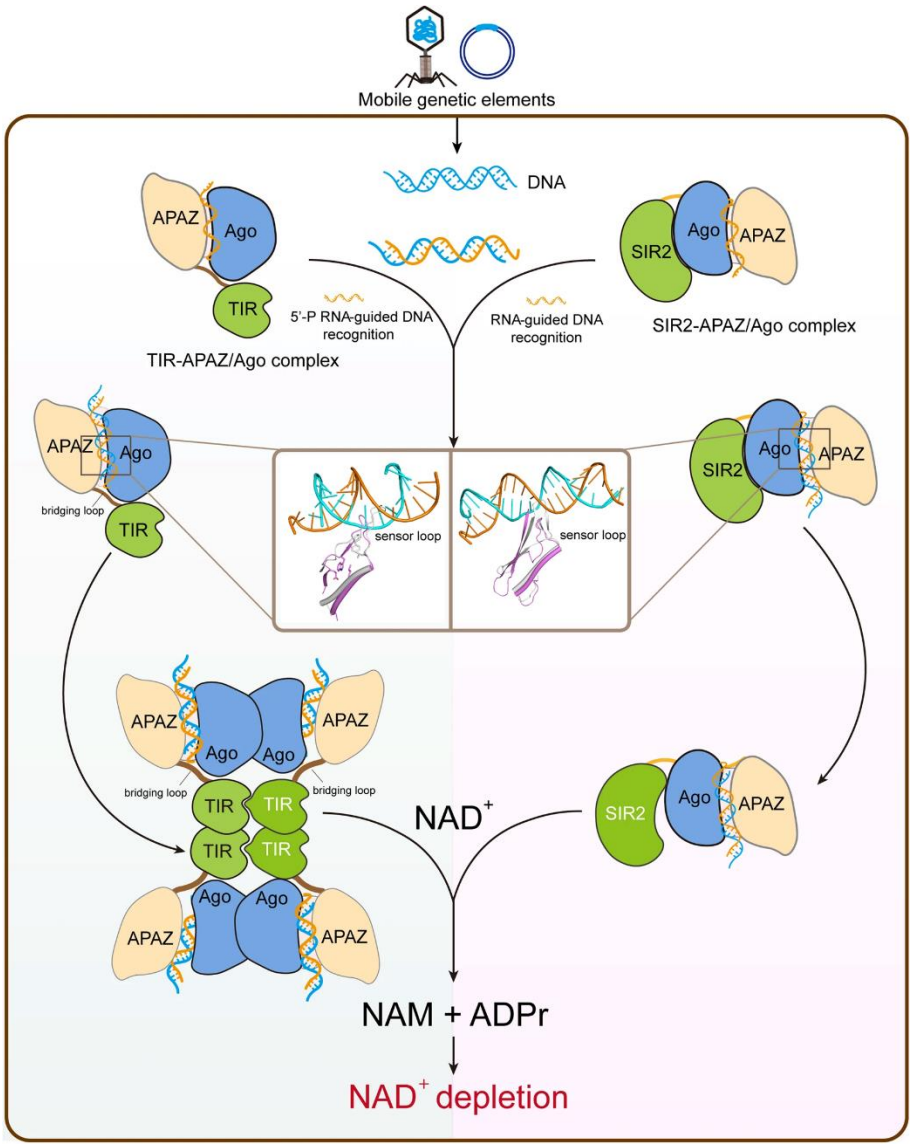


Figure 12. Schematic illustration of TIR-APAZ/pAgo and Sir2-APAZ/pAgo activation. In both systems, a conserved sensor loop in the PIWI domain probes the formation of the gRNA/tDNA heteroduplex, which triggers structural alterations necessary to activate the NADase activity of the TIR and Sir2 domains. While tDNA binding activates the Sir2 domain by increased flexibility, the oligomerization of the TIR domain it the TIR-APAZ/pAgo system is a prerequisite to the formation of the active NADase pocket. Reproduced from Wang *et al*., 2023[116].

The activation of the SPARTA, on the other hand, involves oligomerization and structural rearrangements, involving several steps. Upon

target ssDNA binding, the TIR-APAZ/pAgo-gRNA-tDNA quaternary complex dimerizes via the MID and PIWI domains of the pAgo protein. This is facilitated by the rotation of the PIWI domain away from the MID domain, which makes space for dimerization to occur, while the APAZ domain shifts closer to the guide-target duplex. Interestingly, MapSPARSA bears an 18 aa loop (termed "sensor loop") in the PIWI domain, which protrudes into the duplex binding cleft at ~13 – 15 bp position (Figure 12)[116]. It was found that this loop probes for duplex mismatches at those positions and that mismatches – and ssDNA targets shorter than 15 nt – severely disrupt the NADase activity of this particular SPARSA[10]. This loop retreats away from the guide-target heteroduplex upon its successful formation. All these changes facilitate the tetramerization of the TIR-APAZ/pAgo-gRNA-tDNA complex *via* the TIR domains, leading to a "butterfly" shaped architecture and the formation and stabilization of the composite active site between each pair of the TIR domains (Figure 12)[116,137,138,140]. This is the first demonstration of a functionally-important pAgo oligomerization, although the dimerization without known functional relevance has been described previously (see Results, pp. 70 – 88 and ref. [145]).

Another recently described pAgo from a hyperthermophilic archaeon *S. islandicus* (SiAgo) exhibits a peculiar mechanism of defence by working in tandem with associated proteins SiAga1 and SiAga2 and orchestrates an Abi response (Figure 13)[139]. In this system, SiAga2 is a transmembrane protein capable of binding anionic headgroups of phospholipids in its basic pocket and is the killer effector responsible for the system's antiviral function. Abi mechanisms, employed by various prokaryotic defence systems, involve three stages: (1) a sensor module detecting cues from invading viruses, (2) activation of a toxic effector module, and (3) the effector module inducing cellular dormancy or cell death to halt viral spread. In the case of the SiAgo system, cells undergo a series of events after viral infection, including membrane depolarization, genomic DNA loss, and membrane integrity disruption, ultimately culminating in cell death. A proposed mechanism of SiAgo system action is this (Figure 13): (1) normally, in the cytoplasm, the apo SiAgo-Aga1 complex loiters or perhaps transiently binds Aga2 without triggering a signal, (2) upon viral infection, increased abundance of RNA and DNA substrates arising from viral genome replication and transcription allows SiAgo-Aga1 to obtain ample guides and – for the guide-loaded complex – targets, (3) target binding leads to active recruitment of SiAgo-Aga1 to Aga2 and its activation, (4) activation of Aga2 leads to membrane depolarization

and loss of membrane integrity, which proves fatal for the infected cell. It's worth noting that transmembrane proteins are a common feature in various defence systems such as CBASS, retrons, Thoeris, Zorya, type III CRISPR-Cas, bacterial gasdermins, and others, where they serve (or are predicted to serve) as toxic effectors[146–152]. Further, it's interesting to highlight that apart from SiAga2, the genetic contexts of pAgos do include other membrane proteins[4].

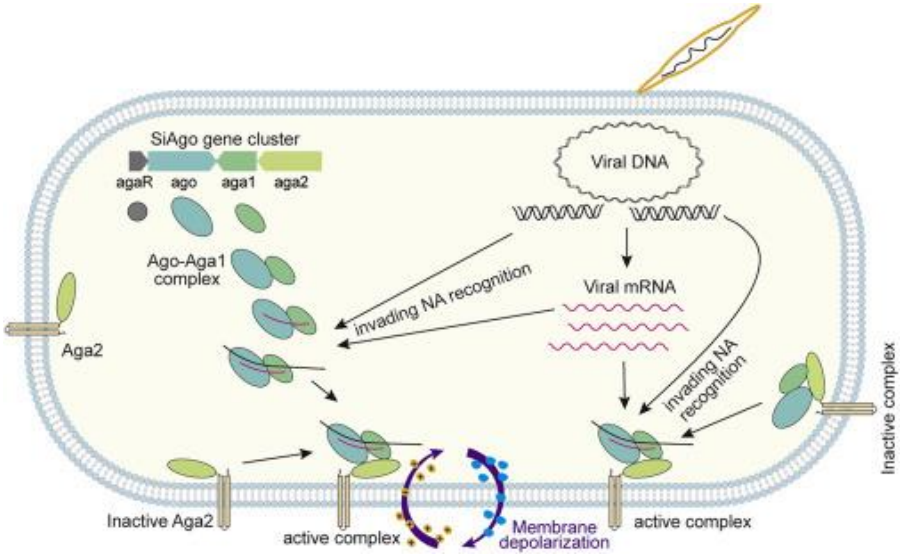


Figure 13. A suggested mechanism of SiAgo activation and action in host defence. Reproduced from Zeng *et al*., 2022[139].

As we can see, Argonaute proteins offer a window into the evolutionary connections between prokaryotes and eukaryotes. Prokaryotes, the precursors to more complex life forms, house prokaryotic Argonautes. These pAgos are the ancestors of eukaryotic Argonautes, emphasizing the evolutionary lineage between the two. What's intriguing about pAgos is their remarkable diversity. While they share core traits with eAgos, pAgos exhibit a spectrum of mechanisms and unexpected partnerships. Some have specialized in defending against RNA or DNA, often forming unique collaborations with other proteins. It's a natural experiment, where countless variations have evolved over millions of years, each tailored to its ecological niche.

However, despite our progress in understanding these molecular players, many pAgos remain mysteries. Their functions and associations are still

largely undiscovered, a testament to the vastness of the microbial world, where countless life forms remain hidden from scientific inquiry.

As we explore deeper into this microbial realm, we may unearth more secrets, shedding light on the hidden aspects of our biological heritage. This leads us to our next topic: an intriguing case involving another pAgo from *Archaeoglobus fulgidus*. While AfAgo has been studied previously, many of its most captivating aspects have remained concealed, waiting for us to uncover.

1.3 *Archaeoglobus fulgidus* Argonaute and its companion

AfAgo is a 427 amino acid 49.2 kDa prokaryotic Argonaute protein found in the hyperthermophilic archaeon *Archaeoglobus fulgidus*. To date, it has been one of the most well-studied and well structurally characterized prokaryotic Argonautes (we shall see about that!)[23–26]. So much so, that it has been extensively used as a model of a well-defined Argonaute for eAgos and other pAgos, especially in studies of Ago-NA interactions, delving into the formation of specific protein-nucleic acid contacts, identification of critical amino acids in MID and PIWI domains in guide strand binding and the importance of the seed sequence, and whatever else one might gain insight on from using such a marvellous benchmark[13,96,110,153–156].

The AfAgo protein gene is situated within a genetic island in the *A. fulgidus* DSM 4304 genome. This island has a lower GC content, approximately 31%, compared to the higher GC content of the surrounding genome sequences and the entire genome (~50% and ~53%, respectively), suggesting that the region containing the AfAgo gene was acquired by *A. fulgidus via* HGT from an unknown host[157]. AfAgo is composed of two major domains, the N-terminal MID (residues 38 – 167), and the C-terminal PIWI (residues 168 – 427)[23]. The MID domain specifically binds the 5′-phosphorylated end of the presumed guide DNA/RNA strand and also makes contacts with the complementary target DNA/RNA strand[24–26]. The PIWI domain makes contacts to both guide and target DNA/RNA strands but is catalytically inactive due to mutations in the RNase H-like catalytic centre. Lacking a PAZ domain, AfAgo is unable to bind the 3′-end of the guide. Phylogenetically, it clusters with long-B pAgos, however, due to the lack of a PAZ domain, some call it truncated long-B or a pseudo-short pAgo[4,5,10].

The companion protein of AfAgo is nowhere to be found in DSM 4304, however. All that is there is a short 172 aa protein, which is in no way similar to other known proteins. This is, without further investigation, a dead end, it seems, at least in current literature. All the information we have reviewed thus far establishes the foundation for our future exploration of the functions, mechanisms, and structures of AfAgo and its companion. Subsequent investigations will delve into these aspects, revealing intriguing and somewhat unexpected insights into these proteins within the context of pAgos.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1. Nucleic acids

DNA and RNA nucleotides listed in Appendix 1 were purchased form Metabion. Plasmid vectors, encoding His_6-AfAgo-N and Af1318 under a single promoter, scfAfAgo were purchased from Twist Bioscience and are listed in Appendix 2.

2.1.2. Chemicals, commercial reagent kits and proteins

All chemicals used in this study were purchased from Fluka, Roth, Thermo Fischer Scientific (and branches), Sigma-Aldrich, Carl Roth, Invitrogen, Lucigen, Lexogen, and Agilent and were of the highest grade available. Perkin Elmer supplied the radioactive nucleotides.

All routine commercial kits and proteins were used according to the manufacturer's instructions. These include: "GeneJET PCR Purification Kit", "GeneJET Plasmid Miniprep Kit", Phusion™ DNA polymerase, FastAP thermosensitive alkaline phosphatase, T4 polynucleotide kinase (PNK), RNase A, "Fast Digest" restriction enzymes, bovine serum albumin (BSA), Ampligase®, lysozyme, Roti-phenol/chloroform/isoamyl alcohol, glycogen, Small RNA-seq Library Prep kit. Speciality kits and materials are indicated in corresponding particular methods.

2.1.3. Bacterial media and strains

Escherichia coli strain DH5α [F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ^-] was used for cloning procedures.

E. coli strain BL21(DE3) B F^- *ompT gal dcm lon hsdS*$_B$($r_B^- m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S) was used for protein expression.

2.2 Methods

2.2.1. Protein cloning and expression

The gene encoding WT AfAgo was amplified from Archaeoglobus fulgidus genomic DNA by PCR and cloned into a pETDuet vector, yielding a construct with an N-terminal $(His)_6$ tag (N-terminal protein sequence MGSSHHHHHHSQDP (1.63 kDa) followed by 1 – 427 aa of WT AfAgo sequence). The deletion in the dimerization mutant AfAgoΔ was introduced via overlap extension PCR by using two primer pairs, MZ-385/MZ-875 and MZ-383/MZ-876 (Appendix 1) for the N- and C-terminal fragments flanking the region to be deleted, respectively. The two PCR products, possessing a 49 bp overlap, were then used as a template for subsequent PCR with the MZ-383/MZ-385 primers, yielding the full-length fragment, which was then cloned into a pETDuet vector. Successful *E. coli* DH5α transformants were selected by colony PCR and vector construction was validated by sequencing. All other vectors carrying protein genes were obtained using whole gene synthesis and cloning service provided by Twist Bioscience.

All proteins were expressed in *E. coli* strain BL21(DE3). Cells were grown in LB broth in the presence of ampicillin at 37 °C. To obtain WT AfAgo or AfAgoΔ, cell culture was grown to A_{600} of 0.5, then the incubation temperature was lowered to 16 °C, IPTG was added to 0.1 mM, cells incubated for ~16 hours at 16 °C. and harvested by centrifugation. Expression of fAfAgo (from a single vector) and scfAfAgo differs from the described above in this manner: protein expression was induced with 0.2% L-arabinose and conducted for 4 hours at 37 °C.

To obtain fAfAgo- and scfAfAgo-bound nucleic acids, vectors pBAD_TwinStrep-fAfAgo, pBAD_TwinStrep-AfAgo or pBAD_TwinStrep-AfAgo + pCDFDuet_His-AfAgo-N were used. Note that fAfAgo components are expressed from two different vectors. Cells were grown at 37 °C in LB medium in the presence of ampicillin (pBAD constructs) or ampicillin and streptomycin (pBAD + pCDF constructs) until an A_{600} value of 0.7 was reached. Subsequently, expression was induced by adding 0.2% w/v L-arabinose with pBAD vectors or 0.5 mM IPTG and 0.2% L-arabinose with pBAD + pCDF vectors, and cells were harvested after 4 h.

In all cases, after protein expression, the cells were harvested by centrifugation.

2.2.2. Protein purification

Harvested cells expressing $(His)_6$-tagged WT AfAgo or the dimerization mutant AfAgoΔ were disrupted by sonication in lysis buffer 1 (20 mM Tris-HCl (pH 8.0 at 25 °C), 500 mM NaCl, 5 mM 2-mercaptoethanol) with 2 mM phenylmethylsulfonyl fluoride (PMSF), incubated for 20 min at 50 °C and cell debris was removed by centrifugation at 48,400×g for 1 hour. The supernatant was loaded onto a HiTrap™ chelating HP column charged with Ni^{2+} (GE Healthcare) and eluted with a linear gradient (15 – 500 mM) of imidazole in lysis buffer 1. The fractions containing protein were pooled, diluted to 0.2 M of NaCl with a buffer containing 20 mM Tris-HCl (pH 8.0 at 25 °C), 10% (v/v) glycerol, 5 mM 2-mercaptoethanol and incubated for 1 h at 37 °C with 1 mM ethylenediaminetetraacetic acid (EDTA) and RNase A/T1 (ThermoFisher Scientific) (1:100). Next, the protein solution was centrifuged at 48,400×g for 30 minutes, the supernatant containing RNA-free AfAgo was loaded onto a HiTrap™ Heparin HP column (GE Healthcare) and eluted using a 0.2 – 1.0 M NaCl gradient. Finally, the protein was run through the HiLoad™ 16/600 Superdex™ 200 column (GE Healthcare) in lysis buffer 1, supplemented with NaCl to 1 M. AfAgo with bound RNA was purified as above, omitting RNase treatment.

E. coli cells expressing fAfAgo and scfAfAgo were disrupted by sonication in lysis buffer 2 (20 mM Tris-HCl (pH 8.0 at 25 °C), 200 mM NaCl, 2 mM PMSF, 5 mM 2-mercaptoethanol), heated for 30 min at 70 °C and cell debris was removed by centrifugation. The collected supernatant was treated with EDTA and RNase A/T1 as described above. Next, the protein solution was centrifuged to remove any precipitate. His_6-AfAgo-N and His_6-fAfAgo proteins were purified by chromatography through HisTrap™ HP chelating and HiTrap™ Heparin HP columns (Cytiva). scfAfAgo protein was purified by chromatography through HiTrap™ Heparin HP and HiLoad™ Superdex™ 200 columns (Cytiva).

All purified proteins were of >90% homogeneity as judged by SDS-PAGE. After purification, proteins were dialysed against a buffer containing 20 mM Tris-HCl (pH 8.0 at 25 °C), 500 mM NaCl, 1 mM dithiothreitol (DTT) and 50% v/v glycerol and stored at -20 °C in. The identity of the purified proteins was confirmed by mass spectrometry. Protein concentrations were determined from A_{280} measurements using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam/.

2.2.3. Nucleic acid purification and analysis

Protein-bound RNA was purified using phenol isolation from an RNase-untreated protein preparation. AfAgo-bound RNA was purified from the sonically lysed preparation, described above. For fAfAgo- and scfAfAgo-bound nucleic acids, the cells expressing proteins as described above were disrupted by incubating 1 h at 30 °C in lysis buffer, containing 20 mM Tris-HCl (pH 8.0 at 25 °C), 100 mM NaCl, 2 mM PMSF, 5 mM 2-mercaptoethanol, 3 mg/ml lysozyme (ThermoFisher Scientific, cat#89833). The AfAgo-nucleic acid complex was purified using StrepTactin (pBAD constructs) or Histrap and StrepTactin (pBAD + pCDF constructs) columns, and all buffer solutions contained 100 mM NaCl.

To extract nucleic acids co-purified with the AfAgo complexes, 1 mL of Roti-phenol/chloroform/isoamyl alcohol (Carl-Roth cat#A156) was added to the 1 mL of purified protein-NA fractions in 5PRIME Phase Lock Gel tubes (Quantabio cat#733-2477). The upper aqueous phase was isolated and 0.1 volume of 1 M sodium acetate, 3 volumes of 100% ethanol and 10 μL glycogen (ThermoFisher cat#R0561) were added. This mixture was vortexed briefly and incubated at -20 °C for 20 hours. Samples were centrifuged for 20 min and the supernatant was removed from the pellet. The pellet was washed with cold (-20 °C) 70% (v/v) ethanol. The pellets containing the co-purified nucleic acids were dried for 20 min at room temperature and resuspended in 30 μL nuclease-free water.

Co-purified nucleic acids were dephosphorylated with FastAP Thermosensitive Alkaline Phosphatase (ThermoFisher cat# EF0651) and labelled with [γ-^{32}P] ATP (PerkinElmer) using T4 polynucleotide kinase (PNK) (ThermoFisher cat#EK0031), according to manufacturer's instructions. Labelled nucleic acids were incubated with nucleases (ThermoFisher DNase I cat#18047019, RNase A/T1 cat# EN0551) for 60 min at 37 °C. After nuclease treatment, samples were mixed with RNA Gel Loading Dye (ThermoFisher cat# R0641), heated for 5 min at 95 °C and resolved on 20% denaturing (8 M urea) polyacrylamide gels in 1× TBE buffer (Invitrogen cat#15581-028). The Decade™ Marker System (Ambion cat#AM7778) molecular weight marker was used for RNA size identification. Radioactivity was captured from gels using phosphor screens and imaged using a Typhoon FLA 7000 laser scanner (GE Healthcare).

2.2.4. RNA sequencing and analysis

Plasmids with the cloned protein genes from *Archaeoglobus fulgidus* DSM 4304 (GenBank accession nos. NP_070147.1 and NC_000917.1, respectively) and reconstructed upstream gene based on *Archaeoglobus fulgidus* DSM 8774 (GenBank accession nos. AIG98198.1 and CP006577.1, respectively) were used in this work. RNA samples without an additional PNK treatment were converted to DNA libraries using a Small RNA-Seq Library Prep Kit (Lexogen cat#052). The concentration and quality of libraries were measured with a Qubit Fluorometer (ThermoFisher) and 2100 Bioanalyzer (Agilent). Libraries were sequenced using the Illumina MiniSeq™ sequencing platform with single-end reads and 150 bp read length. RNA sequencing was performed as described previously[158]. The raw reads were first processed by trimming adapter sequences using AdapterRemoval v2.3.043[159]. Reads then were aligned to the reference genome (*E. coli* strain K12 substrain BL21(DE3) genome NCBI: NZ_CP081489.1) and the additional pBAD-AfAgo, pBAD_TwinStrep-fAfAgo, pBAD_TwinStrep-AfAgo, pCDFDuet_His-AfAgo-N plasmids with BWA-MEM v0.7.1744[160]. In order not to filter out shorter reads during the alignment process, aligned reads with MAPQ values greater or equal to 15 were chosen. After the alignment, only the aligned reads were retrieved from the alignment file using the "bam2fastq" program from the SAMtools v1.746 toolkit[161]. FastQC v0.11.845 [162] was used for read quality control. The processed reads were analysed using a Unix "awk" filter to extract RNA sequences and a Perl program that counted the occurrence of each RNA base in the first 50 positions of the reads. A custom script (fragmentation-bias.jl) (GitHub: https://github.com/agrybauskas/argonaute-bound-rna-manuscript) in combination with Weblogo v3.7.447 [163] was used to produce nucleotide frequency plots. Gene enrichment analysis was performed with bedtools v2.26.048 [164] and FPKM_count.py v4.0.0 of the RSeqQC package[165]. The raw RNA reads are deposited to the Sequence Read Archive under the BioProject accession numbers PRJNA763829 and PRJNA978552.

2.2.5. Crystallization and structure determination

AfAgo used for crystallization was pre-treated with RNase A/T1 Mix (ThermoFisher Scientific) as described above. Complexes of AfAgo with DNA were prepared by mixing the protein solution in the storage buffer with an equimolar amount of oligoduplex in the presence of 2 mM DTT and 5 mM

$MgCl_2$. AfAgo-N was concentrated to 9.2 mg/ml in a buffer containing 20 mM Tris-HCl (pH 8.0 at 25 °C), 100 mM NaCl and 2 mM $MgCl_2$. The fAfAgo complex was prepared by mixing AfAgo in the storage buffer supplemented with 5 mM $MgCl_2$, with single-stranded DNA oligonucleotide GS-851 followed by an equimolar amount of AfAgo-N protein.

Glycerol was removed using NAP columns (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 7.5 at 25 °C), 150 mM NaCl, 5 mM $MgCl_2$ and 2 mM DTT. Complexes were concentrated by ultrafiltration. The concentration of the complexes used in crystallization trials was in the range of 90 – 120 μM (as monomer). Crystallization experiments were prepared by mixing the protein solution with equal volumes of crystallization buffers in sitting drops. Crystals were grown in a cold room (4 – 8 °C). Prior to flash cryo-cooling to 100 K, crystals were washed in the cryo-protection buffers. Protein crystallization and cryo-protection buffers are indicated in Appendix 3 and Appendix 4.

Four datasets were collected at EMBL P14 and P13 beamlines on the PETRA III ring of the DESY synchrotron in Hamburg (Germany). The datasets were processed by XDS [166] followed by POINTLESS[167], AIMLESS and TRUNCATE[168], and by CCP4 software[169]. The structures were solved by molecular replacement using MOLREP v11.6.04 [170] with PDB entries 1YTU and 2W42 as models. Structures were refined with REFMAC v5.8.0230 [171] and PHENIX v1.13 [172] and remodelled using COOT v0.8.9.1[173]. Phases for fAfAgo complex with DNA were obtained by molecular replacement using MOLREP with AfAgo protein (PDB ID: 2BGG) as an initial model. Initial phasing produced the electron density for the AfAgo subunit, phosphorylated 5′-end of the DNA chain and magnesium ion bound at the C-terminus. After a few rounds of remodelling in COOT and refinement by PHENIX, the model was improved significantly, and the second DNA chain was partially inserted. The addition of nucleic acid improved the phases and allowed to model a fragment of AfAgo-N protein containing an α helix, which was used as an initial model for phasing of the *P*1 AfAgo-N dataset. Thus, AfAgo-N was modelled by passing the partial model between both crystal structures improving mutually step by step. Both AfAgo-N structures were refined by REFMAC.

The crystallization dataset parameters, data collection and refinement statistics, and PDB accession codes are presented in Appendix 3 and Appendix 4.

2.2.6. SEC-MALS and mass photometry

The samples of WT AfAgo or AfAgoΔ in the absence of nucleic acids (final protein concentration 1.0 mg/mL or 20 μM in terms of monomer, injected volume 2.0 mL) were separated using a HiLoad 16/60 Superdex 200 preparatory grade column (GE Healthcare). The column was equilibrated with a buffer containing 100 mM Tris-HCl (pH 8.0 at 25 °C), 1000 mM NaCl and 0.04% (w/v) NaN_3, the flow rate was 0.5 mL/min. Samples of AfAgo with 5′-phosphorylated MZ-1289 DNA (final concentration 5 μM protein monomer and 5 μM DNA oligoduplex) were separated on a Superdex 200 Increase 10/300 GL column equilibrated with a buffer containing 15 mM Tris-HCl (pH 8.0 at 25 °C), 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM DTT and 0.04% (w/v) NaN_3, the flow rate was 0.4 mL/min. The light scattering signals were monitored on a miniDawn® TREOS® II detector, concentrations of apo-protein samples were measured using an Optilab® T-rEX refractive index detector (Wyatt Technologies), concentrations of protein-DNA complexes were measured using both refractive index and UV absorption (Waters 2487 UV detector) readings. Data were analysed in Astra software (Wyatt Technologies) using dn/dc values of 0.185 g/mL and 0.170 g/mL for protein and DNA, respectively. Scattering data of protein-DNA complexes were analysed using the "protein conjugate" method in Astra; the required DNA and apo-protein UV extinction coefficients were determined experimentally using the "UV extinction from RI peak" method and the DNA-only and apo-protein samples.

Mass photometry measurements of fAfAgo were performed by Tomás de Garay at Refeyn Ltd, using the Refeyn OneMP system. Before measurement, protein stock solutions were diluted to 20 nM in a buffer, containing 20 mM Tris-HCl (pH 8.0 at 25 °C) and 500 mM NaCl.

2.2.7. Small-angle X-ray scattering

Small-angle scattering data were collected at the P12 EMBL beamline on the PETRA III ring of the DESY synchrotron in Hamburg, Germany[174], equipped with a Pilatus6M ("Dectris") detector located at 3 m distance. The X-ray wavelength was 0.124 nm. Data collection and principal structure parameters are detailed in Appendix 5. WT AfAgo and AfAgoΔ complexes with DNA (MZ-1289) were transferred to sample buffer (20 mM Tris-HCl (pH 7.5 at 25 °C), 5 mM $MgCl_2$, 150 mM NaCl and 2 mM DTT) using Illustra NAP columns (GE Healthcare). The complex of AfAgo with

5′-phosphorylated oligoduplex MZ-1288 and AfAgo-N was prepared as described in section 2.2.5 and concentrated to 59 μM. The AfAgo+MZ-1289 was concentrated to 175 μM. 100 μl of the sample was applied on the Superdex200 Increase 10/300 column (GE Healthcare) equilibrated with the "Low salt" buffer (20 mM Tris-HCl (pH 7.5 at 25 °C), 5 mM $MgCl_2$, 150 mM NaCl, 2 mM DTT). SEC run was performed using the Wyatt-MALLS-DLS system (Agilent, Wyatt[175]) directly connected to the P12 beamline. Frames collected throughout the complete SEC run (flow rate of 0.5 ml/min, 3000 frames, 0.995 s each) were analysed with CHROMIXS[176], and frames corresponding to the peak were averaged with ATSAS v.2.8.4 software [177] and converted into absolute scale by PRIMUS v. 3.0.2 (r12592)[178]. Samples of similarly prepared scfAfAgo complexes with 14 bp (MZ-1288) and 11 bp (MZ-864) were concentrated in a stepwise manner to concentrations of 1 – 3.7 mg/ml in the "Low salt" buffer. A sample of apo scfAfAgo was measured in the same buffer with 0.5 M NaCl. All samples were centrifuged at the maximal speed before data collection. The capillary of the automated sample changer ("Arinax") used in batch measurements was held at 20 °C. The SAXS data are presented in Figure 19, Figure 20, Figure 35b, c, and Figure 37, Table 2. 40 frames (0.05 sec each) were collected from each sample and 80 frames of the corresponding buffer were processed by automatic beamline procedure [174] and converted into absolute scale[177]. Ab initio shape determination was carried out by generating 20 independent DAMMIF v.3.0.2 (r12592) [179] and GASBOR v.2.3i (r12592) models using parameterized scattering curves created by GNOM v.5.0 (r10552)[180]. Models were clustered by DAMCLUST [181] and models forming a cluster were averaged by DAMAVER [182] and used as a starting model for an additional run of DAMMIN[183].

SAXS measurements performed with a range of AfAgoΔ concentrations (1 – 10 mg/ml) showed significant protein aggregation. The pseudo-chain dummy residues models of the complex generated by GASBOR [184] were superimposed with crystallographic dimers of AfAgo as well as with the monomeric AfAgo-DNA complex using SUPCOMB [185] applying stepwise shift (5 Å) along the principal axis of the model as described by Tamulaitis *et al.*, 2014[186].

The SAXS data were compared to crystal structures using CRYSOL v.2.8.3[187]. Particle volume and M_w estimations were performed using several methods (Table 2 and references therein).

2.2.8. Preparation of nucleic acid substrates

Synthetic DNA and RNA oligonucleotides (Appendix 1) were purchased from Metabion. For EMSA experiments, DNA and RNA oligonucleotides were 5′-labelled using [γ^{32}P] ATP (Perkin Elmer) and PNK (ThermoFisher). For unlabelled 5′-phosphorylation, regular ATP (ThermoFisher) was used. Further, duplexes were prepared by briefly heating a mixture of complementary oligonucleotides at a molar ratio of 1:1.2 to 95 °C and annealing to room temperature over 4 hours: MZ-1480 to MZ-1481 for dsRNA, MZ-1447 to MZ-1455 for dsDNA, and MZ-1480 to MZ-1455 for RNA/DNA heteroduplex.

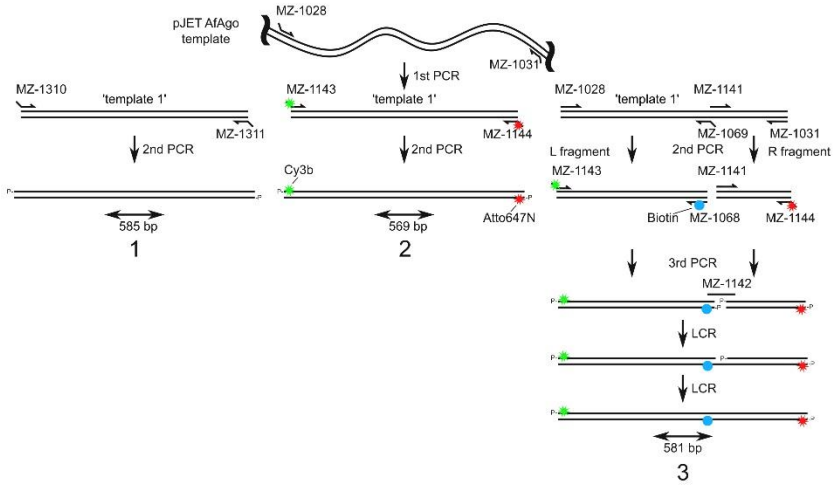


Figure 14. Synthesis scheme for the DNA fragments for single-molecule experiments. Reproduced from Golovinas *et al*., 2021[145].

The synthesis scheme for DNA fragments intended for smFRET and AFM studies is outlined in Figure 14. For smFRET, a 569 bp and a 581 bp DNA constructs were designed, which were labelled with a pair of FRET fluorophores, Cy3B and Atto647N, each attached to thymine bases 3 nt away from the respective DNA termini *via* C6 linkers (Figure 14). Positions of the FRET labels were selected such that upon binding of both DNA ends by an AfAgo dimer, the distance between the label attachment sites (irrespective of the AfAgo dimerization mode), would be favourable for FRET (Figure 15), and that attached labels do not interfere with AfAgo binding to DNA (Figure 16).

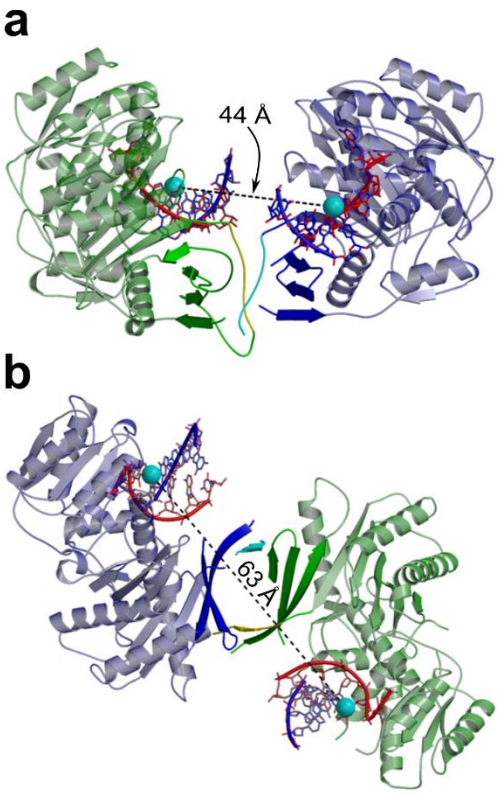


Figure 15. Expected positions of fluorescent labels upon formation of the looped complex. The figure is based on PDB ID 1YTU (a, "closed"), 2W42 (b, "open"), spheres mark fluorophore attachment sites. Protein monomers are coloured green and blue, and DNA guide and target strands are red and blue respectively. Reproduced from Golovinas *et al*., 2021[145].

First, a DNA fragment was amplified from a pJET plasmid template containing an AfAgo gene fragment using oligonucleotides MZ-1028 and MZ-1031. The PCR product was then used as a template (dubbed "template 1") in subsequent reactions. Fragment "1" used for AFM studies was made by PCR from "template 1", using oligonucleotides MZ-1310 and MZ-1311, which were treated with T4 polynucleotide kinase (PNK) prior to amplification, to yield a 585 bp fragment. Fragment "2" was amplified from "template 1" with oligonucleotides MZ-1143 and MZ-1144, bearing Cy3B (green star) and Atto647N (red star) dyes, respectively, on the third base from the 5′-end, yielding 569 bp DNA. Fragment "3" was synthesised in two steps. Firstly, respective fragments flanking the biotinylation site (dubbed "L fragment" and "R fragment") were amplified by PCR from "template 1", using primer pairs MZ-1028 and MZ-1069 for the "L fragment", and MZ-1031 and MZ-1141 for the "R fragment". Secondly, each of the two fragments was used

as templates for subsequent PCRs. “L fragment” was amplified using MZ-1143 and MZ-1068, the latter bearing the biotin (blue circle) on 22 bases from its 5′-end. “R fragment” was amplified using primers MZ-1141 and MZ-1144. The two fragments were then purified using a GeneJET PCR purification kit (ThermoFisher Scientific) and treated with PNK while mixed in equal amounts to a total concentration of 6 mM. The phosphorylation mix was subsequently ligated by Ampligase® (Lucigen). All full-length DNA fragments were subsequently purified from an agarose gel using a runVIEW system (Cleaver Scientific, UK), precipitated with sodium acetate/isopropanol, washed with 75% (v/v) ethanol, and resuspended in water. The binding of the resulting DNA was verified using EMSA, labelling the DNA with 5′-^{32}P using PNK (Figure 16). The DNA concentration in the binding reactions was 1 nM, and the final AfAgo concentrations were 0, 0.2, 0.5, 1, 2, 5, 10, and 50 nM. The binding buffer was 40 mM Tris-acetate (pH 8.4 at 25 °C) with 1 mM EDTA (TAE, Invitrogen cat#24710-030), supplemented with 5 mM $Mg(OAc)_2$, 0.1 mg/ml BSA, 1 mM DTT, 10% (v/v) glycerol. Running buffer – TAE supplemented with 5 mM $Mg(OAc)_2$.

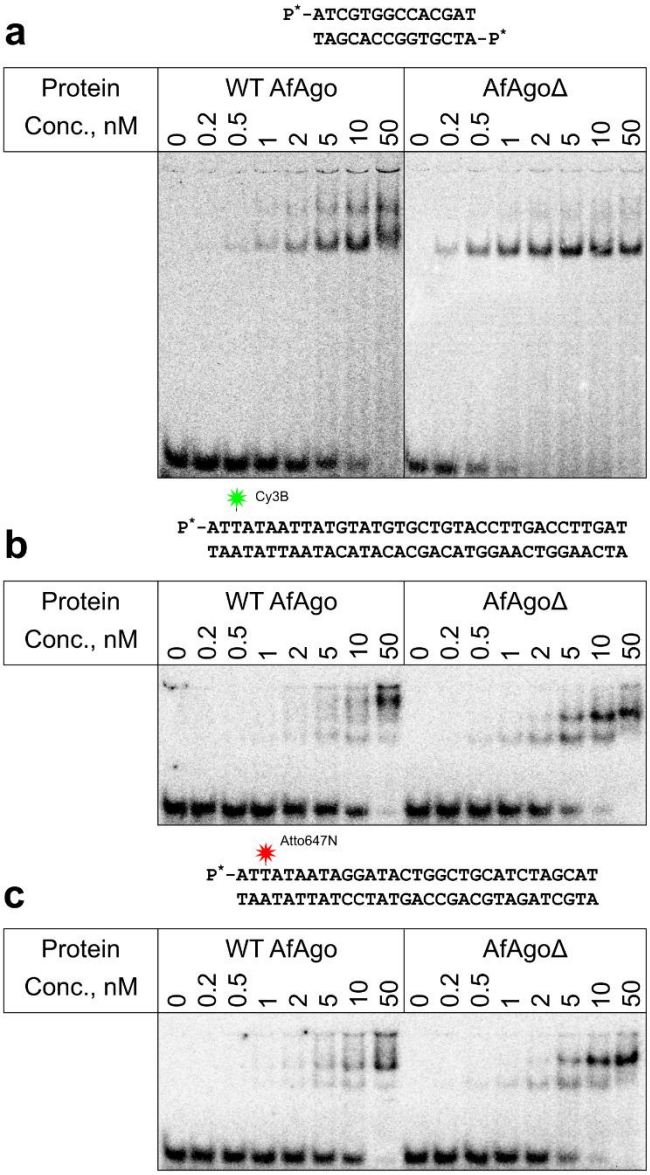


Figure 16. AfAgo binds the synthetic smFRET and SAXS substrates. 5′-^{32}P-labelled DNA substrates were: a self-complementary oligoduplex MZ-952 (panel a); a Cy3B-modified oligoduplex MZ-1443/MZ-1026 (panel b); an Atto647N-modified oligoduplex MZ-1144/MZ-1027 oligoduplex (panel c). The fluorescently modified oligoduplexes carried the 5′-^{32}P label only on the modified strand. The DNA concentration in the binding reactions was 1 nM, final protein concentrations are shown above each lane. Reproduced from Golovinas *et al*., 2021[145].

2.2.9. EMSA experiments

For EMSA, nucleic acid-free AfAgo was diluted to the 2× final concentration in binding buffer (TAE, Invitrogen), containing 40 mM Tris-acetate (pH 8.4 at 25 °C), 1 mM EDTA, supplemented with 100 mM KOAc, 0,1 mg/ml bovine serum albumin, 1 mM DTT, and 10% glycerol. Nucleic acid substrates used for binding studies were MZ-1480 and MZ-1698 – 1708 for ssRNA and MZ-1447 for ssDNA. These were diluted such that the final NA-protein mixture would contain 1 nM 5′-^{32}P-labelled and 4 nM unlabelled 5′-P-NA. The binding reaction mixture was incubated for 10 minutes at room temperature (25 °C) and loaded onto an 8% PAAG (29:1 acrylamide/bis-acrylamide) prepared with TAE buffer, supplemented with 100 mM KOAc. Additional experiments with MZ-1480 and MZ-1707 were conducted by incubating the binding reaction mixture for 10 minutes at 70 °C. Electrophoresis was run at room temperature in all cases. To study the RNA-guided nucleic acid targeting mechanism, AfAgo was pre-mixed with MZ-1480, a 5′-P-ssRNA guide, at 0.4:0.8 μM ratio of AfAgo:guide, incubated for 10 min at room temperature and diluted to 2× final binding reaction concentration in the same binding buffer as above. Diluted target NAs were added to the reaction mixture at a 1:1 volumetric ratio to a final reaction concentration of 5 nM (1 nM 5′-^{32}P + 4 nM 5′-P), the mixture was incubated for 10 min at room temperature and analysed as described above. Target NAs used were MZ-1556 and MZ-1557 as 8- and 4-nucleotide complementary RNA targets, respectively, and MZ-1560 and MZ-1561 as analogous DNA targets, respectively. For the heparin-supplemented reactions, heparin was pre-mixed with the target NAs before adding them to the binding reaction mixture so that the final heparin concentration in the binding reaction was 100 ng/μl.

For (sc)fAfAgo, EMSA reaction mixtures were prepared in 1× TAE buffer, supplemented with 5 mM $Mg(OAc)_2$, 0,1 mg/ml BSA, 1 mM DTT, 10% glycerol and contained 0.01 nM of 5′-^{32}P labelled oligonucleotide substrate and increasing concentrations of protein: 0, 0.05, 0.1, 0.2, 0.5, 1, 2, 5 nM for single-stranded substrates or 0, 0.1, 0.5, 1, 2, 5, 10, 20 nM for double-stranded substrates. To assess how binary pAgo:guide complex recognizes various targets, the binary complex was formed by mixing 400 nM of pAgo with 800 nM of 5′-P guide and incubating at room temperature (25° C) for 10 min. Binding reactions were prepared in the same buffer and contained 0.01 nM of 5′-^{32}P labelled target as well as increasing concentrations of the binary

complex: 0, 0.005, 0.01, 0.02, 0.1, 0.5 nM. Control mixes were prepared by mixing either 0.5 nM of pAgo without the guide (C_E) or 1 nM of guide (C_g) with 0.01 nM of radiolabelled target. Reaction mixtures were incubated at room temperature for 1 h prior to resolution on 8% polyacrylamide gels (29:1 acrylamide/bis-acrylamide in 1× TAE supplemented with 5 mM $Mg(OAc)_2$). Electrophoresis was performed in electrophoresis buffer (1× TAE, 5 mM $Mg(OAc)_2$).

Radiolabelled substrates were detected and quantified using a phosphor screen laser scanner Typhoon (Amersham). The results were analysed with OptiQuant 03.00, ImageQuant TL v8.2.0.0, and OriginPro 8.1 software. The K_d was calculated from the following equation:

$$S_{NB} = A1 + \frac{\frac{100}{S_0}\left(S_0\frac{100-A1}{100} - E_0 - K_d + \sqrt{\left(S_0\frac{100-A1}{100} + E_0 + K_d\right)^2 - 4S_0E_0\frac{100-A1}{100}}\right)}{2} \quad (Eq.1)$$

where S_{NB} – unbound substrate, nM; S_0 – initial substrate concentration, nM; E_0 – initial protein complex concentration, nM; K_d – dissociation constant, A1 – nonbinding fraction of substrate, %.

2.2.10. Single-molecule FRET setup

The setup used in single-molecule experiments for fluorescence data acquisition has been described in depth by Golovinas *et al*. [145] and the principal optomechanical layout of the experiment is shown in Figure 17. The experimental technique for fluorescence burst data acquisition of single diffusing molecules using ALEX (alternating laser excitation) was based on Kapanidis *et al*.[188]. A custom single-molecule fluorescence microscopy setup built around a commercially available Nikon Eclipse Ti-U inverted microscope, equipped with 60× 1.2 WI Plan Apo VC objective (Nikon) for both excitation and signal detection, was used. Excitation was carried out with 25 mW 532 nm and 635 nm diode-pumped solid-state and diode lasers (Crystalaser), respectively and two APD-based single-photon counting modules (Tau-SPAD-50, PicoQuant) were used for signal detection. Laser excitation light was reflected off a dichroic mirror (zt532/635rpc-XT, Chroma), and a quadruple-band interference filter (FF01-446/510/581/703, Semrock) was used to filter off the fluorescence signal. ALEX was implemented by modulating the intensity of the 532 nm laser with a mechanical chopper (MC200B, Thorlabs) and synchronously directly TTL-modulating the intensity of the 635 nm laser. ALEX half-period was 50 μs.

An FPGA module (PCIe-7851R, National Instruments) and a custom Labview (National Instruments) program were used for recording fluorescence photon arrival times and implementing ALEX.

For diffusing molecules, laser excitation was focused 50 μm above the sample chamber glass surface. Excitation was carried out with intensities of 30 μW for the 532 nm, and 20 μW for the 635 nm beam. The size of the confocal pinhole was 75 μm. The duration of each measurement was 10 min.

Measurements of single surface-immobilized molecules with the excitation in the TIR mode were performed by exploiting an alternative functionality of the same setup, as described previously[189]. In short, a 100× 1.4 Oil Plan Apo VC (Nikon) objective was used on the microscope, and the fluorescence signal was split by a dichroic mirror (T640lpxr-UF2, Chroma), while the different spectral channels were projected on the same EMCCD (DU-897ECS0-UVB, Andor).

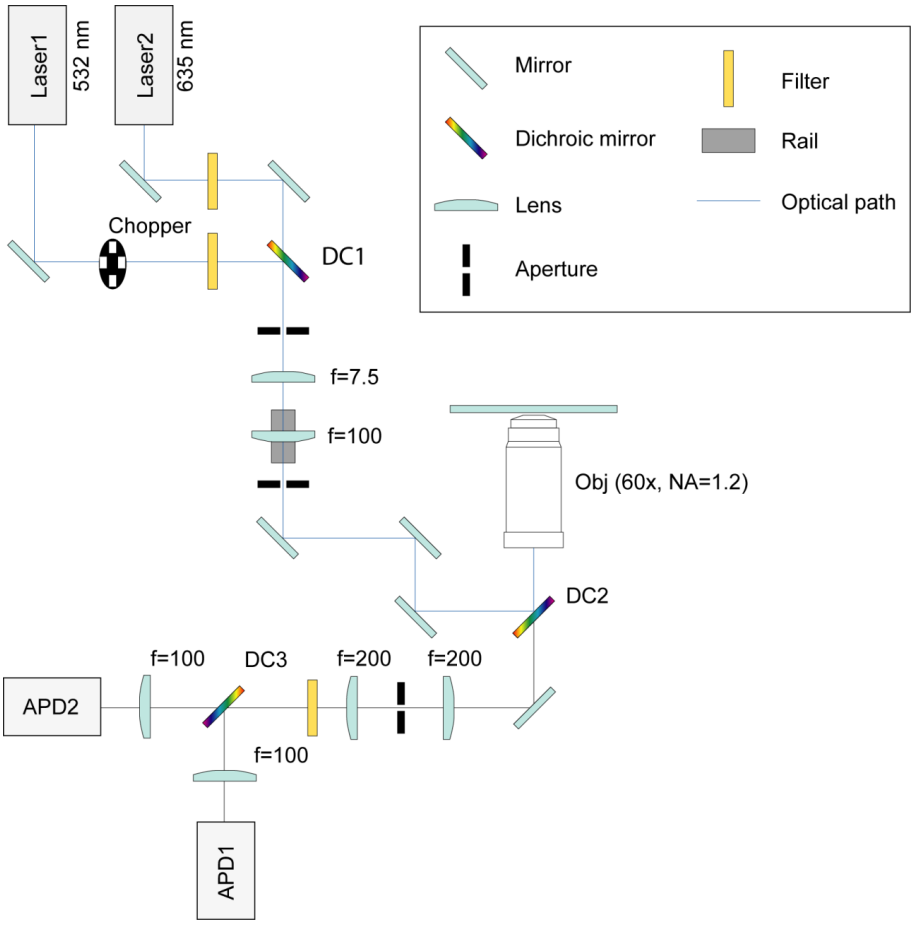


Figure 17. Optical scheme of custom single-molecule fluorescence microscopy setup used to record fluorescence bursts of single diffusing molecules in this study. APD – avalanche photodiode; f – focal distance; NA – numerical aperture. Reproduced from Golovinas *et al*., 2021[145].

2.2.11. Sample cell preparation

Measurements of FRET bursts were performed in a chambered coverglass well (155411, Nunc™ Lab-Tek™, Thermo Scientific). The binding reaction volume was 200 µl. The reaction buffer for AfAgo and AfAgoΔ (RB1) was 33 mM Tris-acetate (pH 7.9 at 20 °C), 66 mM KOAc, 5 mM $Mg(OAc)_2$, and 0.1 mg/ml BSA. The reaction buffer for scfAfAgo measurements (RB2) was 40 mM Tris-acetate (pH 8.4 at 20 °C), 5 mM $MgOAc_2$, 0.1 mg/ml BSA. DNA concentration of 17 – 50 pM was used. Measurements at different protein concentrations were carried out by adding small volumes of protein diluted in RB into the reaction. No oxygen-scavenging or triplet-quenching additives

were used. All serial dilutions of proteins and DNA were carried out in Protein LoBind™ 1.5 ml tubes (Eppendorf cat#0030108116).

Measurements of surface-immobilized DNA fragments were performed in a flow cell assembled from a six-channel Sticky-Slide VI 0.4 (Ibidi) and a coverslip functionalized with polyethene glycol derivatives as described in detail in[189]. The flow cell was incubated with 5 µg/ml of Neutravidin (Molecular probes) in RB for 2 min, washed with RB, incubated with 5 pM DNA in RB until the density of the surface-immobilized DNA fragments appeared to be appropriate, and washed with RB. For the measurement, a 20 nM solution of AfAgo in RB supplemented with 1% (w/v) glucose (TCI Europe), 2.5 mM Trolox (Sigma-Aldrich cat#648471), and 15 U/ml glucose oxidase (Sigma-Aldrich) was injected into the cell. Trolox was pre-treated with UV light for 20 min according to Cordes *et al.*[190]. For surface-immobilized molecule measurements of RNA-guided DNA targeting by (sc)fAfAgo, the cell was prepared as above, using RB2. A slight difference was that for (sc)fAfAgo, the cell was injected with 200 µl 10 pM pre-annealed biotinylated anchor ssDNA/target ssDNA duplex (MZ-1656/MZ-1715 for anchor/8 bp-complementary target and MZ-1656/MZ-1752 for anchor/non-complementary target) and flushed immediately with RB2. (sc)fAfAgo-guide complexes were formed by incubating a mixture of protein and MZ-1655 gRNA solutions in RB2 at 1 µM for 15 minutes at 25 °C and were subsequently diluted in RB2 to a working concentration of 10 nM. Measurements were conducted as described above, using a 10 nM solution of (sc)fAfAgo-gRNA in RB2 supplemented with 1% (w/v) glucose, 2.5 mM Trolox, and 15 U/ml glucose oxidase before injecting into the flow cell.

2.2.12. Single-molecule FRET data analysis

Fluorescence burst analysis was performed using the freely available FRETBursts software[191]. The initial burst search parameters were m = 10 photons, and F = 6 times the fluorescence background. The total intensity of a burst from both channels and excitation wavelengths was thresholded to be larger than 40 counts, and this yielded ~3000 bursts from a 10-minute measurement. Each burst was calculated a proximity ratio, E, according to $E = I_d^a/(I_d^a + I_d^d)$, where I_d^a and I_d^d are acceptor and donor intensities upon donor excitation, respectively, and a stoichiometry parameter, S, according to $S = I_d/(I_d + I_a^a)$, where I_d is the total donor and acceptor intensity upon donor excitation, and I_a^a is acceptor intensity upon acceptor excitation. Then 2D E-S histograms of bursts were built. Subsequently, bursts with the stoichiometry

parameter range of $0.2 - 0.9$ were selected to build distributions of the proximity ratio, E, of bursts of DNA molecules labelled with both fluorophores only. E histograms were fit with the sum of two Gaussian functions using unconstrained optimization. Then the ratio of the number of looped and unlooped DNA molecules in the ensemble was calculated as the ratio of the area of the Gaussian of high E with that of low E.

To quantify the looped state duration the E trajectories were idealized using HMM with a two-state model in QuB software[192]. Then, the cumulative histogram of the looped state durations was built from the idealized trajectories. The trajectory edge dwells were not omitted in order to preserve the information on the occurrence of states lasting during the whole trajectory. The exponential factor of a single-exponential fit of the cumulative histogram was 33 ± 1 s. The maximum recorded looped state duration is, however, limited by the duration of the measurement (200 s) and the duration of the fluorescent state of the fluorophores before photobleaching. The value of the exponential factor thus sets the lower limit of the looped state duration.

The experiment of surface-immobilized DNA fragments was done by first recording a short movie with 635 nm excitation to obtain a reference for fluorescent spot identification since the acceptor channel exhibits significantly less fluorescence background than the donor channel. Then a longer actual movie was recorded with the 532 nm excitation. The analysis of the two-spectral channel fluorescence movies was performed using custom software written in Matlab. Briefly, to identify the fluorescent spots, the first 20 frames of the reference and the actual fluorescence movies were averaged, the obtained average images were filtered with the 2D low-pass Gaussian filter 5 pixels large and with the standard deviation of 1 pixel and subtracted the same image filtered with the averaging filter 7 pixels large. The resulting acceptor channel reference image was thresholded with 20 and the donor channel actual image - with 40 counts/pixel. The obtained images were binarized for particle identification. Particles in both binary images were identified and filtered according to the following criteria: 5×5 pixel ROIs (regions of interest) centred on particles' centres of mass had to be non-overlapping, particle area had to be within the range of $5 - 100$ pixels, particle eccentricity not larger than 0.8. The coordinates of a particle in the donor channel corresponding to a particle identified in the acceptor channel of the reference movie were calculated using the spatial transformation structure calculated from an image of surface-immobilized 200 nm fluorescent polystyrene beads (Invitrogen

cat#F8806). For trace extraction, only those particles in the actual movie were considered whose donor coordinates coincided with the transformed coordinates of the acceptor particles in the reference movie within 1.5 pixels. The donor and acceptor intensity traces were extracted using aperture photometry [193] with the background calculated as an average intensity from a 1-pixel-wide annulus around the particle's ROI. The proximity ratio E was calculated according to the same formula as for the fluorescence bursts.

2.2.13. Atomic force microscopy sample preparation and imaging

DNA-protein complexes were formed by incubating the DNA fragment (5 nM) with WT AfAgo, AfAgoΔ (50 nM in terms of monomer) or fAfAgo (50 nM in terms of heterodimer) for 5 min (WT AfAgo and AfAgoΔ) or 20 min (fAfAgo) at room temperature in the Binding Buffer HEPES (33 mM HEPES (pH 7.8 at 25 °C), 66 mM KOAc, 5 mM $Mg(OAc)_2$) in a total volume of 50 µl. Next, the protein-DNA complexes were cross-linked with 2.5% (w/v) glutaraldehyde for 20 min. Glutaraldehyde was then quenched with an excess of the Tris buffer (33 mM Tris-acetate (pH 7.8 at 25 °C), 66 mM KOAc, 5 mM $Mg(OAc)_2$) for WT AfAgo and AfAgoΔ, quenching was omitted for fAfAgo. The resultant reaction solution after 10-fold dilution with Tris buffer was deposited onto modified mica at room temperature as described below.

As a substrate for DNA deposition, freshly cleaved muscovite mica (grade IV, SPI supplies Inc., USA) was incubated in 0.17 mM APS solution (1-(3-aminopropyl)-silatrane 2,8,9-trioxa-5-aza-1-silabicyclo [3.3.3] undecane) for 30 min to prepare functionalized APS-mica, as described previously for the preparation of protein-DNA complexes[194]. 50 µl of DNA-protein complex solution was deposited on APS-mica for 5 min. After incubation the mica surface was immersed in deionized water for 5 min, flushed with excess water, and then dried under a flow of nitrogen. The images were acquired in the air with the Dimension Icon® (Bruker, Santa Barbara, CA) microscope system in tapping mode. Probes with nominal spring constants of ~2 – 40 N/m were used. Typically, the images were collected at a speed of 0.6 Hz and a resolution of 1024 × 1024 pixels, scan size of 2 µm × 2 µm. The image analysis was performed with WSxM (5.0) and NanoscopeAnalysis (1.9) software packages. The protein-DNA complexes were selected with no effect on DNA length. The theoretical length of 585 bp DNA is 195 nm. The recorded DNA length is 192.6 ± 5.9 nm. Thus, only the structures that fulfilled this requirement underwent further analysis.

2.2.14. AFM data analysis

The stoichiometry of WT AfAgo and AfAgoΔ was assessed by volumetric analysis. The protein molecular volume was determined by measuring the height and half-height diameters of two perpendicular cross-sections. The particle was treated as a spherical cap and the volume of each protein particle was calculated according to Equation 2:

$$V_m = \left(\frac{\pi h}{6}\right)(3r^2 + h^2) \qquad (Eq.2)$$

where h is the particle height and r is the radius at the half height[195]. Molecular volume based on molecular weight was calculated using Equation 3:

$$V_c = \left(\frac{M_0}{N_0}\right)(V_1 + dV_2) \qquad (Eq.3)$$

where M_0 is the molecular mass of the protein, N_0 is Avogadro's number, V_1 and V_2 are the partial specific volumes of protein and water (0.74 cm^3g^{-1} and 1 cm^3g^{-1}, respectively), and d is the extent of protein hydration (0.4 mol water/mol protein)[196]. Thus, the calculated volume of WT AfAgo (50.8 kDa) and AfAgoΔ (49.9 kDa) proteins is ~100 nm^3. For analysis, the measured volume data was divided into three populations by their theoretical volume: *monomer* (<150 nm^3), *dimer* (150 – 250 nm^3), *higher-order* (>250 nm^3).

2.2.15. Phylogenetic analysis

AfAgo close homologues were collected using a standard BLAST search against protein sequences in the NCBI Reference Sequence (RefSeq) database (https://www.ncbi.nlm.nih.gov/refseq/). Sequence redundancy was reduced by clustering homologues having 90% or more sequence identity with at least 90% alignment coverage using MMseqs2[197]. Cluster representatives were aligned using MAFFT [198] and those having only fragments within the MID-PIWI region were removed. Following this procedure, 200 sequences (AfAgo and its homologues) were selected for phylogenetic analysis. TtAgo (Long-A pAgo) homologues were collected in the same way and 20 sequences were selected. Joint MSA from both AfAgo and TtAgo homologues (220 sequences) was constructed using an accuracy-oriented MAFFT mode (L-INS-i) and only the MID-PIWI region of the aligned sequences was

retained. The resulting MSA was processed with trimAl [199] to remove columns with excessive gap content followed by the phylogenetic tree inference using FastTree2[200]. The obtained pAgo tree was rooted at the midpoint and annotated using iToL v5[201].

2.2.16. Gene context analysis

To identify putative pAgo operons, the same proteins used for phylogenetic analysis were subjected to the gene neighbourhood analysis using WebFlags[202]. The analyzed neighbourhood consisted of three genes upstream and three genes downstream of the pAgo gene. Proteins encoded by the upstream or downstream genes were further characterized based on their sequence and/or structure.

2.2.17. Protein sequence and structure analyses

Sequence-based homology detection was performed using HHsearch[203]. Potential transmembrane regions were predicted with DeepTMHMM[204]. For structure-based analysis, AlphaFold models were either obtained from the EBI database [205,206] or constructed using the ColabFold implementation of AlphaFold[207]. Structurally related proteins in PDB were identified by searches with Dali [208] and FoldSeek[209]. Structures were analyzed with ChimeraX[210].

3. RESULTS

3.1 AfAgo dimerization

"The rocks crashed and still the Argonauts rowed as the aftershock of waves tossed them up and forward, further out of reach. Jason stood up and let out a barbaric hoot of triumph. <…> The Symplegades never clashed again."[4]

- Stephen Fry, *Heroes*

[4] Image: engraving, *Jason and the Argonauts Sail Through the Symplegades*. "Tableaux du temple des muses", Cornelis Bloemaert II, 1655.

3.1.1. Available crystal structures show a dimeric AfAgo

Inspection of available AfAgo structures revealed that in all structures known so far, AfAgo subunits form substantial homodimerization interfaces, which were not further scrutinized by authors of the corresponding structural studies. The dimerization interface in the AfAgo-dsRNA structure (PDB ID 1YTU) is asymmetric and primarily involves the C-terminal β-strands (residues 296 – 303) from both subunits present in the asymmetric unit that together form a parallel β-sheet and the N-terminal residues from one of the subunits (Figure 18a). The dimer formed in this case is compact (henceforth, a "closed" dimer). In contrast, dimerization interfaces in three other cases (PDB IDs 1W9H, 2BGG, and 2W42) are nearly symmetrical with respect to the secondary structure elements involved (albeit in PDB IDs 2BGG and 2W42 they belong to different protein chains present in the asymmetric unit): the C-terminal β-strands form 8-strand β-barrels, with the sheets from different subunits interacting *via* strands β14 (residues 297 – 302) and β15 (residues 314 – 318, Figure 18b). The resultant dimers are less compact (henceforth, "open" dimers).

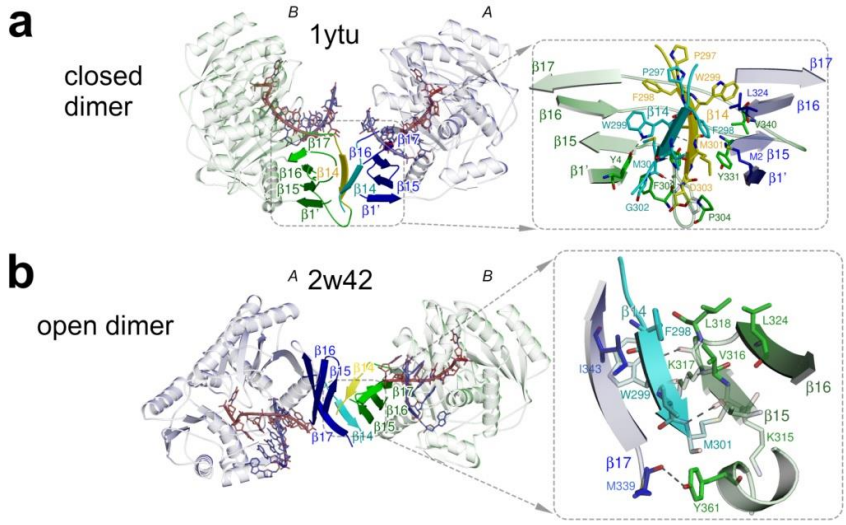
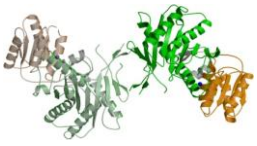
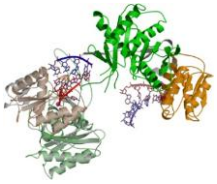
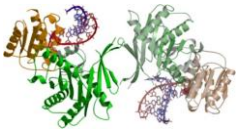


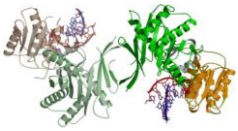
Figure 18. Dimerization of AfAgo. (a-b) Protein subunits are coloured blue (protein chain A) and green (protein chain B). The interface-forming secondary structure elements are highlighted and numbered according to the PDB ID 2W42 assignment made by PDBsum[211]. The "guide" DNA/RNA strands bound by AfAgo are coloured red, "target" strands – blue. Residues 296 – 303 deleted in AfAgoΔ are coloured cyan and yellow. Hydrogen bonds are shown as dashed lines. (a) AfAgo complex with dsRNA (PDB ID 1YTU, both protein chains as present in the asymmetric unit), the "closed" dimer[25]. (b) AfAgo complex with dsDNA (PDB ID 2W42) [24] – the "open" dimer. β-strands from both subunits assemble into a closed

β-barrel structure, with an intersubunit interface formed by β14 and β15 strands of neighbouring subunits. Reproduced from Golovinas *et al*., 2021[145].

The solvent-accessible surface areas buried at the dimerization interfaces in both "open" and "closed" dimers are classified as "significant" by the PISA server (https://www.ebi.ac.uk/pdbe/pisa/pistart.html, [212]; Table 1). This observation prompted us to test the oligomeric state, the possible dimerization mode, and the mechanism of nucleic acid binding of AfAgo in solution using various biochemical and biophysical techniques. For that purpose, two variants of AfAgo were used: the full-length wild-type protein (henceforth, WT AfAgo), and a dimerization mutant AfAgo lacking the 296 – 303 amino acid residues responsible for the majority of dimerization contacts in both the "closed" and "open" homodimers (Figure 18a, b).

Table 1. Dimerization interfaces as analyzed by PISA (PDBe PISA v1.52 [20/10/2014]). *The dimerization interface in PDB ID 1W9H is essentially identical to interfaces in PDB IDs 2BGG and 2W42. The lower CSS score arises from PISA giving lower scores to interfaces generated by symmetry operators (as in the case of PDB ID 1W9H, which contains a single AfAgo subunit per asymmetric unit) than to interfaces formed between different subunits present in the asymmetric unit (the dimers in PDB IDs 2BGG and 2W42 are formed by 2 AfAgo monomers present in the asymmetric unit). Reproduced from Golovinas *et al*., 2021[145].

PDB ID	Dimer: open/ closed	Image	CSS Complex Formation Significance Score	ΔiG P-values	PISA: dimerization surface, Å^2 (buried in the interface)
1W9H	open		0.108 *	0.004	731
1YTU	closed		1	0	908
2BGG	open		1	0.001	601

PDB ID	Dimer: open/ closed	Image	CSS Complex Formation Significance Score	ΔiG P-values	PISA: dimerization surface, Å^2 (buried in the interface)
2W42	open		1	0.002	748

3.1.2. SEC-MALS data

First, the oligomeric state of WT AfAgo and AfAgoΔ proteins was tested using size exclusion chromatography – multi-angle light scattering (SEC-MALS). It was found that WT AfAgo elutes from the SEC column as a polydisperse peak (Figure 19a), with the M_w values ranging from 91.7 kDa at the left-hand side of the peak (close to the theoretical M_w of WT AfAgo homodimer, 101.6 kDa) to 59.5 kDa at the right-hand side (still considerably higher than M_w of a WT AfAgo monomer, 50.8 kDa). AfAgoΔ formed a far broader irregular peak, covering M_w values from 87 kDa (close to M_w of AfAgoΔ dimer, 99.8 kDa) to ~49 kDa (close to AfAgoΔ monomer, 49.9 kDa). This allows the conclusion that WT AfAgo indeed forms homodimers, which under current experimental conditions (~10 μM concentration in the sample, ~1.5 μM concentration on the column) are relatively unstable and partially dissociate into monomers. Deletion of the 296-303 residues in the AfAgoΔ protein further decreased the stability of the dimer, in line with their proposed role in dimerization (Figure 19b). Intriguingly, the differences between WT AfAgo and AfAgoΔ oligomeric states were more pronounced in their DNA-bound forms (WT AfAgo-DNA and AfAgoΔ-DNA, respectively): the majority of WT AfAgo-DNA eluted as a 2:2 protein:DNA complex, while the major peak of AfAgoΔ-DNA matched a 1:1 protein:DNA complex (Figure 19**)**.

3.1.3. SAXS measurements

To characterize the conformation of AfAgo in solution, small-angle X-ray scattering (SAXS) measurements using WT AfAgo-DNA and AfAgoΔ-DNA complexes were performed. The data were analysed in two ways: (i) the *ab initio* shapes of the complexes in solution were calculated and superimposed

with the X-ray AfAgo structures, and (ii) the theoretical scattering data was calculated for the crystallized DNA-bound AfAgo monomer, "open" (PDB ID: 2W42 and 1W9H) and "closed" (PDB ID: 1YTU) dimers, and compared to experimental SAXS scattering data of AfAgo-DNA and AfAgoΔ-DNA (Figure 19). The "closed" AfAgo dimer fits the AfAgo-DNA SAXS data better than the "open" dimer, as judged from the real space fit and the χ^2 (Figure 19c) parameters, implying that in solution WT AfAgo predominantly forms a "closed" dimer. As expected, the AfAgo monomer gave the best fit to the AfAgoΔ-DNA SAXS data (Figure 19c, right column). The SAXS molecular weights calculated for WT AfAgo-DNA (between 94.2 and 106.9 kDa, Table 2) agreed with the expected mass of the dimer complexed with dsDNA (119 kDa). The SAXS M_w for the AfAgoΔ-DNA (between 55.4 and 67.9 kDa) confirmed its monomeric state.

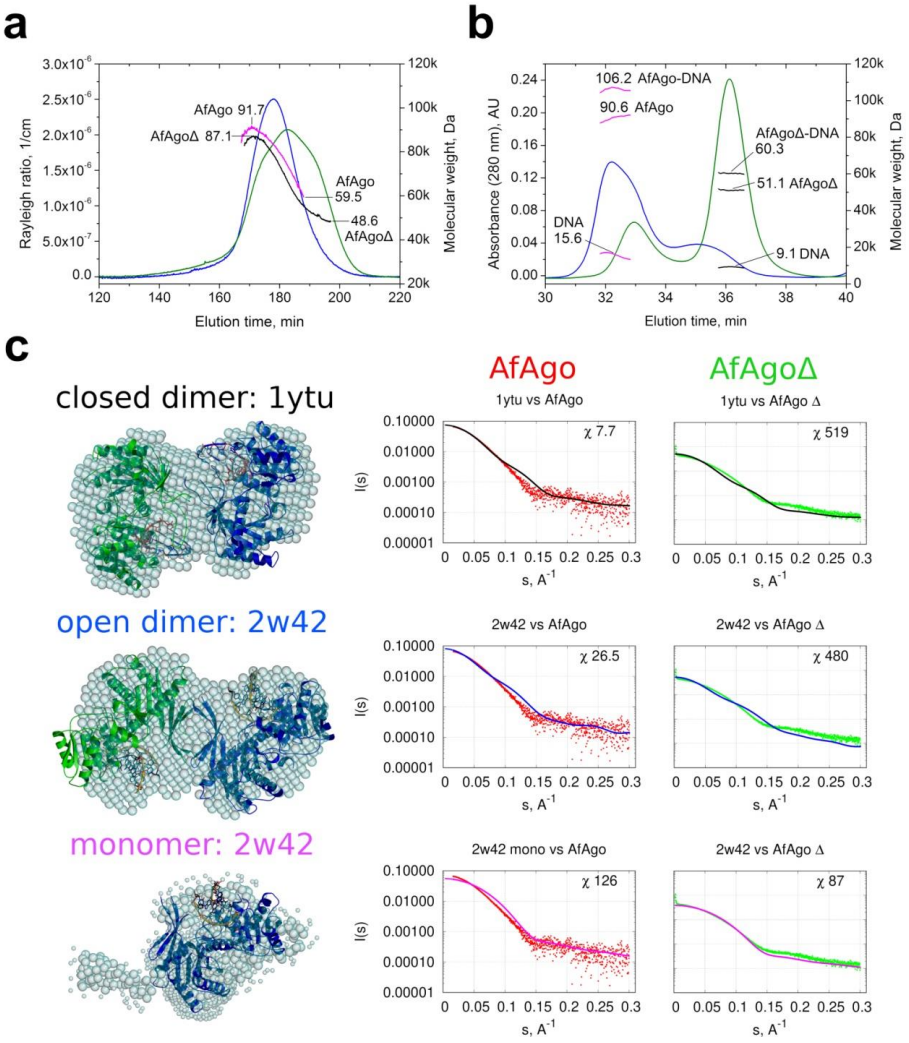


Figure 19. SEC-MALS and SAXS analysis of apo-AfAgo and AfAgo-DNA complexes. (a) SEC-MALS analysis of WT AfAgo and dimerization mutant AfAgoΔ unbound to nucleic acids. The light scattering data (blue for WT AfAgo, green for AfAgoΔ) is shown along with the calculated M_w values (magenta for WT, black for mutant). The highest and lowest M_w values calculated for each protein are indicated. Theoretical M_w of WT AfAgo monomer is 50.8 kDa, theoretical M_w of AfAgoΔ monomer is 49.9 kDa. (b) SEC-MALS analysis of AfAgo-DNA and AfAgoΔ-DNA complexes. The UV absorption data of AfAgo-DNA (blue) and AfAgoΔ-DNA (green) is shown along with the M_w values of full complexes, the protein component, and the DNA component (magenta for AfAgo-DNA sample and black for AfAgoΔ-DNA sample, respectively). The theoretical M_w of a 2:2 WT AfAgo:DNA complex is 119 kDa (2×50.8 + 2×8.7 kDa), theoretical M_w of a 1:1 AfAgoΔ-DNA complex is 58.6 kDa (49.9 + 8.7 kDa). (c) SAXS data of WT AfAgo complex with MZ-1289 DNA (red points) and the dimerization mutant AfAgoΔ with MZ-1289 DNA (green points) are compared with the scattering curves generated from the "closed" dimer with dsRNA (PDB ID: 1YTU, black

curves), “open” dimer (PDB ID: 2W42, blue curves) and AfAgo-DNA complex (PDB ID: 2W42, magenta curves) by CRYSOL. Corresponding AfAgo structures are shown in the second column superimposed with the dummy atom models generated using the SAXS data of the AfAgo complex with MZ-1289 oligoduplex. Reproduced from Golovinas *et al*., 2021[145].

Table 2. Molecular mass determination from SAXS data using various methods. All molecular masses are given in kDa. Reproduced from Golovinas *et al*., 2021[145].

Sample			WT AfAgo+MZ-1289	AfAgoΔ+MZ-1289
Expected M_w (protein + DNA), kDa			119	58.6
Method	Reference	Sofware	MWcalc	
Absolute scale	[213]	PRIMUS 2.8.4 (r10552)	99.7	55.4
Qp			102.7	58.5
Bayes			94.2	56.9
Size&Shape			100.0	67.9
Porod volume/1.6	[181]	DATPOROD, ATSAS 2.8.4 (r10552)	98.8	67.9
SAXSMoW	[214]	SAXSMoW v2.1 http://saxs.ifsc.usp.br/	106.9 (integrated to I0/I(qmax)=102.25)	67.4 (integrated to I0/I(qmax)=102.25)
SEC MW		CHROMIXS ATSAS 2.8.4 (r10552)	103.8	n.a.

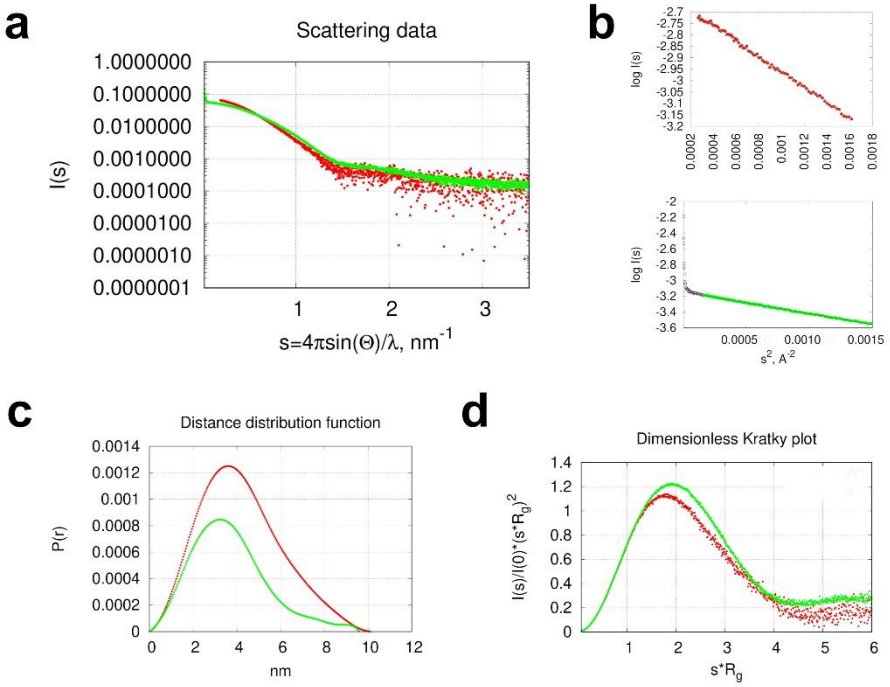


Figure 20. SAXS data of AfAgo+MZ-1289 (red curves) and monomeric mutant AfAgoΔ+MZ-1289 (green curves) complexes. (a) Scattering curves. (b) Guinier plots, log I(s) vs. s^2 of the data at small s values. (c) Pair distance distribution functions. (d) Dimensionless Kratky representation of scattering data $(I(s)/I(0)\times(s\times R_g)^2$ vs. $s\times R_g)$. All curves have similar shapes typical for folded proteins[215]. Reproduced from Golovinas *et al*, 2021[145].

3.1.4. Direct visualization of AfAgo-induced DNA loops by AFM

The dimeric state of WT AfAgo was observed in X-ray structures, SEC-MALS, and SAXS measurements, i.e., techniques that all require relatively high (micromolar and higher) protein and DNA concentrations. This raises a question if AfAgo dimerisation and the ability to simultaneously interact with two nucleic acid molecules are relevant in solution at far lower protein and DNA concentrations. To address this question, AFM was used to examine AfAgo interactions with long DNA molecules.

For direct visualization of protein-DNA complexes, AfAgo and DNA (a 585 bp blunt-end PCR fragment with 5′-phosphorylated termini) were deposited on APS-mica and imaged using tapping AFM. A typical AFM image of AfAgo-DNA complexes is shown in Figure 21. Several types of protein-DNA complexes, shown as enlarged insets in Figure 21, were observed: (i) linear DNA with a protein molecule bound to one DNA end; (ii)

linear DNA with protein molecules bound to both DNA ends; (iii) ring-shaped (looped) DNA. Other species, including naked DNA, or more complex structures, involving, e.g., protein bound to two DNA fragments, were also observed but were not quantified. Analysis of protein volumes in the AfAgo-DNA complexes revealed a broad distribution of sizes, ranging from ~60 nm^3 (lower value than expected for an AfAgo monomer, ~100 nm^3), to above 270 nm^3 (higher value than expected for an AfAgo dimer, ~200 nm^3), albeit the average particle size was considerably smaller in the case of the AfAgoΔ mutant (Figure 22).

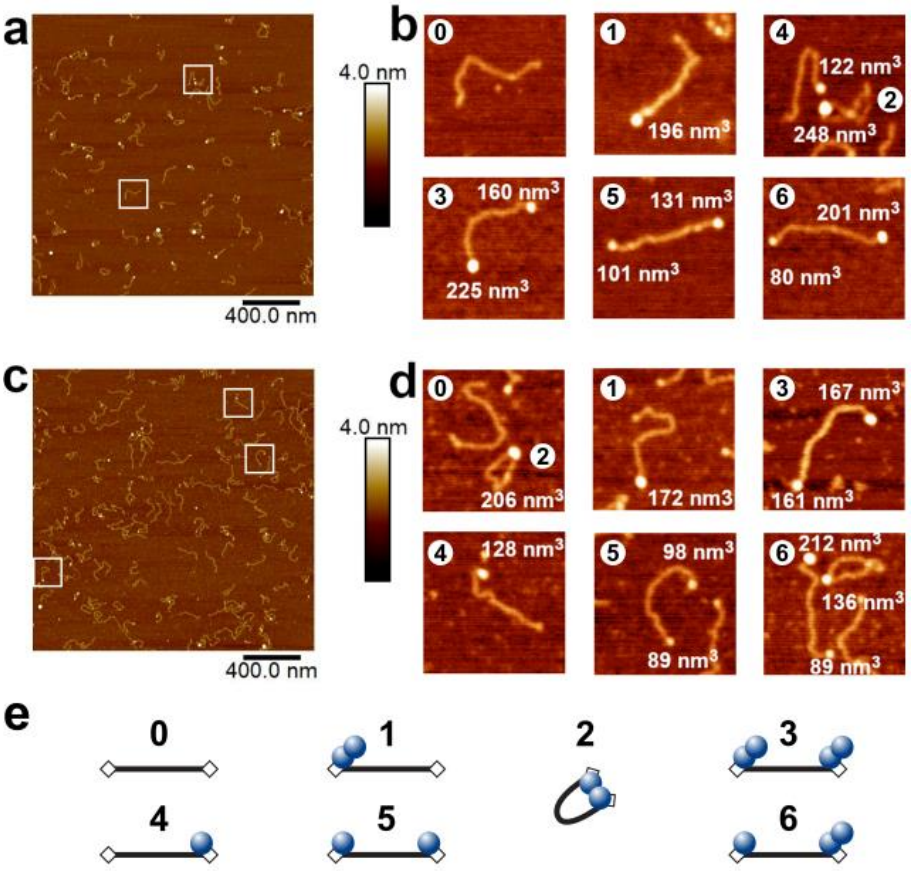


Figure 21. Visualization of AfAgo-induced DNA loops by AFM. Representative AFM images and 4-fold enlarged views of WT AfAgo-DNA (a, b) and AfAgoΔ-DNA (c, d) complexes adsorbed to APS-mica acquired in the air are shown, along with calculated protein volumes. The area of each image in the (a, c) panels is 4 μm^2, the scale bar is 400 nm, the Z range is 4 nm; and the Z range of images in (b, d) panels is 3.0 nm. Regions marked by white squares in panels (a, c) indicate several of the observed protein-DNA complexes enlarged in panels (b, d). Based on protein volumes, AfAgo-DNA complexes shown in panels (b, d) are

assigned to different protein-DNA stoichiometries and arrangements (numbered from 0 to 6) that are schematically depicted in panel (e): “0” – naked DNA; “1” – AfAgo dimer bound to one DNA end; “2” – AfAgo dimer forming a DNA loop; “3” – two AfAgo dimers on different DNA ends; “4” – an AfAgo monomer on one DNA end; “5” – two AfAgo monomers on different DNA ends; “6” – a monomer and a dimer on different DNA ends. Reproduced from Golovinas *et al*., 2021[145].

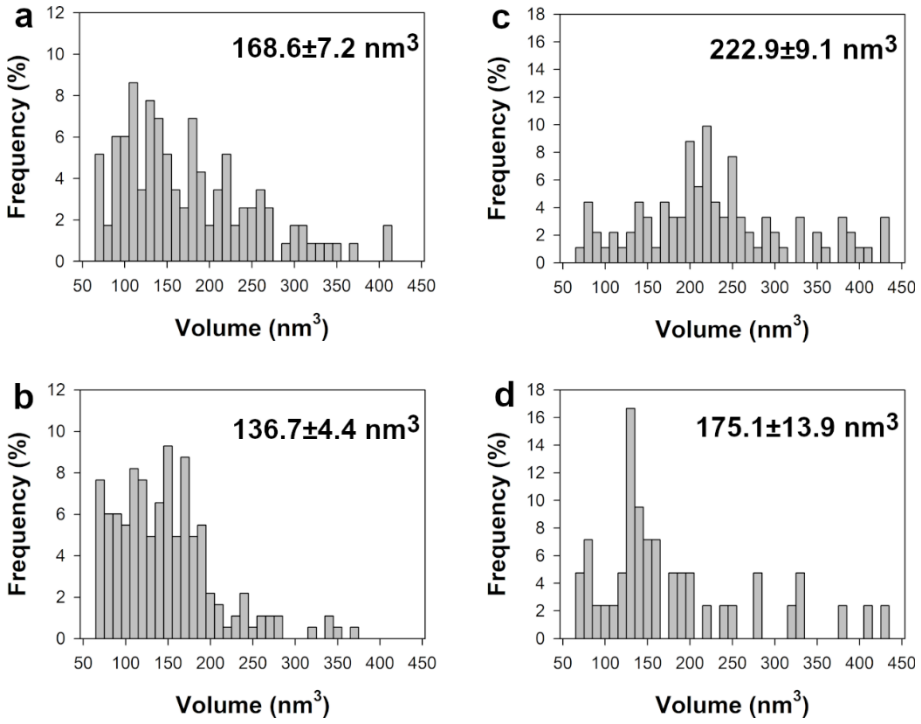


Figure 22. Volumetric analysis of WT AfAgo (panels a and c) and AfAgoΔ (panels b and d) proteins in protein-DNA complexes observed by AFM. Volumes of proteins bound to ends of unlooped DNA molecules for WT AfAgo (n = 118) and AfAgoΔ (n = 183) complexes are presented in panels a and b, respectively. Volumes of proteins bound to looped DNA for WT AfAgo (n = 95) and (d) AfAgoΔ (n = 44) complexes are shown in panels c and d. Mean ± S.E.M. values for each set are shown. Reproduced from Golovinas *et al*., 2021[145].

DNA length in the analysed protein-DNA complexes also showed considerable variation, with average values close to 195 nm, the theoretical length of 585 bp DNA. However, no correlation was observed between calculated DNA length and protein volumes in the complexes, confirming that DNA made no systematic contribution to the measured protein sizes (Figure 23).

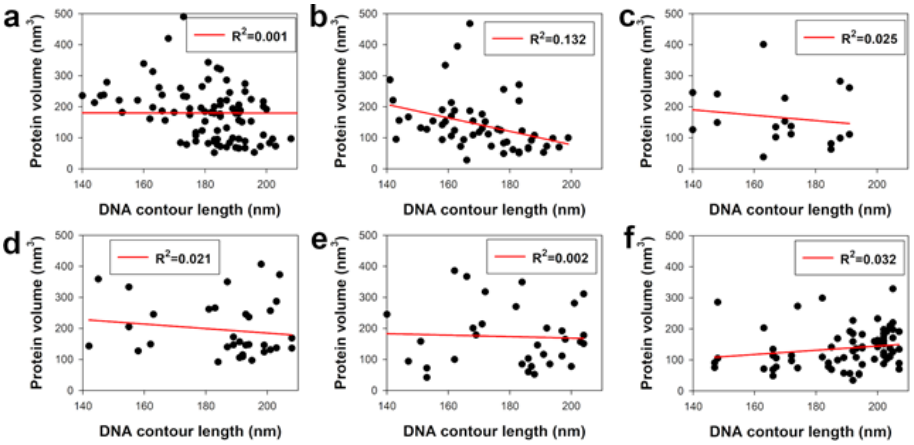


Figure 23. Correlation between the bound protein volume and the DNA contour length in AfAgo-DNA complexes as measured by AFM. WT AfAgo-DNA complexes: (a) protein bound to looped DNA (n = 90), (b) protein bound to one end of linear DNA (n = 53), and (c) protein bound to both ends of linear DNA (n = 18); AfAgoΔ complexes: (d) protein bound to looped DNA (n = 34), (e) protein bound to one end of linear DNA (n = 32), and (f) protein bound to both ends of linear DNA (n=72). Reproduced from Golovinas *et al*., 2021[145].

Notably, the relative distribution of different complexes varied dramatically for WT AfAgo and the dimerization mutant AfAgoΔ (Table 3). The ring-shaped DNA-protein complexes are the dominant species observed with WT AfAgo (51% or 95 out of 187 complexes, 47 of them containing a dimeric protein). A minor fraction of DNA molecules had either protein bound to one end (35%, 66 out of 187 complexes, 33 of them monomers and 22 dimers) or to both ends (13%, 26 out of 187, 6 of them two monomers, 4 of them two dimers, 13 one monomer and one dimer). In the case of AfAgoΔ, the majority of complexes had protein bound either to both DNA ends (34%, or 58 out of 169, 20 of them – two monomers, 8 – two dimers, 25 – one monomer and one dimer, Table 3) or to one end (40%, or 58 out of 169, 38 of them monomers and 24 dimers). A much smaller fraction (26%, or 44 out of 187) were ring-shaped structures. It is assumed that ring-shaped DNA molecules are primarily formed by dimeric WT AfAgo bound to both termini of the DNA fragment, in a similar manner as observed in the X-ray structures. A prominent decrease in ring-shaped DNA in the AfAgoΔ samples is consistent with its impaired dimerization. The remaining looped complexes are likely formed due to the residual ability of AfAgoΔ to form dimers, though inadvertent cross-linking of DNA-bound AfAgoΔ monomers with glutaraldehyde during sample preparation cannot be excluded (see Methods,

p. 67 for details). Cross-linking may also account for the presence of higher-order AfAgo oligomers observed by AFM (Table 3).

Table 3. AfAgo-DNA complexes observed by AFM (See Figure 22). Reproduced from Golovinas *et al*., 2021[145].

Complex / Protein	DNA loops, %	Linear	
		Protein bound to one end, %	*Protein bound to both ends, %*
WT AfAgo	51% (n=95)	35% (n=66)	13% (n=26)
	*Monomer (n=20) Dimer (n=47) Other (n=28)	Monomer (n=33) Dimer (n=22) Other (n=11)	Monomer-monomer (n=6) Dimer-dimer (n=4) Monomer-dimer (n=13) Other (n=3)
AfAgoΔ	26% (n=44)	40% (n=67)	34% (n=58)
	Monomer (n=24) Dimer (n=12) Other (n=8)	Monomer (n=38) Dimer (n=24) Other (n=5)	Monomer-monomer (n=20) Dimer-dimer (n=8) Monomer-dimer (n=25) Other (n=5)

*The calculated volume of WT AfAgo (50.8 kDa) and AfAgoΔ (49.9 kDa) proteins is ~100 nm^3. Thus, the measured protein volume data from AFM images was divided into three populations by their theoretical volume: monomer (<150 nm^3), dimer (150 – 250 nm^3), and other (>250 nm^3). For details, see Figure 22.

3.1.5. WT AfAgo induces DNA loops in solution

To further characterize AfAgo-DNA interactions at nanomolar concentrations, AfAgo-DNA interactions were probed using single-molecule Förster resonance energy transfer (smFRET). If AfAgo homodimer simultaneously interacts with two ends of the same DNA molecule, the induced DNA loops can be monitored as a change in FRET efficiency between dyes tethered close to DNA ends (Figure 24a). Utilization of a single dual-labelled DNA substrate (rather than two short DNA duplexes carrying different fluorescent labels) increases the probability of AfAgo interaction with both DNA ends at low reactant concentrations required for the single-molecule setup.

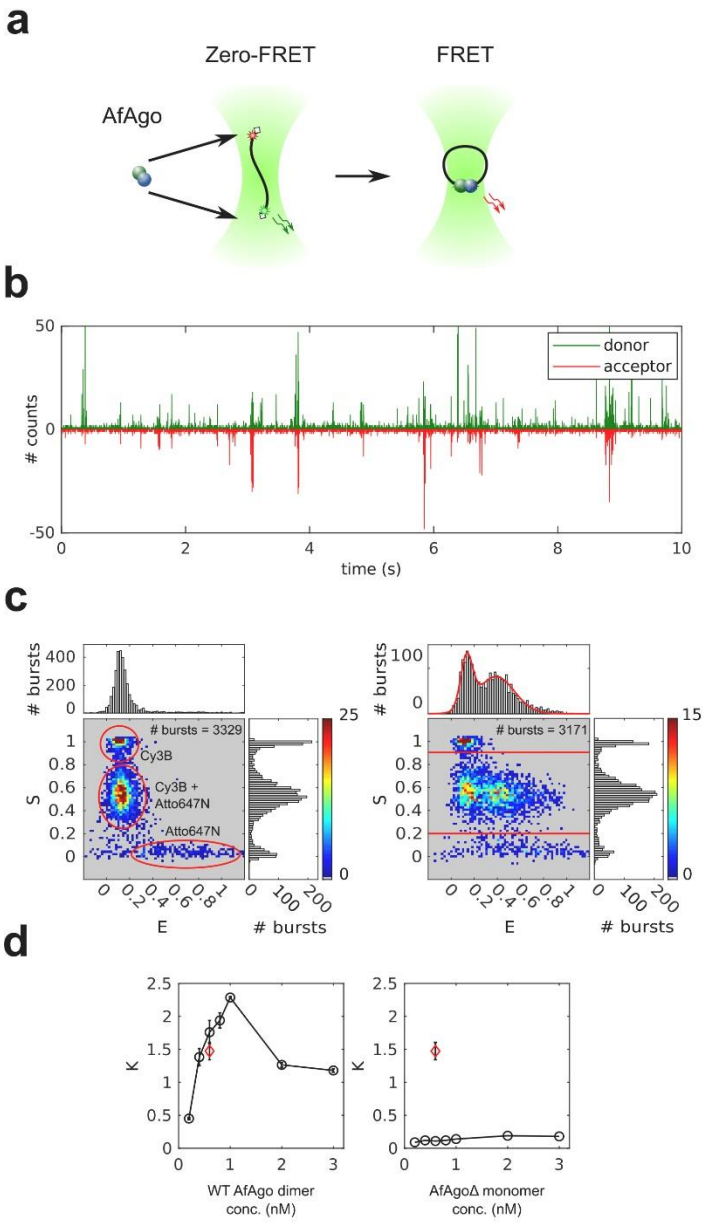


Figure 24. Single-molecule studies of AfAgo-DNA interactions in solution. (a) A schematic overview of the single-molecule assay. Left, free DNA; right, WT AfAgo-DNA (blue and green circles) complex in a looped state. (b) Fluorescence intensity trace with 1 ms time bin of 25 pM DNA with 2 nM AfAgo. Red: inverted acceptor fluorescence upon donor excitation, green: donor fluorescence upon donor excitation. (c) Left – E-S histogram of DNA alone. The top and side axes contain, respectively, one-dimensional E (proximity ratio) and S (donor/acceptor stoichiometry) histograms of all bursts. Denoted are areas representing donor-only DNA, acceptor-only DNA, and dual-labelled DNA. Right – E-S histogram of DNA with

2 nM AfAgo. The one-dimensional E histogram on top is derived from bursts with S = 0.2 – 0.9, designated by horizontal lines in the E-S histogram. The red curve is a two-Gaussian fit to the data that gave positions of the Gaussian maxima on the E-axis (0.13 ± 0.01 and 0.39 ± 0.02). (d) Left - dependence of the ratio of looped and unlooped DNA molecules (parameter K) on WT AfAgo concentration (open circles). Right – the dependence of K on the AfAgoΔ concentration (open circles). Red diamonds in both graphs represent the competition experiment performed with 1.2 nM WT AfAgo and 0.6 nM AfAgoΔ. All data points are average values of three measurements ±1 standard deviation. Reproduced from Golovinas *et al*., 2021[145].

AfAgo interaction with the DNA fragment was monitored by analyzing the fluorescence bursts of single diffusing DNA fragments (Figure 24b). As described in Methods (p. 65), for each DNA molecule the stoichiometry parameter S was calculated, which is close to 0.5 for DNA molecules labelled with both fluorophores, approximately 0 for the acceptor-only DNA and close to 1.0 for the donor-only DNA, and the proximity ratio E, which is expected to be higher for looped DNA molecules with the FRET dyes brought into close proximity than for unlooped DNA molecules.

The E-S histogram of DNA alone (Figure 24c, left) exhibits a prominent population with low E and intermediate S values, which corresponds to dual-labelled unlooped (zero-FRET) DNA molecules. The two minor populations observed in the histogram correspond to donor-only (low E/high S) and acceptor-only (high E/low S) DNA fragments.

The E-S histogram of DNA in the presence of WT AfAgo exhibits an additional population (intermediate S and intermediate E, Figure 24c, right), which presumably represents DNA molecules looped by WT AfAgo. The fraction of looped and unlooped DNA molecules was quantified by fitting a sum of two Gaussian functions to the 2D histogram of E values of dual-labelled molecules (Figure 24c, right), and finding the areas under the Gaussian with a relatively high E centre (representing looped DNA) and a Gaussian with a near-zero E centre (representing unlooped DNA). DNA looping efficiency K was then defined as the ratio of the two areas.

The ratio K at different WT AfAgo concentrations was measured (Figure 24d). It increased monotonously with increasing WT AfAgo concentration until it reached the maximum value of 2.5 (corresponds to about 70% of looped DNA molecules) at 2 nM WT but decreased as the protein concentration was increased further.

A similar set of single-molecule experiments was performed with the dimerization mutant AfAgoΔ. As shown in Figure 24d, the ratio K at all AfAgoΔ concentrations tested was close to zero, indicating that AfAgoΔ was unable to induce DNA loops. The lack of DNA looping was not due to impaired DNA binding, as shown by EMSA (Figure 16). Moreover, AfAgoΔ competes with WT AfAgo for DNA ends, as the K value observed in a competition experiment performed with equal concentrations of WT AfAgo dimer and AfAgoΔ monomer was considerably lower than in an experiment with WT AfAgo alone (Figure 24d). Taken together, efficient DNA looping observed with WT AfAgo and impaired looping by the dimerization interface mutant AfAgoΔ provide further support for the ability of WT AfAgo dimer to simultaneously bind two DNA ends in solution.

3.1.6. Dynamics of WT AfAgo-induced DNA loops

To explore the dynamics of the WT AfAgo-induced DNA looping events, total internal reflection fluorescence (TIRF) microscopy was employed to perform single-molecule FRET experiments on surface-immobilized DNA (Figure 25). For that purpose, a DNA fragment was constructed, that was essentially identical to the one used for single-molecule studies in solution, except that it carried a biotin 386 bp away from the donor end for surface immobilization (Figure 14). After verifying that WT AfAgo induces loops on this substrate in solution (Figure 26b), biotinylated DNA was immobilized on the surface and fluorescence movies in the absence or the presence of WT AfAgo were recorded (Figure 26a). From each frame of the movie donor and acceptor intensities for individual DNA fragments were extracted, selecting trajectories with anti-correlated changes of the donor and acceptor intensities (indicating the occurrence of FRET), and calculations of the time courses of the proximity ratio, E, were performed. An example of such a trajectory is presented in Figure 25c. In a control with no AfAgo, no DNA fragments exhibiting FRET could be found (Figure 25b).

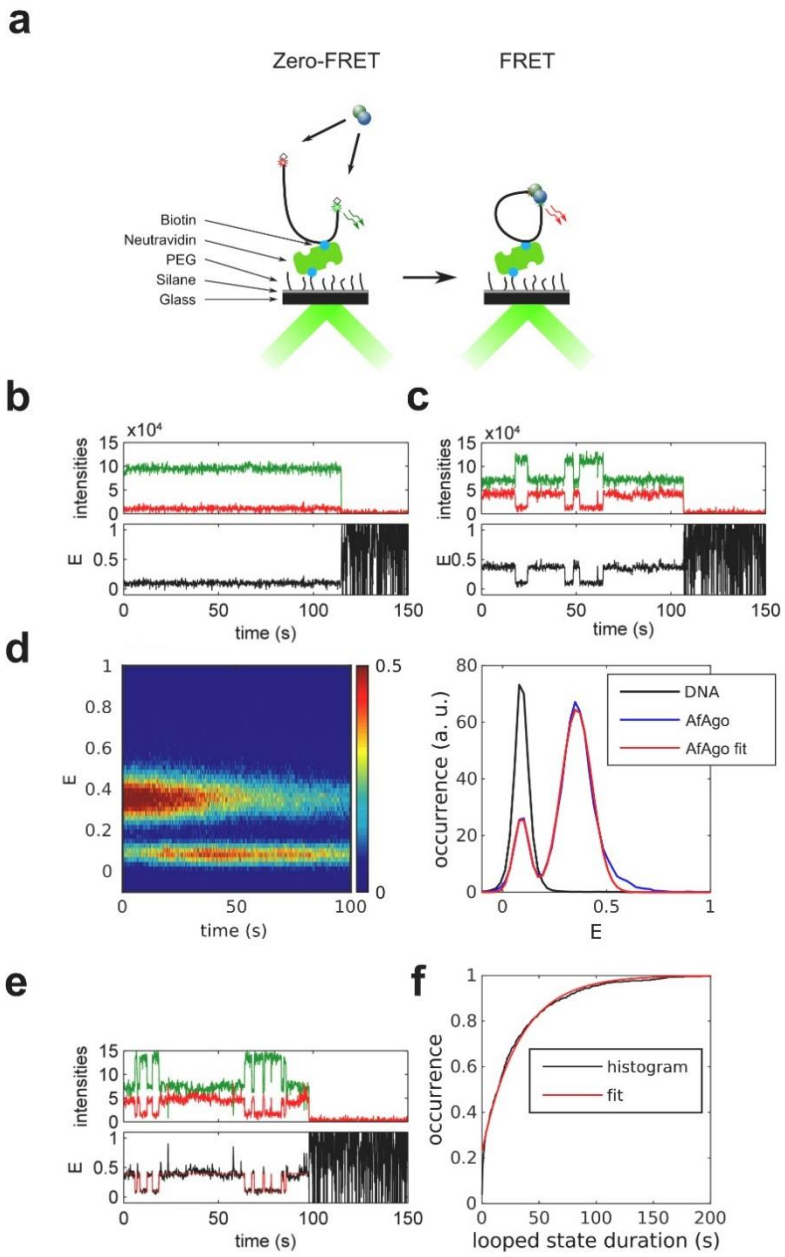


Figure 25. Dynamics of WT AfAgo-induced DNA loops. (a) A schematic overview of the single-molecule assay using TIRF microscopy. (b, c) Trajectories of donor (green) and acceptor (red) intensity and corresponding proximity ratio, E, of individual DNA fragments without (b) and with 20 mM WT AfAgo (c). (d) Left – an image of 287 pooled time traces of the proximity ratio, E, from the measurement with 20 mM of WT AfAgo. The image is normalized to the maximum image intensity. Right – a section of the image in the left integrated

over the first 10 s shown with the two-Gaussian fit. The positions of the Gaussian maxima are 0.09 ± 0.01 and 0.36 ± 0.01. For comparison, a trace- and time-averaged section from the measurement of 227 traces on bare DNA is shown. (e) An example of trajectories of donor (green) and acceptor (red) intensity and corresponding proximity ratio, E, with HMM idealization of an individual DNA fragment with 20 mM WT AfAgo. (f) Cumulative histogram of the looped state durations from 287 E traces with a single-exponential fit with the exponential factor of 33 ± 1 s. Reproduced from Golovinas *et al*., 2021[145].

The single-molecule population and time-averaged E values exhibit two peaks with maxima at 0.09 and 0.36, corresponding to the unlooped and looped DNA molecules, respectively (Figure 25d, right). These E values are also in good agreement with the E values obtained from the measurement in solution (Figure 24c).

A superficial inspection of E trajectories of individual DNA fragments revealed that their looping dynamics are rather diverse. There exist trajectories with the looped state lasting the whole measurement, whereas other trajectories are more dynamic with a number of transitions between the looped and unlooped states (Figure 25c, Figure 26c, d). The looped E state also exhibits more subtle dynamics (Figure 26d), which is attributed to the conformational flexibility of AfAgo at the dimerization interface.

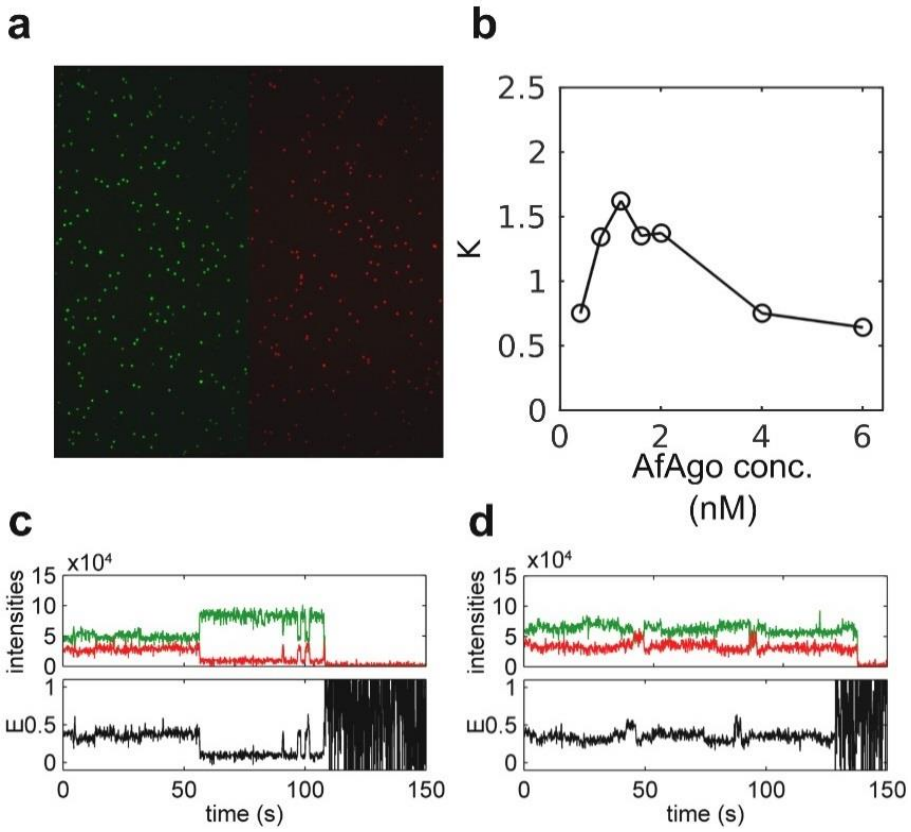


Figure 26. Single-molecule experiments. (a) A fluorescence image of surface-immobilized DNA fragments. It is an average of 20 frames in a fluorescence movie. The left part (green) is the donor image upon donor excitation, and the right part (red) is the acceptor image upon acceptor excitation. (b) The ratio of the number of looped and unlooped DNA molecules, K, depends on the concentration of the AfAgo for the biotinylated DNA fragment in solution. (c, d) Examples of different dynamics of DNA looping by AfAgo in TIRF experiments. Reproduced from Golovinas *et al*., 2021[145].

3.2 AfAgo interactions with nucleic acids

"It is very easy to answer many of these fundamental biological questions; you just look at the thing!"[5]

- Richard Feynman,
There's plenty of room at the bottom: An invitation to enter a new field of physics

[5] Image: *The Anatomy Lesson of Dr Nicolaes Tulp*, Rembrandt, 1632.

3.2.1. AfAgo binds specific nucleic acids *in vivo*

When overexpressed in *E. coli*, AfAgo co-purifies with tightly bound nucleic acids, predominantly RNA (Figure 27a). This interaction is disrupted only at NaCl concentrations exceeding 1.0 M, implying a tight association. The length of the AfAgo-bound RNA varies from a few dozen to a few hundred nucleotides (Figure 27a), with sequencing data showing that most reads fall between 14 and 30 nucleotides (Figure 27B). Sequencing of AfAgo-bound RNA revealed that most successfully mapped RNAs (73%) are derived from the AfAgo expression vector (Figure 27c, Table 4), while a smaller fraction (27%) was derived from the *E. coli* genome. Surprisingly, AfAgo had a strong preference for two 5′-terminal RNA nucleotides, A at the first and U at the second position (occupancies 0.862 and 0.846, respectively), and a discernible preference for U at the third position (occupancy 0.476, Figure 27d). Thus, AfAgo, like many eAgos and long pAgos, has intrinsic specificity for the 5′-terminus of the bound nucleic acid.

Table 4. Summary of AfAgo-bound RNA sequencing analysis. Reproduced from Manakova *et al*., 2023[216].

	Aligned reads, No.	**Aligned reads, %**
Total aligned reads	2358927	58.87
***E. coli* genome**	639706	27.12
pETDuet-1 plasmid	1719221	72.88
AfAgo	611314	25.91
Af1317	320563	13.59
f1 ori	42407	1.8
AmpR	157749	6.69
ColE1 ori	45582	1.93
ROP	3680	0.16
lacI	71647	3.04

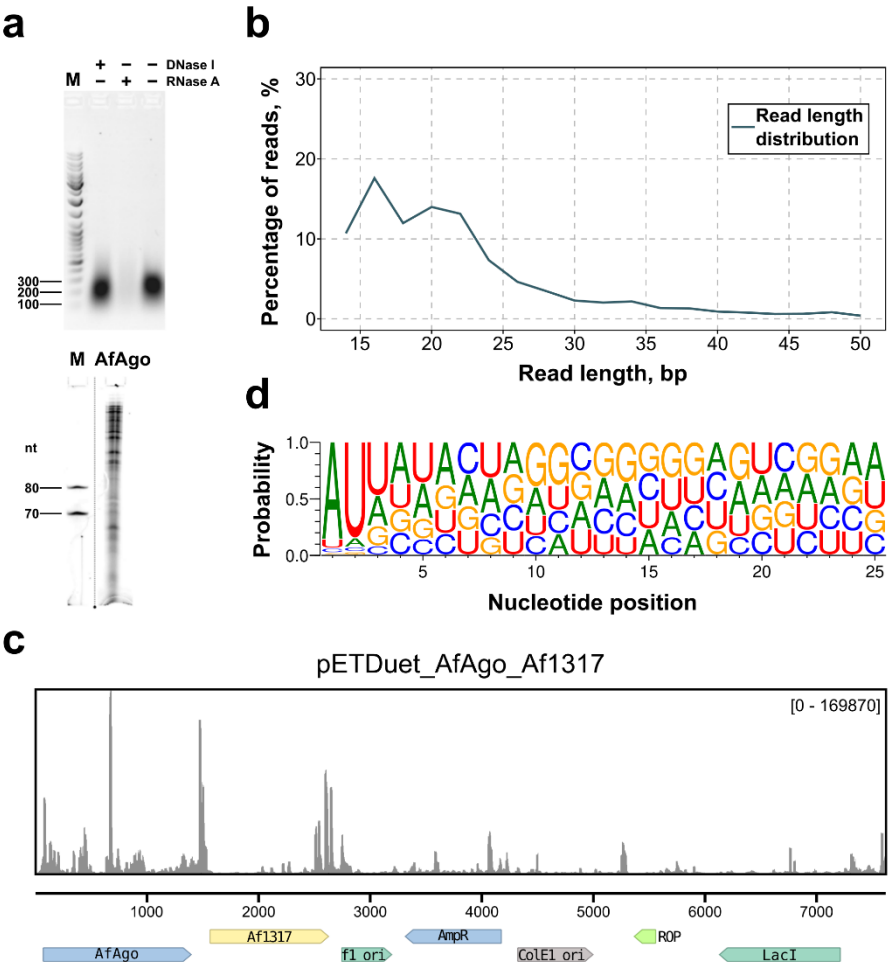


Figure 27. Analysis of *in vivo* (*E. coli*) AfAgo-bound nucleic acids. (a) Top – Digestion of AfAgo nucleic acids with DNAse I and RNase A. Bottom – Size analysis of AfAgo-bound RNA. (b) Read length distribution of sequenced AfAgo-bound nucleic acids. (c) Sequencing read alignments to the AfAgo expression vector. 73% of all reads map to the expression vector, compared to 27% mapping to the *E. coli* genome. (d) Small RNAs co-purified with AfAgo show 5′-AUU bias. Reproduced from Manakova *et al*., 2023[216].

3.2.2. AfAgo specificity for nucleic acid substrates *in vitro*

Previous studies suggested that AfAgo has a strong preference for single- and double-stranded DNA over single- or double-stranded RNA[25]. However, these studies were performed using double-stranded DNA with 5′-C and 5′-T terminal nucleotides, neither of which, according to the present

analysis of *in vivo*-bound nucleic acids, is optimal for AfAgo binding. To re-evaluate AfAgo affinity to nucleic acids, EMSA was employed, using synthetic single-stranded (ss) RNA and DNA oligonucleotides containing phosphorylated 5′-AUU and 5′-ATT termini, respectively (Appendix 1). The experiments revealed that under these experimental conditions (see Methods, p. 61), AfAgo preferentially binds ssRNA over ssDNA. Binding of ssDNA was detected only at exceedingly high (>0.5 μM) AfAgo concentrations (Figure 28). Next, to determine the specificity of AfAgo for the terminal bases, a set of ssRNA oligonucleotides with varying 1st, 2nd, and 3rd 5′-terminal nucleotides were employed (Table 5, Figure 28). AfAgo showed a preference for the 5′-AUU-containing ssRNA (Table 5, Figure 28), while substitution of each of the three 5′-terminal nucleotides of the preferred 5′-AUU ssRNA (1st A, 2nd and 3rd U) reduced the binding affinity (Figure 28, Table 5). This shows that AfAgo is capable of discriminating the first three 5′-terminal nucleotides of bound ssRNA and has a preference for a 5′-AUU RNA sequence *in vitro*.

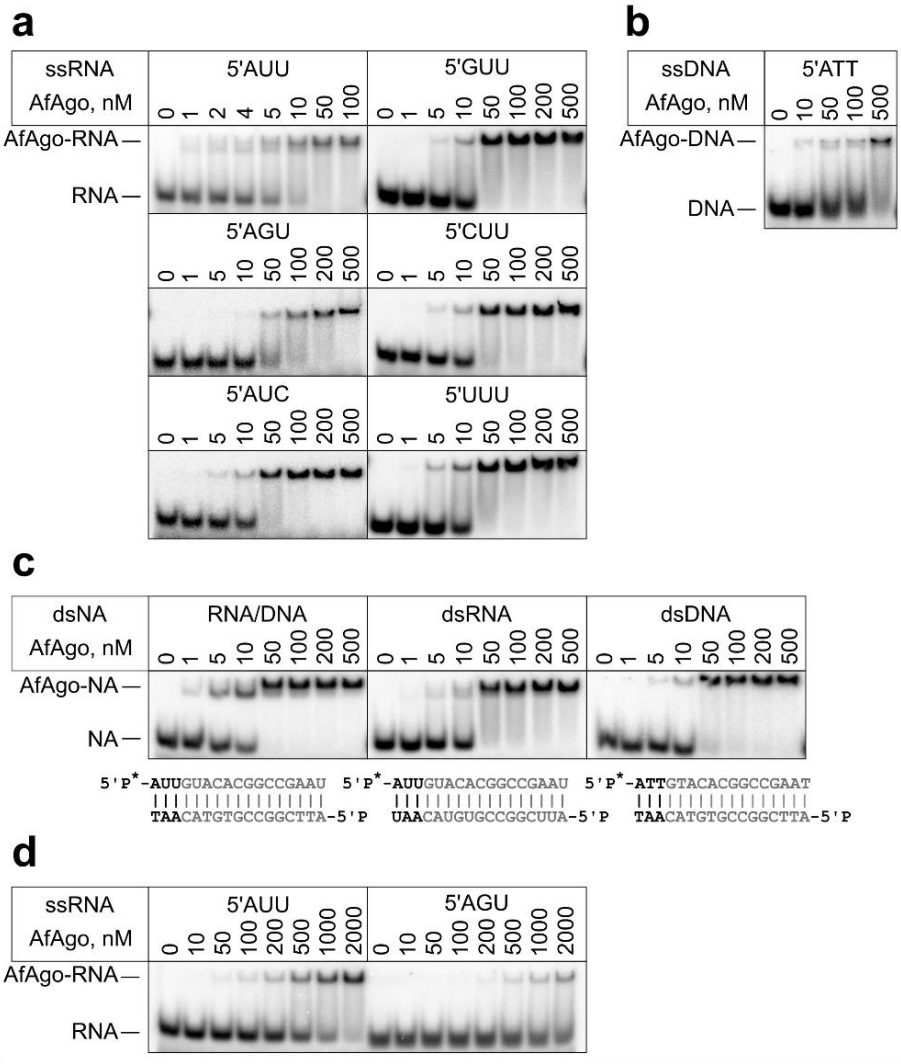


Figure 28. AfAgo interactions with nucleic acids *in vitro*. EMSA experiments were performed with 5 nM total 5′-P-ssRNA (a), ssDNA (b), and dsNA (c), with varying concentrations of AfAgo, indicated above each lane. Double-stranded nucleic acid sequences and structure are depicted schematically below each gel, with 5′-terminal bases of the guide and 3′-terminal bases relevant to AfAgo base recognition highlighted in black, remaining strands in grey. 5′-^{32}P-labelled strand indicated with an asterisk. Calculated K_d values are provided in Table 5. (d) AfAgo interactions with ssRNA *in vitro* at elevated temperature. EMSA experiments were performed with 5 nM total 5′-P-ssRNA, complexes pre-formed at 70 °C, and gels were run at room temperature. While the general binding affinity of ssRNAs decreased, however, the specificity of AfAgo towards a 5′-AUU over 5′-AGU ssRNA remains. Reproduced and modified from Manakova *et al.*[216].

Argonaute proteins usually use a nucleic acid guide to search for and bind a complementary target. Binding experiments described above suggest that AfAgo may use ssRNA guides with 5′-AUU terminal nucleotides. Thus, in the next set of experiments, EMSA was employed to test RNA-guided DNA and RNA targeting by AfAgo (Figure 29). It was found that AfAgo pre-loaded with a guide RNA (gRNA) specifically binds DNA and RNA targets, showing higher affinity to DNA targets. To further probe discrimination of DNA *vs.* RNA targets by the AfAgo-gRNA complex, the target binding reaction was supplemented with heparin, a competitor of nucleic acid binding (Figure 29b).

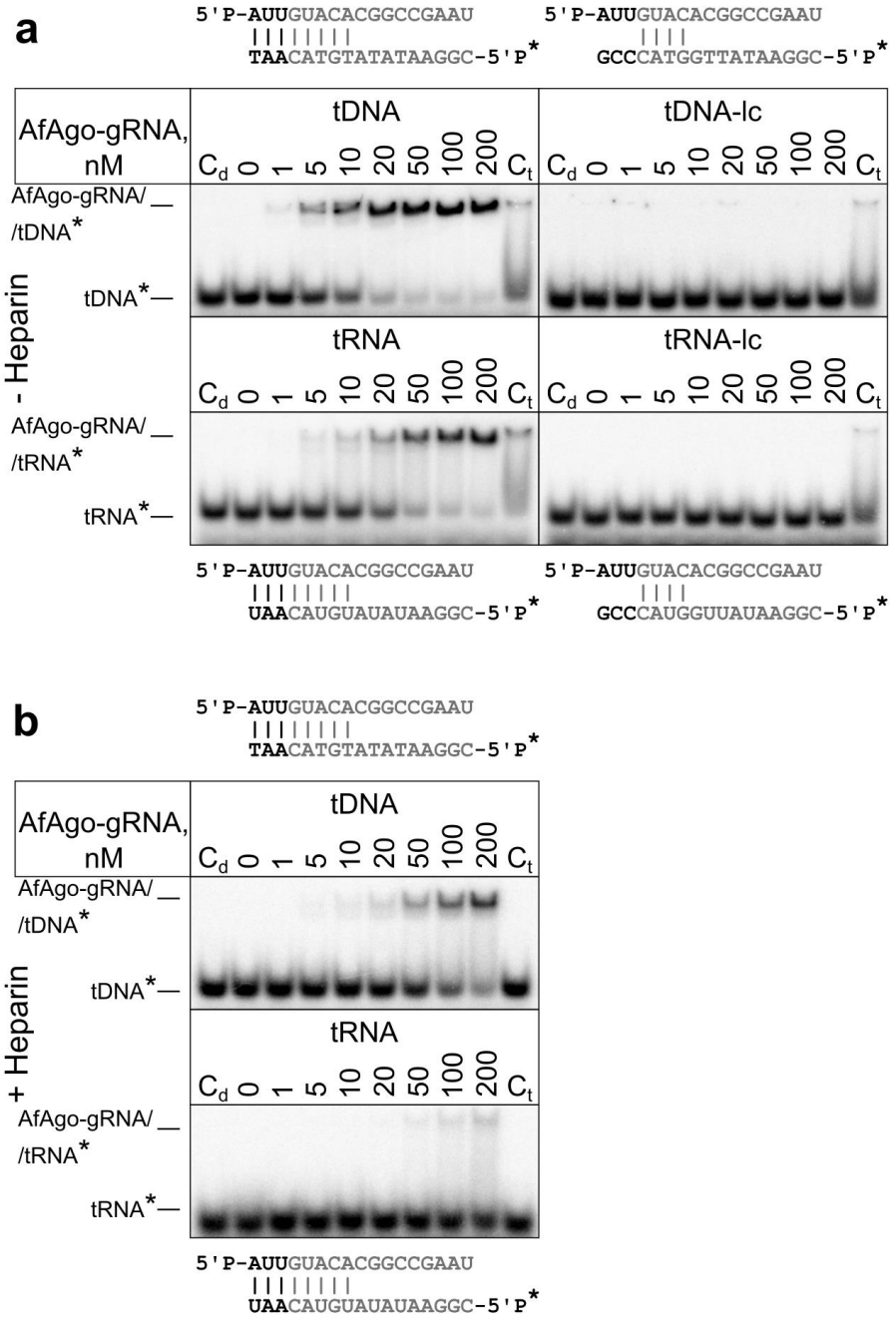


Figure 29. AfAgo RNA-guided NA targeting mechanism and double-stranded nucleic acid binding probed using EMSA. (a) Titration of labelled target ssDNA (top) and ssRNA (bottom) with a pre-formed AfAgo-guide RNA complex (1:2 ratio, AfAgo concentrations indicated above each lane) for either 8 nt complementary (left) or 4 nt complementary (lc –

"low complementarity", right) targets. A schematic of guide-target complementarity is shown adjacent to each respective gel, with 5′-terminal bases of the guide and 3′-terminal bases relevant to AfAgo base recognition highlighted in black, the remaining strands in grey. 5′-^{32}P-labelled target strands are denoted with an asterisk. C_d – duplex control, where guide and target were mixed in the absence of AfAgo at a ratio equivalent to lane "200". C_t – target control, where RNA-free AfAgo was mixed with the target at a ratio equivalent to lane "200". (b) Experiment equivalent to (a), left, conducted in the presence of 100 ng/μl heparin. Reproduced and modified from Manakova *et al*.[216].

Results show that under these conditions AfAgo-gRNA complex shows an even stronger preference for DNA targets over RNA targets, similar to related long-B and short pAgos (Table 5)[9,10,14]. Experiments with pre-formed RNA/RNA, RNA/DNA and DNA/DNA duplexes (Table 5, Figure 28) were also consistent with the mechanism where AfAgo uses ssRNA as a guide for recognition of ssDNA targets.

Table 5. Apparent K_d of different tested nucleic acid substrates determined using EMSA. K_d of a pre-formed AfAgo-guide complex with ssRNA and ssDNA target oligonucleotide were determined for one putative optimal ssRNA guide and two targets complementary to the guide within the "seed" region. K_d was calculated from experimental results where heparin was omitted. Values are means ± standard deviation of three independent replicates. Reproduced from Manakova *et al*., 2023[216].

Oligonucleotide	5′-terminus	K_d, nM
ssRNA	AUU	3.8 ± 0.1
	GUU	42 ± 2.1
	CUU	28 ± 2.7
	UUU	15 ± 2.8
	AGU	84 ± 2.9
	AUC	16 ± 2.1
ssDNA	ATT	236 ± 35
RNA/DNA	AUU	6.1 ± 0.04
dsDNA	ATT	37 ± 17
dsRNA	AUU	15 ± 0.9
Nucleic acid binding by AfAgo-gRNA complex		
Guide	Target	K_d, nM
ssRNA	ssDNA	7.5 ± 0.4
	ssRNA	33 ± 4.2

3.2.3. Structural basis for the terminal base pair recognition

To obtain structural insights into the mechanism of 5′-terminus recognition, crystallizing AfAgo with various RNA guides carrying 5′-AUU and RNA/DNA heteroduplexes was attempted, as well as with various DNA oligoduplexes. Crystallization attempts with RNA-containing duplexes were

not successful, but crystal structures of AfAgo in complex with self-complementary 14 bp DNA-DNA oligoduplexes carrying a 5′-P-ATT terminus were successfully solved, which is analogous to the optimal ssRNA terminus 5′-AUU (PDB ID 6T5T and 6TUO, respectively) and a 5′-P-ATC terminus, which is analogous to a suboptimal ssRNA terminus 5′-AUC (PDB ID 6XUP and 6XU0) (Appendix 3, Figure 30). Although AfAgo interactions with the 5′-ATT DNA oligoduplexes revealed in the current structures may not directly translate to possible interactions with the 5′-AUU RNA guide, these data provide valuable insights into the base-specific interactions formed by AfAgo and the guide-target duplex.

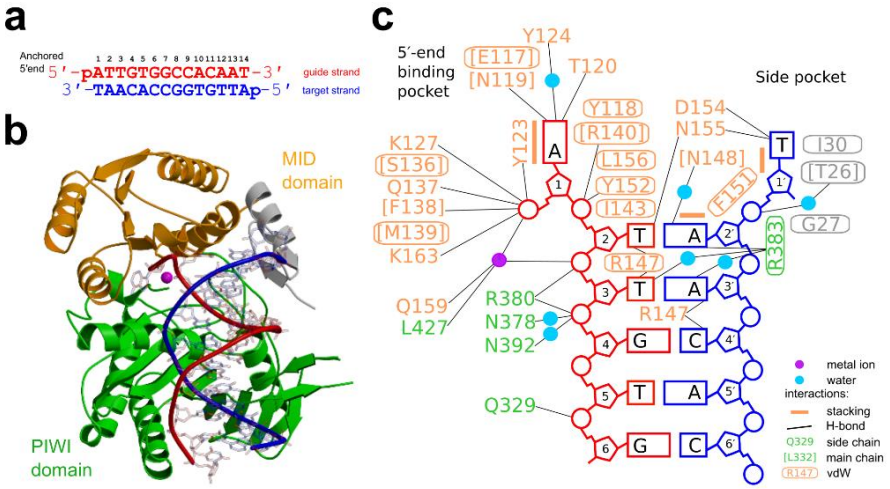


Figure 30. Structure of the AfAgo-DNA complex. (a) The 5′-ATT DNA oligoduplex used for crystallization. (b) The overall structure of the AfAgo-DNA complex. The backbone of DNA strands is coloured as in (a). DNA bases are transparent. The Mg^{2+} ion involved in the coordination of the 5′-phosphate of the guide strand is shown as a magenta sphere. (c) Schematic representation of AfAgo contacts with the DNA. Reproduced from Manakova *et al.*, 2023[216].

The AfAgo complex with the suboptimal 5′-P-ATC DNA formed crystals of *P*1 symmetry that contained two DNA-bound protein subunits per asymmetric unit. The main conformations of both subunits are nearly identical, the most significant difference being the main chain conformation of the loop formed by residues 144 – 149 (Figure 31a). Electron density for both DNA strands is good only for the first 5 – 6 bp, i.e. the part that makes direct contacts with the protein; the remaining part of the DNA duplex points into the solution and is disordered. The overall structure of the AfAgo protein superimposes closely with the previously published AfAgo complexes with

RNA and DNA [24–26] (RMSD 1 – 1.64 Å when overlaid by residues 11 – 427).

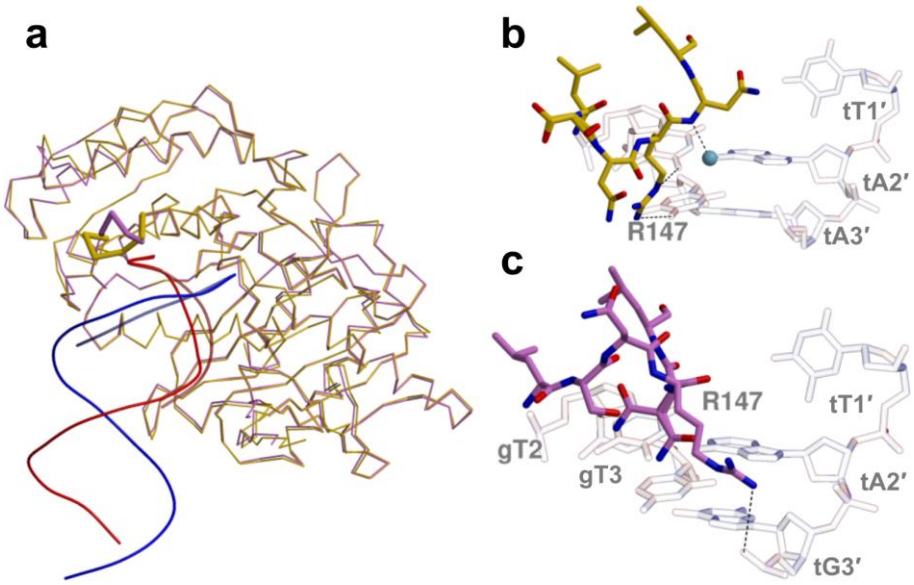


Figure 31. Different conformation of the loop 144-149 in crystal structures of AfAgo. (a) AfAgo protein chains from the complex with 5′-ATT (PDB ID 6T5T, yellow) and with 5′-ATC (PDB ID 6XUP chain A, magenta) are shown as traces. Guide and target DNA strands are red and blue, respectively. (b) The conformation of the loop 144-149 in the structure AfAgo-5′ATT, PDB ID 6T5T. (c) Loop 144 – 149 from the A protein chain in the crystal structure AfAgo-5′ATC, PDB ID 6XUP. Reproduced from Manakova *et al*., 2023[216].

The AfAgo complexes with two variants of the 5′-P-ATT DNA oligoduplex crystallized in the $P22_12_1$ space group and contained a single DNA-bound protein subunit per asymmetric unit (Appendix 3). The distal end of bound DNA in this crystal form was fixed by crystal packing against the neighbouring protein subunit, thereby helping to model the full-length oligoduplex. The guide strand anchors to AfAgo *via* its 5′-phosphate group, which is accommodated in the conserved MID domain binding pocket[24–26], where it makes direct contacts with Lys127, Ser136, Gln137, Phe138, Met139 and Lys163, and Mg^{2+}-mediated contacts with Gln159 and the C-terminal Leu427 (Figure 30c). The gA1:tT1′ base pair (the first guide strand adenine and the complementary target strand thymine) is disrupted, with the bases flipped into separate protein pockets. The flipped gA1 base is inserted into the MID domain pocket, where it is fixed by stacking between Tyr123 and Tyr118, base-specific H-bonds to the main chain N atom of Asn119 and the hydroxyl group of Thr120, and a water-mediated H-bond to the hydroxyl

of Tyr124 (Figure 30c, Figure 32a). The tT1′ base of the target strand is displaced into the “side” pocket (Figure 30c, Figure 32b) formed by helices 26-36 (linker domain) and 149-163 (MID domain), where it stacks against Phe151 and forms H-bonds with Asp154 and Asn155. It should be noted that Asn155 can be modelled in two orientations, one of which is fixed by an H-bond between the OD1 atom to the main chain amide of Phe382. In this case, the ND2 atom of Asn155 is capable of H-bonding to the tT1′ and gT2 bases (Figure 30).

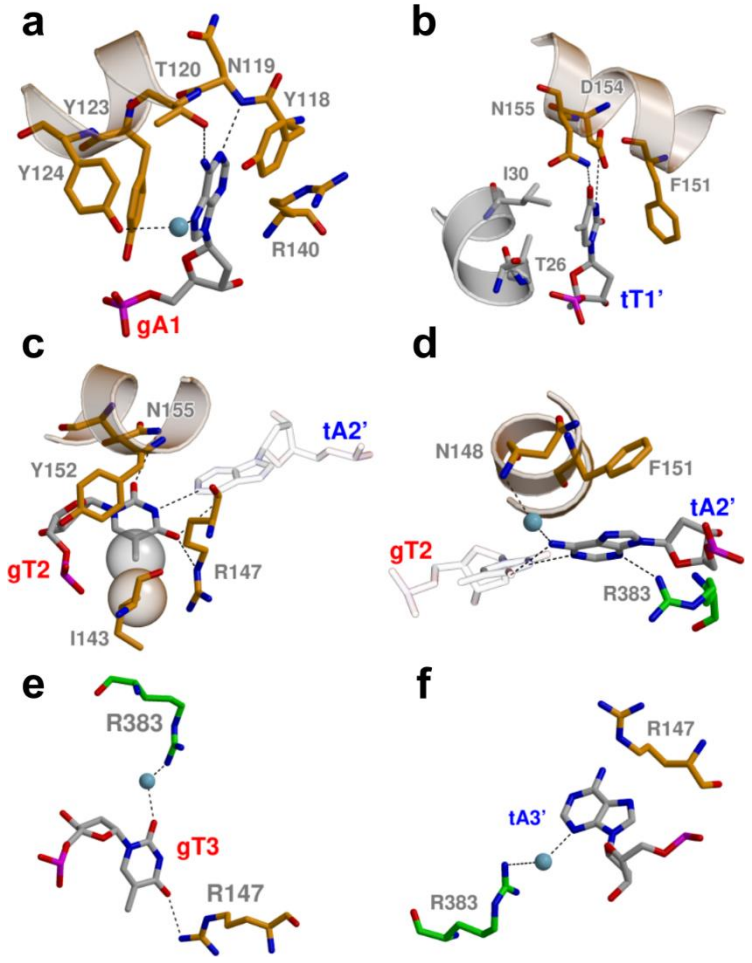


Figure 32. AfAgo interaction with the first three base pairs of the 5′-ATT DNA duplex. gA1 (a) and tT1′ (b) in their respective pockets. (c, d) Recognition of gT2 and tA2′ of the second base pair. (e, f) Interactions with gT3 and tA3′ of the third base pair. Reproduced from Manakova *et al*., 2023[216].

The base pairs following the gA1:tT1′ of the bound DNA oligoduplex remain undisrupted. AfAgo makes direct contacts with the 2nd and 3rd base pairs. Most extensive sequence-specific contacts are made by the gT2 base of the second base pair, including the contact of the thymine methyl group to Ile143, and base-specific H-bonds of thymine O2 and O4 atoms to Asn155 and Arg147, respectively (Figure 32c). The tA2′ adenine from the complementary strand stacks against Phe151, and makes a water-mediated H-bond to the main chain of Asn148 from the major groove side and a direct H-bond to Arg383 from the minor groove side (Figure 32d). Bases from the gT3-tA3′ make two water-mediated H-bonds to the protein, Arg383 in the minor groove side, and Arg147 in the major groove side (Figure 32e, f). In the AfAgo crystal structures with 5′-ATC oligoduplex, two slightly different patterns of the interaction were observed with the 2nd and the 3rd base pairs due to two different conformations of the loop 144-149 that includes residues Arg147 and Asn148 (Figure 31b, c), one of which is nearly identical to that found in AfAgo crystal structures with 5′-ATT oligoduplexes. Taken together, the structures presented in this work reveal the structural details of AfAgo base-specific interactions with three terminal base pairs of the bound guide/target duplex.

3.3 AfAgo and AfAgo-N form a heterodimer

"Terrible was their strength and swiftness; <…> At last Zeus hit upon an expedient. Let us cut them in two, he said; then they will only have half their strength, and we shall have twice as many sacrifices. He spake and split them as you might split an egg with an hair."[6]

- Plato, *Symposium*

[6] Image: *Hommage à Apollinaire*, Marc Chagall, 1911 – 1912.

AfAgo (or Af1318) protein gene is located in a genetic island of the *A. fulgidus* DSM 4304 genome with a lower GC content (~31%) than the average GC content for the flanking genome sequences (~50%) or the whole genome (~53%, Figure 33a)[157]. Such difference in GC content indicates a horizontal gene transfer of the AfAgo-encoding region from an unknown host to *A. fulgidus*. The AfAgo ORF overlaps with the ORFs of two hypothetical proteins, suggesting that proteins encoded by all three operon-forming genes may act together (Figure 33a). The downstream protein possesses a putative transmembrane (TM) region and shows sequence similarities (score 93.5, EMBOSS Needle) to the membrane effector SiAga2 of the antiviral SiAgo system from archaeon *Sulfolobus islandicus*[139]. In contrast, the upstream 172 amino acid protein shows no similarities to proteins of known function. Surprisingly, a nearly identical AfAgo-encoding genetic island found in the genome of a closely related strain *A. fulgidus* DSM 8774 (Figure 33a) contains a two-bp insertion that eliminates a STOP codon in the upstream gene. The protein encoded by the resultant longer (250 aa) open reading frame shows sequence similarities (score 42.5, EMBOSS Needle) to SiAga1, a protein that forms a functional heterodimer with SiAgo (Figure 33a)[139].

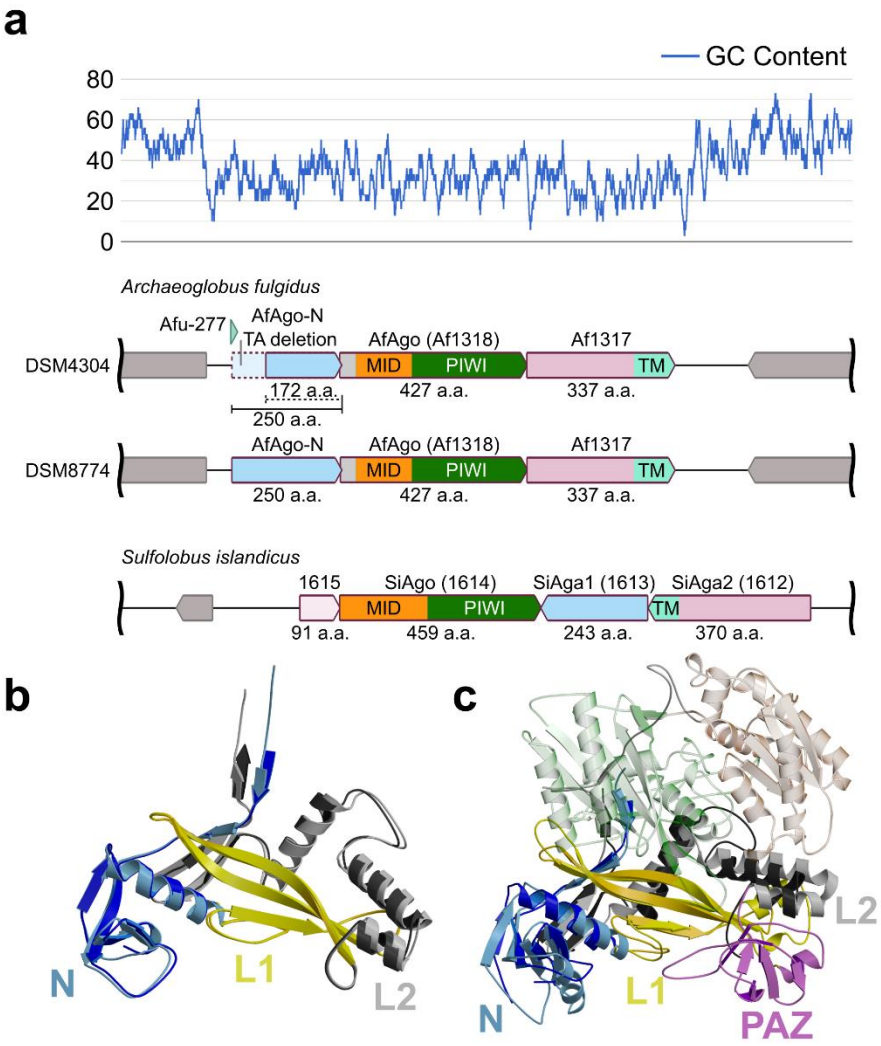


Figure 33. (a) Top – GC content of *A. fulgidus* DSM 4304 genome region of interest, containing AfAgo and related genes. Bottom – genes located in the corresponding region of interest in *A. fulgidus* DSM 4304 and DSM 8774, along with a region of appropriate length from *Sulfolobus islandicus*, containing the SiAgo-Aga1-Aga2 system. (b) Comparison of AfAgo-N crystal structures coloured like in Figure 35a. *P*1 structure is of darker shades. (c) AfAgo-N crystal structure (*P*1, 1.9 Å, darker colours) superimposed on the long Ago, RsAgo (PDB ID: 5AWH, coloured like in A). MID and PIWI domains of RsAgo are shown transparently. Reproduced from Manakova *et al.*, 2024[217].

Notably, the analysis of small non-messenger RNA (snmRNA) from *A. fulgidus* DSM 4304 identified a snmRNA named Afu-277, which is 80 nucleotides in length and located upstream of the AfAgo gene[218]. Though Afu-277 is remotely reminiscent of eukaryotic H/ACA small nucleolar RNAs

(snoRNAs), it could not be unambiguously assigned to *bona fide* H/ACA RNAs. It is proposed here that Afu-277 is merely a fragment of the longer version of mRNA transcribed from the upstream gene. The 2 bp deletion in *A. fulgidus* DSM 4304 strain introduces a STOP codon and promotes degradation of the untranslated 3′-region of mRNA that remains unprotected from RNases[219]. On the other hand, the presence of Afu-277 RNA confirms active transcription of the AfAgo-encoding region in the *A. fulgidus* DSM 4334 host. This prompted testing if the full-length upstream protein forms a functional SiAgo/SiAga1-like heterodimeric complex with AfAgo[139].

3.3.1. The AfAgo operon encodes an N-L1-L2 domain protein

To study the structure and function of the full-length upstream protein, the restored full-length open reading frame, with the missing TA nucleotides inserted, was cloned into an *E. coli* expression vector and the corrected protein was subsequently purified by liquid chromatography (Figure 34). Its structure was successfully solved by X-ray crystallography in two symmetry groups: *P*1 (resolution 1.85 Å) and $P3_221$ (resolution 1.4 Å) (see Methods, p. 53 and Appendix 4 for details). Two full-length upstream protein chains found in the asymmetric unit of the *P*1 structure, and a single chain found in the asymmetric unit of the $P3_221$ structure overlap with an RMSD of 0.85 Å (determined using PDBe Fold[220], Figure 33b). The structurally closest proteins to the full-length upstream protein determined by DALI search (as a search model, a high-resolution structure $P3_221$ 1.4 Å was used) are long-B RsAgo (PDB ID: 6D8P, Z-score 12.1) and long-A CbAgo (PDB ID: 6QZK, Z-score 10.1), confirming that it is a structural equivalent of N, L1 and L2 domains that are typical for the long pAgos.

3.3.2. AfAgo and AfAgo-N form a heterodimer

To test if AfAgo and AfAgo-N form a functional complex, a pBAD expression vector was engineered where both AfAgo-N and AfAgo genes were placed under a single P_{BAD} promoter (Appendix 2) and co-expressed both proteins in *E. coli*. AfAgo-N carrying an N-terminal His_6-tag co-purified on the Ni^{2+}-affinity column with AfAgo, confirming that AfAgo-N and AfAgo proteins form a stable complex (Figure 34b – d). Mass photometry experiments indicated that the predominant population of AfAgo-N/AfAgo particles have a M_w that corresponds to the theoretical M_w of an AfAgo-N/AfAgo heterodimer (79.7 kDa) (Figure 34a).

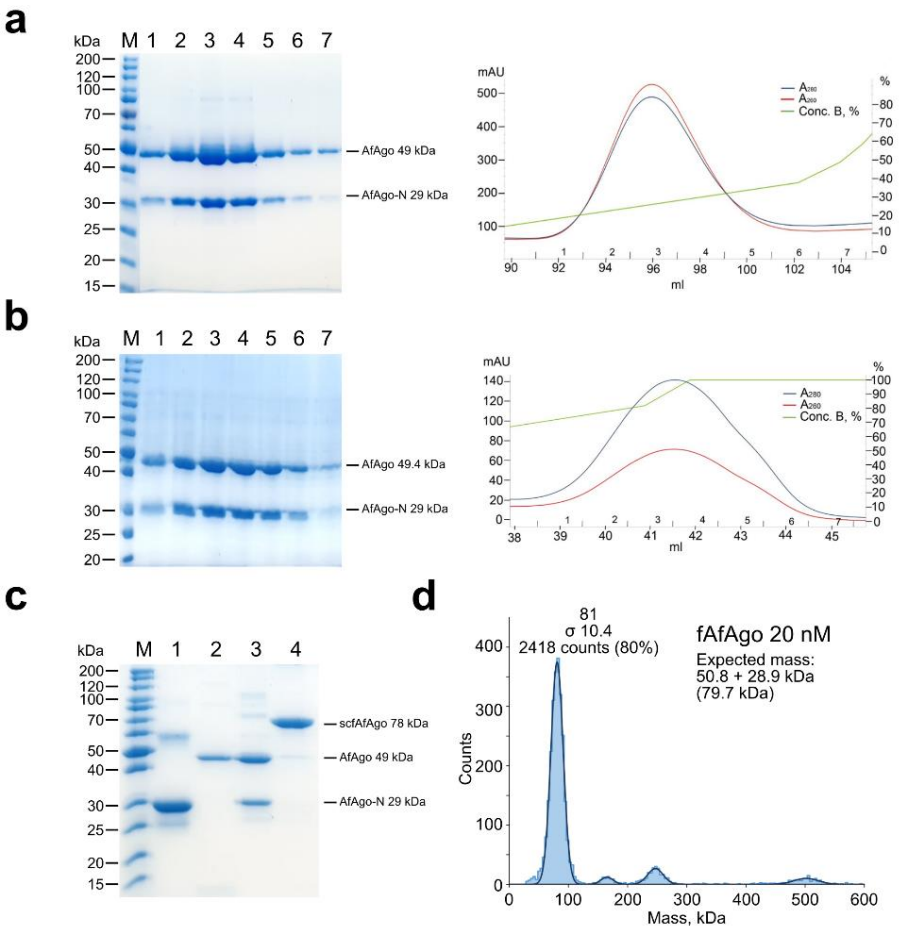


Figure 34. Purification and mass photometry of (sc)fAfAgo. (a) SDS-PAGE of fractions collected (left) after purification by chromatography through HisTrap HP chelating column; numbering corresponds to numbers of the collected fractions indicated below the curves in the chromatogram (right). (b) SDS-PAGE of fractions collected after purification by chromatography through HiTrap Heparin HP column (left); numbering corresponds to numbers of the collected fractions indicated below the curves in the chromatogram (right). (c) SDS-PAGE of purified proteins: M – marker, 1 – the full-length upstream protein (AfAgo-N), 2 – AfAgo, 3 – the heterodimeric complex of AfAgo-N and AfAgo (fAfAgo), 4 – single chain fAfAgo (scfAfAgo). (d) Mass photometry histograms of fAfAgo protein. fAfAgo exhibits a clear distribution centred around 81 kDa (80% of particles), which is close to the theoretical M_w of the heterodimeric complex. Reproduced from Manakova *et al.*, 2024[217].

Next, crystals of the AfAgo-N/AfAgo heterodimer (henceforth – full AfAgo or fAfAgo) were obtained in the presence of a 15 nt 5′-phosphorylated single-stranded DNA oligonucleotide and solved a 2.5 Å structure (Figure 35a). Unfortunately, attempts to crystallize fAfAgo with dsRNA and

RNA/DNA heteroduplexes were not successful in producing diffracting crystals. The structure contains a single fAfAgo heterodimer (the final model covers all AfAgo and 18 – 245 AfAgo-N residues) and fragments of two ssDNA oligonucleotides per asymmetric unit. Even though the oligonucleotide used for crystallization was not self-complementary, its two strands formed a short duplex fragment via base-pairing of the self-complementary stretches 2 – 5 (5′-P-ATCGACCAGGCTACG, Figure 36), forming a guide/target-like heteroduplex, which was accommodated in the nucleic acid binding cleft of the fAfAgo heterodimer. An additional Watson-Crick pair was also observed in the hairpin-like structure at the 3′-end of the DNA oligonucleotide occupying the target strand position (Figure 36).

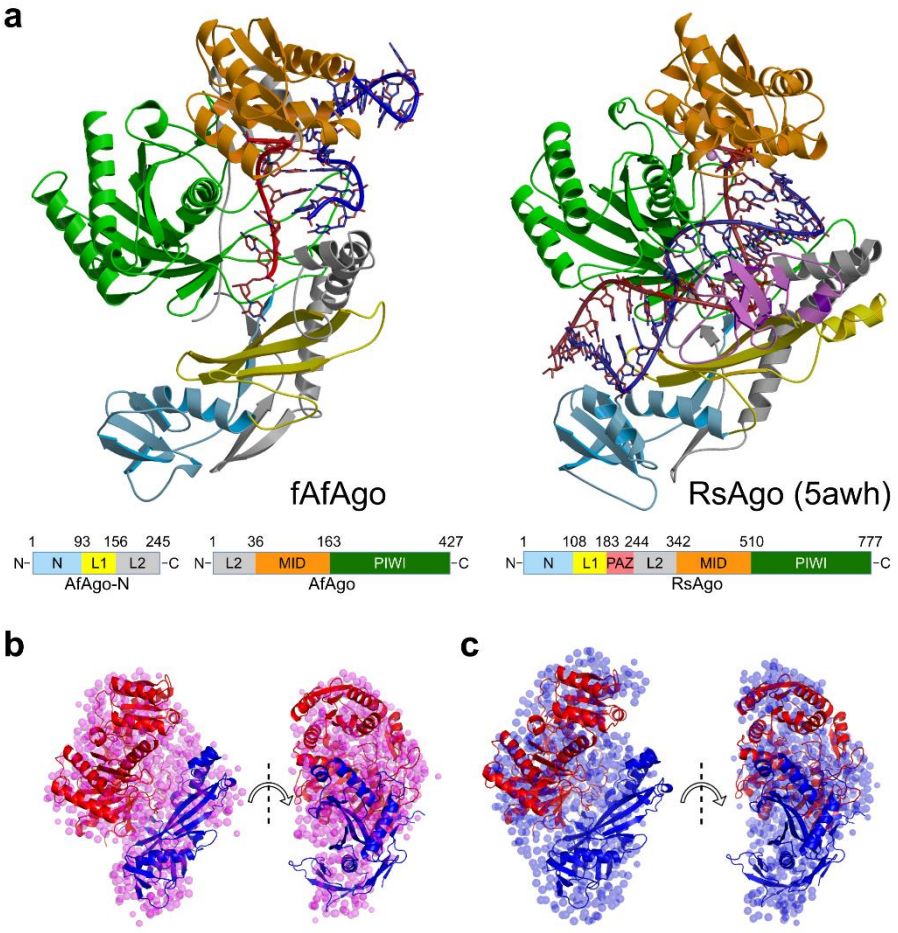


Figure 35. Structural comparison of fAfAgo, scfAfAgo, and RsAgo. (a) Top – crystal structure of fAfAgo-DNA complex (left) compared to RsAgo (right, PDB ID: 5AWH). Guide NA strands are coloured red, and target DNA strands – blue. Bottom – domain architecture of

AfAgo-N and AfAgo. The domain architecture of RsAgo (PDB ID: 5AWH) is shown for comparison. (b) Shape reconstruction based on SAXS data (magenta coloured spheres) of fAfAgo complex with 14 bp DNA (SASBDB ID: SASDRX8) compared with the crystal structure of fAfAgo (AfAgo – red, AfAgo-N – blue). (c) Shape reconstruction based on SAXS data (blue coloured spheres) of scfAfAgo complex with 14 bp DNA (SASBDB ID: SASDRY8) compared with the crystal structure of fAfAgo (AfAgo subunit is coloured red, AfAgo-N – blue). Reproduced from Manakova *et al.*, 2024[217].

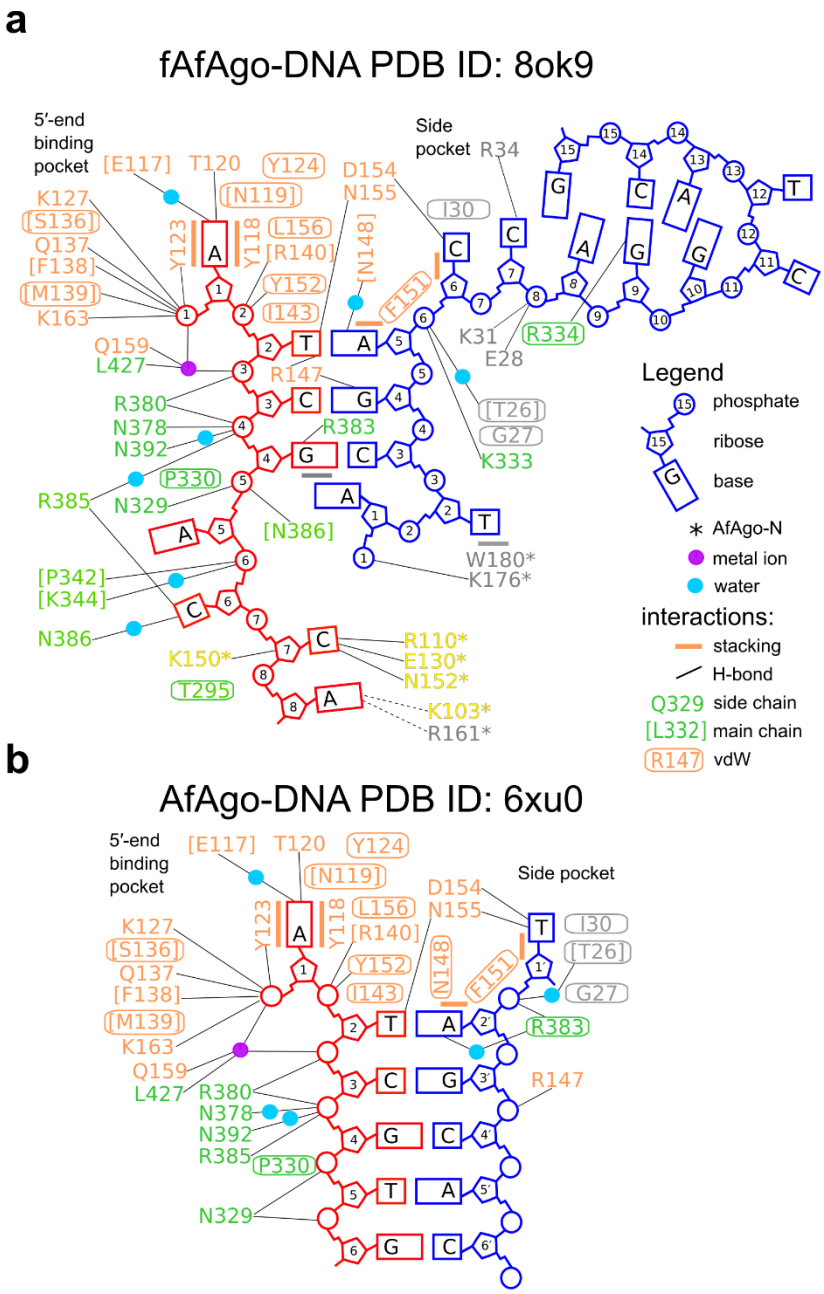


Figure 36. Conservative contacts of AfAgo with 5′-end of the guide (red) and 3′-end of the target (blue) strands. Residues are colour-coded as in Figure 35a: L1 – yellow; L2 – grey; MID – orange; PIWI – green. (a) Scheme of fAfAgo contacts with DNA. The 5′-end base of the guide is displaced into a binding pocket three guide-target base pairs form a short duplex in the crystal structure. An additional Watson-Crick pair was also observed in the hairpin-like structure at the 3′-end of the DNA oligonucleotide occupying the target strand position. (b)

Scheme of contacts with DNA in crystal structure PDB ID 6XU0, protein chain A. Reproduced and modified from Manakova *et al.*, 2024[217].

The contact surface area between two proteins in the complex, as calculated by the PISA server[212], is 1380 $Å^2$, and includes 11 H-bonds, indicative of a stable complex. The AfAgo heterodimer can be aligned with the full-length Argonaute RsAgo (Figure 35a). As predicted, AfAgo superimposes with MID and PIWI domains of RsAgo, whereas AfAgo-N superimposes with the N-terminal domain and the L1-L2 linker domains. Only the RsAgo PAZ domain has no structural counterpart in the fAfAgo complex (Figure 33c, Figure 35a)[26]. The phosphorylated 5′-end of the guide DNA strand is bound in a pocket formed by the MID and PIWI domains, as previously observed for an individual AfAgo protein (Figure 36)[25]. The first base of the guide strand is flipped out of the duplex into the pocket and is stacked between two tyrosine side chains (Y118 and Y123). Similar base-specific contacts were observed with the first and second guide bases like in crystal structures of AfAgo with dsDNA (PDB IDs: 6T5T, 6TUO, 6XU0, 6XUP[216]). The first four guide nucleotides (5′-ATCG) in fAfAgo structure (5′-ATCG) and recent structures of AfAgo with dsDNA (PDB ID: 6XU0 and 6XUP[216]) coincide. Nucleotides $2-4$ form a short duplex with complementary bases of the target-like strand, but the first target nucleotide flipped in the protein pocket in the AfAgo complex is C instead of T, which would be complementary to the first A of the guide strand. Interactions with the second guide and target bases in fAfAgo are similar to the AfAgo dimer structures with dsDNA[216]. In fAfAgo, the bridging of the first target base with the second guide base via N155, which is described in Manakova *et al*., 2023[216], is absent (Figure 36). In all structures, the duplex strands are separated by the insertion of Y152 and F151 aromatic residues.

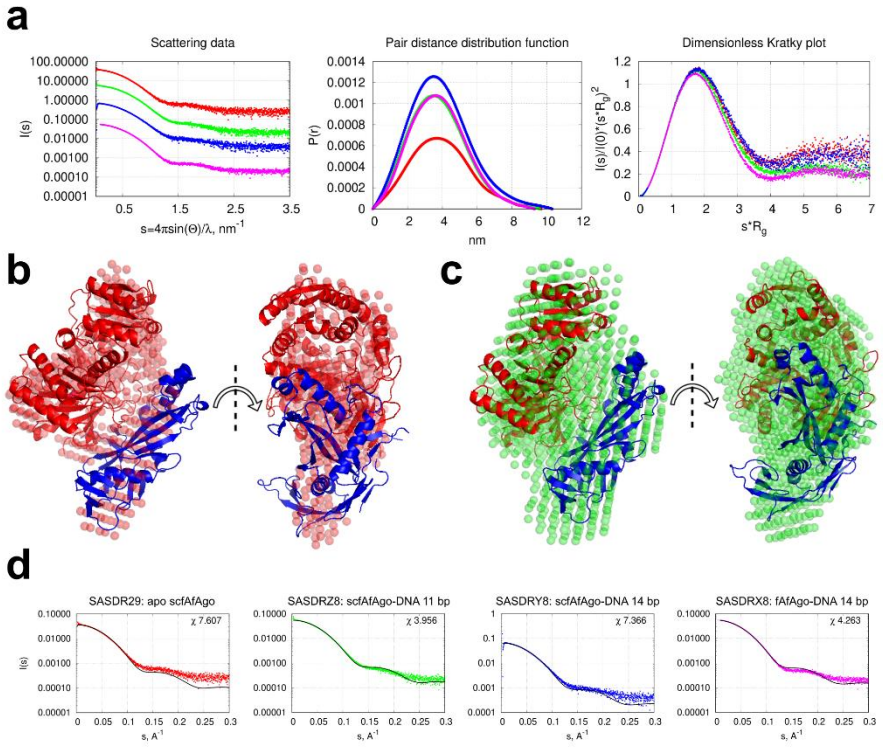


Figure 37. SAXS of (sc)fAfAgo. (a) SAXS data of scfAfAgo (apo – red, SASBDB ID: SASDR29), complex with 11 bp DNA – green (SASBDB ID: SASDRZ8), complex with 14 bp DNA MZ-1288 – blue (SASBDB ID: SASDRY8) and SEC-SAXS data of fAfAgo complex with 14 bp DNA (MZ-1288) – magenta (SASBDB ID: SASDRX8). Scattering curves are displaced along the *y*-axis for clarity. (b) Apo scfAfAgo (SASBDB ID: SASDR29) shape reconstruction from SAXS data (red spheres). (c) SAXS shape reconstruction of scfAfAgo with 11 bp DNA oligoduplex (SASBDB ID: SASDRZ8, green spheres). SAXS shapes superimposed with the crystal structure of fAfAgo (AfAgo subunit coloured red, AfAgo-N – blue). (d) Comparison of SAXS data with scattering function of fAfAgo complex crystal structure calculated by CRYSOL (black curve). DNA was excluded from the calculation. SASBDB IDs are indicated. SAXS experimental data are coloured as indicated in (a). Reproduced and modified from Manakova *et al.*, 2024[217].

The overall double-stranded nucleic acid position in the crystal structures obtained previously with an isolated AfAgo protein, for example [26] PDB ID: 6T5T [216], differs from that observed in the fAfAgo complex (Figure 38a). In AfAgo PDB ID: 6T5T, only the 5′-end of the DNA duplex contacts AfAgo, the remaining part pointing away from the protein. In contrast, the fAfAgo heterodimer forms a channel for guide-target strands that is similar to the channel present in other long pAgos, e.g., RsAgo (PDB ID: 5AWH[13], 6D8P[16]), which directs the bound double-stranded nucleic acid (RNA/DNA

heteroduplex in RsAgo) towards the MID domain. Despite the very short length of duplex-like DNA observed in the fAfAgo structure, its overall position matches guide/target strands in the RsAgo complex[13,16].

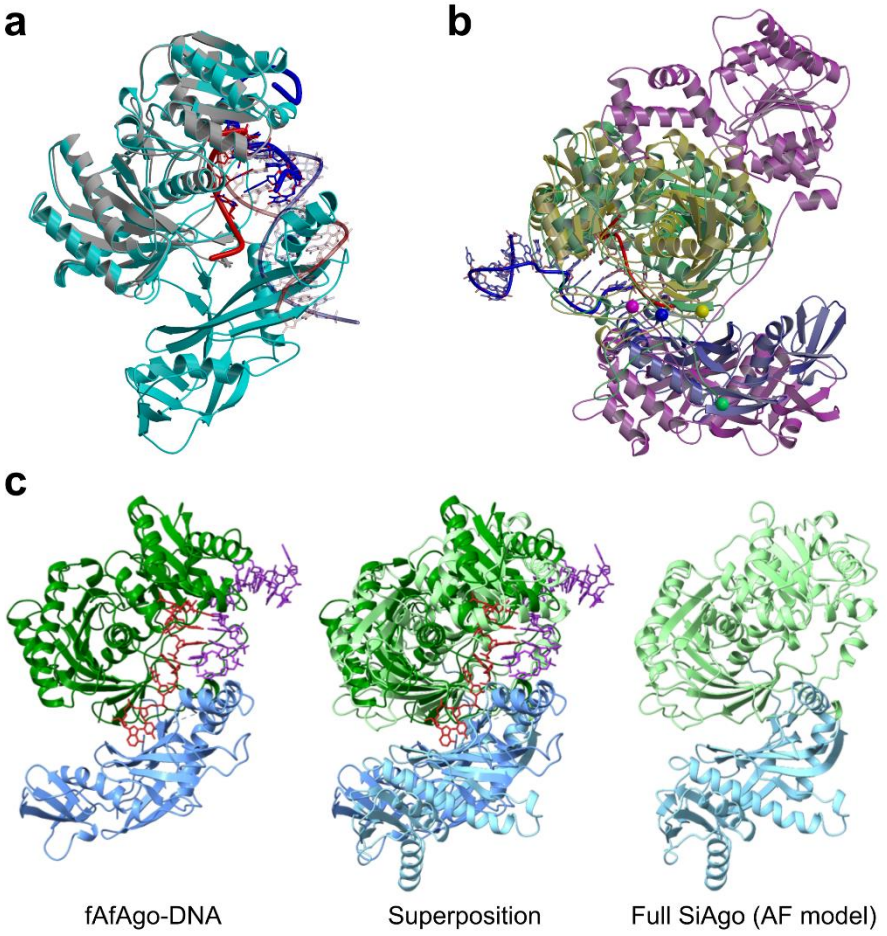


Figure 38. Structural comparison of fAfAgo with GsSir2/Ago and full SiAgo. (a) fAfAgo complex forms a channel for guide-target strands. The crystal structure of AfAgo with dsDNA (cyan, PDB ID: 6T5T) is compared with the fAfAgo complex (grey). Guide strands are red, target strands are blue. DNA bound in 6T5T structure is shown by thick lines. (b) N- and C-termini of both proteins in the fAfAgo complex are located close to each other. A similar situation was observed in the AlphaFold model of GsSir2/Ago[9]. GsSir2/Ago is coloured magenta (Sir2-APAZ protein) and green (Ago protein). The N-terminal residue of GsAgo is shown as a green sphere, magenta sphere corresponds to the C-terminus of the GsSir2-APAZ protein. The distance between these residues is 30 Å, but it should be noted that the N-terminal part of the GsAgo protein is modelled in an elongated conformation, whereas in the real protein, it could be folded in another way. In the fAfAgo complex, the AfAgo protein is coloured yellow, N-terminal residue of AfAgo is shown by a yellow sphere. AfAgo-N protein is coloured

blue, its C-terminal atom is shown as a blue sphere. The distance between these two points is 12 Å. (c) Comparison of full AfAgo bound to DNA with AlphaFold model of full SiAgo (SiAgo-Aga1 heterodimer). The N-terminal region is coloured light blue, MID-PIWI – green. Structural similarity between full AfAgo and full SiAgo based on Dali comparison of individual chains:
N-lobe: 185 residue pairs superimposed, resulting RMSD=4.6 Å, sequence identity=13%
MID-PIWI: 381 residue pairs superimposed, resulting RMSD=3.1 Å, sequence identity=16%.
Reproduced and modified from Manakova *et al.*, 2024[217].

In the fAfAgo structure, the AfAgo-N subunit sterically masks the AfAgo surface that was previously implicated in AfAgo homodimerization[145], making AfAgo oligomerization across this surface unlikely. This is in line with AFM smFRET data. Results show that, unlike the previously studied AfAgo, fAfAgo does not form looped complexes with dsDNA (Figure 39, Figure 40).

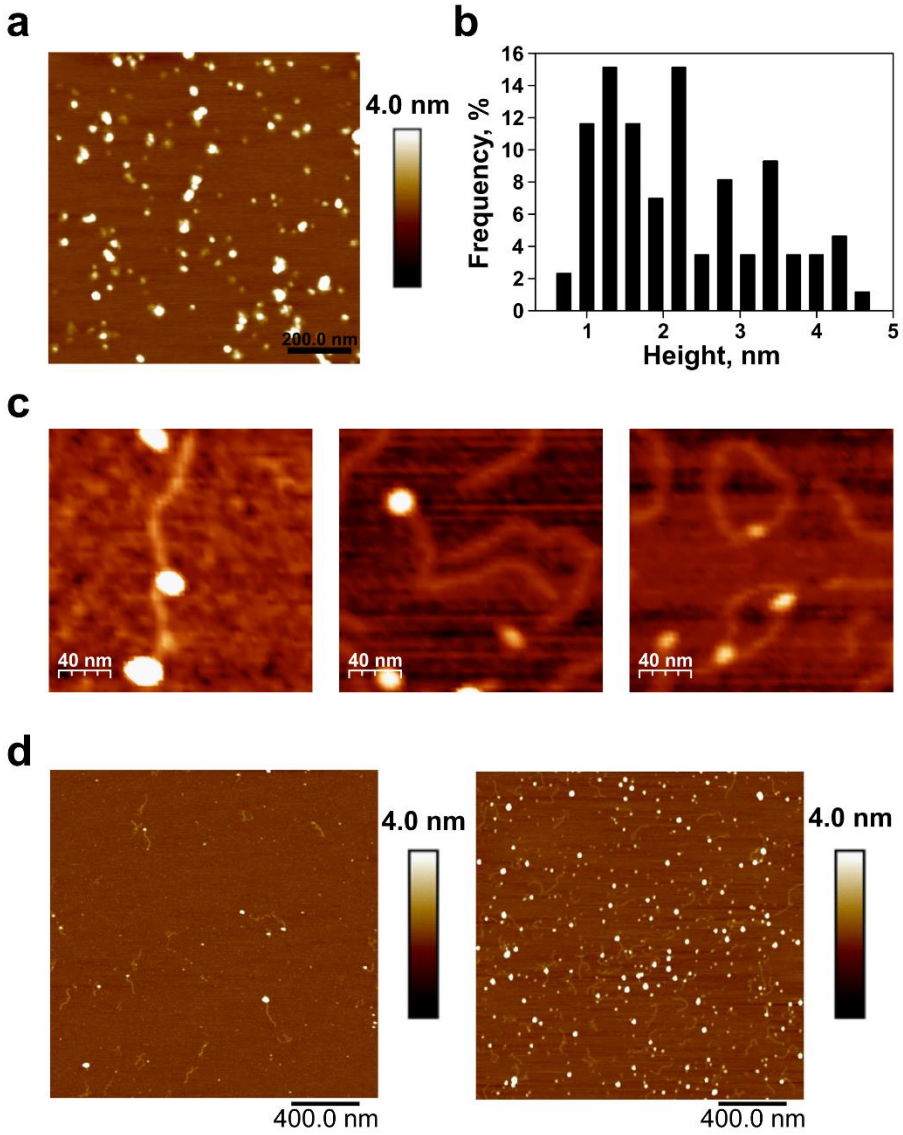


Figure 39. Representative AFM topography image (a) and height distribution (b) of unbound fAfAgo protein complexes (n = 86) adsorbed on freshly cleaved mica. (c) Representative images of protein complex:DNA structures (from left to right): fAfAgo bound on random DNA strand locations; fAfAgo bound to the end of DNA strand; ring-shaped structures. (d) AFM topography image of adsorbed fAfAgo:DNA complexes on APS-modified mica prepared by quenching (left) and not quenching (right) the crosslinker. Area of interest 2 μm × 2 μm. Reproduced from Manakova *et al.*, 2024[217].

Interestingly, in the AlphaFold model and in recently published cryo-EM structures of GsSir/Ago and MapTIR/Ago[9,116,137], the C terminus of the

APAZ protein is adjacent to the N-terminus of the Ago subunit (Figure 38b), implying that AfAgo and AfAgo-N evolved from a single polypeptide. A similar arrangement of the respective termini is also observed in the structural model of the *bona fide* short pAgo GsAgo with its effector protein GsSir2 (Figure 38b)[9]. Moreover, structures of fAfAgo complexes described here also demonstrate a similar arrangement of the respective termini, suggesting that polypeptide splitting at a similar structural/functional region occurred independently in the phylogenetically distant long-B (exemplified by fAfAgo) and short (exemplified by GsAgo) pAgos.

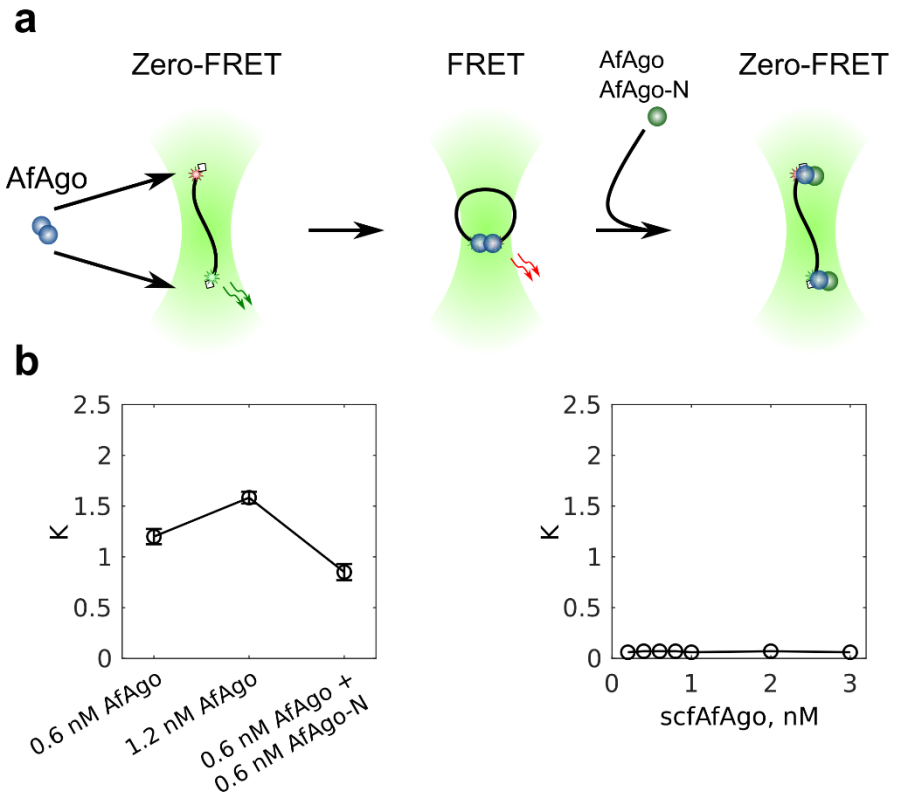


Figure 40. Competitive single-molecule FRET results with dsDNA in solution. (a) Schematic representation of suggested protein-DNA interactions. (b) Supplementing 0.6 nM of AfAgo-N to high-efficiency FRET reaction (DNA + 0.6 nM AfAgo) lowers FRET efficiency, presumably by disrupting the looped AfAgo-DNA complex formed by AfAgo dimerization (left). scfAfAgo does not form dimers and, therefore, looped dsDNA complexes (right). Reproduced from Manakova *et al.*, 2024[217].

Polypeptide splitting is consistent with the fact that some short pAgos, consisting only of MID and inactive PIWI domains, can be fused into a single

polypeptide with an effector protein containing an APAZ(N-L1-L2) domain[9]. To study the single-chain fAfAgo variant, in which the AfAgo-N and AfAgo proteins are fused into a single polypeptide (scfAfAgo), the synthesized gene encoding the AfAgo-N/AfAgo fusion was expressed in *E. coli* and the scfAfAgo protein was purified (Appendix 2, Figure 34). SAXS measurements confirmed that fAfAgo heterodimer and scfAfAgo adopt a similar shape in solution (Figure 35b, c, Figure 37). In addition, scfAfAgo, unlike AfAgo, does not form looped complexes with dsDNA at all (Figure 40), since it, like fAfAgo, has a hidden homodimerization surface.

3.3.3. fAfAgo uses an RNA guide to bind a DNA target

Next, the nucleic acids bound to fAfAgo, scfAfAgo and AfAgo *in vivo* were characterized. To obtain AfAgo-bound nucleic acids, fAfAgo-NA, scfAfAgo-NA and AfAgo-NA complexes were purified from *E. coli* transformed with pBADHisA_TwinStrep-AfAgo with pCDFDuet_His-AfAgo-N, pBAD_TwinStrep-scfAfAgo and pBAD_TwinStrep-AfAgo expression vectors, extracted NAs and subjected to sequencing. Subsequent analysis revealed that the proteins are associated with small RNAs, with fAfAgo predominantly associating with molecules of the length of 29 nt, scfAfAgo – 20 nt, and AfAgo with 15 nt with a 5′-phosphate (Figure 41a and b). The decreased RNA length in AfAgo may be due to the ability of the protein to protect only the 5′-end of the NA, meanwhile, the fAfAgo heterodimer forms a channel for NAs and can protect longer RNAs (Figure 38a). fAfAgo-, scfAfAgo-, and AfAgo-associated small RNAs show a pronounced tendency to bear the 5′-AU dinucleotide (Figure 41b), implying that the proteins use small RNAs containing the 5′-AU dinucleotide as guides, which is in line with the previous results that AfAgo preferentially uses gRNA to recognize complementary tDNA[216].

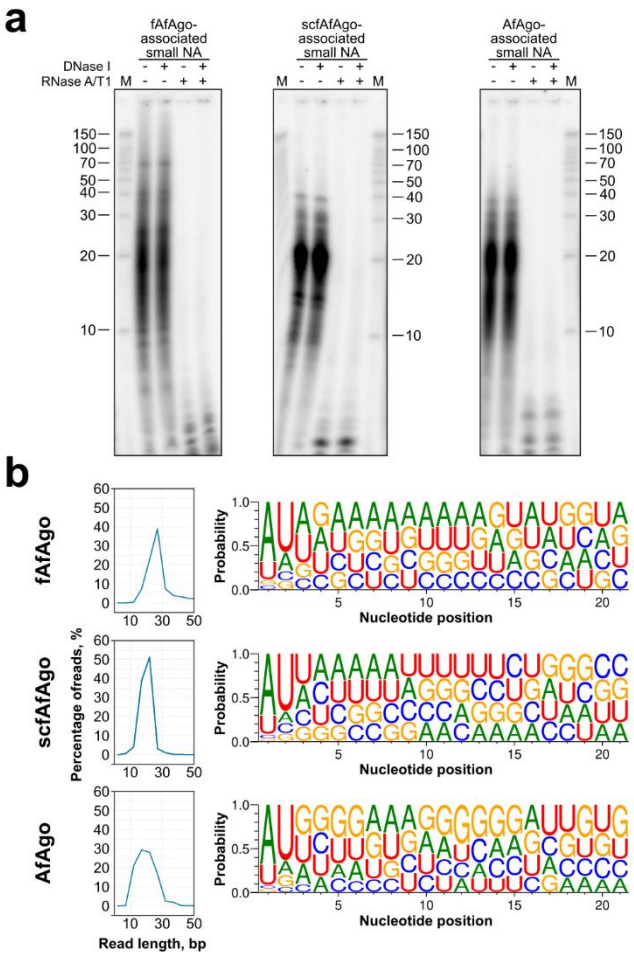


Figure 41. Nucleic acid binding by fAfAgo, scfAfAgo and AfAgo *in vivo*. (a) fAfAgo, scfAfAgo and AfAgo co-purifies with small RNAs. Nucleic acids that co-purified with fAfAgo, scfAfAgo and AfAgo were [γ-^{32}P] ATP radiolabelled and treated with DNase I or RNase A/T1, or both, and resolved on a denaturing polyacrylamide gel. M, RNA ladder Decade Marker System (Ambion). (b) Length distribution (left) of small RNA co-purified with fAfAgo, scfAfAgo and AfAgo as determined by sequencing. Small RNAs associated with all AfAgo forms show 5′-AU preference (right). Reproduced from Manakova *et al.*, 2024[217].

Therefore, the nucleic acid binding properties of fAfAgo, scfAfAgo, and AfAgo were characterized *in vitro*. To determine whether nucleic acid binding preferences are preserved throughout all variants of AfAgo, EMSA experiments with single- and double-stranded substrates were performed (Figure 42). It is evident that both fAfAgo and scfAfAgo bind single-stranded nucleic acids with greater affinity than AfAgo. Binding affinity to double-stranded substrates by all 3 proteins is less variable but in all cases, a notable

preference for dsDNA and RNA/DNA heteroduplex substrates over dsRNA is observed.

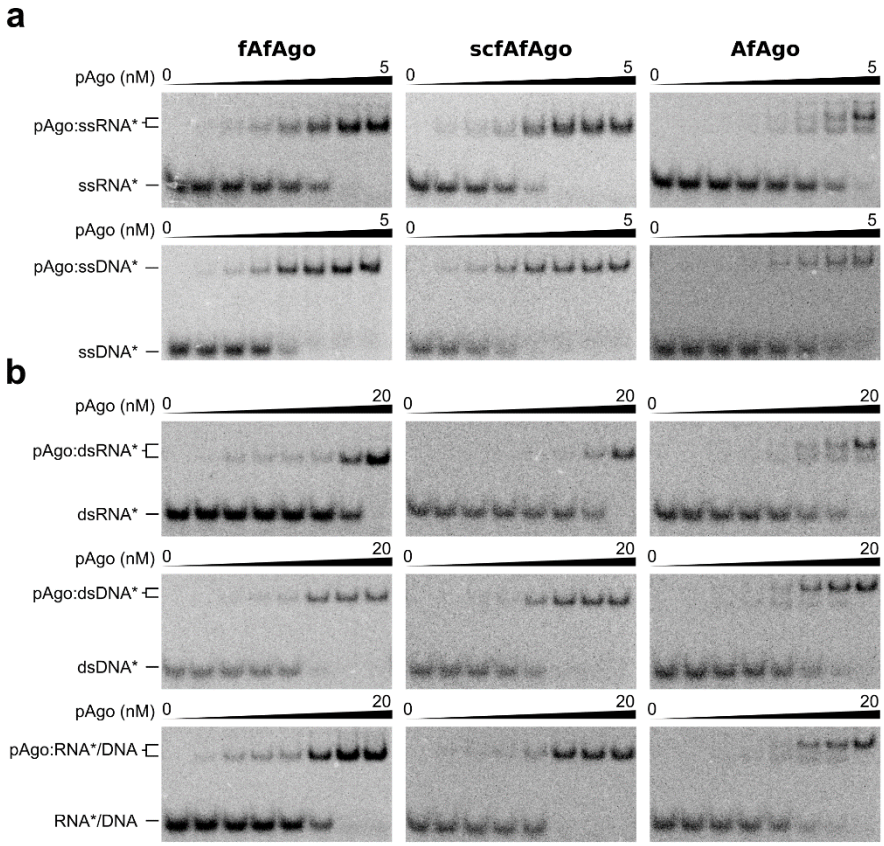


Figure 42. EMSA results. (a) Single-stranded substrate (ssRNA and ssDNA) binding by fAfAgo, scfAfAgo and AfAgo. Protein concentrations in lanes: 0, 0.05, 0.1, 0.2, 0.5, 1, 2, 5 nM. (b) Double-stranded substrate (dsRNA, dsDNA and RNA/DNA) binding by fAfAgo, scfAfAgo and AfAgo. Protein concentrations in lanes: 0, 0.1, 0.5, 1, 2, 5, 10, 20 nM. Nucleic acid binding by AfAgo results in two binary complex populations, possibly due to homodimerization. Reproduced from Manakova *et al.*, 2024[217].

To obtain a more accurate representation of nucleic acid binding by pAgos, which are usually first loaded with a single-stranded guide strand that is subsequently used to recognize the complementary target strand, a variation of EMSA was performed, where instead of apo-protein, the concentration of pAgo preloaded with the guide strand (pAgo:guide binary complex) was varied. It was found that for both fAfAgo and scfAfAgo the optimal combination is an RNA guide and a complementary DNA target since all other

guide-target combinations did not result in full target binding (Figure 43a). In line with previous results, AfAgo does not exhibit such clear selectivity, but its target binding is evidently weaker in comparison to fAfAgo and scfAfAgo[216]. The affinity of all tested binary pAgo:gRNA complexes to DNA targets also depends on the guide:target sequence complementarity, as low complementarity tDNA strands yielded no detectable ternary complexes (Figure 43b). Taken together, the EMSA studies suggest that fAfAgo and scfAfAgo are both more efficient at the gRNA-guided binding of complementary tDNA than the standalone AfAgo[216].

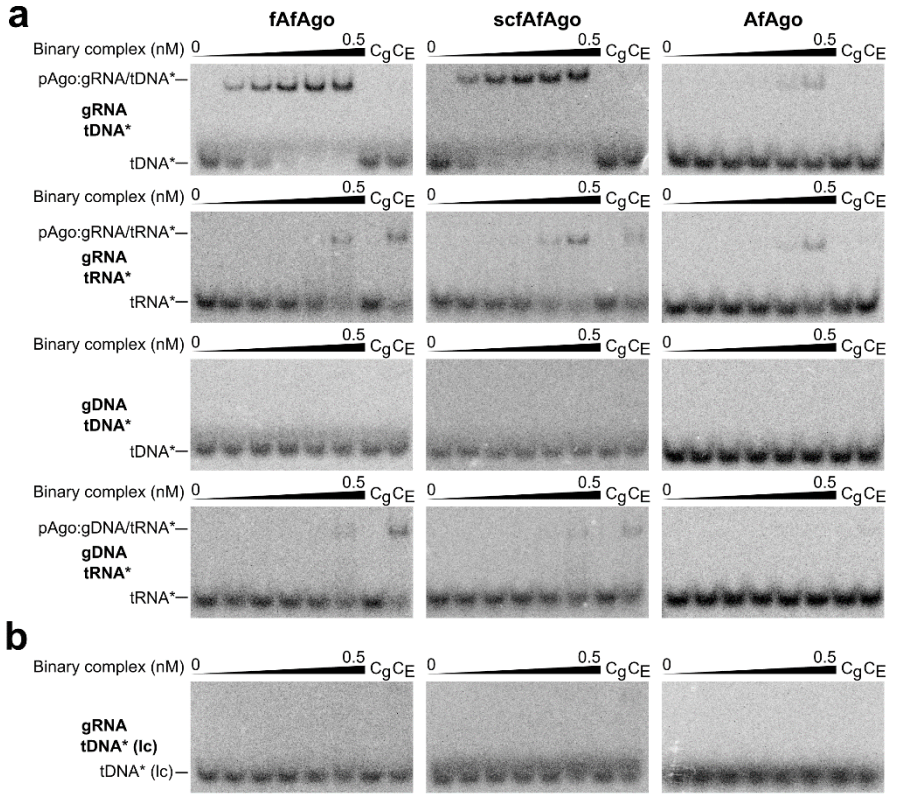


Figure 43. RNA-guided NA targeting. (a) Binding of 5′-^{32}P labelled DNA and RNA targets by fAfAgo, scfAfAgo and AfAgo binary complexes preloaded with 5′-P RNA or DNA guides (molar pAgo:guide ratio 1:2). Targets used are complementary to 1 – 8 nt of the guide. (b) Binding of 5′-^{32}P labelled low complementarity DNA target by fAfAgo, scfAfAgo and AfAgo binary complexes preloaded with 5′-P RNA or DNA guides (molar pAgo:guide ratio 1:2). Low complementarity DNA target is only complementary to 4 –7 nt of the guide. All experiments depicted were carried out in the same binary complex concentration range: 0, 0.005, 0.01, 0.02, 0.1, 0.5 nM. C_g control contains the highest guide concentration used (1 nM)

and radiolabelled target, C_E control contains the highest pAgo concentration used (0.5 nM) and radiolabelled target. Reproduced from Manakova *et al.*, 2024[217].

To investigate the proposed RNA-guided DNA targeting mechanism of AfAgo and variants using smFRET, a DNA fragment with a biotin label for surface immobilization was utilized (Figure 44a). This DNA fragment was paired with a complementary RNA guide that had been pre-assembled with the AfAgo protein. An acceptor fluorophore on the DNA and a donor fluorophore on the RNA were strategically placed at positions that would bring the two fluorophores into close proximity upon AfAgo-mediated RNA-DNA hybridization, facilitating efficient Förster resonance energy transfer (FRET) between them. Fluorescence movies capturing the interactions between individual surface-immobilized DNA fragments and AfAgo-RNA complexes were recorded under various conditions, including different AfAgo proteins and RNA-DNA complementarities of 8 and 0 base pairs. Figure 44b illustrates representative examples of the obtained FRET pair intensity trajectories along with corresponding trajectories of the proximity ratio E. In these trajectories, correlated and sudden changes in donor and acceptor intensities were observed, indicating the association and dissociation events between the AfAgo-RNA complex and the DNA fragment. When using DNA with 0 base pair complementarity, the periods of interaction are brief, suggesting weaker binding. Importantly, a control experiment involving DNA interaction with RNA alone did not yield trajectories showing such coordinated changes in fluorophore intensity (Table 6).

Table 6. Single-molecule FRET data. % interacting – a percentage of detected traces with high acceptor intensity. Fraction high E – a fraction of time with high acceptor intensity, calculated from all registered trajectories. fAfAgo-gRNA and scfAfAgo-gRNA display heightened FRET efficiencies and longer FRET durations with the 8 bp complementary target, compared to the non-complementary target and to gRNA alone, hinting at RNA-guided DNA targeting. RsAgo, used as a control, displays no such activity under the conditions tested. Reproduced from Manakova *et al.*, 2024[217].

Sample	% interacting	Fraction high E
AfAgo-gRNA 1 nM with 8 bp complementary target	6	0.027
AfAgo-gRNA 10 nM with 8 bp complementary target	34	0.110
AfAgo-gRNA 1 nM with non-complementary target	1	0.005
AfAgo-gRNA 10 nM with non-complementary target	8	0.021

Sample	% interacting	Fraction high E
fAfAgo-gRNA 1 nM with 8 bp complementary target	28	0.066
fAfAgo-gRNA 10 nM with 8 bp complementary target	52	0.107
fAfAgo-gRNA 1 nM with non-complementary target	8	0.003
fAfAgo-gRNA 10 nM with non-complementary target	11	0.016
scfAfAgo-gRNA 1 nM with 8 bp complementary target	37	0.283
scfAfAgo-gRNA 10 nM with 8 bp complementary target	49	0.375
scfAfAgo-gRNA 1 nM with non-complementary target	15	0.016
scfAfAgo-gRNA 10 nM with non-complementary target	22	0.089
gRNA 10 nM with 8 bp complementary target	0	0.006
RsAgo-gRNA 50 nM with 8 bp complementary target	-	-

To evaluate the average values of the various E levels observed in individual trajectories, the selected E trajectories were aggregated and the resulting average within the initial 10 seconds was calculated (Figure 44c). The analysis of single-molecule populations and the time-averaged E revealed prominent EFRET peaks at 0.5, 0.25, and 0.25 for AfAgo, fAfAgo, and scfAfAgo, respectively. For comparison, corresponding E distributions for DNA fragments with zero complementarity to RNA were also included. In these distributions, the high EFRET peak is noticeably less pronounced than in the corresponding distributions for the 8 base pair complementary DNA fragment, suggesting that RNA complementarity to DNA enhances their interaction. Additionally, considering that DNA interaction with RNA alone could not be detected, these findings strongly suggest that AfAgo facilitates RNA-DNA interactions.

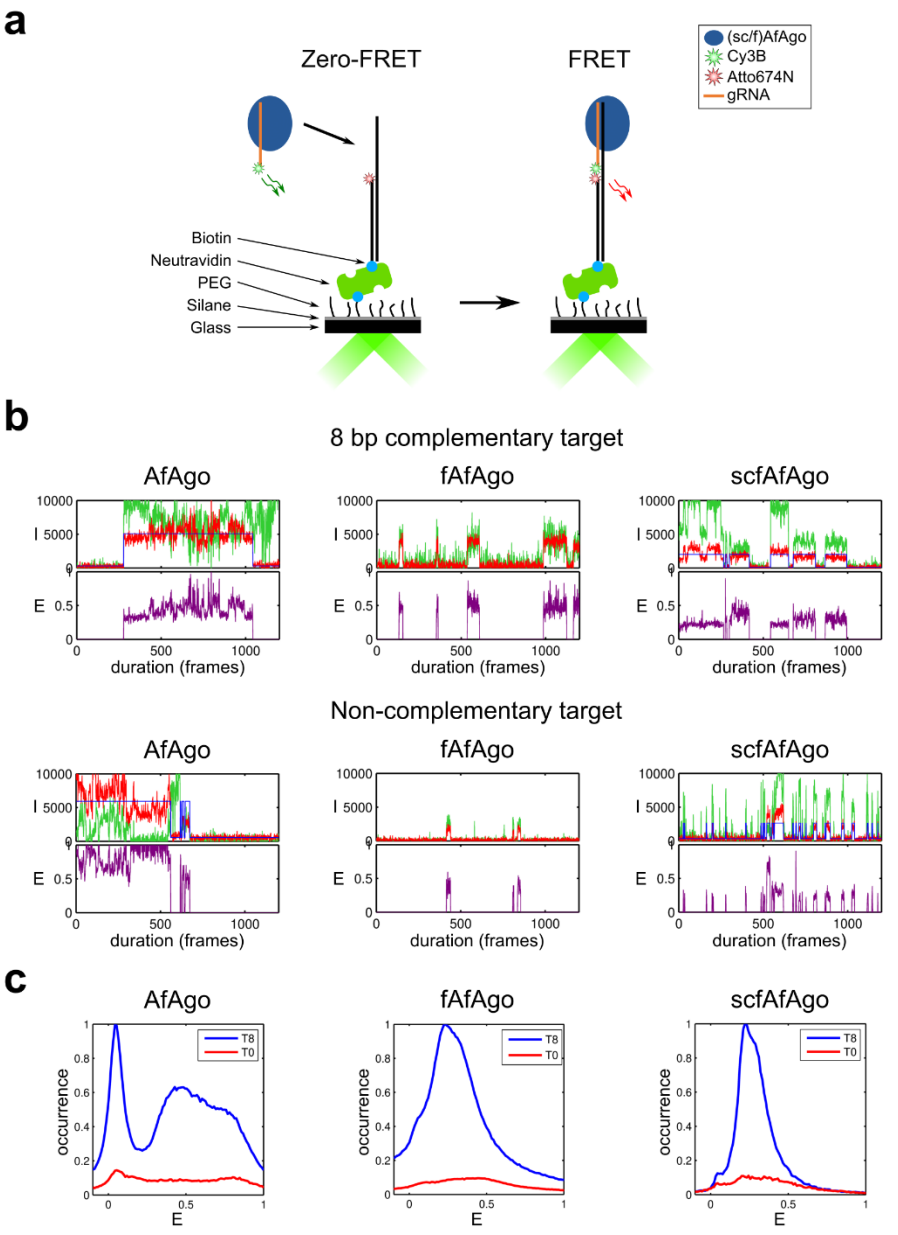


Figure 44. Single-molecule FRET studies of gRNA-loaded AfAgo, fAfAgo, and scfAfAgo interactions with surface-immobilized target DNA. (a) Schematic representation of the experimental setup. (b) Representative traces of donor (green) and acceptor (red) intensities and the corresponding proximity ratio E for 8 bp-complementary (top) and non-complementary (bottom) DNA targets. (c) Distributions of FRET efficiency for 8 bp-complementary (blue) and non-complementary (red) DNA targets. The distributions are scaled

so that the ratio of their areas is equal to the ratio of the relative occurrences of high EFRET. Reproduced from Manakova *et al.*, 2024[217].

3.3.4. AfAgo homologues include both split and single-chain pAgos with various degrees of PAZ reduction

To better understand the relationship between AfAgo and other long-B pAgos, a non-redundant set of AfAgo homologues that have associated genomic DNA sequences was collected. The reasoning was that other so-called "truncated" long-B pAgos, similarly to AfAgo, may also have an upstream gene encoding for the missing N-terminal part. A further aim was also to explore whether long-B pAgos are part of putative operons similar to AfAgo, which is associated not only with the upstream AfAgo-N but also with a downstream gene (Af1317), encoding an uncharacterized protein.

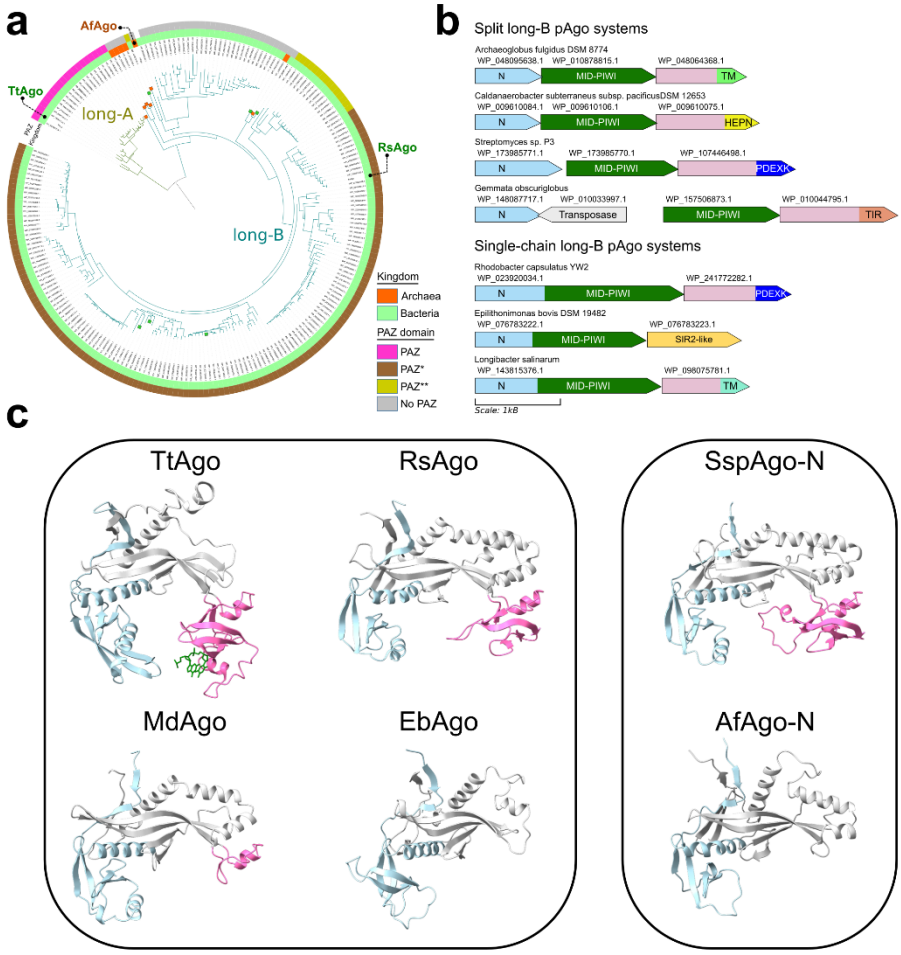


Figure 45. Loss of PAZ domain in pAgos. (a) Phylogenetic tree of 220 pAgo proteins constructed with Fasttree and rooted arbitrarily (midpoint root). Coloured squares on branch leaves represent split pAgos (N and MID-PIWI regions are separate proteins). Branches for long-A and long-B pAgos in the tree are shown in different colours. Archaeal and bacterial proteins are indicated on the inner circle using orange and green colours respectively. The presence and the type of the PAZ domain is indicated on the outer circle: magenta, classical PAZ as in long-A TtAgo; brown, incomplete PAZ domain (PAZ*) as in RsAgo; yellow, remnants of PAZ domain (PAZ**); grey, PAZ domain is absent. (b) Composition of representative split and single-chain pAgo systems taken from the MID-PIWI phylogenetic tree. Coloured tags represent genes coding for N-domains of split pAgos (light blue), MID-PIWI (dark green) and putative effectors (a mixture of pink and other colours). Putative effectors have a common alpha-helical region (pink) followed by the effector domain. TM, putative transmembrane domain; HEPN, nuclease domain; PDEXK, PD-(D/E)XK nuclease domain, TIR, TIR domain, Sir2-like, Sir2 family protein. Protein IDs are indicated above each gene. (c) Examples of the PAZ domain present/absent in the structures of both single-chain and split pAgos. PAZ domain or its remnant is shown in magenta, the N-domain is coloured light

blue, and the L1 and L2 regions are coloured grey. In the case of single-chain pAgos, the MID-PIWI region is removed for clarity. Left – single-chain long-A pAgo: TtAgo, *Thermus thermophilus* Ago (PDB ID: 3DLH); single-chain long-B pAgos: RsAgo, *Rhodobacter sphaeroides* Ago (PDB ID: 5AWH); MdAgo, *Maribacter dokdonensis* Ago (WP_074671526.1, AF structural model AF-A0A1H4M537-F1); EbAgo, *Epilithonimonas bovis* Ago (WP_076783222.1, AF structural model AF-A0A1U7PYU4-F1). Right – split long-B pAgos: SspAgo-N, *Streptomyces* sp. P3 Ago-N (WP_173985771.1) structure derived from the full Ago heterodimeric AF model (WP_173985771.1 + WP_173985770.1); AfAgo-N, *A. fulgidus* Ago-N subunit (structure determined in this study). Reproduced from Manakova *et al.*, 2024[217].

First, a phylogenetic tree was constructed that included closely related AfAgo homologues (long-B pAgos) and a small set of close TtAgo homologues representing long-A pAgos (Figure 45a). It was found that "truncated" long-B pAgos are present in both archaea and bacteria and are not confined to a single clade. Next, analysis of the genome neighbourhood revealed that similarly to AfAgo, these "truncated" pAgos upstream have a gene encoding the N-terminal part of a full-length long-B pAgo (Figure 45b). Thus, it appears that at least some of the "truncated" long-B pAgos represent split pAgo systems and that this splitting occurred multiple times independently. To further substantiate this finding, AlphaFold structural models were constructed for some of the split pAgo representatives and it was found that, similarly to full AfAgo, they form a heterodimeric structure.

Previously, it was observed that long-A pAgos (e.g., TtAgo) have a canonical PAZ domain, whereas long-B pAgos such as RsAgo have a reduced PAZ (often annotated as PAZ*), which lacks the structural elements for binding the 3′-end of the guide strand[4,13]. On the other hand, AfAgo-N characterized here lacks the PAZ domain altogether. To find out whether the presence of the PAZ domain depends on the type of a long-B pAgo (split or a single-chain), corresponding regions were explored using sequence alignments and structural models. Interestingly, it was found that both single-chain and split long-B pAgos may either have or lack the PAZ* domain (Figure 45c). Cases, where the PAZ* domain has undergone further reduction so that only a single α-helix is remaining, have also been found (Figure 45c). Taken together, these observations indicate that long-B pAgos, regardless of whether they are split or not, may have various levels of PAZ degradation including its complete loss.

Gene neighbourhood analysis additionally revealed a conserved association of long-B pAgos (both split and single-chain) with a protein-coding gene located immediately downstream (Figure 45b). Two major types

of these associated proteins were found. The first type, represented by Af1317, corresponds to a fusion of a conserved α-helical domain with an additional functional domain such as putative transmembrane (TM), TIR, HEPN or PD-(D/E)XK domain. The second type corresponds to a Sir2-like protein similar to the ThsA protein from the Thoeris antiphage defence system[115]. This observation suggests that catalytically inactive long-B pAgos may function with these associated proteins by regulating activities of their effector domains that are often toxic.

4. DISCUSSION

4.1 AfAgo dimerization

All characterized long Argonaute proteins interact with their RNA and/or DNA targets as monomers, binding a single copy of each guide and target nucleic acids. Surprisingly, as demonstrated here, AfAgo, a prokaryotic Argonaute from the hyperthermophilic archaeon *Archaeoglobus fulgidus*, follows a different mechanism, which involves homodimerization and simultaneous interaction with two guide-target nucleic acid duplexes.

The first finding is that AfAgo is a homodimer in all previously solved X-ray structures, including apo-protein, and complexes with RNA and DNA (Table 1). Two types of AfAgo dimerization interfaces formed by the C-terminal β-sheets are observed in the structures. Both types of interfaces bury a comparable solvent-accessible surface area (Table 1), but result in a distinct arrangement of AfAgo subunits relative to one another, which are termed "closed" and "open" dimers (Figure 18a and b, respectively)[145]. The "closed" type of AfAgo homodimer, formed when the interface involves both the N-terminal residues and the C-terminal β-strands (Figure 18a), provides a better fit to the SAXS data, suggesting that it is the major type of DNA-bound WT AfAgo dimer present in solution (Figure 19c). It remains to be determined if the alternative "open" dimer observed in several structures (Figure 18b and Table 1) was induced by crystal packing, or rather it is an alternative less abundant arrangement of AfAgo subunits that co-exists in solution at equilibrium with the "closed" form. As expected, the removal of the β-strands located at the intersubunit interface (variant AfAgoΔ) impaired AfAgo dimerization (Figure 19).

Simultaneous binding of WT AfAgo homodimer to both ends of a linear DNA fragment, one DNA end per AfAgo monomer, would result in a DNA loop. The formation of such looped DNA molecules upon incubation with WT AfAgo was directly visualized using AFM (Figure 21). As shown in Table 3, ring-shaped AfAgo-DNA complexes constitute the majority of all protein-DNA complexes detected. A considerable decrease in the fraction of looped DNA complexes in the case of the dimerization interface mutant AfAgoΔ (Table 3) provides further proof that DNA looping is indeed mediated by the dimeric form of AfAgo protein.

To further characterize WT AfAgo-DNA interaction in solution, single-molecule FRET measurements were performed (Figure 24) using a DNA fragment labelled with fluorescent dye close to DNA ends. The design of the DNA substrate ensured that binding of WT AfAgo dimer to both DNA ends would bring the fluorophores into close proximity, resulting in FRET. Comparison of donor/acceptor channel records for free DNA and DNA with either WT AfAgo or dimerization-impaired AfAgoΔ confirmed that only WT AfAgo efficiently forms DNA loops, yet again implying that DNA looping is mediated by dimeric AfAgo.

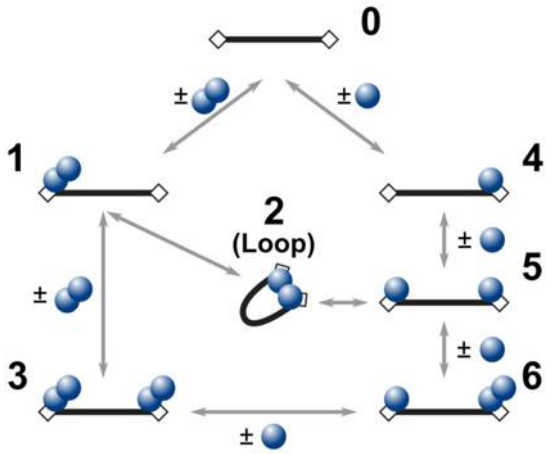


Figure 46. Kinetic schemes depicting possible reaction pathways between a two-target site DNA fragment and WT AfAgo. Black bars represent DNA, rectangular boxes – AfAgo-binding targets (phosphorylated DNA ends), and blue circles – AfAgo monomers. The numbering of various protein-DNA assemblies matches the numbering of complexes observed by AFM (Figure 21). Note that species 1 and 6 may be formed *via* alternative pathways not depicted in the scheme (species 1 may be formed when two monomers associate consecutively with the same DNA end; species 6 may be formed when a monomer and a dimer associate with different DNA ends). Reproduced from Golovinas *et al.*, 2021[145].

Assuming that WT AfAgo in its apo-form is an unstable dimer in solution (Figure 19a), at least two types of mechanisms can be proposed for the formation of the WT AfAgo dimer / looped DNA complex, one involving apo-AfAgo homodimers (Figure 46, left), the other involving apo-AfAgo monomers (Figure 46, right). In the first scenario, the reaction may proceed *via* (i) association of free DNA (species "0") with a single WT AfAgo dimer, which binds to one DNA end (species "1"); (ii) capture of the second DNA terminus by the DNA-bound AfAgo dimer in an intramolecular reaction, resulting in the looped complex (species "2"); (iii) alternatively, association

of the second WT AfAgo dimer with the unoccupied second DNA end of species “1” leads to species “3”, which is no longer capable of loop formation. Such a mechanism was demonstrated for many proteins capable of DNA looping, including restriction endonucleases [189,221,222] and transposases[223–226]. In the alternative scenario, DNA looping involves (i) binding of a single AfAgo monomer to the first DNA end (species “4”, Figure 46); (ii) binding of the second AfAgo monomer to the second DNA end (species “5”); (iii) association of two DNA-bound monomers into the looped complex “2”; (iv) association of additional DNA-unbound AfAgo monomers with DNA-bound AfAgo subunits, a process that occludes loop formation (species “6” Figure 46). Both these reaction pathways predict that at elevated protein concentrations, the number of looped complexes should decrease due to the simultaneous binding of separate AfAgo dimers to both DNA ends (species “3”). Single-molecule FRET experiments in solution support this prediction. Indeed, the amount of looped DNA molecules increases until an optimal protein concentration is reached (~2 nM in the current experimental setup, Figure 24d), but declines upon further increase in WT AfAgo concentration. Even though all species depicted in Figure 46 were detected using AFM (Figure 21), the relative contribution of pathways involving WT AfAgo monomers and dimers into the overall DNA looping reaction remains to be established.

Single-molecule measurements on immobilized DNA allowed to assess the dynamic properties of WT AfAgo-induced DNA loops. It was found that (i) the DNA loops induced by WT AfAgo are relatively stable, with the lower limit estimate for the loop duration exceeding 30 s (Figure 26c); (ii) the proximity ratio E of the looped complexes changes over time, suggesting intrinsic dynamics of the AfAgo dimer attributable to the flexible dimerization interface (Figure 26d).

4.2 AfAgo interactions with nucleic acids

Argonaute proteins use an RNA or DNA guide strand for specific recognition of RNA or DNA target strands[2,27]. Correct base-pairing between the two strands triggers target strand cleavage (catalytically active eAgos and long pAgos involved in antiviral defence) or recruitment of partner proteins (catalytically inactive Agos). However, the function and action mechanisms of most prokaryotic Ago proteins, in particular catalytically inactive “short” pAgos, remain largely unknown, with only several examples

described (see pp. 32 – 34 and pp. 43 – 45, and also ref. [5]). In this work, guide and target strand preferences of a truncated long-B [4] prokaryotic Argonaute AfAgo from a hyperthermophilic archaeon *A. fulgidus* were analyzed and its sequence specificity to the 5′-terminal nucleotides of the guide strand, and the complementary fragment of the target strand was revealed.

First, the results presented in this thesis show that AfAgo *in vivo* tightly interacts with nucleic acids, preferentially short RNA fragments with 5′-terminal AUU sequences. Co-purification of pAgo proteins with RNA was observed before, e.g., for RsAgo, which had a preference for 5′-UY RNA[14]. The tight interaction of AfAgo with RNA seemingly contradicts previous studies[25,26], in which authors described preferential binding of AfAgo to single- and double-stranded DNA, but not RNA. Presumably, this discrepancy arose due to non-optimal 5′-terminal RNA and DNA sequences (5′-U or 5′-C) used, as it was found that *in vitro* AfAgo also preferentially binds RNA with a 5′-AUU terminus and substitutions at the 5′-terminus reduce affinity (Figure 28, Table 5). This is a clear indication that AfAgo uses ssRNA as the guide strand and is capable of base-specific interactions with its 5′-terminus.

Further, ssDNA and ssRNA target binding activity *in vitro* of the AfAgo-gRNA complex was demonstrated, which similarly to many other pAgos [27,112] displays a notable preference for ssDNA targets over ssRNA. This implies that *in vivo* AfAgo may also use gRNA to target tDNA. While *A. fulgidus* is a hyperthermophilic organism, most of the experiments were performed at room temperature (25° C, which is not uncommon in the field[127,227]). However, this should not invalidate the conclusions related to AfAgo preferences for ssRNA and ssDNA as the optimal guide and target strands, respectively. Indeed, pre-incubation of the AfAgo-ssRNA binding reaction mixtures at elevated temperatures (70 °C) prior to EMSA did not alter the ability of AfAgo to discriminate the 5′-terminal ssRNA sequences, albeit it decreased the observed AfAgo binding affinities to all ssRNA variants (Figure 28d), presumably due to the lack of chaperones and other protein-stabilizing factors normally present in host cells.

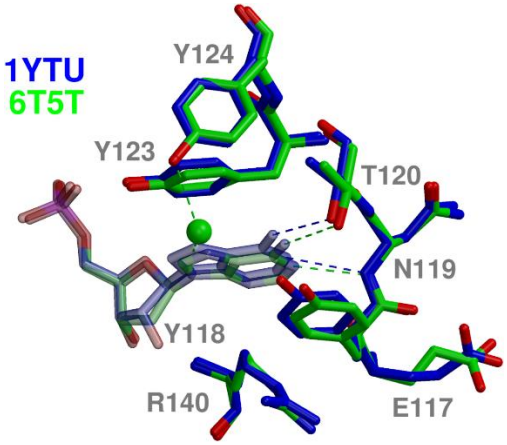


Figure 47. Binding of g1A base in 6T5T (AfAgo-dsDNA, 5′-P-ATT) and 1YTU (AfAgo-dsRNA, 5′-P-AGA). The water molecule from 6T5T is shown as a green sphere. H-bonds are shown as dashed lines. Reproduced from Manakova *et al.*, 2023[216].

Four crystal structures of AfAgo bound to DNA-DNA oligoduplexes with the 5′-AT terminal sequences have also been solved, which mimic the 5′-AU terminus of the *in vivo* bound RNAs. Two structures were obtained with different optimal-like 5′-P-ATT oligoduplexes (PDB ID 6T5T and 6TUO, respectively) and two structures with a suboptimal-like 5′-ATC oligoduplex (PDB ID 6XUP and 6XU0). Although interactions with 5′-ATT oligoduplexes may not directly translate to possible interactions with 5′-AUU RNA guides, where AfAgo might adopt a slightly different conformation and potentially interact with the RNA guide less strongly, structural data suggest that AfAgo employs base-specific readout of the terminal nucleotides of the bound guide and target strands (Figure 30, Figure 32). This interpretation is consistent with the previously published structures of AfAgo bound to the non-optimal DNA or RNA duplexes (5′-P-TTC, PDB ID 2W42 [24] and 5′-P-UUC, PDB ID 2BGG[26], respectively), and to the near-optimal RNA-RNA duplex (5′-terminal sequence 5′-P-AGA, PDB ID 1YTU[25]).

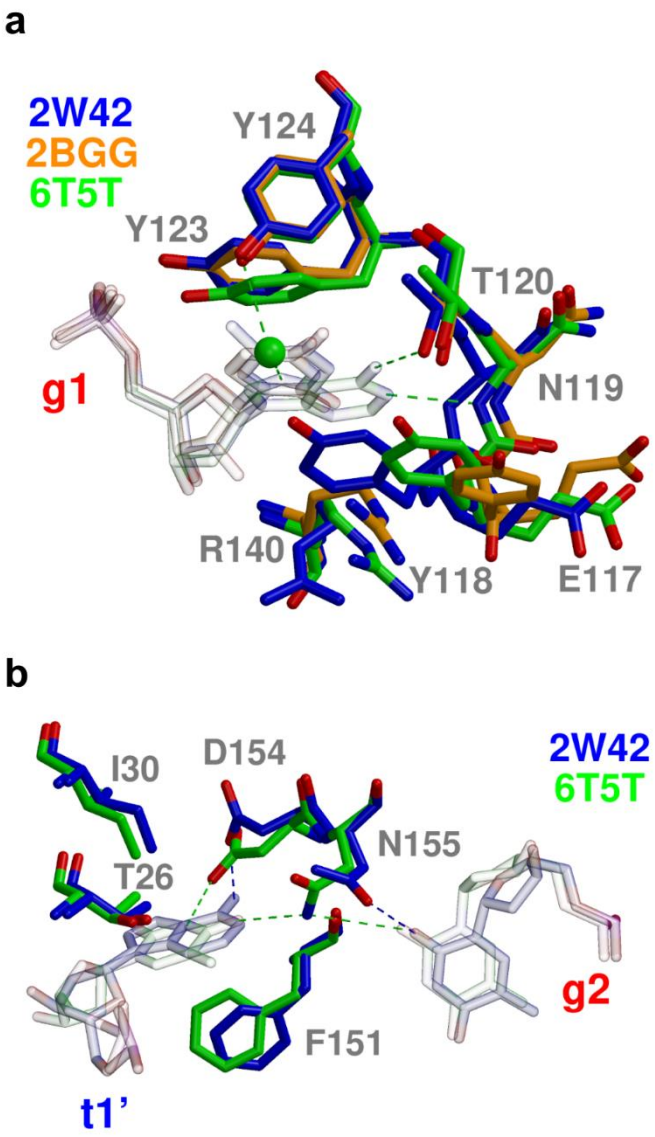


Figure 48. Base-specific contacts in AfAgo-NA complexes. (a) Comparison of the first guide base in crystal structures of AfAgo 2W42 (AfAgo-dsDNA, 5′-P-TTC), 2BGG (AfAgo-dsRNA, 5′-P-UUC) and 6T5T (AfAgo-dsDNA, 5′-P-ATT). (b) Binding of gT2 and t1′ bases in the "side" pocket in 2W42 and 6T5T. Reproduced from Manakova *et al.*, 2023[216].

The most extensive base-specific contacts are made to the 5′-terminal guide strand adenine (gA1) and the complementary target strand thymine tT1′, which is disrupted, and the bases are placed into separate protein pockets. As shown in Figure 47, interactions with gA1 observed in the new structures with

duplex DNA are very similar to those observed in the RNA-bound structure 1YTU [25] (bases at other positions of these structures differ and therefore can not be directly compared). This similarity of base-specific contacts observed with gRNA [25] and DNA [216] indicates that the terminus of the gDNA/tDNA duplex used in this study may provide an adequate mimic for the optimal gRNA/tDNA heteroduplex. Similar disruption of the equivalent base pair was also observed in the AfAgo structures with non-optimal 5′-terminal nucleotides (PDB ID 2W42 [24] and 2BGG[26]). In this case, the flipped gT/U1 base in the 5′-end binding pocket is unable to form adenine-specific contacts observed in the crystal structures presented in this work, including H-bonds with gA1 base made by Asn119 main chain N, Thr120 OH and water-mediated H-bond between Y124 OH group and N7 atom of gA1 (Figure 48). Since in all available structures of AfAgo with RNA the unpaired t1′ base does not enter the "side" pocket, only the "side" pocket interactions of tT1′ in the structures 6XUP, 6XU0, 6T5T, 6TUO can be compared with tA1′ in PDB ID 2W42 (Figure 48b)[24]. In the structures presented in this work, the Asn155 side chain interacts simultaneously with both t1′ (tT1′) and g2 (gT2) bases, and tT1′ makes an additional H-bond with Asp154. In 2W42[24], the tA1′ base in the "side" pocket makes an H-bond with the side chain of Asp154, but the conformation of Asn155 is not suitable for interaction with tA1′. Base-specific contacts formed by the bases of the second guide strand nucleotide gT2 and its complementary target strand nucleotide tA2′ are less numerous (Figure 32), but still sufficient for discrimination against alternative base pairs.

Specific recognition of both guide and target strand nucleotides distinguishes AfAgo from previously characterized Argonaute proteins, which limit specific recognition of terminal nucleotides either to the guide strand (e.g., RsAgo, PDB ID 6D8P[16]) or to the target strand[122]. Another unique feature of AfAgo is that it is a homodimeric protein that can bring together two copies of the guide-target duplex[145]. In crystal structures presented in this work the same dimerization mode (in 6T5T and 6TUO dimer is formed by a crystallographic symmetry operator) was observed, raising further questions regarding possible AfAgo functions *in vivo*.

4.3 Heterodimer formation by AfAgo and AfAgo-N

As one of the first and the best structurally characterized Ago proteins, AfAgo has long been used as a structural model to study Agos and Ago-NA

interactions[13,23–26,96,153–156]. However, AfAgo differs significantly from most other pAgos in several key ways. Firstly, it comprises only MID and inactive PIWI domains, resembling typical short pAgos. However, it is phylogenetically closer to long-B pAgos, earning it a classification as a truncated long-B or pseudo-short pAgo[4,5]. Secondly, unlike other long pAgos that function as monomers, AfAgo forms dimers capable of binding both ends of a DNA fragment to create a looped complex[145]. Thirdly, in contrast to other structurally characterized pAgos, which typically recognize only the terminal nucleotide of the guide and/or target, AfAgo exhibits specific interactions with three nucleotides in both the guide and target strands[216]. Finally, when conducting structural analyses of AfAgo complexes with nucleic acids, it becomes apparent that only the double-stranded terminus of the guide-target duplex makes contact with the MID domain, while the remaining portion extends away from the protein. In this regard, AfAgo stands apart from other pAgo-NA complexes where the guide-target duplex is tightly bound within the protein's nucleic acid binding groove.

Results show that the unique AfAgo properties listed above, by and large, derive from the fact that previous studies treated AfAgo as an isolated standalone protein, neglecting hypothetical proteins encoded in the same operon. Indeed, it is demonstrated here that AfAgo forms a heterodimeric complex with a reconstituted protein encoded upstream of AfAgo in the same operon of *A. fulgidus* DSM 4304 strain. The upstream protein (AfAgo-N) is the structural equivalent of the N-L1-L2 domains of long pAgo proteins. In this way, the fAfAgo heterodimeric complex structurally resembles a long PAZ-less pAgo. fAfAgo is most similar to the long-B RsAgo, in which the PAZ domain is smaller than in other long pAgos[4]. The fAfAgo heterodimer, like other long pAgos, forms a deep groove for interaction with the guide-target duplex that is absent in the standalone AfAgo. In comparison to AfAgo, fAfAgo makes fewer specific contacts with the terminal nucleobases of bound NA (Figure 36). This could be explained by the fact that standalone AfAgo, being only a fragment of a larger functional complex, requires all specific and non-specific contacts possible to stabilise its interaction with the guide-target duplex. As a result, it co-purifies from *E. coli* bound to 5′-AUU RNA guides, and in the complex with a guide-target duplex recognizes three terminal nucleobases in both guide and target strands[216]. In contrast, AfAgo-N and AfAgo together form a groove for nucleic acid binding, which enables numerous non-specific interactions with NA, thus reducing the contribution of specific contacts to the overall affinity of fAfAgo to guide and target

strands. This is consistent with EMSA results, as the fAfAgo-gRNA complex forms much tighter complexes with target DNA than the standalone AfAgo (Figure 43).

In the fAfAgo structure, AfAgo-N interacts with the AfAgo surface through which standalone AfAgo forms a homodimer, preventing AfAgo homodimerization. However, it remains an open question if in the native hyperthermophilic host *A. fulgidus*, there is a dynamic equilibrium between fAfAgo heterodimer and AfAgo homodimer depending on cellular and environmental conditions and if both AfAgo homodimer and fAfAgo heterodimer have distinct roles *in vivo*.

Structurally, fAfAgo heterodimer is similar to *bona fide* short pAgos, which form heterodimeric complexes with effector APAZ(N-L1-L2) proteins, except that it lacks the effector domain (Figure 38)[9]. In both AfAgo and short pAgo heterodimeric complexes, the C-termini of the proteins containing the N-L1-L2 domains are located close to the N-termini of the Ago proteins (Figure 38). In addition to heterodimeric short pAgos, there also exist short pAgos, which together with their upstream encoded effector proteins form a single functional polypeptide. It is therefore likely that single polypeptide pAgos were split into two proteins over the course of evolution. Apparently, such splitting occurred independently among long-B pAgos (exemplified by the AfAgo system studied here) and short pAgos (e.g., JomAgo, PgAgo[5]). To compare the native split fAfAgo to its hypothetical single-chain predecessor, scfAfAgo was constructed, in which the AfAgo-N and AfAgo proteins are fused into a single polypeptide. In all assays, scfAfAgo behaved similarly to fAfAgo (Figure 43), providing no explanation for the potential benefits of the native split variant. Given the co-existence of both single-chain and heterodimeric short pAgos, one may assume that there is no significant functional difference between the active proteins composed of one or two polypeptide chains.

Structural and functional characterization of the restored full AfAgo complex coupled with the computational analysis of its homologues indicates that long-B Agos may come in two major flavours, typical single-chain pAgos exemplified by RsAgo and split pAgos such as full AfAgo. In the case of full AfAgo, the N-L1-L2 and MID-PIWI regions are represented by separate proteins forming a heterodimeric complex. In this regard, full AfAgo and other split long-B Agos are reminiscent of short pAgos. The latter also assemble into functional heterodimeric complexes composed of an APAZ-

containing protein and a MID-PIWI protein. Although previously it was thought that APAZ is analogous to PAZ, by now it has become clear that APAZ corresponds to the N-L1-L2 Ago region. In other words, both AfAgo and short pAgos represent split Ago systems. The difference between a short pAgo and AfAgo is that the N-L1-L2 subunit of the former is often fused with an effector domain. On the other hand, gene neighbourhood analysis of AfAgo and its homologues revealed that they also have a putative effector protein encoded downstream of Ago. This suggests that although long-B pAgos are not fused to effectors, pAgos likely physically associate with them and regulate their activity. Other common features shared by long-B and short pAgos are the catalytically inactive PIWI domain and the lack of the canonical PAZ domain. The N-L1-L2 subunit in short pAgos studied to date lacks the PAZ domain entirely as does full AfAgo. Other long-B pAgos are also either PAZ-less or have a reduced PAZ domain lacking the pocket for binding the 3′-end of the guide strand. Based on these observations, a likely scenario for the emergence of both long-B and short pAgos from long-A pAgos is proposed (Figure 49). In this scenario, following the inactivation of the PIWI domain, the PAZ binding pocket for the 3′-end of the guide strand is no longer needed and is either reduced or lost altogether. As the PIWI domain becomes inactive, pAgo can no longer perform nucleic acid cleavage by itself. Apparently, a new functionalization emerges when long-B pAgos associate with an effector, which can either be a separate protein encoded in the same operon or as a fusion with the N-terminus of pAgo. Splitting of pAgos into two halves (the N-L1-L2 and the MID-PIWI subunits) appears to have occurred multiple times independently because both split and single-chain forms are present in long-B and short pAgos. Thus, the proposed scenario explains the observed diversity of short and long-B pAgos and the repurposing of these pAgos as regulators of toxic effectors that are unleashed by invading foreign nucleic acids.

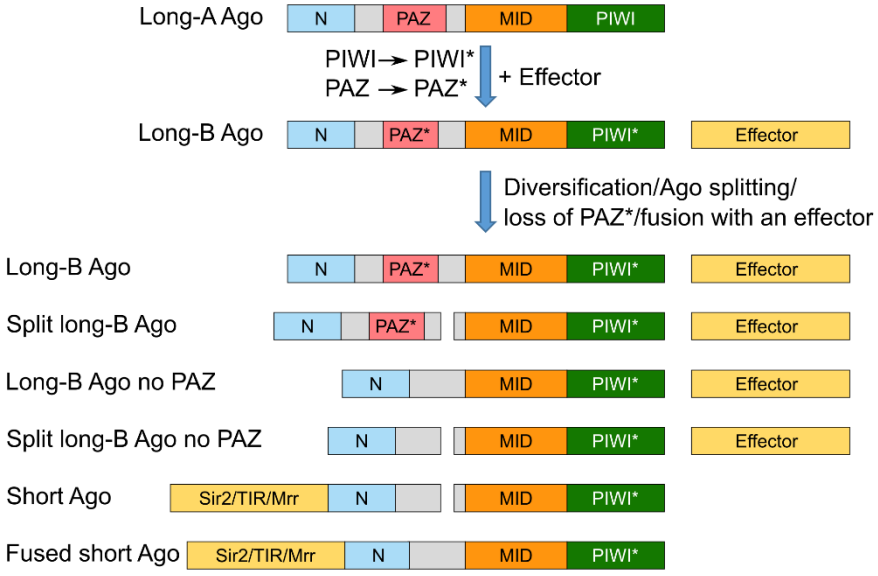


Figure 49. Proposed scenario for the emergence of diverse prokaryotic Argonautes. Following mutations in the PIWI domain and truncation of PAZ, long-A pAgo becomes catalytically inactive, but still able to bind the nucleic acid template (long-B pAgo). Long-B pAgos underwent diversification in several ways: (1) splitting pAgo into two proteins, N-Ago and MID-PIWI-Ago, (2) reduction or complete loss of the PAZ* domain, (3) association with a functional effector either as a separate protein or as a fusion with the N-terminal region of Ago. Reproduced from Manakova *et al.*, 2024[217].

In summary, the results presented here demonstrate that the reconstructed AfAgo operon from *A. fulgidus* DSM 4304 is analogous to the antiviral system SiAgo from *S. islandicus* (apart from the putative regulator) (Figure 33a). The AfAgo-N protein is similar to SiAga1, which forms a heterodimeric complex with SiAgo. Therefore, the fAfAgo heterodimer can be regarded as a structural equivalent of the SiAgo/SiAga1 complex (Figure 38c). Hopefully, further structural and functional studies of AfAgo operon proteins, in particular, the association of fAfAgo heterodimer with the protein encoded downstream AfAgo in the same operon, will shed light on the potential function and mechanism of the AfAgo system.

CONCLUSIONS

1. **Homodimeric Complexes and Looped dsDNA Structures**: AfAgo forms homodimeric complexes and notably generates looped double-stranded DNA structures during *in vitro* examination.
2. **Specificity for 5′-Terminal AUU Guide Strand Sequence**: structural data and *in vivo* investigations reveal that AfAgo exhibits a strong specificity for the 5′-terminal AUU sequence within the guide RNA – a novel finding in precise mechanisms for recognizing guide and target strands by AfAgo.
3. **RNA-Guided ssDNA Targeting**: AfAgo lacks demonstrated catalytic activity but shows RNA-guided single-stranded DNA (ssDNA) targeting. This aspect has not been previously observed for AfAgo.
4. **Formation of Heterodimeric Complexes**: AfAgo forms heterodimeric complexes with AfAgo-N, resulting in a complex known as fAfAgo. Notably, this complex exhibits significant resemblances to long PAZ-less pAgos. This discovery adds complexity to the functional mechanisms of (f)AfAgo.
5. **RNA-Guided ssDNA Targeting by (sc)fAfAgo**: both fAfAgo and the fused scfAfAgo exhibit RNA-guided DNA targeting with a higher affinity than stand-alone AfAgo.

LIST OF PUBLICATIONS

This dissertation is based on the following original publications:

- **Golovinas, E.**[‡], Rutkauskas, D., Manakova, E., Jankunec, M., Silanskas, A., Sasnauskas, G. & Zaremba, M., Prokaryotic Argonaute from *Archaeoglobus fulgidus* interacts with DNA as a homodimer, Scientific Reports 11, 4518 (2021). https://doi.org/10.1038/s41598-021-83889-4.

Personal contributions: performed smFRET experiments together with D.R., performed EMSA experiments and analysed the data, prepared DNA fragments analysed SEC-MALS data, prepared figures and contributed to the writing of the manuscript.

- Manakova, E.[†], **Golovinas, E.**[†], Pocevičiūtė, R., Sasnauskas, G., Grybauskas, A., Gražulis, S. & Zaremba, M., Structural basis for sequence-specific recognition of guide and target strands by the *Archaeoglobus fulgidus* Argonaute protein, Scientific Reports 13, 6123 (2023). https://doi.org/10.1038/s41598-023-32600-w

Personal contributions: performed biochemistry experiments together with R.P. and analysed the data, prepared figures and contributed to the writing of the manuscript.

- Manakova, E.[†], **Golovinas, E.**[†], Pocevičiūtė, R.[†], Sasnauskas, G., Silanskas, A., Rutkauskas, D., Jankunec, M., Zagorskaitė, E., Jurgelaitis, E., Grybauskas, A., Venclovas, Č. & Zaremba, M., The missing part: the *Archaeoglobus fulgidus* Argonaute forms a functional heterodimer with an N-L1-L2 domain protein, Nucleic Acids Research 1–16 (2024). https://doi.org/10.1093/nar/gkad1241.

Personal contributions: performed biochemistry experiments together with R.P. and analysed the data, prepared figures and contributed to the writing of the manuscript.

[†] – these authors are to be considered joined first authors.

[‡] – this author is to be considered the first author.

CONFERENCE PRESENTATIONS

- EMBO | EMBL Symposium: The Non-Coding Genome, Poster presentation "RNA-guided DNA Binding Specificity of a Prokaryotic Argonaute Protein from *Archaeoglobus fulgidus*", **E. Golovinas**, E. Manakova, G. Sasnauskas, S. Gražulis, V. Šikšnys and M. Zaremba. Virtual, October 13-15, 2021.

- EMBO | EMBL Symposium: The Complex Life of RNA, Poster presentation "Prokaryotic Argonaute from *Archaeoglobus fulgidus* Interacts with DNA as a Homodimer", **E. Golovinas**, D. Rutkauskas, E. Manakova, M. Jankunec, A. Šilanskas, G. Sasnauskas, M. Zaremba. Virtual, October 7-9, 2020.

- International conference FEBS3+ 2019, Poster presentation "Specificity of the Argonaute Protein from *Archaeoglobus fulgidus* to the 5′-end of the Guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, V. Šikšnys, Riga, Latvia, June 17-19, 2019.

- International conference for students of physics and natural sciences Open Readings 2019, Poster presentation "Specificity of the Argonaute Protein from *Archaeoglobus fulgidus* to the 5′-end of the Guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, Virginijus Šikšnys, Vilnius, Lithuania, March 19-22, 2019.

- International conference of life sciences The COINS 2019. Poster presentation "Specificity Of The Argonaute Protein From *Archaeoglobus fulgidus* to the 5′-end of the Guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, V. Šikšnys, Vilnius, Lithuania, February 26-28, 2019.

REFERENCES

1. Bohmert, K. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.* **17**, 170–180 (1998) DOI: 10.1093/emboj/17.1.170.

2. Olina, A. V., Kulbachinskiy, A. V., Aravin, A. A. & Esyunina, D. M. Argonaute Proteins and Mechanisms of RNA Interference in Eukaryotes and Prokaryotes. *Biochem.* **83**, 483–497 (2018) DOI: 10.1134/S0006297918050024.

3. Kuhn, C. D. & Joshua-Tor, L. Eukaryotic Argonautes come into focus. *Trends Biochem. Sci.* **38**, 263–271 (2013) DOI: 10.1016/j.tibs.2013.02.008.

4. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The Expanded Universe of Prokaryotic Argonaute Proteins. *MBio* **9**, 1–20 (2018) DOI: 10.1128/mBio.01935-18.

5. Koopal, B., Mutte, S. K. & Swarts, D. C. A long look at short prokaryotic Argonautes. *Trends Cell Biol.* **33**, 605–618 (2023) DOI: 10.1016/j.tcb.2022.10.005.

6. Kuzmenko, A., Oguienko, A., Esyunina, D., Yudin, D., Petrova, M., Kudinova, A., Maslova, O., Ninova, M., Ryazansky, S., Leach, D., Aravin, A. A. & Kulbachinskiy, A. DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* **587**, 632–637 (2020) DOI: 10.1038/s41586-020-2605-1.

7. Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J. J. & Van Der Oost, J. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014) DOI: 10.1038/nature12971.

8. Jolly, S. M., Gainetdinov, I., Jouravleva, K., Zhang, H., Strittmatter, L., Bailey, S. M., Hendricks, G. M., Dhabaria, A., Ueberheide, B. & Zamore, P. D. Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. *Cell* **182**, 1545-1559.e18 (2020) DOI: 10.1016/j.cell.2020.07.036.

9. Zaremba, M., Dakineviciene, D., Golovinas, E., Zagorskaitė, E., Stankunas, E., Lopatina, A., Sorek, R., Manakova, E., Ruksenaite, A., Silanskas, A., Asmontas, S., Grybauskas, A., Tylenyte, U., Jurgelaitis, E., Grigaitis, R., Timinskas, K., Venclovas, Č. & Siksnys, V. Short prokaryotic Argonautes provide defence against incoming mobile genetic elements through NAD+ depletion. *Nat. Microbiol.* **7**, 1857–

1869 (2022) DOI: 10.1038/s41564-022-01239-0.

10. Koopal, B., Potocnik, A., Mutte, S. K., Aparicio-Maldonado, C., Lindhoud, S., Vervoort, J. J. M., Brouns, S. J. J. & Swarts, D. C. Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. *Cell* **185**, 1471-1486.e19 (2022) DOI: 10.1016/j.cell.2022.03.012.

11. Garb, J., Lopatina, A., Bernheim, A., Zaremba, M., Siksnys, V., Melamed, S., Leavitt, A., Millman, A., Amitai, G. & Sorek, R. Multiple phage resistance systems inhibit infection via SIR2-dependent NAD+ depletion. *Nat. Microbiol.* **7**, 1849–1856 (2022) DOI: 10.1038/s41564-022-01207-8.

12. Swarts, D. C., Hegge, J. W., Hinojo, I., Shiimori, M., Ellis, M. A., Dumrongkulraksa, J., Terns, R. M., Terns, M. P. & Van Der Oost, J. Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res.* **43**, 5120–5129 (2015) DOI: 10.1093/nar/gkv415.

13. Miyoshi, T., Ito, K., Murakami, R. & Uchiumi, T. Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nat. Commun.* **7**, 1–12 (2016) DOI: 10.1038/ncomms11846.

14. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. *Mol. Cell* **51**, 594–605 (2013) DOI: 10.1016/j.molcel.2013.08.014.

15. Lisitskaya, L., Petushkov, I., Esyunina, D., Aravin, A. & Kulbachinskiy, A. Recognition of double-stranded DNA by the Rhodobacter sphaeroides Argonaute protein. *Biochem. Biophys. Res. Commun.* **533**, 1484–1489 (2020) DOI: 10.1016/j.bbrc.2020.10.051.

16. Liu, Y., Esyunina, D., Olovnikov, I., Teplova, M., Kulbachinskiy, A., Aravin, A. A. & Patel, D. J. Accommodation of Helical Imperfections in Rhodobacter sphaeroides Argonaute Ternary Complexes with Guide RNA and Target DNA. *Cell Rep.* **24**, 453–462 (2018) DOI: 10.1016/j.celrep.2018.06.021.

17. Liu, Q., Guo, X., Xun, G., Li, Z., Chong, Y., Yang, L., Wang, H., Zhang, F., Luo, S., Cui, L., Zhao, P., Ye, X., Xu, H., Lu, H., Li, X., Deng, Z., Li, K. & Feng, Y. Argonaute integrated single-tube PCR system enables supersensitive detection of rare mutations. *Nucleic Acids Res.* 1–14 (2021) DOI: 10.1093/nar/gkab274.

18. Wang, F., Yang, J., He, R., Yu, X., Chen, S., Liu, Y., Wang, L., Li, A., Liu, L., Zhai, C. & Ma, L. PfAgo-based detection of SARS-CoV-2. *Biosens. Bioelectron.* **177**, 112932 (2021) DOI: 10.1016/j.bios.2020.112932.

19. Song, J., Hegge, J. W., Mauk, M. G., Chen, J., Till, J. E., Bhagwat, N., Azink, L. T., Peng, J., Sen, M., Mays, J., Carpenter, E. L., Van Der Oost, J. & Bau, H. H. Highly specific enrichment of rare nucleic acid fractions using Thermus thermophilus argonaute with applications in cancer diagnostics. *Nucleic Acids Res.* **48**, 1–15 (2020) DOI: 10.1093/nar/gkz1165.

20. Filius, M., Cui, T. J., Ananth, A. N., Docter, M. W., Hegge, J. W., Oost, J. Van Der & Joo, C. High-Speed Super-Resolution Imaging Using Protein-Assisted DNA- PAINT. *Nano Lett.* **20**, 2264–2270 (2020) DOI: 10.1021/acs.nanolett.9b04277.

21. Chang, L., Sheng, G., Zhang, Y., Shao, S., Wang, Y. & Sun, Y. AgoFISH : cost-effective in situ labelling of genomic loci based on DNA-guided dTtAgo. *Nanoscale Horizons* **4**, 918–923 (2019) DOI: 10.1039/c9nh00028c.

22. Kropocheva, E. V., Lisitskaya, L. A., Agapov, A. A., Musabirov, A. A., Kulbachinskiy, A. V. & Esyunina, D. M. Prokaryotic Argonaute Proteins as a Tool for Biotechnology. *Mol. Biol.* (2022) DOI: 10.1134/S0026893322060103.

23. Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23**, 4727–4737 (2004) DOI: 10.1038/sj.emboj.7600488.

24. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the Seed-Target Recognition Step in RNA Silencing by a PIWI/MID Domain Protein. *Mol. Cell* **33**, 204–214 (2009) DOI: 10.1016/j.molcel.2008.12.012.

25. Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T. & Patel, D. J. Structural basis for 5′ -end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**, 666–670 (2005) DOI: 10.1038/nature03514.

26. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663–666 (2005) DOI: 10.1038/nature03462.

27. Lisitskaya, L., Aravin, A. A. & Kulbachinskiy, A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nat. Commun.* **9**, 5165 (2018) DOI: 10.1038/s41467-018-07449-7.

28. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009) DOI: 10.1038/nrg2504.

29. Denli, A. M. & Hannon, G. J. RNAi: an ever-growing puzzle. *Trends Biochem. Sci.* **28**, 196–201 (2003) DOI: 10.1016/S0968-0004(03)00058-6.

30. Hannon, G. J. RNA interference. *Nature* **418**, 244–51 (2002) DOI: 10.1038/418244a.

31. Zamore, P. D. & Haley, B. Ribo-gnome: The Big World of Small RNAs. *Science* **309**, 1519–1524 (2005) DOI: 10.1126/science.1111444.

32. Montgomery, M. K. & Fire, A. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet.* **14**, 255–8 (1998) DOI: 10.1016/s0168-9525(98)01510-8.

33. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–11 (1998) DOI: 10.1038/35888.

34. Guo, S. & Kemphues, K. J. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–20 (1995) DOI: 10.1016/0092-8674(95)90082-9.

35. Matzke, M. A. & Matzke, A. How and Why Do Plants Inactivate Homologous (Trans)genes? *Plant Physiol.* **107**, 679–685 (1995) DOI: 10.1104/pp.107.3.679.

36. Baulcombe, D. C. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol. Biol.* **32**, 79–88 (1996) DOI: 10.1007/BF00039378.

37. Tabara, H., Grishok, A. & Mello, C. C. RNAi in C. elegans : Soaking in the Genome Sequence. *Science* **282**, 430–431 (1998) DOI: 10.1126/science.282.5388.430.

38. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970

(1991) DOI: 10.1002/j.1460-2075.1991.tb04966.x.

39. Grishok, A., Tabara, H. & Mello, C. C. Genetic Requirements for Inheritance of RNAi in C. elegans. *Science* **287**, 2494–2497 (2000) DOI: 10.1126/science.287.5462.2494.

40. The Nobel Prize in Physiology or Medicine 2006. *Press release* https://www.nobelprize.org/prizes/medicine/2006/press-release/ (2006).

41. McCaffrey, A. P., Meuse, L., Pham, T.-T. T., Conklin, D. S., Hannon, G. J. & Kay, M. A. RNA interference in adult mice. *Nature* **418**, 38–39 (2002) DOI: 10.1038/418038a.

42. Zhang, X. D. *Optimal high-throughput screening: practical experimental design and data analysis for genome-scale RNAi research*. (Cambridge University Press, 2011).

43. Dayeh, D. M., Cantara, W. A., Kitzrow, J. P., Musier-Forsyth, K. & Nakanishi, K. Argonaute-based programmable RNase as a tool for cleavage of highly-structured RNA. *Nucleic Acids Res.* **46**, 1–13 (2018) DOI: 10.1093/nar/gky496.

44. Hutvagner, G. & Simard, M. J. Argonaute proteins: Key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32 (2008) DOI: 10.1038/nrm2321.

45. Drinnenberg, I. A., Weinberg, D. E., Xie, K. T., Mower, J. P., Wolfe, K. H., Fink, G. R. & Bartel, D. P. RNAi in Budding Yeast. *Science* **326**, 544–550 (2009) DOI: 10.1126/science.1176945.

46. Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., Patel, D. J. & Van Der Oost, J. The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014) DOI: 10.1038/nsmb.2879.

47. Tolia, N. H. & Joshua-Tor, L. Slicer and the Argonautes. *Nat. Chem. Biol.* **3**, 36–43 (2007) DOI: 10.1038/nchembio848.

48. Zagoskin, M. V, Wang, J. & Davis, R. E. Small RNA pathways in the nematode Ascaris in the absence of piRNAs. *Nat. Commun.* **13**, 1–19 (2022) DOI: 10.1038/s41467-022-28482-7.

49. Jin, S., Zhan, J. & Zhou, Y. Argonaute proteins: structures and their endonuclease activity. *Mol. Biol. Rep.* (2021) DOI: 10.1007/s11033-021-06476-w.

50. Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C. G., Tolia, N. H., Joshua-Tor, L., Mitani, S., Simard, M. J. & Mello, C. C. Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. *Cell* **127**, 747–757 (2006) DOI: 10.1016/j.cell.2006.09.033.

51. Nakanishi, K. Anatomy of four human Argonaute proteins. *Nucleic Acids Res.* 1–21 (2022) DOI: 10.1093/nar/gkac519.

52. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science* **305**, 1434–1437 (2004) DOI: 10.1126/science.1102514.

53. Rivas, F. V., Tolia, N. H., Song, J.-J., Aragon, J. P., Liu, J., Hannon, G. J. & Joshua-Tor, L. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* **12**, 340–349 (2005) DOI: 10.1038/nsmb918.

54. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* **11**, 576–577 (2004) DOI: 10.1038/nsmb777.

55. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* **19**, 145–151 (2012) DOI: 10.1038/nsmb.2232.

56. Wu, J., Yang, J., Cho, W. C. & Zheng, Y. Argonaute proteins: Structural features, functions and emerging roles. *J. Adv. Res.* **24**, 317–324 (2020) DOI: 10.1016/j.jare.2020.04.017.

57. Hauptmann, J., Dueck, A., Harlander, S., Pfaff, J., Merkl, R. & Meister, G. Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nat. Struct. Mol. Biol.* **20**, 814–817 (2013) DOI: 10.1038/nsmb.2577.

58. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001) DOI: 10.1101/gad.862301.

59. Elbashir, S. M. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J.* **20**, 6877–6888 (2001) DOI: 10.1093/emboj/20.23.6877.

60. Gorski, S. A., Vogel, J. & Doudna, J. A. RNA-based recognition and targeting: Sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* **18**, 215–228 (2017) DOI: 10.1038/nrm.2016.174.

61. Künne, T., Swarts, D. C. & Brouns, S. J. J. Planting the seed: Target recognition of short guide RNAs. *Trends Microbiol.* **22**, 74–83 (2014) DOI: 10.1016/j.tim.2013.12.003.

62. Bobadilla Ugarte, P., Barendse, P. & Swarts, D. C. Argonaute proteins confer immunity in all domains of life. *Curr. Opin. Microbiol.* **74**, 102313 (2023) DOI: 10.1016/j.mib.2023.102313.

63. Azlan, A., Dzaki, N. & Azzam, G. Argonaute: The executor of small RNA function. *J. Genet. Genomics* **43**, 481–494 (2016) DOI: 10.1016/j.jgg.2016.06.002.

64. Sheu-Gruttadauria, J. & MacRae, I. J. Structural Foundations of RNA Silencing by Argonaute. *J. Mol. Biol.* **429**, 2619–2639 (2017) DOI: 10.1016/j.jmb.2017.07.018.

65. MacRae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D. & Doudna, J. A. Structural Basis for Double-Stranded RNA Processing by Dicer. *Science* **311**, 195–198 (2006) DOI: 10.1126/science.1121638.

66. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single Processing Center Models for Human Dicer and Bacterial RNase III. *Cell* **118**, 57–68 (2004) DOI: 10.1016/j.cell.2004.06.017.

67. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281–297 (2004) DOI: 10.1016/S0092-8674(04)00045-5.

68. Leitão, A. L. & Enguita, F. J. A Structural View of miRNA Biogenesis and Function. *Non-Coding RNA* **8**, 10 (2022) DOI: 10.3390/ncrna8010010.

69. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006) DOI: 10.1038/nature04917.

70. Stuwe, E., Tóth, K. F. & Aravin, A. A. Small but sturdy: Small RNAs in cellular memory and epigenetics. *Genes Dev.* **28**, 423–431 (2014) DOI: 10.1101/gad.236414.113.

71. Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A. & Hannon, G. J. An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science* **322**, 1387–1392 (2008) DOI: 10.1126/science.1165171.

72. Aravin, A. a, Hannon, G. J. & Brennecke, J. The Piwi-piRNA Pathway

Provides an Adaptive Defense in the Transposon Arms Race. *Science* **318**, 761–764 (2007) DOI: 10.1126/science.1146484.

73. Yamanaka, S., Siomi, M. C. & Siomi, H. PiRNA clusters and open chromatin structure. *Mob. DNA* **5**, 1–12 (2014) DOI: 10.1186/1759-8753-5-22.

74. Seto, A. G., Kingston, R. E. & Lau, N. C. The Coming of Age for Piwi Proteins. *Mol. Cell* **26**, 603–609 (2007) DOI: 10.1016/j.molcel.2007.05.021.

75. Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J. & Tuschl, T. The Small RNA Profile during Drosophila melanogaster Development. *Dev. Cell* **5**, 337–350 (2003) DOI: 10.1016/S1534-5807(03)00228-4.

76. Xiao, Y. & Ke, A. PIWI Takes a Giant Step. *Cell* **167**, 310–312 (2016) DOI: 10.1016/j.cell.2016.09.043.

77. Czech, B. & Hannon, G. J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem. Sci.* **41**, 324–337 (2016) DOI: 10.1016/j.tibs.2015.12.008.

78. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258 (2011) DOI: 10.1038/nrm3089.

79. Malone, C. D. & Hannon, G. J. Small RNAs as Guardians of the Genome. *Cell* **136**, 656–668 (2009) DOI: 10.1016/j.cell.2009.01.045.

80. Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3′ End Formation of PIWI-Interacting RNAs In Vitro. *Mol. Cell* **43**, 1015–1022 (2011) DOI: 10.1016/j.molcel.2011.07.029.

81. Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T. & Patel, D. J. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* **19**, 405–419 (2005) DOI: 10.1016/j.molcel.2005.07.011.

82. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell* **115**, 209–216 (2003) DOI: 10.1016/S0092-8674(03)00801-8.

83. Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N. & Zamore, P. D. Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell*

115, 199–208 (2003) DOI: 10.1016/S0092-8674(03)00759-1.

84. Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A Protein Sensor for siRNA Asymmetry. *Science* **306**, 1377–1380 (2004) DOI: 10.1126/science.1102755.

85. Ghildiyal, M., Xu, J., Seitz, H., Weng, Z. & Zamore, P. D. Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* **16**, 43–56 (2010) DOI: 10.1261/rna.1972910.

86. Kawamata, T. & Tomari, Y. Making RISC. *Trends Biochem. Sci.* **35**, 368–376 (2010) DOI: 10.1016/j.tibs.2010.03.009.

87. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005) DOI: 10.1101/gad.1284105.

88. Leuschner, P. J. F., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006) DOI: 10.1038/sj.embor.7400637.

89. Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., Chen, S., Liu, Q. & Zhang, H. Structure of C3PO and mechanism of human RISC activation. *Nat. Struct. Mol. Biol.* **18**, 650–657 (2011) DOI: 10.1038/nsmb.2032.

90. Robb, G. B. & Rana, T. M. RNA Helicase A Interacts with RISC in Human Cells and Functions in RISC Loading. *Mol. Cell* **26**, 523–537 (2007) DOI: 10.1016/j.molcel.2007.04.016.

91. Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. Structure of yeast Argonaute with guide RNA. *Nature* **486**, 368–374 (2012) DOI: 10.1038/nature11211.

92. Schirle, N. T. & MacRae, I. J. The Crystal Structure of Human Argonaute2. *Science* **336**, 1037–1040 (2012) DOI: 10.1126/science.1221551.

93. Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J. & Joshua-Tor, L. The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell* **150**, 100–110 (2012) DOI: 10.1016/j.cell.2012.05.017.

94. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute Divides Its RNA Guide into Domains with Distinct Functions and RNA-Binding Properties. *Cell* **151**, 1055–1067 (2012) DOI: 10.1016/j.cell.2012.10.036.

95. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science* **346**, 608–613 (2014) DOI: 10.1126/science.1258040.

96. Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. & Patel, D. J. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008) DOI: 10.1038/nature07666.

97. Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T. & Patel, D. J. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754–761 (2009) DOI: 10.1038/nature08434.

98. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29 (2009) DOI: 10.1186/1745-6150-4-29.

99. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009) DOI: 10.1099/mic.0.023960-0.

100. Graver, B. A., Chakravarty, N. & Solomon, K. V. Prokaryotic Argonautes for in vivo biotechnology and molecular diagnostics. *Trends Biotechnol.* 1–13 (2023) DOI: 10.1016/j.tibtech.2023.06.010.

101. Koonin, E. V. Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: Common ancestry vs convergence. *Biol. Direct* **12**, 1–14 (2017) DOI: 10.1186/s13062-017-0177-2.

102. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J. Bacteriol.* **193**, 6039–6056 (2011) DOI: 10.1128/JB.05535-11.

103. Shabalina, S. & Koonin, E. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol.* **23**, 578–587 (2008) DOI: 10.1016/j.tree.2008.06.005.

104. Lisitskaya, L., Kropocheva, E., Agapov, A., Prostova, M., Panteleev, V., Yudin, D., Ryazansky, S., Kuzmenko, A., Aravin, A. A., Esyunina, D. & Kulbachinskiy, A. Bacterial Argonaute nucleases reveal different modes of DNA targeting in vitro and in vivo. *Nucleic Acids Res.*

2022.09.09.507302 (2023) DOI: 10.1093/nar/gkad290.

105. Kaya, E., Doxzen, K. W., Knoll, K. R., Wilson, R. C., Strutt, S. C., Kranzusch, P. J. & Doudna, J. A. A bacterial Argonaute with noncanonical guide RNA specificity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4057–4062 (2016) DOI: 10.1073/pnas.1524385113.

106. Doxzen, K. W. & Doudna, J. A. DNA recognition by an RNA-guided bacterial Argonaute. *PLoS One* **12**, e0177097 (2017) DOI: 10.1371/journal.pone.0177097.

107. Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. *Nucleic Acids Res.* **47**, 5822–5836 (2019) DOI: 10.1093/nar/gkz379.

108. Kropocheva, E., Kuzmenko, A., Aravin, A. A., Esyunina, D. & Kulbachinskiy, A. A programmable pAgo nuclease with universal guide and target specificity from the mesophilic bacterium Kurthia massiliensis. *Nucleic Acids Res.* **1**, 1–28 (2021) DOI: 10.1093/nar/gkab182.

109. Olina, A., Kuzmenko, A., Ninova, M., Aravin, A. A., Kulbachinskiy, A. & Esyunina, D. Genome-wide DNA sampling by Ago nuclease from the cyanobacterium Synechococcus elongatus. *RNA Biol.* **17**, 677–688 (2020) DOI: 10.1080/15476286.2020.1724716.

110. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008) DOI: 10.1038/nature07315.

111. Willkomm, S., Makarova, K. S. & Grohmann, D. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. *FEMS Microbiol. Rev.* **42**, 376–387 (2018) DOI: 10.1093/femsre/fuy010.

112. Hegge, J. W., Swarts, D. C. & Van Der Oost, J. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nat. Rev. Microbiol.* **16**, 5–11 (2018) DOI: 10.1038/nrmicro.2017.73.

113. Kim, S., Jung, Y. & Lim, D. Argonaute system of Kordia jejudonensis is a heterodimeric nucleic acid-guided nuclease. *Biochem. Biophys. Res. Commun.* **525**, 755–758 (2020) DOI: 10.1016/j.bbrc.2020.02.145.

114. Ofir, G., Herbst, E., Baroz, M., Cohen, D., Millman, A., Doron, S., Tal, N., Malheiro, D. B. A., Malitsky, S., Amitai, G. & Sorek, R. Antiviral activity of bacterial TIR domains via immune signalling molecules.

Nature **600**, 116–120 (2021) DOI: 10.1038/s41586-021-04098-7.

115. Ka, D., Oh, H., Park, E., Kim, J.-H. & Bae, E. Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD+ degradation. *Nat. Commun.* **11**, 2816 (2020) DOI: 10.1038/s41467-020-16703-w.

116. Wang, X., Li, X., Yu, G., Zhang, L., Zhang, C., Wang, Y., Liao, F., Wen, Y., Yin, H., Liu, X., Wei, Y., Li, Z., Deng, Z. & Zhang, H. Structural insights into mechanisms of Argonaute protein-associated NADase activation in bacterial immunity. *Cell Res.* **33**, 699–711 (2023) DOI: 10.1038/s41422-023-00839-7.

117. Morehouse, B. R., Govande, A. A., Millman, A., Keszei, A. F. A., Lowey, B., Ofir, G., Shao, S., Sorek, R. & Kranzusch, P. J. STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020) DOI: 10.1038/s41586-020-2719-5.

118. Essuman, K., Summers, D. W., Sasaki, Y., Mao, X., DiAntonio, A. & Milbrandt, J. The SARM1 Toll/Interleukin-1 Receptor Domain Possesses Intrinsic NAD+ Cleavage Activity that Promotes Pathological Axonal Degeneration. *Neuron* **93**, 1334-1343.e5 (2017) DOI: 10.1016/j.neuron.2017.02.022.

119. Wan, L., Essuman, K., Anderson, R. G., Sasaki, Y., Monteiro, F., Chung, E. H., Nishimura, E. O., DiAntonio, A., Milbrandt, J., Dangl, J. L. & Nishimura, M. T. TIR domains of plant immune receptors are NAD+-cleaving enzymes that promote cell death. *Science* **365**, 799–803 (2019) DOI: 10.1126/science.aax1771.

120. Horsefield, S., Burdett, H., Zhang, X., Manik, M. K., Shi, Y., Chen, J., Qi, T., Gilley, J., Lai, J. S., Rank, M. X., Casey, L. W., Gu, W., Ericsson, D. J., Foley, G., Hughes, R. O., Bosanac, T., Von Itzstein, M., Rathjen, J. P., Nanson, J. D., Boden, M., Dry, I. B., Williams, S. J., Staskawicz, B. J., Coleman, M. P., Ve, T., Dodds, P. N. & Kobe, B. NAD+ cleavage activity by animal and plant TIR domains in cell death pathways. *Science* **365**, 793–799 (2019) DOI: 10.1126/science.aax1911.

121. Zander, A., Willkomm, S., Ofer, S., Van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., Albers, S. V., Werner, F. & Grohmann, D. Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nat. Microbiol.* **2**, (2017) DOI: 10.1038/nmicrobiol.2017.34.

122. Swarts, D. C., Szczepaniak, M., Sheng, G., Chandradoss, S. D., Zhu,

Y., Timmers, E. M., Zhang, Y., Zhao, H., Lou, J., Wang, Y., Joo, C. & van der Oost, J. Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Mol. Cell* **65**, 985-998.e6 (2017) DOI: 10.1016/j.molcel.2017.01.033.

123. Liu, Y., Li, W., Jiang, X., Wang, Y., Zhang, Z., Liu, Q., He, R., Chen, Q., Yang, J., Wang, L., Wang, F. & Ma, L. A programmable omnipotent Argonaute nuclease from mesophilic bacteria Kurthia massiliensis. *Nucleic Acids Res.* 1–12 (2021) DOI: 10.1093/nar/gkaa1278.

124. Olina, A., Agapov, A., Yudin, D., Sutormin, D., Galivondzhyan, A., Kuzmenko, A., Severinov, K., Aravin, A. A. & Kulbachinskiy, A. Bacterial Argonaute Proteins Aid Cell Division in the Presence of Topoisomerase Inhibitors in Escherichia coli. *Microbiol. Spectr.* 2022.09.13.507849 (2023) DOI: 10.1128/spectrum.04146-22.

125. Xun, G., Liu, Q., Chong, Y., Guo, X., Li, Z., Li, Y., Fei, H., Li, K. & Feng, Y. Argonaute with stepwise endonuclease activity promotes specific and multiplex nucleic acid detection. *Bioresour. Bioprocess.* **8**, 46 (2021) DOI: 10.1186/s40643-021-00401-6.

126. Willkomm, S., Oellig, C. A., Zander, A., Restle, T., Keegan, R., Grohmann, D. & Schneider, S. Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. *Nat. Microbiol.* **2**, 1–7 (2017) DOI: 10.1038/nmicrobiol.2017.35.

127. Guo, X., Sun, Y., Chen, L., Huang, F., Liu, Q. & Feng, Y. A Hyperthermophilic Argonaute From Ferroglobus placidus With Specificity on Guide Binding Pattern. *Front. Microbiol.* **12**, 1–10 (2021) DOI: 10.3389/fmicb.2021.654345.

128. Sheng, G., Gogakos, T., Wang, J., Zhao, H., Serganov, A., Juranek, S., Tuschl, T., Patel, D. J. & Wang, Y. Structure/cleavage-based insights into helical perturbations at bulge sites within T. thermophilus Argonaute silencing complexes. *Nucleic Acids Res.* **45**, 9149–9163 (2017) DOI: 10.1093/nar/gkx547.

129. Lisitskaya, L., Shin, Y., Agapov, A., Olina, A., Kropocheva, E., Ryazansky, S., Aravin, A. A., Esyunina, D., Murakami, K. S. & Kulbachinskiy, A. Programmable RNA targeting by bacterial Argonaute nucleases with unconventional guide binding and cleavage specificity. *Nat. Commun.* **13**, 4624 (2022) DOI: 10.1038/s41467-022-32079-5.

130. Hunt, E. A., Evans, T. C. & Tanner, N. A. Single-stranded binding

proteins and helicase enhance the activity of prokaryotic argonautes in vitro. *PLoS One* **13**, e0203073 (2018) DOI: 10.1371/journal.pone.0203073.

131. Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D. C., van der Oost, J., Patel, D. J. & Wang, Y. Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci.* **111**, 652–657 (2014) DOI: 10.1073/pnas.1321032111.

132. Willkomm, S., Zander, A., Gust, A. & Grohmann, D. A prokaryotic twist on Argonaute function. *Life* **5**, 538–553 (2015) DOI: 10.3390/life5010538.

133. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **162**, 84–95 (2015) DOI: 10.1016/j.cell.2015.06.029.

134. Hegge, J. W., Swarts, D. C., Chandradoss, S. D., Cui, T. J., Kneppers, J., Jinek, M., Joo, C. & Van Der Oost, J. DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute. *Nucleic Acids Res.* **47**, 5809–5821 (2019) DOI: 10.1093/nar/gkz306.

135. McKie, S. J., Neuman, K. C. & Maxwell, A. DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis. *BioEssays* **43**, 1–19 (2021) DOI: 10.1002/bies.202000286.

136. Swarts, D. C. Prokaryotic Argonautes Function beyond Immunity by Unlinking Replicating Chromosomes. *Cell* **182**, 1381–1383 (2020) DOI: 10.1016/j.cell.2020.08.037.

137. Ni, D., Lu, X., Stahlberg, H. & Ekundayo, B. Activation mechanism of a short argonaute-TIR prokaryotic immune system. *Sci. Adv.* **9**, 1–11 (2023) DOI: 10.1126/sciadv.adh9002.

138. Shen, Z., Yang, X., Xia, S., Huang, W., Taylor, D. J., Nakanishi, K. & Fu, T. Oligomerization-mediated activation of a short prokaryotic Argonaute. *Nature* **621**, 154–161 (2023) DOI: 10.1038/s41586-023-06456-z.

139. Zeng, Z., Chen, Y., Pinilla-Redondo, R., Shah, S. A., Zhao, F., Wang, C., Hu, Z., Wu, C., Zhang, C., Whitaker, R. J., She, Q. & Han, W. A short prokaryotic Argonaute activates membrane effector to confer antiviral defense. *Cell Host Microbe* **30**, 1–14 (2022) DOI:

10.1016/j.chom.2022.04.015.

140. Guo, M., Zhu, Y., Lin, Z., Yang, D., Zhang, A., Guo, C. & Huang, Z. Cryo-EM structure of the ssDNA-activated SPARTA complex. *Cell Res.* **33**, 731–734 (2023) DOI: 10.1038/s41422-023-00850-y.

141. Koonin, E. V. & Zhang, F. Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *BioEssays* **39**, 1–9 (2017) DOI: 10.1002/bies.201600186.

142. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020) DOI: 10.1146/annurev-virology-011620-040628.

143. Li, S., Manik, M. K., Shi, Y., Kobe, B. & Ve, T. Toll/interleukin-1 receptor domains in bacterial and plant immunity. *Curr. Opin. Microbiol.* **74**, 102316 (2023) DOI: 10.1016/j.mib.2023.102316.

144. Essuman, K., Milbrandt, J., Dangl, J. L. & Nishimura, M. T. Shared TIR enzymatic functions regulate cell death and immunity across the tree of life. *Science* **377**, 1–21 (2022) DOI: 10.1126/science.abo0001.

145. Golovinas, E., Rutkauskas, D., Manakova, E., Jankunec, M., Silanskas, A., Sasnauskas, G. & Zaremba, M. Prokaryotic Argonaute from Archaeoglobus fulgidus interacts with DNA as a homodimer. *Sci. Rep.* **11**, 4518 (2021) DOI: 10.1038/s41598-021-83889-4.

146. Johnson, A. G., Wein, T., Mayer, M. L., Duncan-Lowey, B., Yirmiya, E., Oppenheimer-Shaanan, Y., Amitai, G., Sorek, R. & Kranzusch, P. J. Bacterial gasdermins reveal an ancient mechanism of cell death. *Science* **375**, 221–225 (2022) DOI: 10.1126/science.abj8432.

147. Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G. & Sorek, R. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018) DOI: 10.1126/science.aar4120.

148. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat. Microbiol.* **11**, (2020) DOI: 10.1038/s41564-020-0777-y.

149. Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y. & Sorek, R. Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-1561.e12 (2020) DOI: 10.1016/j.cell.2020.09.065.

150. Mestre, M. R., González-Delgado, A., Gutiérrez-Rus, L. I., Martínez-Abarca, F. & Toro, N. Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. *Nucleic Acids Res.* **48**, 12632–12647 (2020) DOI: 10.1093/nar/gkaa1149.

151. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci.* **115**, E5307–E5316 (2018) DOI: 10.1073/pnas.1803440115.

152. Shah, S. A., Alkhnbashi, O. S., Behler, J., Han, W., She, Q., Hess, W. R., Garrett, R. A. & Backofen, R. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR- cas gene cassettes reveals 39 new cas gene families. *RNA Biol.* **16**, 530–542 (2019) DOI: 10.1080/15476286.2018.1483685.

153. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010) DOI: 10.1038/nature09039.

154. Frank, F., Hauver, J., Sonenberg, N. & Nagar, B. Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* **31**, 3588–3595 (2012) DOI: 10.1038/emboj.2012.204.

155. Boland, A., Tritschler, F., Heimstädt, S., Izaurralde, E. & Weichenrieder, O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep.* **11**, 522–527 (2010) DOI: 10.1038/embor.2010.81.

156. Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M. W. & Ladurner, A. G. A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nat. Struct. Mol. Biol.* **14**, 897–903 (2007) DOI: 10.1038/nsmb1302.

157. Klenk, H., Clayton, R. A., Tomb, J., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Venter, J. C., *et al.* The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* **390**, 364–370 (1997) DOI: 10.1038/37052.

158. Heidrich, N., Bauriedl, S., Barquist, L., Li, L., Schoen, C. & Vogel, J. The primary transcriptome of Neisseria meningitidis and its interaction with the RNA chaperone Hfq. *Nucleic Acids Res.* **45**, 6147–6167

(2017) DOI: 10.1093/nar/gkx168.

159. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016) DOI: 10.1186/s13104-016-1900-2.

160. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009) DOI: 10.1093/bioinformatics/btp324.

161. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009) DOI: 10.1093/bioinformatics/btp352.

162. Andrews, S. & others. FastQC: a quality control tool for high throughput sequence data. (2010).

163. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004) DOI: 10.1101/gr.849004.

164. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12.1-11.12.34 (2014) DOI: 10.1002/0471250953.bi1112s47.

165. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012) DOI: 10.1093/bioinformatics/bts356.

166. Kabsch, W. XDS. *Acta Crystallogr. Sect. D* **D66**, 125–132 (2010) DOI: 10.1107/S0907444909047337.

167. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006) DOI: 10.1107/S0907444905036693.

168. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011) DOI: 10.1107/S0907444910045749.

169. Collaborative Computational Project, N. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D Biol.*

Crystallogr. **50**, 760–763 (1994) DOI: 10.1107/S0907444994003112.

170. Vagin, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).

171. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).

172. Afonine, P. V, Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).

173. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

174. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M. & Svergun, D. I. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443 (2015) DOI: 10.1107/S160057671500254X.

175. Graewert, M. A., Da Vela, S., Gräwert, T. W., Molodenskiy, D. S., Blanchet, C. E., Svergun, D. I. & Jeffries, C. M. Adding Size Exclusion Chromatography (SEC) and Light Scattering (LS) Devices to Obtain High-Quality Small Angle X-Ray Scattering (SAXS) Data. *Crystals* **10**, 975 (2020) DOI: 10.3390/cryst10110975.

176. Panjkovich, A. & Svergun, D. I. CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018) DOI: 10.1093/bioinformatics/btx846.

177. Franke, D., Petoukhov, M. V, Konarev, P. V, Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M. & Svergun, D. I. ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017) DOI: 10.1107/S1600576717007786.

178. Manalastas-Cantos, K., Konarev, P. V, Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V, Molodenskiy, D. S., Panjkovich, A., Mertens, H.

D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I. & Franke, D. ATSAS 3.0 : expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355 (2021) DOI: 10.1107/S1600576720013412.

179. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009) DOI: 10.1107/S0021889809000338.

180. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

181. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012) DOI: 10.1107/S0021889812007662.

182. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003) DOI: 10.1107/S0021889803000268.

183. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).

184. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001) DOI: 10.1016/S0006-3495(01)76260-1.

185. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001) DOI: 10.1107/S0021889800014126.

186. Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A. O., Dickman, M. J., Horvath, P. & Siksnys, V. Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Mol. Cell* **56**, 506–517 (2014) DOI: 10.1016/j.molcel.2014.09.027.

187. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

188. Kapanidis, A. N., Lee, N. K., Laurence, T. A., Doose, S., Margeat, E. & Weiss, S. Fluorescence-aided molecule sorting: Analysis of

structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci.* **101**, 8936–8941 (2004) DOI: 10.1073/pnas.0401690101.

189. Tutkus, M., Sasnauskas, G. & Rutkauskas, D. Probing the dynamics of restriction endonuclease NgoMIV-DNA interaction by single-molecule FRET. *Biopolymers* **107**, e23075 (2017) DOI: 10.1002/bip.23075.

190. Cordes, T., Vogelsang, J. & Tinnefeld, P. On the Mechanism of Trolox as Antiblinking and Antibleaching Reagent. *J. Am. Chem. Soc.* **131**, 5018–5019 (2009) DOI: 10.1021/ja809117z.

191. Ingargiola, A., Lerner, E., Chung, S. Y., Weiss, S. & Michalet, X. FRETBursts: An open source toolkit for analysis of freely-diffusing Single-molecule FRET. *PLoS One* **11**, 1–27 (2016) DOI: 10.1371/journal.pone.0160716.

192. Nicolai, C. & Sachs, F. Solving ion channel kinetics with the QuB software. *Biophys. Rev. Lett.* **8**, 191–211 (2013) DOI: 10.1142/S1793048013300053.

193. Holden, S. J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O. J. & Kapanidis, A. N. Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophys. J.* **99**, 3102–3111 (2010) DOI: 10.1016/j.bpj.2010.09.005.

194. Lyubchenko, Y. L., Shlyakhtenko, L. S. & Ando, T. Imaging of nucleic acids with atomic force microscopy. *Methods* **54**, 274–283 (2011) DOI: 10.1016/j.ymeth.2011.02.001.

195. Schneider, S. W., Lärmer, J., Henderson, R. M. & Oberleithner, H. Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Pflugers Arch. Eur. J. Physiol.* **435**, 362–367 (1998) DOI: 10.1007/s004240050524.

196. Edstrom, R. D., Meinke, M. H., Yang, X. R., Yang, R., Elings, V. & Evans, D. F. Direct visualization of phosphorylase-phosphorylase kinase complexes by scanning tunneling and atomic force microscopy. *Biophys. J.* **58**, 1437–1448 (1990) DOI: 10.1016/S0006-3495(90)82489-9.

197. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017) DOI: 10.1038/nbt.3988.

198. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment

software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013) DOI: 10.1093/molbev/mst010.

199. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009) DOI: 10.1093/bioinformatics/btp348.

200. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010) DOI: 10.1371/journal.pone.0009490.

201. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021) DOI: 10.1093/nar/gkab301.

202. Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021) DOI: 10.1093/bioinformatics/btaa788.

203. Hildebrand, A., Remmert, M., Biegert, A. & Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins Struct. Funct. Bioinforma.* **77**, 128–132 (2009) DOI: 10.1002/prot.22499.

204. Hallgren, J., Tsirigos, K. D., Damgaard Pedersen, M., Juan, J., Armenteros, A., Marcatili, P., Nielsen, H., Krogh, A. & Winther, O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* (2022) DOI: 10.1101/2022.04.08.487609.

205. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021) DOI: 10.1038/s41586-021-03819-2.

206. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N.,

Kohli, P., Kleywegt, G., Birney, E., Hassabis, D. & Velankar, S. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022) DOI: 10.1093/nar/gkab1061.

207. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022) DOI: 10.1038/s41592-022-01488-1.

208. Holm, L. Using Dali for Protein Structure Comparison. in *Methods in Molecular Biology* vol. 2112 29–42 (2020). DOI: 10.1007/978-1-0716-0270-6_3.

209. van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J. & Steinegger, M. Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* (2023) DOI: 10.1038/s41587-023-01773-0.

210. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021) DOI: 10.1002/pro.3943.

211. Laskowski, R. A., Jabłonska, J., Pravda, L., Vareková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2017) DOI: 10.1002/pro.3289.

212. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007) DOI: 10.1016/j.jmb.2007.05.022.

213. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

214. Fischer, H., De Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010) DOI: 10.1107/S0021889809043076.

215. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P. & Fieschi, F. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J.*

Struct. Biol. **169**, 45–53 (2010) DOI: 10.1016/j.jsb.2009.08.009.

216. Manakova, E., Golovinas, E., Pocevičiūtė, R., Sasnauskas, G., Grybauskas, A., Gražulis, S. & Zaremba, M. Structural basis for sequence-specific recognition of guide and target strands by the Archaeoglobus fulgidus Argonaute protein. *Sci. Rep.* **13**, 6123 (2023) DOI: 10.1038/s41598-023-32600-w.

217. Manakova, E., Golovinas, E., Pocevičiūtė, R., Sasnauskas, G., Silanskas, A., Rutkauskas, D., Jankunec, M., Zagorskaitė, E., Jurgelaitis, E., Grybauskas, A., Venclovas, Č. & Zaremba, M. The missing part: the Archaeoglobus fulgidus Argonaute forms a functional heterodimer with an N-L1-L2 domain protein. *Nucleic Acids Res.* 1–16 (2024) DOI: 10.1093/nar/gkad1241.

218. Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J. & Hüttenhofer, A. Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7536–7541 (2002) DOI: 10.1073/pnas.112047299.

219. Deneke, C., Lipowsky, R. & Valleriani, A. Effect of ribosome shielding on mRNA stability. *Phys. Biol.* **10**, 046008 (2013) DOI: 10.1088/1478-3975/10/4/046008.

220. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268 (2004) DOI: 10.1107/S0907444904026460.

221. Lushnikov, A. Y., Potaman, V. N., Oussatcheva, E. A., Sinden, R. R. & Lyubchenko, Y. L. DNA strand arrangement within the SfiI-DNA complex: Atomic force microscopy analysis. *Biochemistry* **45**, 152–158 (2006) DOI: 10.1021/bi051767c.

222. Zaremba, M., Sasnauskas, G. & Siksnys, V. The link between restriction endonuclease fidelity and oligomeric state: A study with Bse634I. *FEBS Lett.* **586**, 3324–3329 (2012) DOI: 10.1016/j.febslet.2012.07.009.

223. Pathania, S., Jayaram, M. & Harshey, R. M. Path of DNA within the Mu transpososome: Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils. *Cell* **109**, 425–436 (2002) DOI: 10.1016/S0092-8674(02)00728-6.

224. Peterson, G. & Reznikoff, W. Tn5 transposase active site mutations

suggest position of donor backbone DNA in synaptic complex. *J. Biol. Chem.* **278**, 1904–1909 (2003) DOI: 10.1074/jbc.M208968200.

225. Hickman, A. B., Kailasan, S., Genzor, P., Haase, A. D. & Dyda, F. Casposase structure and the mechanistic link between DNA transposition and spacer acquisition by CRISPR-Cas. *Elife* **9**, 1–26 (2020) DOI: 10.7554/eLife.50004.

226. Pouget, N., Turlan, C., Destainville, N., Salomé, L. & Chandler, M. IS911 transpososome assembly as analysed by tethered particle motion. *Nucleic Acids Res.* **34**, 4313–4323 (2006) DOI: 10.1093/nar/gkl420.

227. Zander, A., Holzmeister, P., Klose, D., Tinnefeld, P. & Grohmann, D. supplement Single-molecule FRET supports the two-state model of Argonaute action. *RNA Biol.* **11**, 45–56 (2014) DOI: 10.4161/rna.27446.

228. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci.* **96**, 9459–9464 (1999) DOI: 10.1073/pnas.96.17.9459.

229. Dougherty, D. A. Cation-π Interactions Involving Aromatic Amino Acids. *J. Nutr.* **137**, 1504S-1508S (2007) DOI: 10.1093/jn/137.6.1504S.

230. Zechiedrich, E. L., Khodursky, A. B. & Cozzarelli, N. R. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in Escherichia coli. *Genes Dev.* **11**, 2580–2592 (1997) DOI: 10.1101/gad.11.19.2580.

APPENDICES

Appendix 1. Oligonucleotides used in this work.

Oligonucleotide	Sequence, 5′→3′	Notes
GS-851	ATCGACCAGGCTACG	15 nt 5′-phosphorylated DNA oligonucleotide used for crystallization experiments of fAfAgo
MZ-864	ATTATAATAGG	11 bp 5′-phosphorylated DNA used for SAXS, annealed with MZ-865
MZ-865	CCTATTATAAT	11 bp DNA used for SAXS, annealed with MZ-864
MZ-383	TGATTCTGCAGTTATAGGAACCACGGATTCGTTTGTAATGAGC	
MZ-385	TGATTGGATCCGATGATGGAATATAAAATAGTTGAAAATGGTTTGAC	
MZ-875	GCTATACTTCACTTAAATGAAACTCCTAACAATAGATTTCATCCGTATG	
MZ-876	CCTTCATACGGATGAAATCTATTGTTAGGAGTTTCATTTAAGTGAAGTATAGC	
MZ-952	ATCGTGGCCACGAT	
MZ-1026	ATCAAGGTCAAGGTACAGCACATACATAATTATAAT	
MZ-1027	ATGCTAGATGCAGCCAGTATCCTATTATAAT	
MZ-1028	GTGCTGTACCTTGACCTTGATGAACTGGCGCAACACGTATTG	
MZ-1031	ATACTGGCTGCATCTAGCATACGATCTCAACACTTAATGGTTT	
MZ-1068	ATTCTGGTCTCGGACTCCCATTACCCAAAATGGATGAG	Biotin on T22
MZ-1069	ATTCTGGTCTCGGACTCCCATTACCCAAAATGGATGAG	
MZ-1141	CCTAACAATAGATTTCATCCG	
MZ-1142	GGGTAATGGGAGTCCGAGACCAGAATCCTAACAATAGATTTCATCCGTATGAAGG	

Oligonucleotide	Sequence, 5′→3′	Notes
MZ-1143	ATTATAATTATGTATGTGCTGTACCTTGACCTTGAT	Cy3b on T3, 5′-P
MZ-1144	ATTATAATAGGATACTGGCTGCATCTAGCAT	Atto647N on T3, 5′-P
MZ-1288	ATTGTGGCCACAAT	14 nt 5′-phosphorylated DNA oligonucleotide used for SAXS experiments
MZ-1289	ATTGTACGTACAAT	14 nt 5′-phosphorylated DNA oligonucleotide used for SAXS experiments
MZ-1310	ATTGCTCTACTGTATAATGCTGTGCTGTACCTTGACCTTGAT	PCR primer for single-molecule DNA synthesis
MZ-1311	ATTGCTCTACTGTATAATGCTATACTGGCTGCATCTAGCAT	PCR primer for single-molecule DNA synthesis
MZ-1447	ATTGTACACGGCCGAAT	17 nt DNA oligonucleotide used as ssDNA and gDNA
MZ-1455	ATTCGGCCGTGTACAAT	17 nt DNA oligonucleotide, complementary to MZ-1447 and MZ-1480
MZ-1480	AUUGUACACGGCCGAAU	17 nt RNA oligonucleotide used as ssRNA and gRNA
MZ-1481	AUUCGGCCGUGUACAAU	17 nt RNA oligonucleotide, complementary to MZ-1447 and MZ-1480
MZ-1506	AUUGUGGCCACAAU	14 nt 5′-phosphorylated RNA oligonucleotide used for SAXS experiments

Oligonucleotide	Sequence, 5′→3′	Notes
MZ-1556	CGGAAUAUAUGUACAAU	17 nt RNA oligonucleotide used as tRNA, complementary to g1-g8 bases
MZ-1557	CGGAAUAUUGGUACCCG	17 nt RNA oligonucleotide used as non-complementary tRNA
MZ-1560	CGGAATATATGTACAAT	17 nt DNA oligonucleotide used as tDNA, complementary to g1-g8 bases
MZ-1561	CGGAATATTGGTACCCG	17 nt DNA oligonucleotide used as non-complementary tDNA
MZ-1655	AUUGUACACGAAGGACUGAAU	Guide RNA oligonucleotide for smFRET measurements. Has a Cy3B donor dye on 5′-U21
MZ-1656	GTGGATGCGAACGATTGCTGTGAGAGATCT-Bio	Anchor oligonucleotide for smFRET measurements. Has an Atto647N acceptor dye on 5′-G1 and a biotin on 5′-T30
MZ-1698	GUUGUACACGGCCGAAC	17 nt RNA oligonucleotide analogous to MZ-1480 but 1st base is G

Oligonucleotide	Sequence, 5′→3′	Notes
MZ-1699	UUUGUACACGGCCGAAA	17 nt RNA oligonucleotide analogous to MZ-1480 but 1st base is U
MZ-1706	AUCGUACACGGCCGGAU	17 nt RNA oligonucleotide analogous to MZ-1480 but 3rd base from 3′ is G
MZ-1707	AGUGUACACGGCCGACU	17 nt RNA oligonucleotide analogous to MZ-1480 but 2nd base from 3′ is C
MZ-1708	CUUGUACACGGCCGAAG	17 nt RNA oligonucleotide analogous to MZ-1480 but 1st base from 3′ is G, and 1st base from 5′ is C
MZ-1715	AGATCTCTCACAGCAATCGTTCGCATCCACTTTTTTTTGAATTGAAGATGTGTACAACTTTT	8 bp complementary target for smFRET measurements.
MZ-1752	AGATCTCTCACAGCAATCGTTCGCATCCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	Non-complementary target oligo for smFRET measurements

Appendix 2. Plasmid vectors used in this work.

Plasmid	Details	Source or reference, links
pJET1.2_AfAgo	Bacterial cloning vector with AfAgo gene.	https://benchling.com/s/seq-moTpkScsEXHhLZQ7dg0i?m=slm-zBpueIq0nmDtVeDWmoiN
pETDuet_AfAgo-N	Bacterial expression vector with His_6-TEV-AfAgo-N gene.	https://benchling.com/s/seq-Kyx29GowMA91jJ3yW5fQ?m=slm-bFwXuxpZt21RHU1h4zSD
pETDuet_AfAgo	Bacterial expression vector with AfAgo gene.	https://benchling.com/s/seq-qyb3cWqmCKX4AYIX4sXR?m=slm-xdFtrcDOClc5BXTwpqmX
pBAD_fAfAgo	Bacterial expression vector with His_6-AfAgo-N and AfAgo genes.	https://benchling.com/s/seq-ddaL2rIgYyQdokySjPNy?m=slm-ghmmn1JkAYJFlkMGzTVI
pBAD_scfAfAgo	Bacterial expression vector with a scfAfAgo gene.	https://benchling.com/s/seq-jPXbDQzUSWtNLzKjIwci?m=slm-BKFbZayPJYLuIRNUAjcd

Plasmid	Details	Source or reference, links
pBAD_TwinStrep-scfAfAgo	Bacterial expression vector with a TwinStrep-scfAfAgo gene.	https://benchling.com/s/seq-MmH81sIO3gB25ULKIPGq?m=slm-FyUrrhY5gZpvZ85hSZG7
pBAD_TwinStrep-AfAgo	Bacterial expression vector with a TwinStrep-AfAgo gene.	https://benchling.com/s/seq-NM4ng3rwgUGk5CyapymR?m=slm-boVvzF7S2Mz07qirUlBD
pCDFDuet_His-AfAgo-N	Bacterial expression vector with a His_6-AfAgo-N gene.	https://benchling.com/s/seq-KmezzqulpPWLyKvZ0STd?m=slm-czNO03DmyJ4Kv346IeBA

Appendix 3. X-ray crystallographic data collection and refinement statistics for AfAgo.

Oligoduplex	**ATC**GTGGCCAC**GAT** ||||||||||||||| **TAG**CACCGGTG**CTA**	**ATC**GTGGCCAC**GAT** ||||||||||||||| **TAG**CACCGGTG**CTA**	**ATT**GTGGCCAC**AAT** ||||||||||||||| **TAA**CACCGGTG**TTA**	**ATT**GTACGTAC**AAT** ||||||||||||||| **TAA**CATGCATG**TTA**
Crystallization buffer	50 mM sodium cacodylate (pH 5.5 at 25 °C), 120 mM KCl, 10 mM $MgCl_2$, 7% (w/v) PEG3350, 5% (v/v) glycerol	50 mM sodium cacodylate (pH 6.5 at 25 °C), 40 mM KCl, 10 mM $MgCl_2$, 11% (w/v) PEG3350	50 mM sodium cacodylate (pH 5.5 at 25 °C), 200 mM KCl, 10 mM $MgCl_2$, 5% (w/v) PEG4000, 5% (v/v) glycerol	
Cryo-protection buffer	100 mM sodium cacodylate (pH 5.5 at 25 °C), 200 mM KCl, 10 mM $MgCl_2$, 20% PEG3350 (w/v), 10% (v/v) glycerol			100 mM sodium cacodylate (pH 6.5 at 25 °C), 40 mM KCl, 10 mM $MgCl_2$, 20% (w/v) PEG3350, 20% (v/v) glycerol
Data collection statistics				

Space group	*P*1	*P*1	$P22_12_1$	$P22_12_1$
Cell constants a, b, c, α, β, γ	a=51.80 Å, b=60.87 Å, c=101.72 Å, α=76.56°, β=75.59°, γ=79.39°	a=51.91 Å, b=61.20 Å, c=103.09 Å, α=98.32°, β=104.96°, γ=100.62°	a=52.10 Å, b=99.55 Å, c=109.90 Å, α=β=γ=90°	a=52.01 Å, b=99.63 Å, c=109.80 Å, α=β=γ=90°
Wavelength, Å	0.97630	0.97630	0.97970	0.97970
X-ray source	PETRA III, EMBL C/O DESY, P14		PETRA III, EMBL C/O DESY, P13	
Unique reflections: overall (outer shell)	83713 (4556)	85105 (12355)	63695 (9192)	53664 (3128)
Resolution range, Å	41.50 - 1.90	54.90 - 1.80	99.52 - 1.70	99.63 - 1.80
Completeness: overall (outer shell), %	91.6 (91.2)	91.5 (90.9)	100 (100)	99.9 (100)
Multiplicity: overall (outer shell)	3.8 (3.9)	3.9 (3.9)	10.3 (10.1)	6.5 (6.7)

I/σ: overall (outer shell)	11.2 (1.6)	8.3 (1.7)	20.8 (2.3)	20.8 (2.6)
Rmerge: overall (outer shell), %	4.9 (82.1)	6.1 (63.8)	6.3 (96.9)	4.4 (74.8)
B-factor from Wilson, $Å^2$	36.4	35.0	30.8	29.6
Refinement statistics				
Resolution range, Å	40.72 - 1.90	46.45 - 1.90	54.95 - 1.70	54.90 - 1.80
Reflections: work (non-anomalous)/test	83692 (9394)	85025 (8390)	63601 (6249)	53695 (5011)
Atom number: protein/solvent	7671 (449)	7175 (495)	4325 (349)	4325 (394)
Rcryst (Rfree), %	18.1 (22.4)	17.7 (22.2)	18.3 (21.6)	18.8 (23.1)
RMSD: bond lengths, Å / bond angles, (°)	0.010 / 0.999	0.012 / 1.104	0.012 / 1.148	0.005 / 0.753

Ramachandran: favoured/allowed/ outliers, %	96.3 / 3.7 / 0	96.6 / 3.4 / 0	98 / 8 / 0	97.24 / 2.51 / 0.25
Average B-factors: all atoms/ main chain/ side chain/ solvent/ DNA, $Å^2$	48.0 / 42.5 / 43.9 / 53.0 / 81.2	46.9 / 40.0 / 40.5 / 49.2 / 85.9	42.0 / 33.8 / 39.6 / 46.0 / 74.2	47.0 / 35.3 / 41.3 / 48.6 / 91.1
PDB ID	6XUP	6XU0	6T5T	6TUO

Appendix 4. Crystal data collection and refinement statistics for AfAgo-N and fAfAgo.

Structure	AfAgo-N 1.9 Å	AfAgo-N 1.4 Å	fAfAgo-DNA
Crystallization reservoir solution	Natrix2 #21 (Hampton Research): 40 mM sodium cacodylate (pH 6.0 at 25 °C), 12 mM NaCl, 8 mM KCl, MPD 50% (v/v), 12 mM spermine tetrahydrochloride	50 mM Tris-HCl (pH 8.0 at 25 °C), 50 mM Bicine, 27% (v/v) 2-propanol, 110 mM ammonium acetate	100 mM NaHEPES (pH 7.5 at 25 °C), 2% (v/v) Ethylene glycol, 10% (w/v) PEG3350
Cryo-protection solution	none	80 mM Tris-HCl (pH 8.5 at 25 °C), 160 mM ammonium acetate, 24% (v/v) 2-propanol, 20% (v/v) glycerol	Reservoir solution supplemented with ethylene glycol to 30% (v/v)
Data collection statistics			
Space group	*P*1	$P3_221$	$P2_12_12_1$
Cell constants	a=42.03 Å, b=57.84 Å, c=61.62 Å, α=73.94°, β=89.56°, γ=89.75°	a=75.26 Å, b=75.26 Å, c=94.72 Å, α=β=90°, γ=120°	a=81.71 Å, b=105.89 Å, c=144.35 Å, α=β=γ=90°
Wavelength, Å	0.9768	0.9797	1.0100

PETRA III, EMBL C/O DESY beamline	P14	P13	P14
Unique reflections: overall (outer shell)	44415 (6503)	61624 (3045)	44100 (4576)
Resolution range, Å	55.59 – 1.95	94.72 – 1.40	144.35 – 2.59
Completeness: overall (outer shell), %	93.2 (93.5)	100 (100)	99.9 (100)
Multiplicity: overall (outer shell)	3.6 (3.6)	9.7 (8.4)	13.3 (13.7)
I/σ: overall (outer shell)	15.9 (2.2)	23.8 (2.3)	30.2 (6.6)
Rmerge: overall (outer shell), %	0.033 (0.492)	0.047 (0.908)	0.054 (0.413)
B-factor from Wilson, $Å^2$	31.8	18.7	65
Refinement statistics			
Reflections: work (non-anomalous)/test	39974 / 4433	55371 / 6208	39581 / 4454

Atom number: protein / DNA / solvent	3800 / - / 308	2116 / 373	5970 / 482 / 268
Rcryst (Rfree), %	0.17 / 0.22	0.18 / 0.21	0.17 / 0.23
RMSD: bond lengths, Å / bond angles, (°)	0.007 / 1.509	0.012 / 1.895	0.017 / 2.520
Ramachandran: favoured/allowed/ outliers, %	97% / 3% / 0%	98% / 2% / 0%	96% / 4% / 0%
average B-factors: all atoms/ main chain/ side chain/ DNA / solvent, $Å^2$	37.0 / 33.1 / 39.2 / -/42.3	24.0 / 18.9 / 23.5 / - / 37.7	60.0 / 54.5 / 57.1 / 70.2 / 62.4
PDB ID	8OLD	8OLJ	8OK9

Appendix 5. SAXS data collection and main structural parameters.

Instrument, Detector	P12, pilatus6m	
Detector-to-sample distance, m	3.0	
Wavelength, nm	0.123981	
Measured s range, nm^{-1}	0.0224526-7.3176000	
Number of buffer exposure frames averaged (measured) / frame exposure time	101 (101) / 0.995 sec	76 (80) / 0.195 sec
Number of sample exposure frames averaged (measured) / frame exposure time	24 (24) / 0.995 sec	30 (40) / 0.195 sec
Capillary temperature/ Sample changer temperature	20 °C / Room temperature	20 °C / 10 °C
Data reduction and online characterization	radaver (r11095), databsolute v0.1 (r11095)	
Structural parameters		
Sample	WT AfAgo+MZ-1289, SEC peak	AfAgoΔ+MZ-1289, 4 mg/ml
Guinier points (AUTORG)	1-87	39-132
s range, nm^{-1} (points) used in GNOM	0.0640 – 3.3457 (1-1200)	0.1860 – 3.3457 (60-1200)
Rg, nm (AUTORG/ GNOM)	3.18 ± 0.016/ 3.233 ± 0.005202	2.84 ± 0.03/ 2.879 ± 0.002440
I(0) (AUTORG/ GNOM)	0.0725 ± 0.00011/ 0.07301 ± 0.00008771	0.0428 ± 3.7e-05/ 0.04289 ± 0.00002499
Dmax, nm (DATCLASS/ SHANUM/ GNOM)	11.3/ 10.2/ 10.1	10.9/ 10.5/ 9.6
Porod volume, nm^3 (DATPOROD)	158.03	108.67
SASBDB ID	SASDH39	SASDH49

ACKNOWLEDGEMENTS

<...>
Stands at the sea,
wonders at wondering: I
a universe of atoms
an atom in the universe.

-R.P. Feynman

During this time, here, in this universe, I have had the great fortune of meeting people who made significant contributions to my life and who, in one way or another, helped me become who I am now and embark on a journey of perpetual learning, which so far led to this work you now hold in your hands. I would like to thank each and every one of you, however, the list would become obscenely long. Please, consider it non-exhaustive and in no particular order of importance. To those who did not make it on the list, for the sake of brevity or due to my forgetfulness, please accept my apologies and know that I am forever grateful.

Firstly, I would like to express my sincerest gratitude to my supervisor, Dr Mindaugas Zaremba, for believing in me and pushing me forward, sometimes to my limits. Your keen interest in science and your dedication, in concert with your compassion and sense of humour, helped me surf the zone of proximal development and emerge from the abysmal ocean, called 'the doctoral studies', a bit smarter and wiser, perhaps, and relatively unscathed.

My thanks also go out to my other colleagues, including, but not limited to, Dr Giedrius Sasnauskas, Dr Arūnas Šilanskas, Dr Elena Manakova, Dr Česlovas Venclovas, and Algirdas Grybauskas. Your assistance, consultations, and insights not only helped me grow as a researcher and a person but also allowed this thesis to come to fruition, in no small part through your direct input as co-authors of the studies and papers. I am also immeasurably grateful to the members of the 'PhD lodge': Irmantas Mogila, Dalia Smalakytė, Dr Greta Bigelytė, Jonas Juozapaitis, Dr Andrius Merkys, and Antanas Vaitkus. I appreciate all the help you provided and all the conversations and activities we had both during and outside of work. I consider myself extremely lucky to have made your acquaintance. I am also forever indebted to Reda Pocevičiūtė, whom I supervised through her bachelor's and master's studies. Thank you for trusting me and doing much of

the dirty work. I hope at least a fraction of what I learned from you, I was able to impart.

Further, my gratitude extends to my dear friends Indrė Mackevičiūtė and Lina & Vilius Belkevičiai, the latter two of whom I've known for almost half my life. You've seen me at my best and my worst and stuck with me throughout the years. I don't know where I would be without your love, kindness, support, humour and often level-headed callouts on my utter stupidity.

Of course, I'd like to thank my parents and my late grandparents—without you, there wouldn't be me. Thank you for fostering my curiosity during childhood and listening to my passionate gobbledygook on all things science, even though I know you probably didn't understand much.

Of course, I express my thanks to Dr Povilas Šimonis, Dr Gediminas Drabavičius, and Vytenis Ruzgas. Thirteen years of friendship, brotherhood, long exhaustive discussions, and dabbling in science and philosophy, games and brews—whether tea, coffee or mead—art, projects—individual and collaborative—love and relationships, joys and sorrows. Not only did you make my life more colourful, but you also helped shape me into the person I am today. I am eternally grateful that I have met you.

Finally, my thanks go to the giants I have not met personally, but who nevertheless profoundly impacted my life, philosophy, interests, and mind: Richard P. Feynman, Richard Dawkins, Stephen Hawking, Bertrand Russell, and Jordan B. Peterson. They are a source of inspiration and kindred souls in awe of the universe and the workings of man and mind, their sharp and keen wits and ideas serving as beacons for any apt or prospective thinker alike.

Technical acknowledgements

Thaks to: A. Šilanskas for purifying the proteins, E. Manakova for structural analysis, R. Pocevičiūtė for biochemistry, library prep, and SEC-MALS runs, E. Zagorskaitė for NA purification & library prep, A. Grybauskas for sequencing data analysis, Č. Venclovas for bioinformatics, G. Sasnauskas for help with SEC-MALS, extensive consultations on everything & paper co-authorship, D. Rutkauskas for smFRET, M. Jankunec for AFM.

CURRICULUM VITAE

Department of Protein – DNA Interactions, Institute of Biotechnology,

Life Sciences Center, Vilnius University

7 Saulėtekio Ave, LT- 10257 Vilnius

Phone no. +370 681 65 673

E-mail: edvardas.golovinas@gmc.vu.lt; edvardasg2007@gmail.com

EDUCATION

2014-2016 Master's degree in Biochemistry, Vilnius University

2010-2014 Bachelor's degree in Bioengineering, Vilnius Gediminas Technical University (Currently, Vilnius Tech)

WORK EXPERIENCE

2024-01–present Researcher, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions.

2023-05–present Junior research fellow, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, 2023-2026 Research Council of Lithuania research groups project "Interactions of prokaryotic Argonaute proteins with other antiphage defence systems", P-MIP-23-211.

2022-03–2023-11 Junior research fellow, Vilnius University Life Sciences Centre, EMBL partnership institute at VU LSC, project No. 01.2.2-CPVA-V-716-01-0001.

2021-04–2023-12 Junior research fellow, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions.

2020-05–2022-11 Laboratory technician, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, 2020-

2023 Research Council of Lithuania research groups project “Structural and functional studies of split prokaryotic Argonaute proteins”, S-MIP-20-37.

2018-07–2021-03 Biologist researcher, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions.

2018-01–2022-01 Specialist, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, 2018-2022 Research Council of Lithuania global grant project “Molecular mechanisms of novel bacterial antiviral defence systems”, 09.3.3-LMT-K-712-01-0126.

2017-09–2020-04 Laboratory technician, Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, 2017-2020 Research Council of Lithuania research groups project “Structural and functional studies of prokaryotic Argonaute proteins”, S-MIP-17-61.

2017-01–2017-09 Biotechnologist, Northway Biotechpharma, Vilnius (Lithuania), Downstream process.

2015-10–2016-12 Engineer, Centre for Physical Sciences and Technology, Vilnius (Lithuania). Study of materials, interfaces and model biological systems using spectroscopic methods: sum-frequency generation, IR absorption, Raman spectroscopy.

2015-02–2015-06 Engineer, Centre for Physical Sciences and Technology, Vilnius (Lithuania). Study of materials, interfaces and model biological systems using spectroscopic methods: sum-frequency generation, IR absorption, Raman spectroscopy.

2015-03–2015-05 Intern, Vilnius University Institute of Biochemistry, Vilnius (Lithuania). Study of molecular interactions in solid cholesterol and cholesterol solutions in organic solvents using Raman spectroscopy and IR absorption spectroscopy.

2014-10–2014-12 Intern, Centre for Physical Sciences and Technology, Vilnius (Lithuania). Study of molecular interactions in cholesterol-phospholipid monolayers at air/water interface using sum-frequency generation spectroscopy.

2011-09–2014-01 Laboratory technician, Vilnius Gediminas Technical University, Vilnius (Lithuania).

2013-06–2013-08 Intern, Vilnius University Institute of Biotechnology, Vilnius (Lithuania). Purification and crystallization of CRISPR/Cas proteins expressed in *E. coli* in preparation for X-ray crystallography, performing crystallization screening experiments using an Oryx8 robot.

2013-02–2013-04 Technician, Centre for Physical Sciences and Technology, Vilnius (Lithuania).

2012-10–2013-01 Intern, Centre for Physical Sciences and Technology, Vilnius (Lithuania). Study of self-assembled monolayers of lipid anchor molecules (e.g., WC14, HC18) on a gold electrode using sum-frequency generation spectroscopy.

2012-06–2012-08 Intern, Centre for Physical Sciences and Technology, Vilnius (Lithuania). Study of self-assembled monolayers of 2-mercaptoethanol on gold using sum-frequency generation spectroscopy.

PUBLISHED SCIENTIFIC RESEARCH

- Manakova, E., **Golovinas, E**, Pocevičiūtė, R., Sasnauskas, G., Šilanskas, A., Rutkauskas, D., Jankunec, M., Zagorskaitė, E., Jurgelaitis, E., Grybauskas, A., Venclovas, Č. & Zaremba, M. The missing part: the *Archaeoglobus fulgidus* Argonaute forms a functional heterodimer with an N-L1-L2 domain protein. *Nucleic Acids Res.* 1–16 (2024) DOI: 10.1093/nar/gkad1241

- Manakova, E., **Golovinas, E.**, Pocevičiūtė, R., Sasnauskas, G., Grybauskas, A., Gražulis, S. & Zaremba, M., Structural basis for sequence-specific recognition of guide and target strands by the *Archaeoglobus fulgidus* Argonaute protein, Scientific Reports 13, 6123 (2023) https://doi.org/10.1038/s41598-023-32600-w

- Zaremba, M., Dakineviciene, D., **Golovinas, E.**, Zagorskaitė, E., Stankunas, E., Lopatina, A., Sorek, R., Manakova, E., Ruksenaite, A., Silanskas, A., Asmontas, S., Grybauskas, A., Tylenyte, U., Jurgelaitis, E., Grigaitis, R., Timinskas, K., Venclovas, Č. & Siksnys, V, Short prokaryotic Argonautes provide defence against incoming mobile

genetic elements through NAD^+ depletion, Nature Microbiology 7, 1857-1869 (2022) https://doi.org/10.1038/s41564-022-01239-0.

- **Golovinas, E.**, Rutkauskas, D., Manakova, E., Jankunec, M., Silanskas, A., Sasnauskas, G. & Zaremba, M., Prokaryotic Argonaute from *Archaeoglobus fulgidus* interacts with DNA as a homodimer, Scientific Reports 11, 4518 (2021). https://doi.org/10.1038/s41598-021-83889-4.

CONFERENCES

- EMBO | EMBL Symposium: The Non-Coding Genome, Poster presentation "RNA-guided DNA Binding Specificity of a Prokaryotic Argonaute Protein from *Archaeoglobus fulgidus*", **E. Golovinas**, E. Manakova, G. Sasnauskas, S. Gražulis, V. Šikšnys and M. Zaremba. Virtual, October 13-15, 2021.

- EMBO | EMBL Symposium: The Complex Life of RNA, Poster presentation "Prokaryotic Argonaute from *Archaeoglobus fulgidus* Interacts with DNA as a Homodimer", **E. Golovinas**, D. Rutkauskas, E. Manakova, M. Jankunec, A. Šilanskas, G. Sasnauskas, M. Zaremba. Virtual, October 7-9, 2020.

- International conference Vita Scientia, Poster presentation "Specificity of the Argonaute protein from *Archaeoglobus fulgidus* to the 5′-end of the guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, V. Šikšnys, Vilnius, Lithuania, January 3, 2020.

- International conference FEBS3+ 2019, Poster presentation "Specificity of the Argonaute Protein from *Archaeoglobus fulgidus* to the 5′-end of the Guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, V. Šikšnys, Riga, Latvia, June 17-19, 2019.

- International conference for students of physics and natural sciences Open Readings 2019, Poster presentation "Specificity of the Argonaute Protein from *Archaeoglobus fulgidus* to the 5′-end of the Guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, Virginijus Šikšnys, Vilnius, Lithuania, March 19-22, 2019.

- International conference of life sciences The COINS 2019. Poster presentation "Specificity Of The Argonaute Protein From *Archaeoglobus fulgidus* to the 5′-end of the Guide", M. Zaremba, E. Manakova, **E. Golovinas**, S. Gražulis, V. Šikšnys, Vilnius, Lithuania, February 26-28, 2019.

- 58th International Conference for Students of Physics and Natural Sciences "Open Readings 2015", oral presentation "Study of the Effects of Lipid Monolayer Composition and Phase on Interfacial Water Structure by Sum-frequency Spectroscopy", **E. Golovinas**, L. Abariūtė, G. Valinčius, G. Niaura, Vilnius, March 24–27, 2015.

- International Conference of Lithuanian Society of Chemistry "CHEMISTRY & CHEMICAL TECHNOLOGY 2016", poster "Study of Temperature Influence on Molecular Interactions in Cholesterol Using Raman Spectroscopy", **E. Golovinas**, G. Niaura, G. Valinčius, Vilnius, April 28-29, 2016.

- International Conference of Natural and Life Sciences "The Coins 2015", oral presentation "Study of Lipid Raft Formation in Cholesterol-Lipid Monolayers Using Sum Frequency Generation Spectroscopy", Vilnius, March 3–7, 2015.

- International Conference of Lithuanian Chemical Society "CHEMISTRY & CHEMICAL TECHNOLOGY 2015", poster "Study of Molecular Interactions in Cholesterol-Phospholipid Monolayer Using Sum Frequency Generation Spectroscopy", **E. Golovinas**, L. Abariūtė, G. Valinčius, G. Niaura, Vilnius, January 23, 2015.

- 11th International Conference of Lithuania's Chemists, poster "Analysis of Structure of Self-Assembled Monolayers on a Gold Electrode Using Sum-frequency Generation Spectroscopy", **E. Golovinas**, I. Ignatjev, M. Mickevičius, Vilnius, September 27, 2013.

ORGANIZATIONAL AND TEACHING EXPERIENCE

2018 BioHackathon: Lab Issues – participant mentorship at a Hackathon event, curated by the Vilnius iGEM team.

2018 CRISPR 2018 Vilnius – international conference for CRISPR-Cas systems research; participation in organizational team.

OTHER SKILLS

2023-05-19 Licence for persons responsible for radiation safety in business entities working with non-medical open radioactive sources (renewed).

2022-11-10 EMBO Research integrity workshop, Life Sciences Centre, Vilnius University, Lithuania

2018-10-26 Licence for persons responsible for radiation safety in business entities working with non-medical open radioactive sources.

2018-06-01 Licence to work with non-medical open radioactive sources.

SANTRAUKA

1. SANTRUMPOS

a.r.	aminorūgštis
Af	*Archaeoglobus fulgidus*
Ago	Argonaute
AFM	atominės jėgos mikroskopija (angl. *atomic force microscopy)*
APAZ	PAZ analogas (angl. a*nalogue of PAZ*)
bp	bazių pora
dg	dvigrandinė
EMSA	elektroforezinio judrio poslinkio analizė (angl. *electrophoretic mobility shift assay*)
FRET	fluorescencinė rezonansinė energijos pernaša (angl. *Förster/fluorescence resonance energy transfer*)
MALS	daugiakampė šviesos sklaida (angl. *multi-angle light scattering*)
MID	„Middle“ – struktūrinis Argonaute baltymų domenas
NR	nukleorūgštis
PAZ	PIWI-Argonaute-Zwille domenas
PDB	Protein data bank (baltymų duomenų bankas)
PIWI	angl. *P-element induced wimpy testis*
SAXS	Rentgeno spindulių sklaida mažais kampais (angl. *Small-angle X-ray scattering*)
SEC	ekskliuzinė chromatografija (angl. *size-exclusion chromatography*)
tDNR	taikinio DNR
TIRF	visiškojo vidaus atspindžio fluorescencija (angl. *total internal reflection fluorescence*).
tRNR	taikinio RNR
vDNR	vedlio DNR
vg	viengrandinė
vRNR	vedlio RNR

2. ĮVADAS

Argonaute (Ago) baltymai aptinkami visuose trijuose gyvybės domenuose – bakterijose, archėjose ir eukariotuose. Pirmiausiai atrasti, taigi, geriausiai ištirti yra eukariotiniai Argonaute baltymai (eAgo)[1–3]. Būdami RNR interferencijos (RNRi) ląstelinės mašinerijos branduoliu, eAgo naudoja trumpus RNR vedlius tam, kad nusitaikytų į komplementarius RNR taikinius ir atliktų genų ekspresijos reguliacijos ir šeimininko apsaugos funkcijas[2]. Vis dėlto, prokariotinių Argonaute baltymų (pAgo) struktūrinė ir funkcinė įvairovė yra ženkliai didesnė: filogenetinė analizė parodė, kad pAgo gali būti skirstomi į tris atskiras šakas – ilguosius-A, ilguosius-B ir trumpuosius pAgo[4,5]. Ilgieji pAgo panašūs į eukariotinius savo struktūra ir domenų organizacija: jie sudaryti iš tų pačių pagrindinių struktūrinių domenų – N-galinio, MID, PIWI, PAZ bei dviejų linkerinių domenų L1 ir L2. Tuo tarpu, trumpieji pAgo neturi N-galinės dalies ir yra sudaryti tik iš MID ir PIWI domenų bei neretai yra asocijuoti su kitais efektorinius bei APAZ domenus turinčiais baltymais[5]. Geriausiai ištirta šaka yra ilgieji-A pAgo, kuriems priklauso visos charakterizuotos kataliziškai aktyvios pAgo nuleazės, tačiau sparčiai vystomi ir trumpųjų pAgo tyrimai. Pademonstruota, kad kai kurie trumpieji pAgo atlieka apsaugines funkcijas, gindami šeimininko ląstelę nuo fagų infekcijų ar plazmidžių pasitelkdami abortatyvią infekciją ar degraduodami įsibrovėlį arba atlieka kitas funkcijas, pvz., chromosomų dekatenavimą[6–12]. Ilgųjų-B pAgo šaka, tuo tarpu, susilaukė mažai tyrėjų dėmesio, taigi yra mažai ištirta ir tik vienas jos narys gerai charakterizuotas[13–16]. Viena iš tikėtinų priežasčių yra ta, kad visi žinomi ilgieji-B pAgo yra kataliziškai neaktyvūs, su mutuotomis kanoninio katalizinio centro aminorūgštimis PIWI nukleaziniame domene, kas potencialiai apriboja galimą jų panaudojimą nukleolitinei funkcijai atlikti įvairiuose įrankiuose, kaip tai daro CRISPR-Cas baltymai. Išties, keletas kataliziškai aktyvių pAgo buvo pritaikyti įrankiams kurti[17–19], o kai kurie naudoja ir kataliziškai neaktyvius šių pAgo mutantus[20,21]. Visgi, kataliziškai neaktyvių pAgo praktinė nauda taip pat buvo pademonstruota[10]. Taigi, kaip diskutuota anksčiau[22], tokių pAgo praktinis naudingumas negali būti atmestas. Aukšto lygio įrankių kūrimas, visgi, reikalauja gilaus, fundamentalaus mechanizmų išmanymo. Taigi, būtina išsamiai charakterizuoti tokius potencialius kandidatus bei pašalinti galimus netikslumus, kilusius iš ankstesnių tyrimų rezultatų ar jų trūkumo, potencialiai atskleidžiant naujas, neatskleistas tyrimo objektų savybes bei atveriant kelius naujiems tyrimams.

Pagrindinis šios disertacijos objektas – sutrumpintas ilgasis-B prokariotinis Argonaute baltymas iš *Archaeoglobus fulgidus* ir su juo asocijuotas baltymas. Nors šis pAgo aprašytas senokai, tačiau esantys duomenys yra nepilni ir prieštaringi. **Tikslas** buvo charakterizuoti šį pAgo bei asocijuotą baltymą biochemiškai ir struktūriškai. Šiam tikslui pasiekti buvo išsikelti šie **uždaviniai**:

1. Ištirti AfAgo oligomerinę būseną *in vitro*.
2. Nustatyti AfAgo specifiškumą vedlio bei taikinio nukleorūgštims.
3. Ištirti galimą kompleksų sudarymą tarp AfAgo ir asocijuoto baltymo.
4. Jei kompleksai formuojasi, patikrinti struktūriškai ir biochemiškai ar jie skiriasi nuo pavienio AfAgo.

Mokslinis naujumas ir praktinė vertė.

AfAgo buvo naudojamas kaip modelinis baltymas eukariotinių ir prokariotinių Argonaute baltymų charakterizavimui, nes jis buvo vienas pirmųjų ištirtų struktūriškai tiek pavienis, tiek komplekse su nukleorūgštimis[23–26]. Šis baltymas taip pat yra įdomus atvejis evoliucijoje, nes filogenetinės analizės jį klasifikuoja kaip ilgojo-B tipo pAgo, tačiau jis sudarytas tik iš MID ir PIWI domenų – bruožas, iki šiol laikytas išskirtinis trumpiesiems pAgo baltymams[4,5]. AfAgo turi kataliziškai neaktyvų PIWI domeną ir nėra pademonstruota jokių šio baltymo funkcijų *in vivo*. Ši disertacija pademonstruoja keletą svarbių atradimų. Pirma, visose žinomose AfAgo kristalinėse struktūrose, prieinamose PDB duomenų bazėje, AfAgo yra homodimerinės formos bei pasižymi ženkliu dimerizacijos paviršiumi, kuris ankstesnių tyrimų autorių nebuvo nagrinėtas[23–26]. Pasitelkiant pavienių molekulių fluorescencinę rezonansinę energijos pernašą, mažų kampų Rentgeno spindulių sklaidą, skysčių chromatografiją kombinuotą su daugiakampe šviesos sklaida bei struktūrinę analizę, rezultatai šiame darbe pademonstruoja, kad AfAgo geba formuoti homodimerinius kompleksus bei kilpines dgDNR struktūras *in vitro* – savybė, kuri iki šiol nebuvo stebėta pAgo pasaulyje. Antra, kai kurie pAgo yra specifiški vedlio ar taikinio nukleorūgšties prigimčiai (RNR arba DNR), 5′-galo nukleotidui bei 5′-galo fosforilinimo būsenai bei geba naudoti vedlio grandinę komplementaraus taikinio surišimui[22,27], kas iki šiol nebuvo aiškiai apibrėžta AfAgo atveju. Naudojant kartu su AfAgo išsigryninusių RNR sekoskaitą, rentgenostruktūrinę analizę bei elektroforezinio judrio poslinkio tyrimus buvo pademonstruota, kad AfAgo turi preferenciją vgRNR vedliui, turinčiam 5′-P-AUU galą bei geba naudoti tokį vedlį vgDNR taikinio surišimui *in vitro*.

Galiausiai, iki šiol neaprašytas baltymas AfAgo-N buvo aptiktas viename operone prieš AfAgo *A. fulgidus* DSM 8774 bei atkurtas DSM 4304, kur jis buvo paslėptas delecijos ir susidariusio skaitymo rėmelio poslinkio. Rezultatai parodė, kad šis baltymas, pavadintas AfAgo-N, yra struktūrinis ilgųjų pAgo N-L1-L2 domenų ekvivalentas ir sudaro heterodimerinį kompleksą su AfAgo per tą patį dimerizacijos paviršių, per kurį AfAgo formuoja homodimerus. Susidaręs heterodimerinis kompleksas – fAfAgo – struktūriškai panašus į trumpuosius PAZ domeno neturinčius pAgo. fAfAgo taip pat pasižymi RNR vedamu DNR taikinio surišimo aktyvumu. Šis bruožas išlieka nepakitęs ir į vieną polipeptidą sujungtam fAfAgo – scfAfAgo. Šie atradimai atkleidžia iki šiol nestebėtas pAgo savybes, tokias kaip homodimerizacija bei asociacija su neefektoriniu N-galiniu domenu, specifiškumą daugiau nei vienam 5′-galo nukleotidui bei atveria naujus kelius tyrimams ateityje.

Ginamieji disertacijos teiginiai

1. AfAgo sudaro homodimerinius kompleksus ir kilpines dgDNR struktūras *in vitro*.
2. AfAgo pasižymi specifiškumu 5′-galo AUU sekai vedlio RNR grandinėje *in vivo*.
3. AfAgo pasižymi RNR vedamu vgDNR surišimu *in vitro*.
4. AfAgo sudaro heterodimerinį kompleksą su AfAgo-N, susidarant fAfAgo, kuris panašus į PAZ domeno neturinčius pAgo.
5. fAfAgo bei į vieną polipeptidą sulietas scfAfAgo pasižymi RNR vedamu vgDNR surišimu su didesniu afiniškumu nei pavienis AfAgo.

3. METODAI

Šiame darbe naudoti įvairūs metodai: taikytas elektroforezinio judrio poslinkio metodas sąveikai su nukleorūgštimis tirti; pavienių molekulių fluorescencinė rezonansinė energijos pernaša, atominės jėgos mikroskopija, skysčių chromatografija su daugiakampe šviesos sklaida, mažų kampų Rentgeno spindulių sklaidos metodai leido patikrinti ir vizualizuoti baltymų oligomerines būsenas ir, pirmų dviejų metodų atveju, stebėti sąveikas su nukleorūgštimis; rentgenostruktūrinės analizės būdu įvertinti baltymų-nukleorūgščių kontaktai, lemiantys sąveikos specifiškumą bei tarpbaltyminė sąveika, įvertintas baltymų kumpleksų struktūrinis panašumas su kitais pAgo baltymais; RNR sekoskaita leido nustatyti baltymų specifiškumą kartų išsigryninančių RNR sekoms. Detaliai metodikos aprašytos susijusiose publikacijose bei šioje disertacijoje.

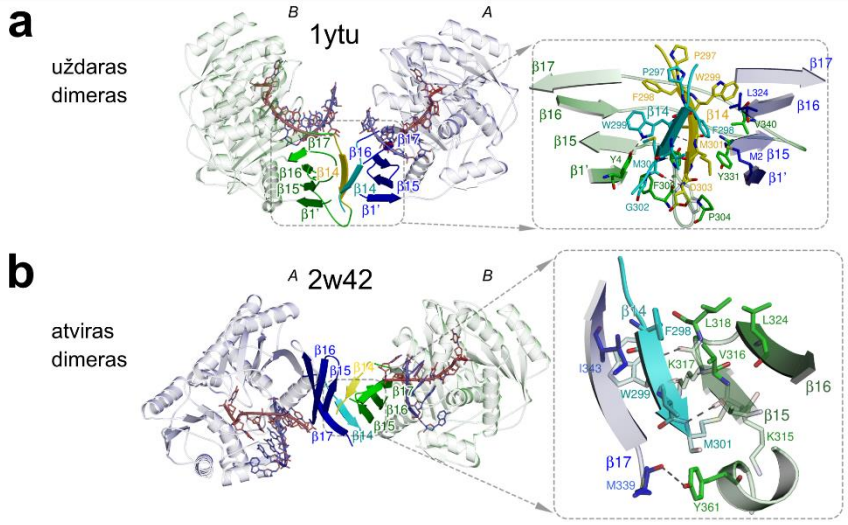
Tyrimams naudoti baltymai buvo išreikšti *E. coli* BL21(DE3) kamiene naudojant indukuojamas baltymų sintezės sistemas ir išgryninti skysčių chromatografijos būdu. Plazmidiniai vektoriai, koduojantys baltymų genus buvo sukonstruoti savarankiškai (detali metodika publikacijose arba disertacijoje) arba užsakyti susintetinti iš Twist Bioscience.

4. REZULTATŲ APTARIMAS

4.1 AfAgo dimerizacija

Visi iki šiol charakterizuoti ilgieji Argonaute baltymai sąveikauja su savo RNR ar/ir DNR taikiniais kaip monomerai, sąveikaudami su viena vedlio ir taikinio nukleorūgščių kopija. Šiame darbe aprašoma tai, kad AfAgo – prokariotinis Argonaute baltymas iš hipertermofilinės archėjos *Archaeoglobus fulgidus*, veikia kitu mechanizmu, sudarydamas homodimerus ir sąveikaudamas vienu metu su dviem vedlio-taikinio nukleorūgščių dupleksais.

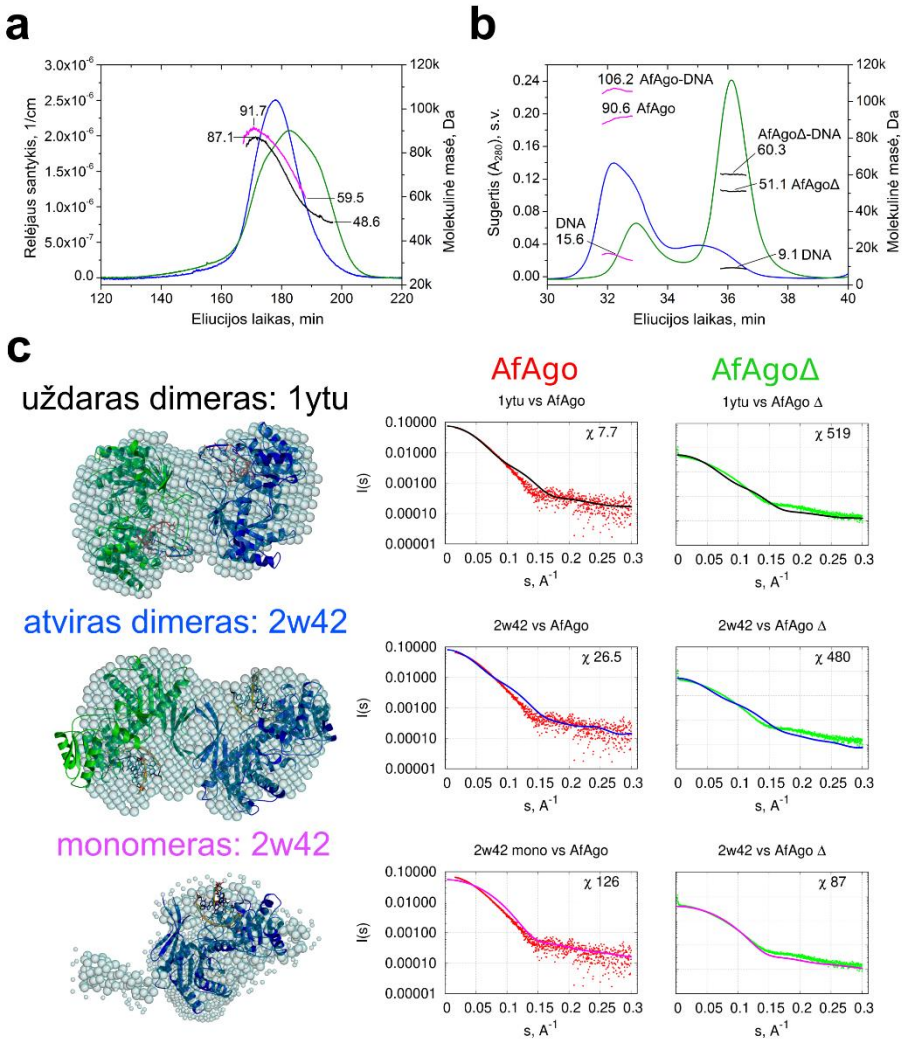
Pirmas atradimas yra tas, kad AfAgo matomas kaip dimeras visose iki šiol išspręstose kristalinėse struktūrose, įskaitant pavienio baltymo bei baltymo kompleksų su RNR ir DNR molekulėmis. Struktūrose stebimi dviejų tipų dimerizacijos paviršiai, sudaryti C-galinių β-lakštų.



1 pav. AfAgo dimerizacija. (a-b) Baltymo subvienetai nuspalvinti mėlynai (baltymo grandinė A) ir žaliai (baltymo grandinė B). Sąveikos paviršių formuojančios antrinės struktūros pažymėtos ir sunumeruotos pagal PDB struktūrą 2W42, priskyrimai atlikti PDBsum[211]. AfAgo surištos vedlio DNR/RNR grandinės nuspalvintos raudonai, o taikinio – mėlynai. Aminorūgščių liekanos 296-303, panaikintos dimerizacijos paviršiaus mutante AfAgoΔ, pavaizduotos žydra ir geltona spalvomis. Vandeniliniai ryšiai pavaizduoti punktyrinėmis linijomis. (a) AfAgo kompleksas su dgRNR (PDB ID 1YTU, abi baltymo grandinės randamos asimetriniame vienete), „uždaros“ konformacijos dimeras[25]. (b) AfAgo kompleksas su dgDNR (PDB ID 2W42, [24]) – „atviras“ dimeras. β-juostos iš abiejų subvienetų suformuoja

uždarą β-statinės struktūrą, o subvienetų sąveikos paviršius sudarytas iš gretimų subvienetų β14 ir β15 juostų. Parengta pagal Golovinas *et al*., 2021[145].

Abiejų tipų dimerizacijos paviršiai slepia panašų tirpikliu pasiekiamą paviršiaus plotą(1 pav.), tačiau nulemia skirtingą AfAgo subvienetų išsidėstymą vienas kito atžvilgiu, kurie buvo pavadinti „uždarais“ ir „atvirais“ dimerais(1 pav.)[145]. „Uždaras“ AfAgo homodimero tipas, susidarantis, kai dimerizacijos paviršių sudaro ir N-galo aminorūgštys, ir C-galo β-lakštai (1 pav., a), geriau atitinka SAXS duomenis, kas rodo, jog tai yra pagrindinė su DNR sąveikaujančio WT AfAgo dimero forma, esanti tirpale (2 pav., c). Vis dėlto, dar lieka nenustatyta, ar alternatyvi „atviro“ dimero forma, stebima keliose struktūrose (1 pav., b), buvo nulemta molekulių susipakavimo kristale, ar tai yra kitokia, rečiau pasitaikanti homodimerinio AfAgo subvienetų sąveikos forma, tirpale egzistuojanti pusiausvyroje su „uždaro“ tipo dimerais. β-lakštų, esančių subvienetų sąveikos paviršiuje, pašalinimas (mutantas AfAgoΔ) ženkliai sugadino AfAgo gebėjimą sudaryti dimerus (2 pav.).



2 pav. Apo-AfAgo ir AfAgo-DNR kompleksų SEC-MALS ir SAXS analizė. (a) WT AfAgo ir dimerizacijos mutanto AfAgoΔ be nukleorūgščių SEC-MALS analizė. Šviesos sklaidos duomenys (mėlyna WT AfAgo, žalia AfAgoΔ) parodyta greta apskaičiuotų M_w verčių (purpurinė – WT, juoda – mutantui), nurodytos didžiausia ir mažiausia apskaičiuota M_w kiekvienam baltymui. Teorinė WT AfAgo monomero masė yra 50,8 kDa, o teorinė AfAgoΔ monomero – 49,9 kDa. (b) AfAgo-DNR ir AfAgoΔ-DNR kompleksų SEC-MALS analizė. UV sugerties duomenys AfAgo-DNR (mėlyna) ir AfAgoΔ-DNR (žalia) pavaizduoti greta pilnų kompleksų, baltymo komponentės bei DNR komponentės (purpurinė AfAgo-DNR kompleksui, juoda AfAgoΔ-DNR kompleksui, atitinkamai) M_w verčių. Teorinė 2:2 WT AfAgo:DNR komplekso masė yra 119 kDa (2×50.8 + 2×8.7 kDa); teorinė 1:1 AfAgoΔ:DNR komplekso masė yra 58,6 kDa (49,9 + 8,7 kDa). (c) WT AfAgo komplekso su MZ-1289 DNR (raudoni taškai) bei dimerizacijos mutanto AfAgoΔ komplekso su MZ-1289 DNR (žali taškai) SAXS duomenys palyginti su sklaidos kreivėmis, generuotomis CRYSOL „uždaram“ dimerui su dgRNR (PDB ID 1YTU, juodos kreivės), „atviram“ dimerui (PDB ID 2W42, mėlynos kreivės)

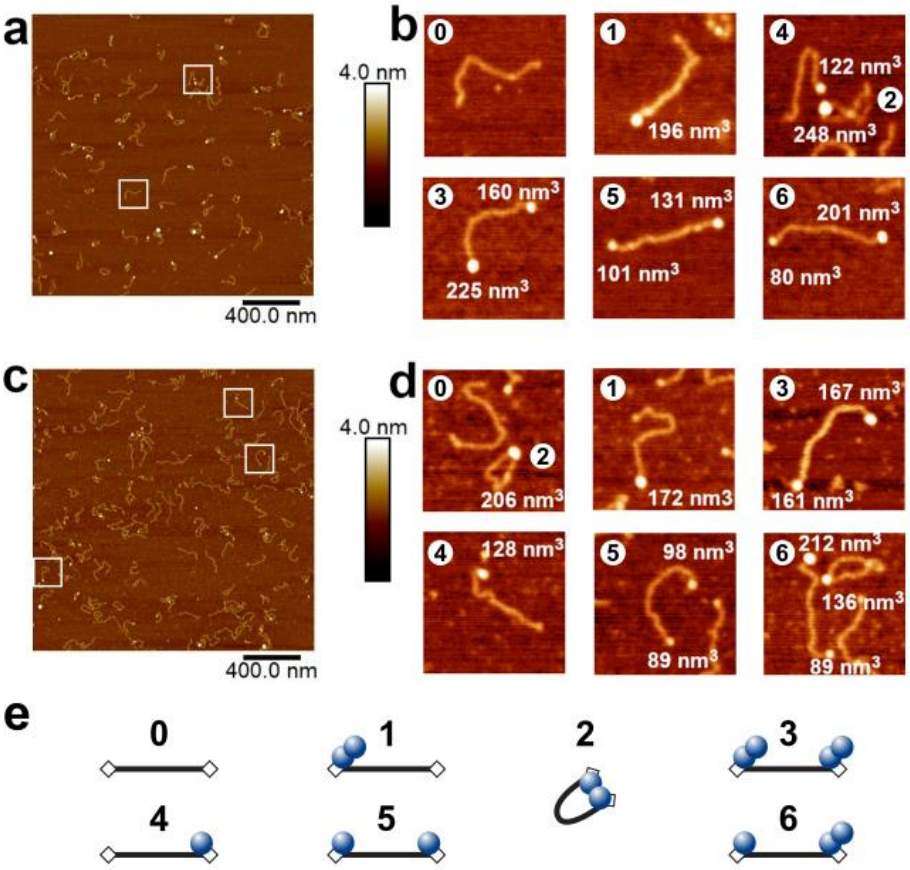
bei AfAgo-DNR kompleksui (PDB ID 2W42, purpurinės kreivės). Antrame stulpelyje pavaizduotos atitinkamos AfAgo struktūros, perdengtos su atominiais modeliais, sugeneruotais iš AfAgo komplekso su MZ-1289 oligodupleksu SAXS duomenų. Parengta pagal Golovinas *et al*., 2021[145].

WT AfAgo homodimerui vienu metu prisijungiant prie abiejų linijinio DNR fragmento galų – po vieną AfAgo monomerą vienam DNR galui – susidaro kilpinė DNR struktūra. Tokių kilpinių DNR struktūrų susidarimas tiesiogiai stebėtas naudojant AFM (3 pav.). Kaip parodyta 1 lentelėje, žiedo formos AfAgo-DNR kompleksai sudaro didžiąją visų aptiktų baltymo-DNR kompleksų dalį. Ženklus kilpinių DNR kompleksų susidarymo sumažėjimas esant dimerizacijos paviršiaus mutantui AfAgoΔ (1 lentelė) rodo, kad DNR kilpų sudarymą iš tikrųjų nulemia dimerinė AfAgo baltymo forma.

1 lentelė. AfAgo-DNR kompleksai, stebėti AFM (3 pav.). Parengta pagal Golovinas *et al*., 2021[145].

Kompl. / Baltym.	DNR kilpos, %	*Linijinė*	
		Baltmas ant vieno galo, %	*Baltymas ant abiejų galų, %*
WT AfAgo	51% (n=95)	35% (n=66)	13% (n=26)
	*Monomeras (n=20) Dimeras (n=47) Kita (n=28)	Monomeras (n=33) Dimeras (n=22) Kita (n=11)	Monomeras-monomeras (n=6) Dimeras-dimeras (n=4) Monomeras-dimeras (n=13) Other (n=3)
AfAgoΔ	26% (n=44)	40% (n=67)	34% (n=58)
	Monomeras (n=24) Dimeras (n=12) Kita (n=8)	Monomeras (n=38) Dimeras (n=24) Kita (n=5)	Monomeras-monomeras (n=20) Dimeras-dimeras (n=8) Monomeras-dimeras (n=25) Kita (n=5)

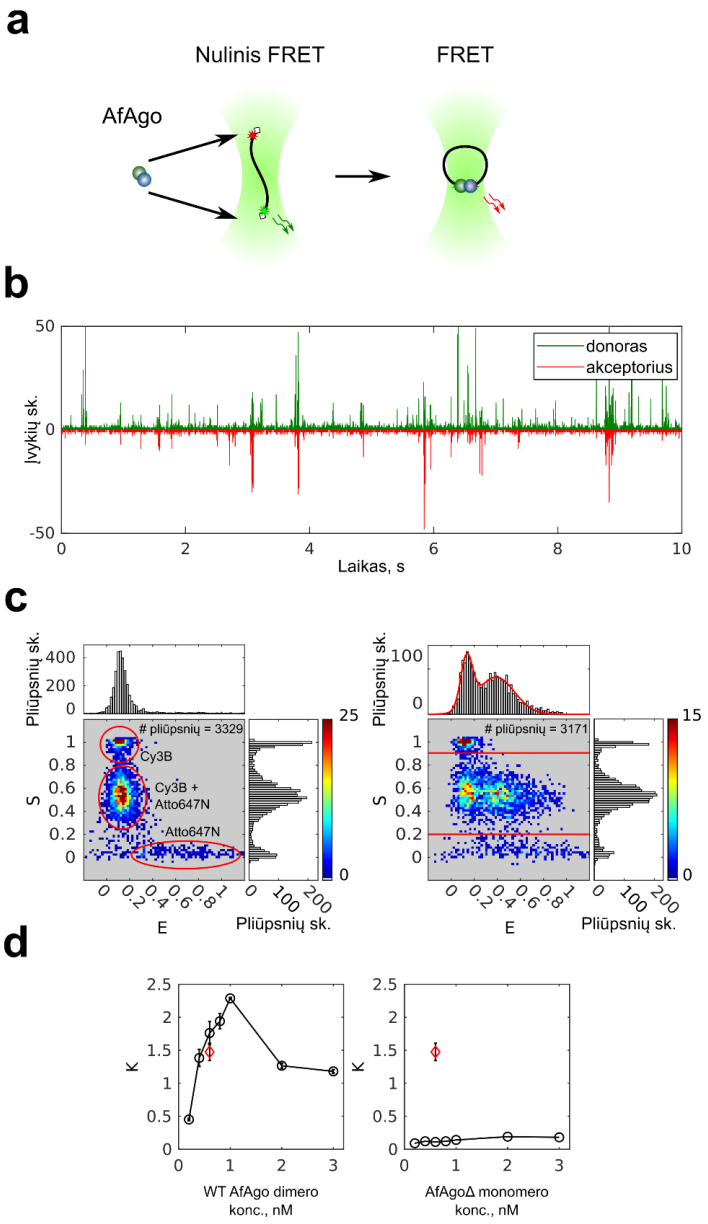
*Apskaičiuotas WT AfAgo (50,8 kDa) ir AfAgoΔ (49,9 kDa) tūriai ~100 nm^3. Pagal tai iš AFM apskaičiuoti baltymų tūriai buvo paskirstyti į tris kategorijas pagal teorinį tūrį: monomeras (<150 nm^3), dimeras (150 – 250 nm^3), ir kiti dariniai (>250 nm^3).



3 pav. AfAgo sudarytų DNR kilpų vizualizacija AFM. Pavaizduoti reprezentatyvūs bei 4 kartus padidinti WT AfAgo-DNR (a, b) bei AfAgoΔ-DNR (c, d) kompleksų, adsorbuotų ant APS funkcionalizuoto žėručio, vaizdai, gauti ore, nurodyti apskaičiuoti baltymų tūriai. Kiekvieno paveikslo plotas (a-c) dalyse yra 4 μm^2, mastelio dydis 400 nm, aukščio skalė 4 nm; aukščio skalė b ir d dalyse yra 3 nm. Regionai, apibrėžti baltais stačiakapiais paveiksluose (a) ir (c) pažymi keletą stebėtų baltymų-DNR kompleksų, pavaizduotų išdidintuose paveiksluose (b) ir (d). Remiantis baltymų tūriais, AfAgo-DNR kompleksai, pavaizduoti paveiksluose b ir d, priskirti skirtingoms baltymo-DNR stechiometrijoms ir išsidėstymams (sunumeruota nuo 0 iki 6), pavaizduotiems paveiksle (e): „0“ – laisva DNR; „1“ – AfAgo dimeras sąveikauja su vienu DNR galu; „2“ – afAgo dimeras formuoja DNR kilpą; „3“ – du AfAgo dimerai ant skirtingų DNR galų; „4“ – AfAgo monomeras ant vieno DNR galo; „5“ – du AfAgo monomerai ant skirtingų DNR galų; „6“ – monomeras ir dimeras ant skirtingų DNR galų. Parengta pagal Golovinas *et al*., 2021[145].

Siekiant nuodugniau ištirti WT AfAgo-DNR sąveikas tirpale, buvo atlikti pavienių molekulių FRET matavimai (4 pav.), naudojant DNR molekulės fragmentą, galuose pažymėtą donorinio ir akceptorinio fluoroforų pora. DNR substratas buvo parinktas toks, kad AfAgo dimerui prisijungus prie abiejų

DNR galų, fluoroforai būtų suartinti erdvėje, kas leistų vykti FRET. Donorinio ir akceptorinio kanalų signalų palyginimas esant laisvai DNR ir DNR su WT AfAgo arba dimerizacijos paviršiaus mutantu AfAgoΔ parodė, kad tik WT AfAgo efektyviai formuoja DNR kilpas, kas leidžia manyti, jog DNR kilpinių struktūrų sudarymą nulemia dimerinis AfAgo.

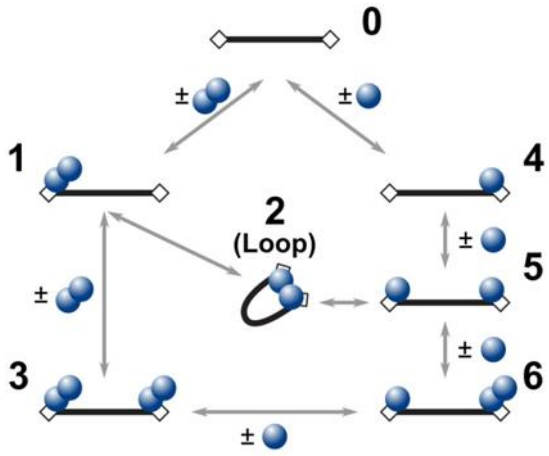


4 pav. AfAgo-DNR sąveikų pavienių molekulių tyrimai tirpale. (a) Schematinis pavienių molekulių tyrimų pavaizdavimas. Kairėje – laisva DNR, dešinėje – AfAgo-DNR (AfAgo – mėlynas ir žalias rutuliai) kilpinės būsenos kompleksas. (b) 25 pM DNR su 2 nM AfAgo fluorescencijos intensyvumo trajektorijos su 1 ms grupavimu. Raudona – invertuota ekceptoriaus fluorescencija esant donoro žadinimui; žalia – donoro fluorescencija žadinant donorą. (c) Kairėje – pavienės DNR E-S histograma. Viršutinė ir šoninė ašys vaizduoja visų pliūpsnių vienmates E (artumo santykio) bei S (donoro/akceptoriaus stechiometrijos) histogramas, atitinkamai. Pažymėti plotai atspindi tik donoru žymėtą DNR, tik akceptoriumi

žymėtą DNR bei abiem fluoroforais žymėtą DNR. Dešinėje – DNR su 2 nM AfAgo E-S histograma. Vienmatė E histograma viršuje kyla iš pliūpsnių, kurių S = 0,2 – 0,9, pavaizdavimui atskirtų horizontaliomis linijomis E-S histogramoje. Raudona kreivė yra histogramos aproksimacija dvigubu Gauso skirstiniu, iš kurio gauti Gauso skirstinių maksimumų centrai E ašyje (0,13 ± 0,01 ir 0,39 ± 0,02). (d) Kairėje – kilpinių ir nekilpinių DNR molekulių santykio (parametras K) priklausomybė nuo WT AfAgo koncentracijos. Dešinėje – parametro K priklausomybė nuo AfAgoΔ koncentracijos. Raudoni keturkampiai abiejuose grafikuose atspindi konkurencijos eksperimentą, atliktą su 1,2 nM WT AfAgo ir 0,6 nM AfAgoΔ. Visi eksperimentiniai taškai yra trijų matavimų vidurkiai ±1 standartinis nuokrypis. Parengta pagal Golovinas *et al*., 2021[145].

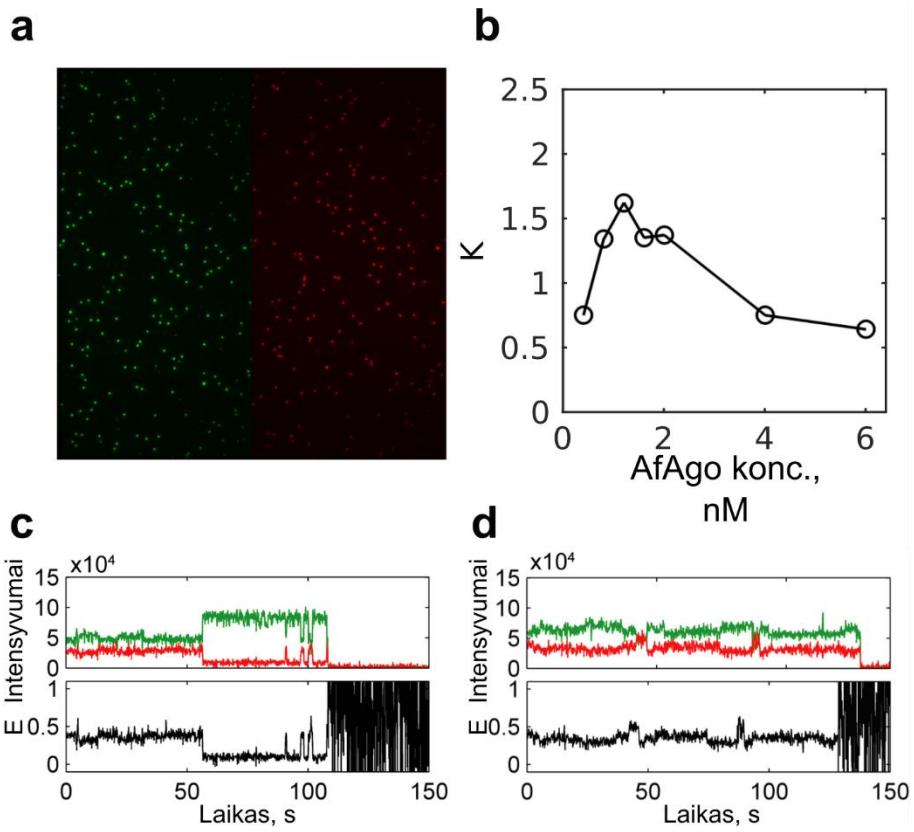
Laikant, kad pavienis WT AfAgo (apo forma) tirpale yra nestabilus dimeras (2 pav., a), yra bent du galimi mechanizmai, kuriais susidaro WT AfAgo dimero bei kilpinės formos DNR kompleksas: vienu atveju dalyvauja apo-AfAgo homodimerai (5 pav., kairėje), kitu atveju – apo-AfAgo monomerai (5 pav., dešinėje). Pirmuoju atveju, sąveikos reakcija gali vykti per (I) laisvos DNR (populiacija „0“) sąveiką su vienu WT AfAgo dimeru, kuris prisijungia prie vieno DNR galo (populiacija „1“); (II) priešingo laisvo DNR galo sąveiką su jau prie kito DNR galo prisijungusiu AfAgo dimeru intramolekulinės reakcijos metu, susidarant kilpiniam kompleksui (populiacija „2“); (III) alternatyviai, antro WT AfAgo dimero prisijungimas prie laisvojo populiacijos „1“ DNR galo sudaro populiaciją „3“, kuri nebegali sudaryti kilpinių struktūrų. Toks mechanizmas buvo pademonstruotas su nemažai baltymų, gebančių sudaryti kilpines DNR struktūras, pvz., su restrikcijos endonukleazėmis [189,221,222] ir transpozazėmis[223–226]. Kitu atveju, DNR kilpinių struktūrų sudarymas vyksta (I) vienam AfAgo monomerui prisijungiant prie vieno DNR galo (populiacija „4“, 5 pav.); (II) antram AfAgo monomerui prisijungiant prie priešingo DNR galo (populiacija „5“); (III) sąveikaujant dviems su DNR galais sąveikaujantiems monomerams, susidarant kilpiniam kompleksui „2“; DNR kilpinių struktūrų susidarymas ribojamas (IV) prie su DNR sąveikaujančių AfAgo monomerų prisijungiant papildomiems su DNR nesąveikaujantiems AfAgo monomerams (populiacija „6“). Abu reakcijos keliai rodo, kad esant aukštesnėms baltymo koncentracijoms, kilpines struktūras sudarančių kompleksų populiacija turėtų mažėti dėl AfAgo dimerų susidarymo priešinguose DNR galuose (populiacija „3“). Pavienių molekulių FRET eksperimentų tirpale rezultatai palaiko šio mechanizmo validumą: kilpines struktūras formuojančių DNR molekulių populiacija didėja iki yra pasiekiama optimali baltymo koncentracija (~2 nM, 4 pav., d), tačiau ima mažėti toliau didinant WT AfAgo koncentraciją. Nepaisant to, kad visų rūšių kompleksų populiacijos, pavaizduotos 5 pav. buvo stebimos AFM (3 pav.), lieka neišaiškinta, kokį santykinį indėlį

kiekvienas reakcijos kelias turi bendroje DNR kilpinių struktūrų susidarymo reakcijoje.



5 pav. Kinetinės schemos, vaizduojančios galimus sąveikos kelius tarp dvi taikinio sritis turinčio DNR fragment ir WT AfAgo. Juoda spalva pavaizduota DNR, baltais stačiakampiais – su AfAgo sąveikaujančios sritys (fosforilinti DNR galai), mėlynais skrituliais – AfAgo monomerai. Įvairių baltymo-DNR sąveikų numeracija atitinka kompleksų, stebėtų AFM, numeraciją 3 pav. Pabrėžtina, kad būsenos 1 ir 6 gali susidaryti alternatyviais keliais, nepavaizduotais schemoje: būsena 1 gali susiformuoti dviems monomerams paeiliui prisijungiant prie vieno DNR galo; būsena 6 gali atsirasti, kai monomeras ir dimeras sąveikauja su skirtingais DNR galais. Parengta pagal Golovinas *et al*., 2021[145].

Pavienių ant paviršiaus imobilizuotų DNR molekulių eksperimentai leido įvertinti WT AfAgo sukeltų DNR kilpinių struktūrų susidarymo dinamines savybes. Išaiškinta, kad (I) šios DNR kilpinės struktūros yra pakankamai stabilios – apatinis kilpinės būsenos trukmės įvertis viršija 30 sek. (6 pav., c); (II) kilpinių kompleksų artumo santykis E kinta laikui bėgant, kas leidžia manyti, jog savita AfAgo dimero dinamika gali kilti dėl lankstaus dimerizacijos paviršiaus (6 pav., d).



6 pav. Pavienių molekulių eksperimentai. (a) Ant paviršiaus imobilizuotų DNR fragmentų fluorescencijos atvaizdas – 20 kadrų fluorescencijos filmo vidurkis. Kairėje (žalia) yra donoro vaizdas žadinant donorą, o dešinėje (raudona) – akceptoriaus vaizdas žadinant akceptorių. (b) Parametro K priklausomybė nuo AfAgo koncentracijos eksperimente su biotinilintomis DNR tirpale. (c, d) Įvairių AfAgo nulemtų DNR kilpų susidarymo dinamikų pavyzdžiai TIRF eksperimentuose. Parengta pagal Golovinas *et al*., 2021[145].

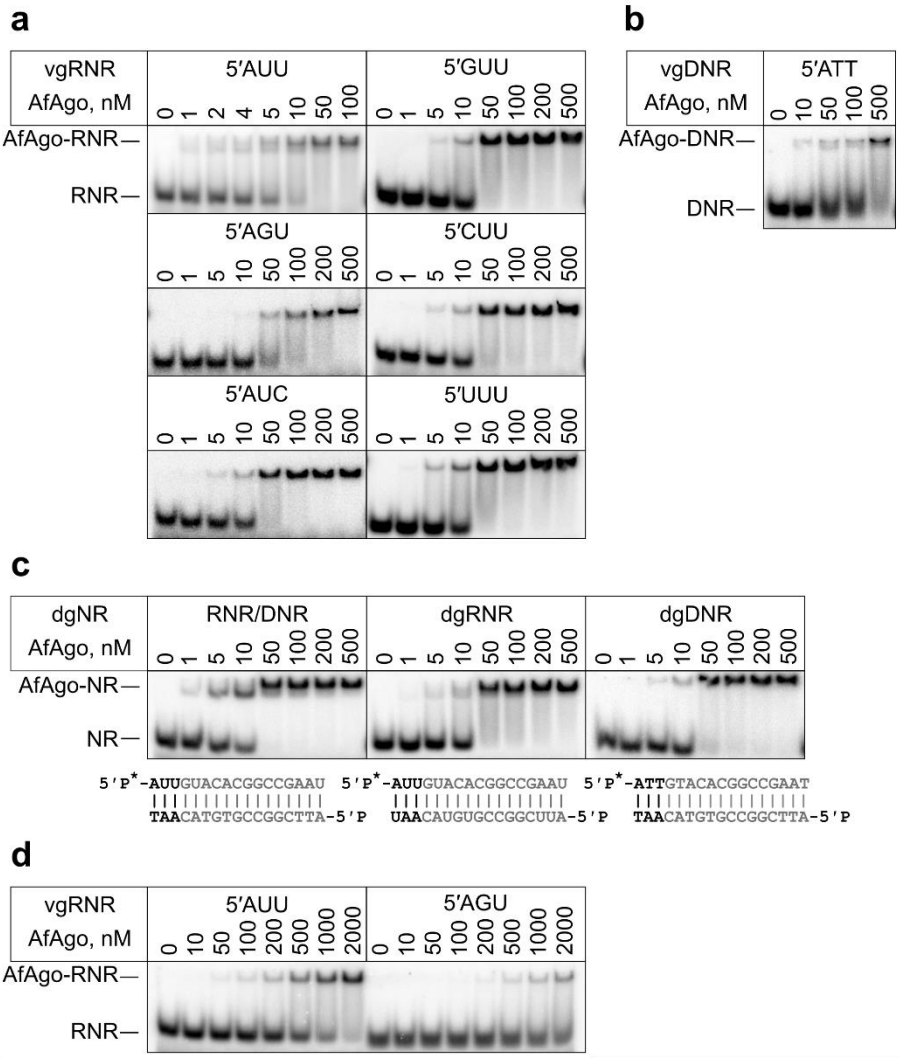
4.2 AfAgo sąveikos su nuleorūgštimis

Argonaute baltymai naudoja RNR arba DNR vedlį specifiniam RNR arba DNR taikinio grandinės surišimui[2,27]. Teisingas bazių porų formavimasis tarp dviejų grandinių nulemia taikinio grandinės kirpimą (esant kataliziškai aktyviems eAgo ar ilgiems pAgo, dalyvaujantiems priešvirusinės apsaugos mechanizmuose) arba papildomų baltymų-partnerių pritraukimą (kataliziškai neaktyviems eAgo). Vis dėlto, daugelio pAgo baltymų, ypač kataliziškai neaktyvių „trumpųjų“ pAgo, funkcijos ir veikimo mechanizmai lieka nežinomi. Šiame darbe tirta sutrumpinto ilgojo-B [4] prokariotinio Argonaute baltymo iš *A. fulgidus* preferencija vedlio ir taikinio nukleorūgštims,

atskleistas šio baltymo specifiškumas vedlio grandinės 5′-galo nukleotidams bei komplementariam taikinio grandinės sekos fragmentui.

2 lentelė. Iš EMSA ekperimentų apskaičiuotos K_d vertės įvairiems patikrintiems nukleorūgščių substratams. Preformuoto AfAgo-vedlio komplekso sąveikos su vgRNR ir vgDNR taikiniais K_d buvo nustatytos naudojant vieną numanomai optimalų vgRNR vedlį ir du taikinius, komplementarius vedliui „sėklos“ (angl. *seed*) regione. K_d apskaičiuotos iš eksperimentų, kuriuose nenaudotas heparinas. Vertės yra trijų nepriklausomų pakartojimų aritmetinis vidurkis ± standartinis nuokrypis. Parengta pagal Manakova *et al*., 2023[216].

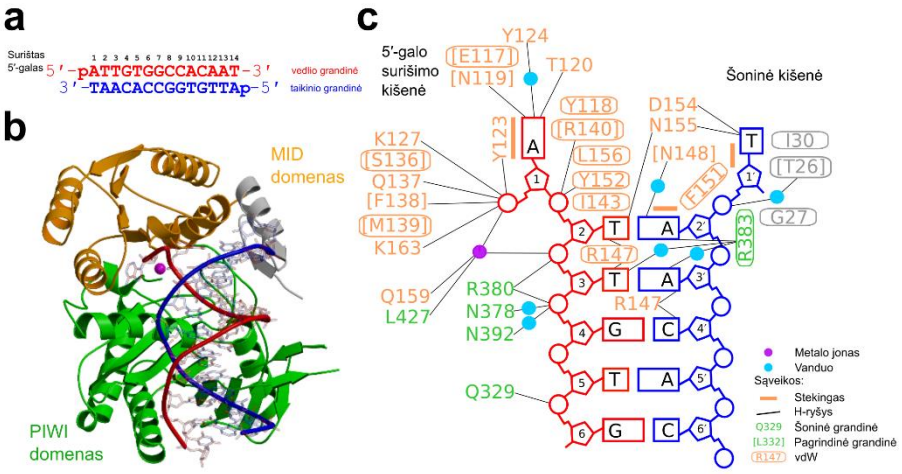
Oligonukleotidas	5′-galas	K_d, nM
vgRNR	AUU	3.8 ± 0.1
	GUU	42 ± 2.1
	CUU	28 ± 2.7
	UUU	15 ± 2.8
	AGU	84 ± 2.9
	AUC	16 ± 2.1
vgDNR	ATT	236 ± 35
RNR/DNR	AUU	6.1 ± 0.04
dgDNR	ATT	37 ± 17
dgRNR	AUU	15 ± 0.9
NR sąveika su AfAgo-vRNR kompleksu		
Vedlys	Taikinys	K_d, nM
vgRNR	vgDNR	7.5 ± 0.4
	vgRNR	33 ± 4.2



7 pav. AfAgo sąveikos su nukleorūgštimis *in vitro*. EMSA eksperimentai atlikti esant 5 nM bendrai vgRNR (a), vgDNR (b) ir dgNR (c) koncentracijai, varijuojant AfAgo koncentraciją (nurodyta virš kiekvieno takelio). Dvigrandininių nukleorūgščių sekos ir struktūros pavaizduotos schematiškai po atitinkamo gelio atvaizdu. AfAgo bazių atpažinimo mechanizmui aktualios vedlio 5′-galo ir taikinio 3′-galo bazės paryškintos juodai, likusi sekos dalis pilka. 5′-^{32}P pažymėta grandinė nurodyta žvaigždute. Apskaičiuotos K_d vertės nurodytos 2 lentelėje. Parengta pagal Manakova *et al*., 2023[216].

Pirma, šioje disertacijoje pateikiami rezultatai rodo, kad AfAgo stipriai sąveikauja su nukleorūgštimis *in vivo*, ypač su trumpais RNR fragmentais, turinčiais AUU seką 5′-gale. pAgo baltymų išsigryninimas kartu su RNR buvo stebėtas anksčiau, pvz., RsAgo, kuris turėjo specifiškumą 5′-UY (Y –

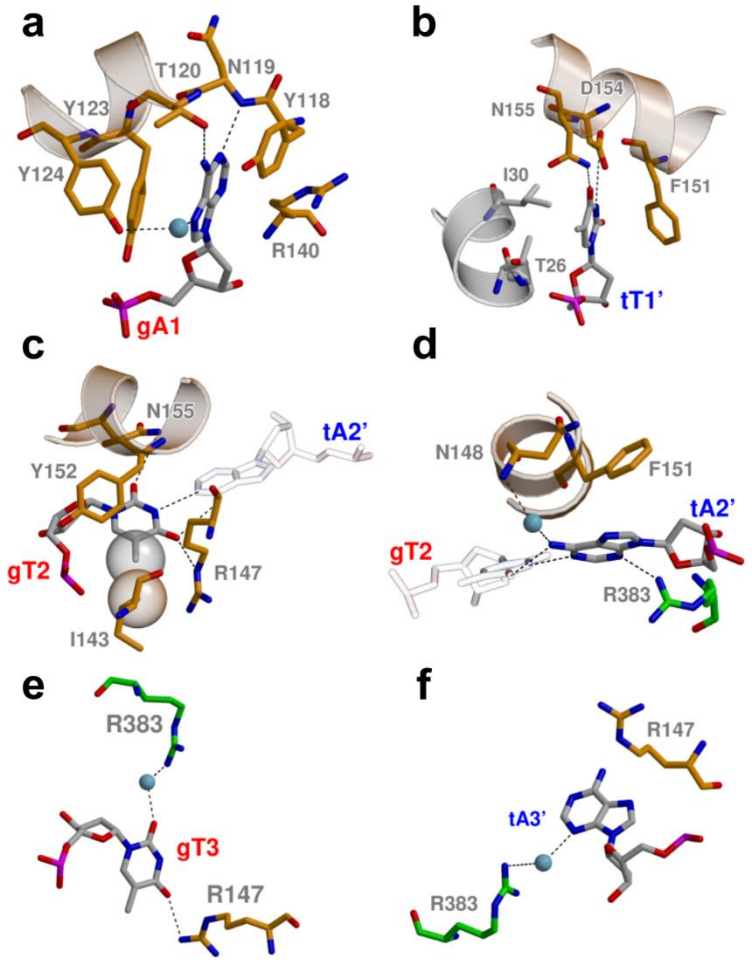
pirimidino bazė) seką turinčiai RNR[14]. Stipri AfAgo sąveika su RNR, panašu, prieštarauja ankstesniems tyrimams[25,26], kuriuose autoriai aprašo atrankią AfAgo sąveiką su viengrandinine bei dvigrandinine DNR, bet ne RNR. Tikėtina, kad šis nesutapimas atsirado dėl to, kad buvo naudotos neoptimalią 5′-galo seką turinčios RNR ir DNR (5′-U ar 5′-C), nes tyrimai parodė, kad *in vitro* AfAgo taip pat turi didesnį polinkį sąveikauti su RNR, turinčia 5′-AUU gala ir bazių pakeitimai 5′-gale sumažina specifiškumą (7 pav., 2 lentelė). Tai yra aiški indikacija, kad AfAgo naudoja vgRNR kaip vedlį bei geba sudaryti bazėms specifines sąveikas su RNR vedlio 5′-galu.



8 pav. AfAgo-DNR komplekso struktūra. (a) 5′-ATT DNR oligodupleksas, naudotas kristalizacijai. (b) bendra AfAgo-DNR komplekso struktūra. DNR grandinių karkasas nuspalvintas kaip paveksle (a). DNR bazės bespalvės, Mg^{2+} jonas, dalyvaujantis vedlio grandinės 5′-fosfato koordinavime, pavaizduotas purpuriniu rutuliu. (c) AfAgo sudaromų kontaktų su DNR schematinis pavaizdavimas. Parengta pagal Manakova *et al*., 2023[216].

Taip pat, šiame darbe pademonstruotas AfAgo-RNR komplekso gebėjimas sąveikauti su vgDNR ir vgRNR taikiniais *in vitro*, kur AfAgo, kaip ir daug kitų pAgo[27,112], pasižymėjo ryškia preferencija vgDNR taikiniams, lyginant su vgRNR taikiniais. Tai rodo, kad *in vivo* AfAgo taip pat galėtų naudoti RNR vedlius nusitaikyti į DNR. Nors *A. fulgidus* yra hipertermofilinis mikroorganizmas ir dauguma eksperimentų atlikti kambario temperatūros sąlygomis (kas nėra retas atvejis šioje tyrimų srityje[127,227]), manoma, jog tai nepaneigia rezultatų ir išvadų apie AfAgo preferenciją vgRNR ir vgDNR kaip optimaliems ir vedliui ir taikiniui, atitinkamai. Tai pagrindžia ir tas faktas, kad inkubuojant AfAgo-vgRNR sąveikos reakcijos mišinius aukštesnėje temperatūroje (70° C) prieš atliekant EMSA neturėjo įtakos

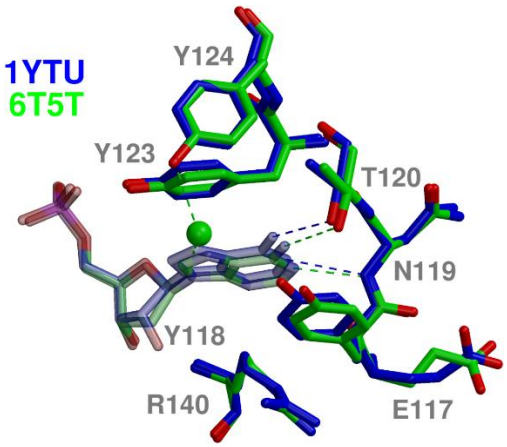
AfAgo gebėjimui atpažinti vgRNR 5′-galo sekas, nors ir buvo stebimas bendras AfAgo sąveikos su substratais afiniškumo sumažėjimas visų vgRNR atveju (7 pav., d). Tai, tikėtina, galėjo vykti dėl to, kad eksperimentinėmis sąlygomis trūko baltymus stabilizuojančių viduląstelinių faktorių, įprastai esančių šeimininko ląstelėse.



9 pav. AfAgo sąveika su pirmomis trimis 5′-ATT DNR duplekso bazių poromis. gA1 (a) ir tT1′ (b) atitinkamose surišimo kišenėse. (c, d) Antrosios bazių poros gT2 ir tA2′ atpažinimas. (e, f) Sąveikos su trečiosios bazių poros gT3 ir tA3′. Parengta pagal Manakova *et al*., 2023[216].

Taip pat buvo išspręstos keturios AfAgo sąveikaujančio su DNR-DNR oligodupleksais kristalinės struktūros, kuriose DNR turi 5′-AT galinę seką, kuri panaši į 5′-AU galą, randamą *in vivo* su AfAgo sąveikaujančiose RNR.

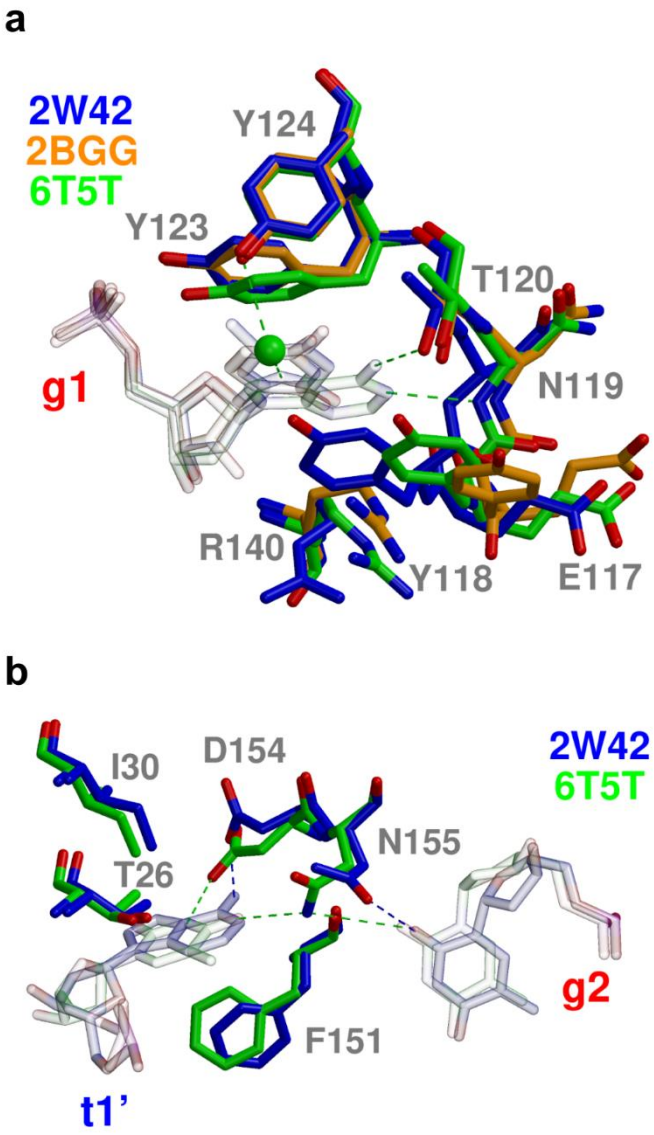
Gautos dvi kristalinės struktūros su skirtingais, į optimalų panašiais 5′-ATT oligodupleksais (PDB ID 6T5T ir 6TUO, atitinkamai) bei dvi struktūros su į suboptimalų panašiais 5′-ATC oligodupleksais (PDB ID 6XUP ir 6XU0). Nors sąveikų su 5′-ATT oligodupleksais negalima tiesiogiai tapatinti su galimomis sąveikomis su 5′-AUU RNR vedliu, kur AfAgo gali įgyti kiek kitas konformacijas ir potencialiai ne taip stipriai sąveikauti su RNR vedliu, mūsų struktūriniai duomenys rodo, kad AfAgo taiko bazėms specifines sąveikas vedlio bei taikinio grandinių galų nukleotidų atpažinimui (8 pav., 9 pav.). Ši interpretacija atitinka anksčiau publikuotas AfAgo struktūras su neoptimaliais ar RNR dupleksais (5′-P-TTC, PDB ID 2W42 [24] ir 5′-P-UUC, PDB ID 2BGG[26], atitinkamai), bei su beveik optimaliu RNR-RNR dupleksu (5′-P-AGA, PDB ID 1YTU[25]).



10 pav. g1A bazės surišimas 6T5T ir 1YTU struktūrose. Vandens molecule 6T5T struktūroje atvaizduota žaliu rutuliu, vandeniliniai ryšiai – punktyrinėmis linijomis. Parengta pagal Manakova *et al*., 2023[216].

Daugiausia bazėms specifinių kontaktų sudaroma su 5′-galo vedlio grandinės adenino baze gA1 (g, angl. *guide* – vedlys) ir jam komplementariu taikinio grandinės timinu tT1′. Ši bazių pora yra suardyta, bazės išsuktos ir surištos atskirose baltymo surišimo kišenėse. Kaip pavaizduota 10 pav., sąveikos su gA1 stebimos naujose struktūrose su DNR dupleksu yra labai panašios į tas, kurios randamos AfAgo-RNR komplekso struktūroje 1YTU [25] (bazės kitose šių struktūrų pozicijose skiriasi, taigi negali būti tiesiogiai lyginamos). Šis bazėms specifinių kontaktų panašumas su vRNR [25] bei vDNR [216] rodo, kad vDNR/tDNR duplekso galas, naudotas šiame tyrime, yra pakankamai tapatus optimaliam vRNR/tDNR heterodupleksui. Panašus ekvivalentiškos bazių poros suardymas taip pat stebėtas AfAgo struktūrose su

neoptimaliais 5′-galo nukleotidais (PDB ID 2W42 [24] ir 2BGG[26]). Šiuo atveju, išsukta gT/U1 bazė 5′-galo surišimo kišenėje negali sudaryti adeninui specifinių kontaktų, stebimų kristalinėse struktūrose aptartose šiame darbe, įskaitant: vandenilinius ryšius su gA1 baze per Asn119 pagrindinės grandinės N atomą bei vandenilinį ryšį tarp Y124 OH grupės ir gA1 N7 atomo, susidarantį per Thr120 OH grupę ir vandens molekulę (11 pav.). Kadangi visose esamose AfAgo struktūrose su RNR nesuporuota t1′ bazė nepatenka į „šoninę“ kišenę, tik tT1′ sąveikos „šoninėje“ kišenėje struktūrose 6XUP, 6XU0, 6T5T, 6TUO gali būti lyginamos su tA1′ sąveikomis struktūroje 2W42[24] (11 pav., b). Šiame darbe pateiktose struktūrose Asn155 šoninė grandinė vienu metu sąveikauja su t1′ (tT1′) ir g2 (gT2) bazėmis, o tT1′ sudaro papildomą vandenilinį ryšį su Asp154. Struktūroje 2W42 [24] tA1′ bazė “šoninėje” kišenėje sudaro vandenilinį ryšį su Asp154 šonine grandine, tačiau Asn155 konformacija netinkama sąveikai su tA1′. Antrasis vedlio sekos nukleotidas gT2 ir jam komplementarus taikinio sekos nukleotidas tA2′ sudaro mažiau bazėms specifinių kontaktų su baltymu (9 pav.), tačiau to pakanka specifinei šių bazių atrankai, lyginant su alternatyviomis bazių poromis.



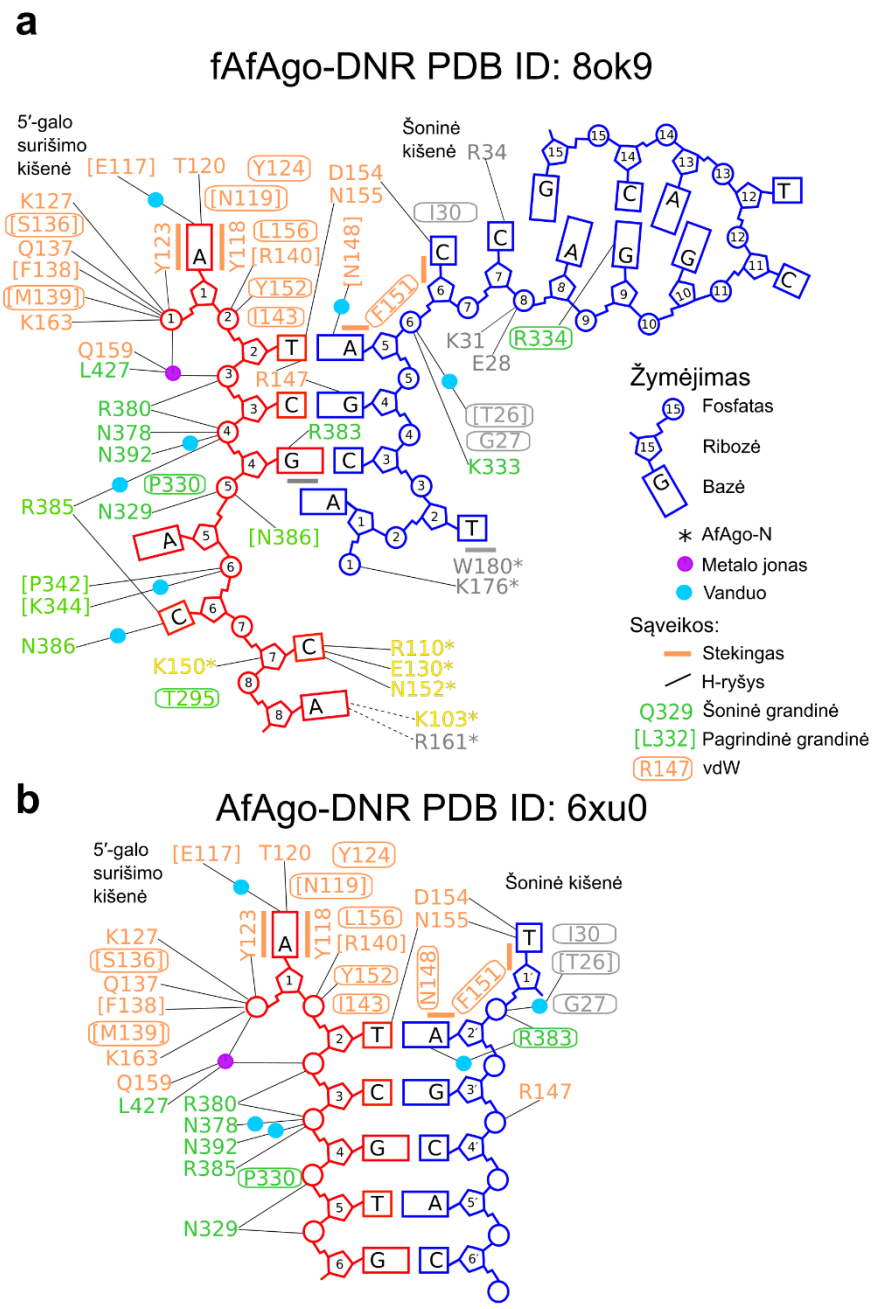
11 pav. Nukleobazėms specifiniai kontaktai AfAgo-nukleorūgščių kompleksuose. (a) Pirmos vedlio bazės kristalinėse AfAgo struktūrose 2W42, 2BGG ir 6T5T palyginimas. (b) gT2 ir t1′ bazių surišimas „šoninėje“ kišenėje 2W42 ir 6T5T struktūrose. Parengta pagal Manakova *et al*., 2023[216].

Specifinis vedlio ir taikinio sekų nukleotidų atpažinimas išskiria AfAgo iš kitų anksčiau aprašytų Argonaute baltymų, kurie apsiriboja tik vedlio (pvz., RsAgo, PDB ID 6D8P[16]) ar tik taikinio [122] grandinės galo nukleotidų atpažinimu. Kita unikali AfAgo savybė yra ta, kad šis baltymas yra

homodimerinis ir geba erdvėje suartinti dvi vedlio-taikinio duplekso kopijas[145]. Šiame darbe pristatytose struktūrose stebimas toks pats dimerizacijos mechanizmas, kas atveria kelią svarstymams apie galimas AfAgo funkcijas *in vivo*.

4.3 AfAgo ir AfAgo-N heterodimerizacija

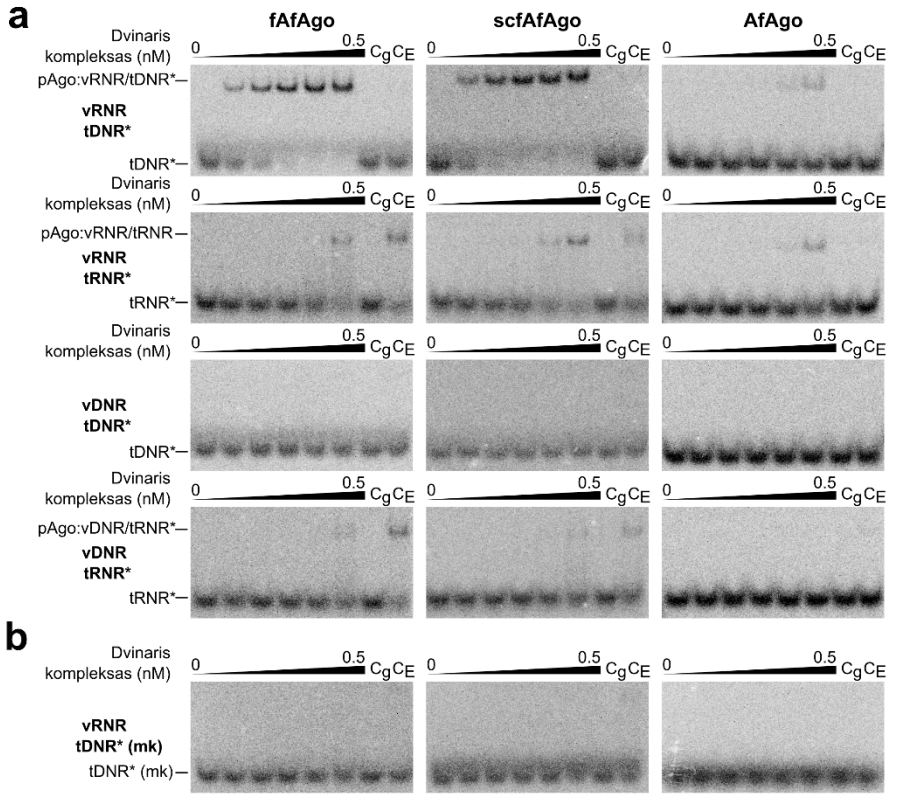
AfAgo iki šiol buvo tyrinėjamas kaip pavienis baltymas, neatižvelgiant į jo genominį kontekstą ir hipotetinius aplinkinius baltymus. Šitaip charakterizuotas AfAgo buvo ilgai naudojamas kaip struktūrinis modelis kitų Ago baltymų ir Ago-NR sąveikų tyrimams[13,23–26,96,153–156]. Vis dėlto, rezultatai rodo, kad AfAgo formuoja heterodimerinį kompleksą su prieš AfAgo geną atkurtame skaitymo rėmelyje koduojamu baltymu, esančiu tame pačiame *A. fulgidus* DSM 4304 kamieno operone. Baltymas – AfAgo-N, koduojamas prieš AfAgo yra struktūriškai ekvivalentiškas ilgųjų pAgo baltymų N-L1-L2 domenams. Taigi, AfAgo-N/AfAgo (fAfAgo) heterodimerinis komplekas struktūriškai panašus į PAZ domeno neturinčius pAgo. fAfAgo panašiausias į ilgąjį-B RsAgo, kurio PAZ domenas yra mažesnis nei kitų ilgųjų pAgo[4]. Kaip ir kiti ilgieji pAgo, fAfAgo heterodimeras formuoja gilų sąveikai su vedlio-taikinio heterodupleksu skirtą griovį, kurio nėra pavieniame AfAgo. Lyginant su AfAgo, fAfAgo sudaro mažiau specifinių kontaktų su surištų nukleorūgščių galų nukleobazėmis (12 pav.). Tai galima paaiškinti tuo, jog pavienis AfAgo yra tik dalis didesniojo funkcinio komplekso, taigi reikalauja visų įmanomų specifinių ir nespecifinių kontaktų, kad stabilizuotų sąveiką su vedlio-taikinio heterodupleksu. Taigi, AfAgo gryninasi iš *E. coli* komplekse su 5′-AUU RNR vedliais, o komplekse su vedlio-taikinio dupleksu atpažįsta tris galines tiek vedlio, tiek taikinio grandžių nukleobazes[216].



12 pav. Konservatyvūs AfAgo kontaktai su vedlio grandies 5′-galu ir taikinio grandies 3′-galu. (a) fAfAgo kontaktų su DNR schema. Vedlio 5′-galo bazė išsukta į surišimo kišenę; trys vedlio-taikinio bazių poros kristale formuoja trumpą dupleksą. Taip pat stebima papildoma Watsono-Cricko pora, susidariusi taikinio grandies poziciją užimančio DNR oligonukleotido 3′-gale esančioje plaukų smiegtuko tipo struktūroje. Vedlio seka pavaizduota raudona, taikinio

– mėlyna spalva. (b) Baltymo-DNR kontaktų schema kristalinėje struktūroje PDB ID 6XU0, baltymo grandinė A. Parengta pagal Manakova *et al*., 2024[217].

Tuo tarpu, AfAgo-N ir AfAgo kartu formuoja griovį nukleorūgščių surišimui, kuris leidžia susidaryti didesniam skaičiui nespecifinių sąveikų su nukleorūgštimis, taip sumažėjant specifinių kontaktų įtakai bendram fAfAgo afiniškumui vedlio ir taikinio grandims. Tą patvirtina ir EMSA rezultatai, kurie rodo, jog fAfAgo-vRNR kompleksas geriau sąveikauja su taikinio DNR nei AfAgo-vRNR (13 pav.).



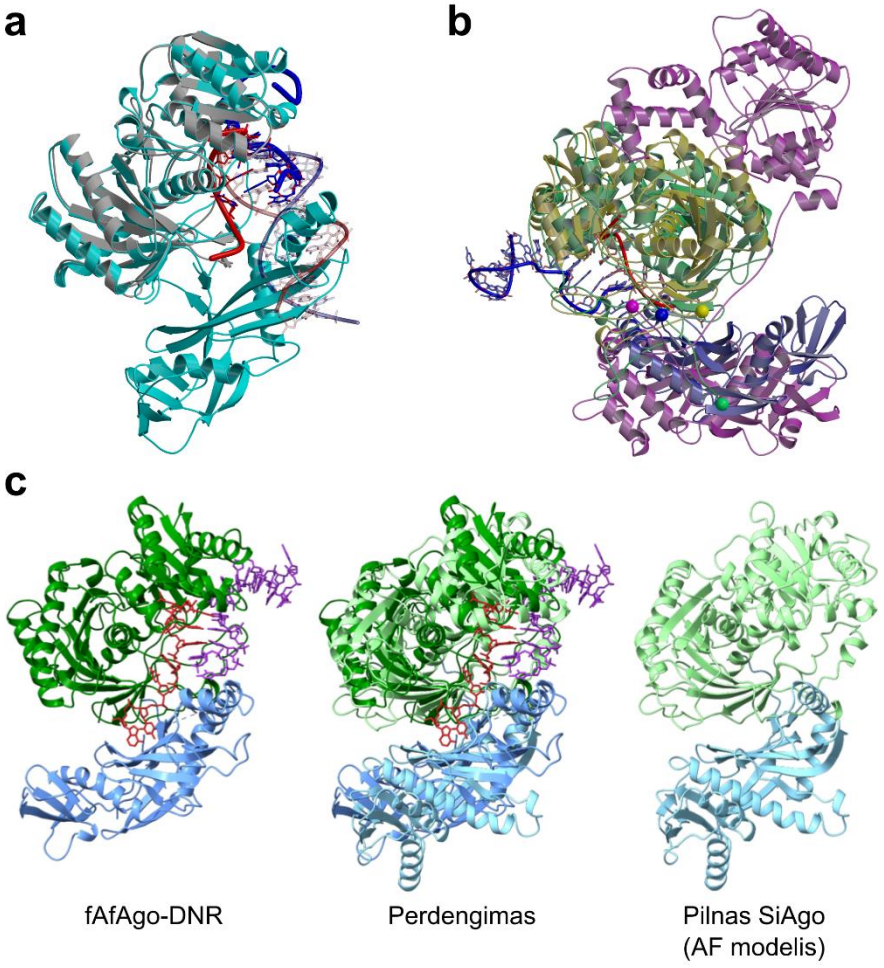
13 pav. RNR vedamas nusitaikymas į nukleorūgštis. (a) 5′-^{32}P žymėtų DNR ir RNR taikinių sąveika su fAfAgo, scfAfAgo ir AfAgo kompleksais su 5′-P RNR bei DNR vedliais (molinis pAgo:vedlio santykis 1:2). Naudoti taikiniai yra komplementarūs vedliui nuo 1-os iki 8-os bazės. (b) Tų pačių pAgo kompleksų su vedlio NR sąveika su mažo komplementarumo (mk) taikiniais. Mažo komplementarumo taikiniai komplementarūs vedliams nuo 4-os iki 7-os bazės. Visi pavaizduoti eksperimentai atlikti tame pačiame dvinario pAgo:vedlio komplekso koncentracijų intervale: 0, 0,005, 0,01, 0,02, 0,1, 0,5 nM. C_g – kontrolė su didžiausia naudota vedlio koncentracija (1 nM) bei žymėtu taikiniu; C_E – kontrolė su didžiausia naudota pAgo koncentracija (0,5 nM) bei žymėtu taikiniu. Parengta pagal Manakova *et al*., 2024[217].

fAfAgo struktūroje AfAgo-N sąveikauja su tuo pačiu AfAgo paviršiumi, per kurį pavienis AfAgo sudaro homodimerą, šitaip eliminuodamas AfAgo heterodimerizaciją. Visgi, lieka neaišku, ar natyviame hipertermofiliniame šeimininke *A. fulgidus* tarp fAfAgo heterodimero ir AfAgo homodimero egzistuoja dinaminė pusiausvyra, priklausanti nuo ląstelės ir aplinkos sąlygų ir ar AfAgo homodimeras bei fAfAgo heterodimeras atlieka unikalias funkcijas *in vivo*.

Struktūriškai fAfAgo heterodimeras panašus į kanoninius trumpuosius pAgo, kurie sudaro heterodimerinius kompleksus su efektoriniais APAZ(N-L1-L2) baltymais, tačiau fAfAgo neturi efektorinio domeno (14 pav.)[9]. Tiek AfAgo, tiek trumpųjų pAgo heterodimeriniuose kompleksuose, N-L1-L2 domenus turinčių baltymų C-galai yra greta Ago baltymų N-galų (14 pav.). be heterodimerinių trumpųjų pAgo egzistuoja ir trumpi pAgo, kurie su prieš jais operone koduojamais efektorinius domenus turinčiais baltymais sudaro vientisą funkcinį polipeptidą. Taigi, tikėtina, jog vieno polipeptido pAgo baltymai evoliucijos eigoje galėjo skilti į du baltymus. Panašu, kad toks skilimas įvyko nepriklausomai ilguosiuose-B pAgo (kaip AfAgo) ir trumpuosiuose pAgo (kaip JomAgo, PgAgo[5]). Siekiant palyginti natyvų perskeltą fAfAgo su jo spėjamu vienos polipeptidinės grandinės pirmtaku, buvo sukonstruotas scfAfAgo baltymas, kuriame AfAgo-N ir AfAgo baltymai sulieti į vieną polipeptidą. Visuose *in vitro* eksperimentuose scfAfAgo elgėsi panašiai į fAfAgo (13 pav.), kas nepadėjo atskleisti galimų natyvaus perskelto baltymo privalumų. Atsižvelgiant į tai, kad egzistuoja tiek vienos polipeptidinės grandinės, tiek perskelti heterodimeriniai trumpieji pAgo, galima manyti, kad nėra ryškaus funkcinio skirtumo tarp aktyvių baltymų, sudarytų iš vienos ar dviejų polipeptidinių grandinių.

Struktūriniai ir bioinformatikiniai atkurto AfAgo komplekso tyrimai rodo, kad ilgieji-B pAgo gali būti iš esmės dviejų rūšių: tipiniai vienos polipeptidinės grandinės pAgo (pvz., RsAgo) bei perskelti pAgo, kaip pilnas fAfAgo. Pastaruoju atveju, N-L1-L2 ir MID-PIWI domenai yra atskiruose baltymuose, sudarančiuose heterodimerinį kompleksą. Šiuo aspektu fAfAgo ir kiti perskelti ilgieji-B pAgo yra panašūs į trumpuosius pAgo – pastarieji taip pat sudaro heterodimerinius funkcinius kompleksus, sudarytus iš APAZ domeną turinčio bei MID-PIWI domenus turinčio baltymo. Nors anksčiau manyta, jog APAZ domenas yra analogiškas PAZ, šiuo metu yra aišku, kad APAZ atitinka N-L1-L2 regioną. Taigi, tiek AfAgo, tiek trumpieji pAgo priklauso perskeltų pAgo sistemoms. Skirumas tarp trumpųjų pAgo ir AfAgo

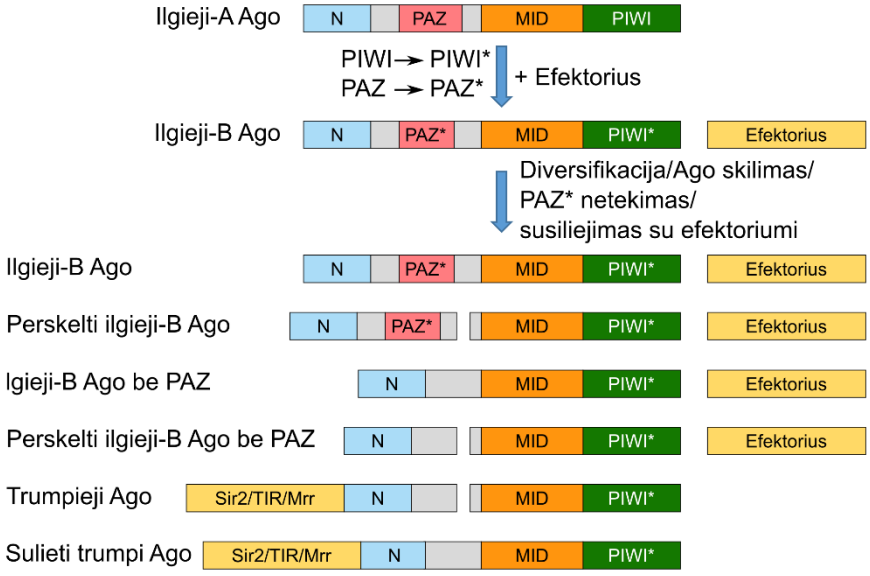
yra tas, kad trumpųjų pAgo ir AfAgo yra tas, kad trumpųjų pAgo N-L1-L2 domenus turintis subvienetas dažnai yra sulietas su efektoriniu domenu į vieną polipeptidą. Visgi, AfAgo ir jo homologų genų aplinkos analizė atskleidė, kad jie taip pat turi spėjamų efektorinių baltymų, koduojamų už Ago. Tai leidžia manyti, kad nors ilgieji-B pAgo nėra sulieti su efektoriniais baltymais, jie, tikėtina, su jais sąveikauja ir reguliuoja jų aktyvumą.



14 pav. Struktūrinis fAfAgo palyginimas su GsSir2/Ago bei pilnu SiAgo. (a) fAfAgo kompleksas formuoja kanalą vedlio-taikinio grandinėms. AfAgo su dgDNR kristalinė struktūra (žydra, PDB ID: 6T5T) palyginta su fAfAgo kompleksu (pilka). Vedlio grandinės pavaizduotos raudonos, taikinio – mėlynos spalvos. DNR struktūrose 2W42 ir 6T5T pavaizduota storomis linijomis. (b) Abiejų fAfAgo komplekso baltymų N- ir C-galai yra greta vienas kitom panašiai, kaip stebėta su GsSir2/Ago[9]. GsSir2/Ago pavaizduotas purpurine (Sir2-APAZ) ir žalia (Ago) spalvomis. GsAgo N-galo aminorūgštis pavaizduota žaliu rutuliu; purpurinis rutulys atitinka GsSir2-APAZ baltymo C-galą. Atstumas tarp šių dviejų aminorūgščių yra 30 Å, tačiau

pažymėtina, kad GsAgo baltymas čia modeliuotas elonguotos konformacijos, tuo tarpu realiame baltyme sankloda gali būti kitokia. fAfAgo komplekse AfAgo baltymas pavaizduotas geltona spalva, o jo N-galo aminorūgštis – geltonu rutuliu. AfAgo-N baltymas pavaizduotas mėlyna spalva, o jo C-galinis atomas – mėlynu rutuliu. Atstumas tarp šių dviejų taškų yra 12 Å. (c) fAfAgo-DNR komplekso palyginimas su pilno SiAgo (SiAgo-Aga1 heterodimero) AlphaFold modeliu. N-galo regionas pavaizduotas šviesiai mėlyna, MID-PIWI – žalia spalva. Pavienių polipeptidinių grandinių palyginimu Dali paremtas struktūrinis fAfAgo ir pilno SiAgo panašumas:
N-skiltis: perdengtos 185 aminorūgščių poros, RMSD=4,6 Å, sekos panašumas 13%
MID-PIWI: perdengta 381 aminorūgščių pora, RMSD=3,1 Å, sekos panašumas 16%. Parengta ir adaptuota pagal Manakova *et al*., 2024[217].

Kiti bruožai, bendri ilgiesiems-B ir trumpiesiems pAgo yra kataliziškai neaktyvus PIWI domenas ir kanoninio PAZ domeno nebuvimas. Iki šiol tirtų trumpųjų pAgo N-L1-L2 subvienetas, kaip ir fAfAgo, visiškai neturi PAZ domeno. Kiti ilgieji-B pAgo PAZ domeno arba neturi, arba turi sumažėjusį PAZ, neturintį vedlio grandies 3′-galo surišimo kišenės. Šie pastebėjimai leidžia siūlyti tikėtiną ilgųjų-B ir trumpųjų pAgo kilmės iš ilgųjų-A pAgo kilmės mechanizmą (15 pav.). Pagal šį mechanizmą, po PIWI domeno inaktyvacijos PAZ domeno vedlio grandies 3′-galo surišimo kišenį tampa nereikalinga ir nunyksta iš dalies arba prarandama visiškai. PIWI domenui tapus neaktyviam, pAgo nebegali savarankiškai atlikti nukleorūgščių fosfodiesterinio ryšio skėlimo reakcijos. Panašu, kad tai atveria kelią atsirasti naujam funkcionalumui, kuomet ilgieji-B pAgo gali asocijuotis su efektoriumi, kuris gali būti tiek atskiras baltymas, koduojamas tame pačiame operone, tiek prilietas pAgo N-gale, susidarant vienam polipeptidui. pAgo skilimas į dvi dalis (N-L1-L2 ir MID-PIWI subvienetus), panašu, įvyko daugelį kartų nepriklausomai, nes egzistuoja tiek perskeltos, tiek vieno polipeptido formos ilgieji-B ir trumpieji pAgo. Taigi, siūlomas mechanizmas paaiškina stebimą trumpųjų ir ilgųjų-B pAgo įvairovę bei šių pAgo naujo funkcionalumo, kaip toksinių efektorių, įjungiamų esant invazinėms nukleorūgštims šeimininko ląstelėje, reguliatorių atsiradimą.



15 pav. Siūlomas įvairių prokariotinių Argonaute baltymų kilmės mechanizmas. Įvykus mutacijoms PIWI domene bei PAZ domeno sutrumpėjimui, ilgieji-A pAgo tampa kataliziškai neaktyvūs, tačiau vis dar geba surišti nukleorūgštis (ilgieji-B pAgo). Ilgieji-B pAgo tuomet diversifikavosi keliais būdais: (1) skylant pAgo į du baltymus – N-Ago ir MID-PIWI-Ago; (2) sumažėjant arba visiškai nunykstant PAZ* domenui; (3) asocijuojantis su funkciniu efektoriumi, kuris yra kaip atskiras baltymas arba yra prilietas pAgo baltymo N-gale. Parengta pagal Manakova *et al.*, 2024[217].

Apibendrinus, šiame darbe pateikti rezultatai rodo, kad atkurtas AfAgo operonas iš *A. fulgidus* DSM 4304 kamieno yra analogiškas priešvirusinei sistemai SiAgo iš *S. islandicus*, išskyrus spėjamą reguliatorių. AfAgo-N baltymas panašus į SiAga1, formuojantį su SiAgo heterodimerinį kompleksą. Taigi, fAfAgo heterodimeras gali būti laikomas struktūriniu SiAgo/SiAga1 komplekso ekvivalentu (14 pav. c). Tikimasi, kad tolimesni struktūriniai ir funkciniai AfAgo operono baltymų, ypač fAfAgo heterodimero sąveikos su operone po jo koduojamu baltymu, tyrimai atskleis galimą AfAgo sistemos funkciją ir veikimo mechanizmą.

4.4 Apibendrinimas

Prokariotiniai Argonaute baltymų tyrimai yra sparčiai besivystanti sritis. pAgo baltymai pasižymi didele funkcine bei konkrečių mechanizmų įvairove bei plačiu operoninės organizacijos bei baltymų-kompanjonų variabilumu[4,5]. Ilgųjų-A pAgo šaka yra ištyrinėta daugiausiai, daug

dėmesio pastaruoju metu susilaukia ir trumpieji pAgo[5]. Ilgųjų-B pAgo šaka tyrinėta mažai, greičiausiai dėl to, kad jos nariai yra katalitiškai neaktyvūs. Šiame darbe nagrinėtas ilgųjų-B šakai priskiriamas AfAgo ir jo operoninis kaimynas AfAgo-N. Atskleista, kad AfAgo, priešingai nei aprašyta anksčiau, ne tik yra selektyvus 5′-AUU galą turinčiai vgRNR, tačiau sudaro specifinius kontaktus ir su atitinkamais vgDNR taikinio nukleotidais. Taip pat, pademonstruotas AfAgo gebėjimas sudaryti homodimerus bei parodytas RNR vedamas nusitaikymas į DNR taikinį.

Tyrimų metu atskleistas operone greta esantis AfAgo-N sąveikauja su AfAgo ir sudaro funkcinį heterodimerą, struktūriškai panašų į PAZ domeno neturinčius pAgo baltymus. Šis heterodimeras taip pat pasižymi RNR vedamu nusitaikymu į vgDNR taikinį bei selektyvumu 5′-AUU galą turintiems vgRNR vedliams, tačiau ne tokį ryškų, kaip pavienis AfAgo.

Bendrai, šiame darbe pristatyti tyrimai praplečia žinias apie galimus pAgo veikimo mechanizmus, operonines organizacijas bei atveria kelią tolimesniems tyrimams. Šiuo metu nėra žinoma, ar AfAgo homodimerizacija turi aiškią funkcinę prasmę, pvz., reguliacinę, tačiau tą sudėtinga patikrinti, nes nėra pademonstruotas AfAgo apsauginis ar kitas funkcinis mechanizmas gyvose modelinėse sistemose kaip *E. coli*. Taip pat, atsiveria kelias tolimesniems viso operono baltymų tyrimams, įskaitant trumpai paminėtą Af1317, turintį spėjamus transmembraninius domenus ir, galbūt, esantį sistemos efektoriumi, kaip tai yra pademonstruota analogiškoje SiAgo sistemoje[139]. Taigi, tolimesnis žingsnis šio objekto tyrimuose ir būtų Af1317 bei fAfAgo sąveikos ir funkcijos tyrimai, kurie suteiktų daugiau žinių apie pAgo funkcinių mechanizmų įvairovę.

5. IŠVADOS

1. **Homodimeriniai kompleksai ir kilpinės dgDNR struktūros**: tiriant *in vitro,* AfAgo sudaro homodimerinius kompleksus ir kilpines dvigrandinės DNR struktūras.
2. **Specifiškumas 5′-galo AUU sekai vedlio grandyje**: struktūriniai duomenys bei tyrimai *in vivo* rodo ryškų specifiškumą vedlio RNR 5′-galo AUU sekai – tai yra naujas atradimas apie tikslius vedlio bei taikinio grandžių atpažinimo mechanizmus.
3. **RNR vedamas DNR taikinio atpažinimas**: nors AfAgo nėra pademonstruotas katalitinis aktyvumas, šis baltymas pasižymi RNR vedamu viengrandinės DNR atpažinimu. Šis mechanizmas iki šiol nebuvo pademonstruotas AfAgo.
4. **Heterodimerinių kompleksų sudarymas**: AfAgo sudaro heterodimerinį kompleksą su AfAgo-N, pavadintą fAfAgo. Šis kompleksas yra panašus į ilguosius PAZ domeno neturinčius pAgo baltymus. Tai papildo (f)AfAgo funkcinių mechanizmų kompleksiškumą.
5. **RNR vedamas (sc)fAfAgo nusitaikymas į DNR**: tiek fAfago, tiek sulietas jo variantas scfAfAgo pasižymi RNR vedamu DNR atpažinimu, su didesniu afiniškumu nei pavienis AfAgo.

NOTES

NOTES

NOTES

NOTES

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 15 egz.