Prokaryotic Argonaute from *Archaeoglobus fulgidus* interacts with DNA as a homodimer

Edvardas Golovinas[1], Danielis Rutkauskas[1,2], Elena Manakova[1], Marija Jankunec[1,3], Arunas Silanskas[1], Giedrius Sasnauskas[1], Mindaugas Zaremba[1,*]

[1] Institute of Biotechnology, Life Sciences Center, Vilnius University, Sauletekio av. 7, LT-10257, Vilnius, Lithuania.
[2] Institute of Physics, Center for Physical Sciences and Technology, Savanoriu 231, LT-02300, Vilnius, Lithuania
[3] Institute of Biochemistry, Life Sciences Center, Vilnius University, Sauletekio av. 7, LT-10257, Vilnius, Lithuania.

* To whom correspondence should be addressed. Tel: +370-5-2234357; Fax: +370-5-2234367; Email: zare@ibt.lt.

ABSTRACT

Background: Argonaute (Ago) proteins are found in all three domains of life. The best characterized group is eukaryotic Argonautes (eAgos), which are the core of RNA interference. The best understood prokaryotic Ago (pAgo) proteins are full-length pAgos. They are monomeric proteins, all composed of four major structural/functional domains (N, PAZ, MID and PIWI) and thereby closely resemble eAgos. It is believed that full-length pAgos function as prokaryotic antiviral systems, with the PIWI domain performing cleavage of invading nucleic acids. However, the majority of identified pAgos are shorter and catalytically inactive (encode just MID and inactive PIWI domains), thus their action mechanism and function remain unknown.

Results: In this work we focus on AfAgo, a short pAgo protein encoded by an archaeon *Archaeoglobus fulgidus*. We find that in all previously solved AfAgo structures, its two monomers form substantial dimerization interfaces involving the C-terminal β-sheets. Led by this finding, we have employed various biochemical and biophysical assays, including single-molecule FRET, SAXS and AFM, to test the possible dimerization of AfAgo. SAXS results confirm that WT AfAgo, but not the dimerization surface mutant AfAgoΔ, forms a homodimer both in the apo-form and when bound to a nucleic acid. Single molecule FRET and AFM studies demonstrate that the dimeric WT AfAgo binds two ends of a linear DNA fragment, forming a relatively stable DNA loop.

Conclusion: Our results show that contrary to other characterized Ago proteins, AfAgo is a stable homodimer in solution, which is capable of simultaneous interaction with two DNA molecules. This finding broadens the range of currently known Argonaute-nucleic acid interaction mechanisms.

KEYWORDS

Protein-DNA interactions, Argonaute, pAgo, AFM, smFRET, SAXS

INTRODUCTION

Argonaute (Ago) proteins are found in all three domains of life (bacteria, archaea, and eukaryotes). The best characterized group is eukaryotic Ago (eAgo) proteins. Being the functional core of RNA interference machinery, eAgos are involved in regulation of gene expression, silencing of mobile genome elements, and defense against viruses. From the structural and mechanistic point of view, all eAgos are very similar, as they all use small RNA molecules as guides for sequence-specific recognition of RNA targets, and are monomeric proteins sharing four conserved functional domains, which are organized in a bilobed structure [1]. The N-terminal lobe consists of the N-domain that separates guide and target strands [2], and the PAZ domain responsible for binding the 3′-terminus of the guide RNA; the C-terminal lobe consists of the MID domain, which binds the 5′-terminus of the guide RNA, and the PIWI domain, an RNase. Upon recognition of the RNA target, eAgos may either cleave it employing the catalytic activity of the PIWI domain, or, especially eAgo proteins that encode catalytically inactive PIWI domains, recruit partner proteins leading to degradation of the target RNA or repression of its translation [3].

Ago proteins are also identified in 9% of sequenced bacterial and 32% archaeal genomes [4, 5]. Unlike eAgos, which exclusively use RNA guides for recognition of RNA targets, different pAgos may use either RNA or DNA guides and/or targets [6], and may also differ in their structural organization. The best understood prokaryotic Ago (pAgo) proteins are the so called full-length pAgos, which are composed of N, PAZ, MID and PIWI domains, and thus closely resemble eAgo proteins. There is mounting evidence that full-length pAgos function as prokaryotic antiviral systems, with the PIWI domain performing cleavage of invading nucleic acids. However, the majority (~60 %) of identified pAgos are shorter (encode just MID and PIWI domains) and are catalytically inactive due to mutations in the PIWI domain. Though similar artificial truncations of eukaryotic Agos preserve most of functionality characteristic to full-length proteins [7–10], the function and mechanism of the naturally-occurring short catalytically inactive pAgos remains unknown [4, 5].

In this work we focus on the short prokaryotic Argonaute AfAgo encoded by a hyperthermophilic archaeon *Archaeoglobus fulgidus* [4, 5]. Like other short pAgos, AfAgo contains a MID and a

63 catalytically inactive PIWI domains (albeit sequence analysis suggests that AfAgo MID and PIWI
64 domains are closer to those found in full-length, rather than most short pAgos [4, 5]). For over a
65 decade it served as a model system for structural and mechanistic studies of Argonaute-nucleic acids
66 interactions [10–12]. It is also one of the first and the best structurally characterized prokaryotic
67 Argonautes, with an apo- and 3 dsDNA/RNA-bound structures currently available [13–16]. However,
68 its biological role, in part due to lack of catalytic activity, remains elusive. Unexpectedly, inspection of
69 these structures revealed that regardless of the crystal form and symmetry, two AfAgo subunits
70 always form a substantial dimerization interface involving N-terminal residues and/or β-strands
71 located close to the C-termini. Using various biochemical and biophysical assays, including single-
72 molecule FRET, small angle X-ray scattering (SAXS), and atomic force microscopy (AFM), we show
73 that AfAgo is indeed a stable homodimer in solution, and is capable of simultaneous interaction with
74 two DNA molecules. This broadens the range of currently known interaction mechanisms involving
75 nucleic acids and Argonaute proteins.

76 **RESULTS**

77 **AfAgo is a homodimer in the available X-ray structures**

78 AfAgo is a 427 amino acid (aa) 49.2 kDa prokaryotic Argonaute protein found in the hyperthermophilic
79 archaeon *Archaeoglobus fulgidus*. To date, four AfAgo structures, both of the apo-form and bound to
80 DNA and RNA duplexes were solved [13–16]. AfAgo monomer is composed of two major domains,
81 the N-terminal MID (residues 38-168), and the C-terminal PIWI (residues 168-427) [13]. The MID
82 domain specifically binds the 5′-phosphorylated end of the presumed guide DNA/RNA strand, and
83 also makes contacts to the complementary target DNA/RNA strand [14–16]. The PIWI domain makes
84 contacts to both guide and target DNA/RNA strands, but is catalytically inactive due to mutations in
85 the RNase H-like catalytic center. Unexpectedly, inspection of available AfAgo structures
86 (Supplementary table S4) revealed that in all structures known so far, AfAgo subunits form
87 homodimers. The dimerization interface in the AfAgo-dsRNA structure (PDB ID 1ytu) is asymmetric,
88 and primarily involves the C-terminal β-strands (residues 296-303) from both subunits present in the
89 asymmetric unit that together form a parallel β-sheet, and the N-terminal residues from one of the
90 subunits (Figure 1A). The dimer formed in this case is compact (henceforth, a ‘closed’ dimer). In
91 contrast, dimerization interfaces in three other cases (PDB IDs 1w9h, 2bgg and 2w42) are nearly
92 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 1B). The resultant dimers are less compact (henceforth, 'open' dimers).

The solvent accessible surface areas buried at the dimerization interfaces, as calculated by the PISA server (https://www.ebi.ac.uk/pdbe/pisa/pistart.html, [17]) and the number of inter-subunit H-bonds (Supplementary table S4) in both 'open' and 'closed' dimers are typical for stable protein dimerization interfaces [18, 19]. This observation prompted us to test the oligomeric state, the possible dimerization mode, and mechanism of nucleic acid binding of AfAgo in solution using various biochemical and biophysical techniques. For that purpose, we used two variants of AfAgo: the full-length wild-type protein (henceforth, WT AfAgo), and a dimerization mutant AfAgo lacking the 296-303 amino acid residues involved in dimerization (henceforth, AfAgoΔ). Both proteins were successfully purified as described in Materials and Methods, albeit the AfAgoΔ variant was prone to aggregation. Stability of both proteins was considerably improved upon addition of a phosphorylated blunt-end DNA oligoduplex MZ-1289 (Supplementary table S1).

SAXS measurements

To determine the solution conformation and oligomeric state of AfAgo, we have performed small angle X-ray scattering (SAXS) measurements using DNA-bound and DNA-free full-length WT AfAgo protein and the DNA-bound dimerization mutant AfAgoΔ. Two types of data analysis were performed: (i) the *ab initio* shapes of the proteins in solution were calculated and superimposed with the X-ray AfAgo structures, and (ii) the theoretical scattering data was calculated for the crystallized AfAgo monomer, 'open' (PDB ID: 2w42 and 1w9h) and 'closed' (PDB ID: 1ytu (with and without bound dsRNA)) dimers, and compared to experimental SAXS scattering data of AfAgo and AfAgoΔ (Figure 1**Error! Reference source not found.**). The 'closed' AfAgo dimer fits the DNA-bound and DNA-free WT AfAgo SAXS data better than the 'open' dimer, as judged from the real space fit and the χ^2 (Figure 1C) parameters that reflect the agreement between scattering functions of corresponding crystal structures and SAXS experiments (Figure 1C). As expected, AfAgo monomer gave the best fit to the AfAgoΔ SAXS data (Figure 1C, right column). The SAXS molecular weights calculated for AfAgo (between 94.2 and 106.9 kDa (108-133 kDa for apo AfAgo, that was prone to aggregation), Supplementary table S3) agreed with the expected mass of the dimer complexed with dsDNA (110.5 kDa). The SAXS MW for

the AfAgoΔ (between 55.4 and 67.9 kDa) confirmed the monomeric state of the dimerization mutant-dsDNA complex.

Direct visualization of AfAgo-induced DNA loops by AFM

AfAgo and DNA were deposited on APS-mica and imaged using tapping AFM. A typical AFM image of AfAgo-DNA complexes is shown in Figure 2. Several types of protein-DNA complexes, shown as enlarged insets in Figure 2, 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. We find that the relative distribution of different complexes varied dramatically for WT AfAgo and the dimerization mutant AfAgoΔ (Table 1). The ring-shaped DNA-protein complexes are the dominant species observed with WT AfAgo (55% or 114 out of 208 complexes). The minor fraction of DNA molecules had either protein bound to one end (29%, 61 out of 208 complexes) or to both ends (16%, 33 out of 208). In the case of AfAgoΔ, the majority of complexes had protein bound to both DNA ends (59%, or 187 out of 319, Table 1), and a much lesser fraction (12%, or 38 out of 319) were ring-shaped structures. Since AfAgoΔ lacks the dimerization interface observed in the structures, the observed small fraction of ring-shaped DNA complexes could be due to sample treatment with glutaraldehyde (see Materials and Methods for details), which could occasionally lead to inadvertent cross-linking of two AfAgoΔ monomers bound to termini of the same DNA molecule, thereby generating looped complexes.

WT AfAgo induces DNA loops in solution

Dimerization of WT AfAgo observed in X-ray structures and SAXS measurements is unique among Argonaute proteins. It raises a question if the AfAgo/DNA stoichiometry observed in the structures, an AfAgo dimer bound to two nucleic acid molecules, is also relevant in solution. To address this question we have examined the interaction of AfAgo with a DNA substrate bearing two binding sites exploiting the method of single-molecule Förster resonance energy transfer (smFRET). If AfAgo binds DNA as a dimer, as observed in the X-ray structures, it should be capable of interacting with both DNA ends simultaneously, inducing a DNA loop, which, in addition to AFM, could be monitored as a change of FRET efficiency between dyes tethered close to DNA ends (Figure 3A). Utilization of a single dual-labeled two-target DNA substrate (rather than two short DNA duplexes carrying different fluorescent

152 labels) increases the probability of AfAgo interaction with both DNA targets at low reactant
153 concentrations required for the single-molecule setup.

154 We have designed a 569 bp DNA construct, which was labelled with a pair of FRET fluorophores,
155 Cy3B and Atto647N, each attached to thymine bases 3 nt away from the respective DNA termini via a
156 C6 linker (Supplementary figure S2). The positions of FRET labels were selected such that upon
157 binding of both DNA ends by an AfAgo dimer, the distance between the label attachment sites
158 (irrespective of the AfAgo dimerization mode), is favorable for FRET (Supplementary figure S4). To
159 promote tighter binding, both DNA ends were phosphorylated, and the 5′-terminal nucleotides were
160 adenines, since AfAgo, like Argonaute CbAgo from *Clostridium butyricum* [20], has a preference for a
161 5′-terminal A (publication in preparation).

162 AfAgo interaction with the DNA fragment was monitored by analyzing the fluorescence bursts of
163 single diffusing DNA fragments (Figure 3B). For each DNA molecule we have calculated two
164 parameters. The first parameter S represents the stoichiometry of different fluorophores present on
165 the DNA, and is equal to the ratio $I_d/(I_d + I_a^a)$, where I_d is the total donor and acceptor intensity upon
166 donor excitation, and I_a^a is acceptor intensity upon acceptor excitation. The relative excitation
167 intensities of the donor and acceptor fluorophores were adjusted such that the stoichiometry
168 parameter S was about 0.5 for DNA molecules labeled with both fluorophores, approx. 0 for the
169 acceptor-only DNA, and close to 1.0 for the donor-only DNA. The second parameter, the proximity
170 ratio E, is equal to $I_d^a/(I_d^a + I_d^d)$, where I_d^a and I_d^d are acceptor and donor intensities upon donor
171 excitation, respectively. It is expected to be higher for looped DNA molecules with the ends brought
172 into close proximity than for unlooped DNA molecules.

173 The E-S histogram of DNA alone (Figure 3C, left) exhibits a prominent population with low E and
174 intermediate S values, which corresponds to dual-labeled unlooped (zero-FRET) DNA molecules. The
175 two minor populations observed in the histogram correspond to donor-only (low E/high S) and
176 acceptor-only (high E/low S) DNA fragments. The E-S histogram of DNA in the presence of WT AfAgo
177 exhibits an additional population (intermediate S and intermediate E, Figure 3C, right), which
178 presumably represents DNA molecules looped by WT AfAgo. For further analysis we chose the dual-
179 labeled DNA molecules (S between 0.2 and 0.9) and built their E histogram. The histogram exhibits
180 peaks of near-zero and high E, corresponding to unlooped and looped DNA molecules, respectively
181 (Figure 3C, right).The fit of this histogram with a sum of two Gaussian functions allowed us to

calculate the ratio K defined as the ratio of the fraction of looped DNA molecules (the area under the Gaussian with the high E center) and unlooped DNA molecules (the area under the Gaussian with a near-zero E center) over the population of the dual-labeled DNA fragments.

We have measured the ratio K at different WT AfAgo concentrations (Figure 3D). 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 1 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 3D, the ratio K at all AfAgoΔ concentrations tested was close to zero, indicating that AfAgoΔ was unable to induce DNA loops. Lack of DNA looping was not due to lack of DNA binding, as shown by electrophoretic mobility shift assay (Supplementary figure S5). 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 3D). Taken together, our results indicate that WT AfAgo dimer is capable of simultaneous interaction with two DNA ends in solution.

Dynamics of WT AfAgo-induced DNA loops

To explore the dynamics of the WT AfAgo-induced DNA looping events, we have used total internal reflection fluorescence (TIRF) microscopy to perform single molecule FRET experiments on surface-immobilized DNA (Figure 4). For that purpose we used a DNA fragment 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 (Supplementary figure S2). After verifying that WT AfAgo induces loops on this substrate in solution (Supplementary figure S6B), we immobilized the biotinylated DNA on a surface and then recorded fluorescence movies in the absence or in the presence of WT AfAgo (Supplementary figure S6A). From each frame of the movie we then extracted donor and acceptor intensities for individual DNA fragments, selected trajectories with anti-correlated changes of the donor and acceptor intensities (indicating the occurrence of FRET) and calculated the time courses of the proximity ratio, E. An example of such a trajectory is presented in Figure 4C. In a control with no AfAgo we could find no DNA fragments exhibiting FRET (Figure 4B).

To assess the average values of different E levels apparent in individual trajectories we pooled the selected E trajectories and averaged the resulting image within the first 10 s (Figure 4D, left). The single-molecule population and time-averaged E exhibits two peaks with maxima at 0.09 and 0.36, corresponding to the unlooped and looped DNA molecules, respectively (Figure 4D, right). These E values are also in good agreement with the E values obtained from the measurement in solution (Figure 3C).

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 (Figure 4C, Supplementary figure S6C, D), exhibiting a number of transitions between the looped and unlooped states. The looped E state also exhibits more subtle dynamics (Supplementary figure S6D) which could probably be attributed to the conformational flexibility of AfAgo at the dimerization interface.

To quantify the looped state duration we first idealized the E trajectories using HMM in QuB software (Figure 4E). Then, from the idealized trajectories we built the cumulative histogram of the looped state durations (Figure 4F). 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 our 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.

DISCUSSION

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, we reveal here that 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.

First, we show that AfAgo is a homodimer in all previously solved X-ray structures, including apo-protein, and complexes with RNA and DNA (Supplementary table S4). 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 (Supplementary table S4), but result in distinct arrangement of AfAgo subunits relative to one another, which we term 'closed' and 'open' dimers (Figure 1 A and B, respectively). The 'closed' type of AfAgo homodimer, formed when the interface involves both the N-terminal residues and the C-terminal β-strands (Figure 1A), provides a better fit to our SAXS data, suggesting that it is the major type of DNA-bound WT AfAgo dimer present in solution (Figure 1C). It remains to be determined if formation of the alternative 'open' dimer observed in several structures (Figure 1B and Supplementary table S4) was influenced 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, removal of the β-strands located at the intersubunit interface (variant AfAgoΔ) renders AfAgo incapable of dimerization (Figure 1).

Simultaneous binding of WT AfAgo homodimer to both ends of a linear DNA fragment would result in a DNA loop. Formation of such looped DNA molecules upon incubation with WT AfAgo was directly visualized using AFM (Figure 2); as shown in Table 1, they constitute the majority of all protein-DNA complexes detected. As expected, very few looped DNA molecules were observed in similar experiments performed with the dimerization mutant AfAgoΔ (Table 1). Thus, AFM provides further proof that homodimeric WT AfAgo is capable of simultaneous interaction with two DNA ends, while the dimerization mutant AfAgoΔ, being a monomer, binds just one DNA end.

To further characterize the interaction of WT AfAgo with DNA in solution, we have performed single-molecule FRET measurements (Figure 3) using a 569 bp DNA fragment labelled with different fluorescent labels (a FRET pair) close to DNA ends. 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-incapable AfAgoΔ confirmed that WT AfAgo forms DNA loops, and that loop formation requires AfAgo dimerization.

At least two mechanisms can be proposed for the formation of the WT AfAgo dimer / looped DNA complex. The first mechanism is valid if apo-WT AfAgo forms a stable dimer. The reaction (Figure 5) proceeds 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 in an intramolecular reaction, resulting in a looped complex (species '2'); (iii) alternatively, association of the second WT AfAgo dimer with the unoccupied target (the second DNA end) of species '1' leads to

species ‘3’, which is no longer capable of loop formation. Such mechanism was demonstrated for many proteins capable of DNA looping, including restriction endonucleases [21–23] and transposases [24–27]. The second mechanism is valid if apo-WT AfAgo is a monomer in solution, and dimerizes only upon binding to DNA. In such a case the following steps lead to a looped complex: (i) a single AfAgo monomer binds the first DNA end (species ‘4’, Figure 5); (ii) the second AfAgo monomer binds the second end (species ‘5’); (iii) two DNA-bound monomers associate forming the looped complex ‘2’ (Figure 5). Such mechanism is less common, albeit it was also demonstrated for some nucleic acid enzymes [28]. Single molecule FRET experiments in solution allowed us to distinguish the two above mechanisms. We have found that the amount of the looped (high-FRET) complex is dependent on AfAgo concentration. It increases until an optimal protein concentration is reached (approx. 1 nM dimer in our experimental setup, Figure 3D), and steeply declines upon further increase in WT AfAgo concentration. Such behavior is expected only for the mechanism involving homodimeric WT AfAgo (species ‘0’, ‘1’, ‘2’, and ‘3’, Figure 5), as excessive protein concentrations favor binding of two WT AfAgo dimers to both DNA ends (species ‘3’), and thereby hinder loop formation. No such loop disruption at increased protein concentrations is expected for the second mechanism involving monomeric apo-WT AfAgo (species ‘0’, ‘4’, ‘5’, and ‘2’), as the species favored at the highest protein concentrations (‘5’) is still capable of loop formation. Taken together, our single-molecule studies in solution favor the DNA looping mechanism involving a stable apo-WT AfAgo dimer (species ‘0’, ‘1’, ‘2’, ‘3’, Figure 5), a finding that is also consistent with SAXS measurements of apo-WT AfAgo (Table 1, Figure 1C). However, we cannot completely rule out mixed mechanisms involving an equilibrium of monomeric and dimeric apo-WT AfAgo forms and DNA simultaneously bound to both monomeric and dimeric AfAgo proteins (species ‘6’, Figure 5).

Single-molecule measurements on immobilized DNA allowed us to assess the dynamic properties of WT AfAgo-induced DNA loops. We find that (i) the DNA loops induced by WT AfAgo are relatively stable, with the lower limit estimate for the loop duration exceeding 30 s (Supplementary figure S6C); (ii) the proximity ratio E of the looped complexes changes over time, suggesting intrinsic dynamics of the AfAgo dimer, presumably attributable to the flexible dimerization interface (Supplementary figure S6D).

CONCLUSIONS

The ability of WT AfAgo to form homodimers and bring two nucleic acid fragments into close proximity, to the best of our knowledge, is unique among Argonaute proteins, and raises additional questions regarding the currently unknown AfAgo function. Simultaneous interaction with two target sites in the case of restriction endonucleases is believed to increase specificity by preventing inadvertent cleavage of lone unmodified target sites [23, 29]. However, since AfAgo has no intrinsic nuclease activity, it cannot be directly involved in host defense against invading nucleic acids, as recently proposed for the catalytically active full-length pAgos [30, 31]. Instead, the ability of WT AfAgo to form stable synaptic complexes with two DNA ends is reminiscent of transposases [24–27], Cas1-Cas2 integrases [32, 33], and (retro)viral integrases [34, 35], which often bring the reactive 3′-OH groups of two DNA ends into proximity of the integration site. AfAgo thus could serve as the recognition module for the integrated DNA fragment; target DNA recognition, binding and catalysis of the integration reactions would require involvement of additional, currently unknown, partner proteins. In order to test this hypothesis, we currently perform structural and functional studies of other AfAgo-like short prokaryotic Argonautes, and also try to identify and characterize their putative partners.

METHODS

Protein 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 followed by 1-427 aa of WT AfAgo sequence). The deletion in the dimerization mutant AfAgoΔ was constructed via overlap extension PCR by using two primer pairs, MZ-385/MZ-875 and MZ-383/MZ-876 (Supplementary table S1) 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. Both proteins were expressed in *E. coli* strain BL21(DE3). Cells were grown in LB broth in the presence of ampicillin at 37 °C. When A600 of the cell culture reached 0.5, the incubation temperature was lowered to 16 °C, 0.1 mM IPTG were added, cells incubated for approx. 16 hours at 16 °C and harvested by centrifugation.

Protein purification

Harvested cells expressing $(His)_6$-tagged WT AfAgo or the dimerization mutant AfAgoΔ were disrupted by sonication in buffer A (20 mM Tris–HCl (pH 8.0 at 25 °C), 500 mM mM NaCl, 5 mM mercaptoethanol) with 2 mM PMSF (phenylmethylsulfonyl fluoride), incubated for 20 min at 50 °C and cell debris was removed by centrifugation at 48,400 x 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 buffer A. 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% glycerol, 5 mM 2-mercaptoethanol and incubated for 1 h at 37 °C with 1 mM EDTA (ethylenediaminetetraacetic acid) and RNase A/T1 (ThermoFisher Scientific) (1:100). Next, the protein solution was centrifuged at 48,400 x 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 buffer A and dialyzed against 20 mM Tris–HCl (pH 8.0 at 25 °C), 500 mM NaCl, 50% glycerol.

Small angle X-ray scattering

Small angle scattering data of WT AfAgo and monomeric mutant were collected at the P12 EMBL beamline on the PETRA III ring of the DESY synchrotron in Hamburg, Germany [36]. Details of data collection and principal structural parameters are presented in Supplementary table S2 and Supplementary figure S1. Protein 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 1,4-dithiothreitol) using Illustra NAP columns (GE Healthcare). Apo AfAgo sample was measured in the same buffer containing 500 mM NaCl.

Dimeric AfAgo complex with MZ-1289 was analyzed by SEC-SAXS with FPLC (Agilent) using Wyatt-MALLS-DLS detection system. The AfAgo+MZ-1289 was concentrated to 175 µM and loaded on the column Superdex 200 Increase 10/300 (GE Healthcare) equilibrated with the sample buffer. Frames collected during the complete SEC run (flow rate 0.5 ml/min, 3000 frames) were analyzed with CHROMIXS [37], frames corresponding to the peak were averaged and processed. *Ab initio* shape determination was carried out by generating 20 independent DAMMIF [38] models using parameterized scattering curves created by GNOM [39] under P2 symmetry restraints. Models were clustered by DAMCLUST [40] and models forming a cluster were averaged by DAMAVER [41] and used as a starting model for an additional run of DAMMIN [42].

SAXS measurements performed with a range of AfAgoΔ concentrations (1-10 mg/ml, both the apo-form and a complex with MZ-1289 DNA), showed significant protein aggregation, particularly pronounced with the apo-AfAgoΔ. The pseudo-chain dummy residues models of the complex generated by GASBOR [43] were superimposed with crystallographic dimers of AfAgo as well as with the monomeric AfAgo-DNA complex using SUPCOMB [44] applying step-wise shift (5 Å) along the principal axis of the model as described in [45].

Comparison of the SAXS data with crystal structures was carried out by CRYSOL [46] (Figure 1). Particle volume and MW estimations were performed using several methods (Supplementary table S3 and references therein).

DNA fragments

DNA fragments were assembled and prepared as depicted in Supplementary figure S2. 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% ethanol and resuspended in water.

AFM sample preparation and imaging

DNA-protein complexes were formed by incubating the DNA fragment (5 nM) with WT AfAgo or AfAgoΔ (concentration in terms of monomer 50 nM) for 5 min at room temperature in the Binding Buffer HEPES (33 mM HEPES (pH 7.8 at 25 °C), 66 mM CH_3COOK, 5 mM $(CH_3COO)_2Mg$,) in a total volume of 50 µl. Next, the protein-DNA complexes were cross-linked with 2.5% 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 CH_3COOK, 5 mM $(CH_3COO)_2Mg$,). The resultant reaction solution after 10-fold dilution with Tris buffer was deposited onto modified mica at room temperature as described below.

Freshly cleaved muscovite mica (grade IV, SPI supplies Inc., USA) was incubated in a mixture of 1-(3-aminopropyl)-silatrane (APS) solution for 30 min to prepare functionalized APS-mica, as described previously for the preparation of protein-DNA complexes [47]. 50 µl of DNA-protein complex solution was deposited on APS-mica for 5 min. After incubation the mica surface was immersed into 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 'DimensionIcon' (Bruker, Santa Barbara, CA) microscope system in tapping

385 mode. Probes with nominal spring constants of ~ 5 or 40 N/m were used. Typically, the images were
386 collected at a speed of 0.6 Hz and a resolution of 1024 × 1024 pixels, scan size 2 µm x 2 µm.

387 **Single-molecule fluorescence microscopy**

388 The overall idea of fluorescence burst data acquisition of single diffusing molecules in alternating laser
389 excitation (ALEX) mode was based on [48]. The principal opto-mechanical layout of the experiment is
390 shown in Supplementary figure S3 and described in Supplementary Methods.

391 The measurement of single surface-immobilized molecules with the excitation in the total internal
392 reflection mode (TIR) was performed on the same setup exploiting its alternative functionality as
393 described previously [21]. Briefly, the objective was 100× 1.4 Oil Plan Apo VC (Nikon), the
394 fluorescence signal was split by T640lpxr-UF2 dichroic mirror (Chroma) and the different spectral
395 channels were projected on the same EMCCD (DU-897ECS0-UVB, Andor).

396 **Sample cell preparation for single molecule measurements**

397 FRET bursts measurements were performed in a chambered coverglass well (155411, Nunc Lab-Tek,
398 Thermo Scientific). The reaction volume was 200 µl. The reaction buffer (RB) was 33 mM Tris-acetate
399 (pH 7.9 at 20 °C), 66 mM CH_3COOK, 5 mM $(CH_3COO)_2Mg$, and 0.1 mg/ml BSA (bovine serum
400 albumin). The DNA concentration was 17-50 pM. Measurements at different protein concentrations
401 were carried out by adding to the reaction small volumes of protein diluted in RB in 'Protein LoBind'
402 1.5 mL tubes (Eppendorf). No oxygen-scavenging or triplet-quenching additives were used.

403 Measurements of surface-immobilized DNA fragments were performed in a flow cell assembled from
404 a six-channel Sticky-Slide VI 0.4 (Ibidi) and a coverslip functionalized with PEG (polyethylene glycol)
405 derivatives as described in detail in [21]. The flow cell was incubated with 5 µg/ml of Neutravidin
406 (Molecular probes) in RB for 2 min, washed with RB, incubated with 5 pM DNA in RB until the density
407 of the surface-immobilized DNA fragments appeared to be appropriate, and washed with RB. For the
408 measurement, 10 nM solution of AfAgo in RB supplemented with 1% glucose (TCI Europe), 2.5 mM
409 Trolox (Sigma-Aldrich), and 15 U/ml glucose oxidase (Sigma-Aldrich) was injected into the cell. Trolox
410 was treated with UV light for 20 min according to Cordes et al. [49]. Single molecule data analysis
411 was performed as described in Supplementary Methods.

412

DECLARATIONS

Funding

This work was supported by the Research Council of Lithuania [S-MIP-17-61 to M.Z.]. The access to P12 was supported by the Horizon 2020 programme of the European Union, iNEXT (H2020 Grant # 653706). Funding for open access charge: the Research Council of Lithuania.

Autors' contributions

AS purified the target proteins. EM performed the SAXS measurements and crystallographic analysis. DR and EG performed smFRET experiments. MJ performed AFM measurements. GS, DR, EM and EG drafted the manuscript. MZ designed and coordinated the study, and critically edited the manuscript. All authors read and approved the final manuscript.

Acknowledgement

Authors gratefully acknowledge prof. Virginijus Siksnys for fruitful discussions and comments on the manuscript. Authors also acknowledge Audra Ruksenaite for mass spectrometry of AfAgo proteins, Tomas Urbaitis for generation of the dimerization mutant AfAgoΔ and initial experiments, and prof. Gintaras Valincius for support. The synchrotron SAXS data was collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We would like to thank Dr. Melissa Graewert for the assistance in using the beamline.

Competing interests

None of the authors have any competing interests.

REFERENCES

1. Song J-J, Smith SK, Hannon GJ, Joshua-Tor L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. Science (80-). 2004;305:1434–7. doi:10.1126/science.1102514.

2. Kwak PB, Tomari Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol. 2012;19:145–51.

3. Pratt AJ, MacRae IJ. The RNA-induced silencing complex: A versatile gene-silencing machine. J Biol Chem. 2009;284:17897–901.

4. Ryazansky S, Kulbachinskiy A, Aravin AA. The Expanded Universe of Prokaryotic Argonaute Proteins. MBio. 2018;9:1–20.

441 5. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V., et al. The evolutionary
442 journey of Argonaute proteins. Nat Struct Mol Biol. 2014;21:743–53.
443 6. Lisitskaya L, Aravin AA, Kulbachinskiy A. Supplementary information - DNA interference and
444 beyond: structure and functions of prokaryotic Argonaute proteins. Nat Commun. 2018;9:1–12.
445 7. Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O. Crystal structure of the MID-
446 PIWI lobe of a eukaryotic argonaute protein. Proc Natl Acad Sci U S A. 2011;108:10466–71.
447 8. Dayeh DM, Kruithoff BC, Nakanishi K. Structural and functional analyses reveal the contributions of
448 the C- and N-lobes of Argonaute protein to selectivity of RNA target cleavage. J Biol Chem.
449 2018;293:6308–25.
450 9. Frank F, Hauver J, Sonenberg N, Nagar B. Arabidopsis Argonaute MID domains use their
451 nucleotide specificity loop to sort small RNAs. EMBO J. 2012;31:3588–95.
452 doi:10.1038/emboj.2012.204.
453 10. Frank F, Sonenberg N, Nagar B. Structural basis for 5′-nucleotide base-specific recognition of
454 guide RNA by human AGO2. Nature. 2010;465:818–22. doi:10.1038/nature09039.
455 11. Till S, Lejeune E, Thermann R, Bortfeld M, Hothorn M, Enderle D, et al. A conserved motif in
456 Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain.
457 Nat Struct Mol Biol. 2007;14:897–903.
458 12. Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O. Crystal structure and ligand
459 binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep. 2010;11:522–7.
460 13. Parker JS, Roe SM, Barford D. Crystal structure of a PIWI protein suggests mechanisms for
461 siRNA recognition and slicer activity. EMBO J. 2004;23:4727–37.
462 14. Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ. Structural basis for 5′ -end-specific
463 recognition of guide RNA by the A. fulgidus Piwi protein. Nature. 2005;434:666–70.
464 15. Parker JS, Roe SM, Barford D. Structural insights into mRNA recognition from a PIWI domain-
465 siRNA guide complex. Nature. 2005;434:663–6.
466 16. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D. Enhancement of the Seed-Target
467 Recognition Step in RNA Silencing by a PIWI/MID Domain Protein. Mol Cell. 2009;33:204–14.
468 doi:10.1016/j.molcel.2008.12.012.
469 17. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. J Mol Biol.
470 2007;372:774–97.

471 18. Janin J, Chothia C. The structure of protein-protein recognition sites. J Biol Chem.
472 1990;265:16027–30. http://www.jbc.org/content/265/27/16027.short.
473 19. Conte L Lo, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol
474 Biol. 1999;285:2177–98. doi:https://doi.org/10.1006/jmbi.1998.2439.
475 20. Hegge JW, Swarts DC, Chandradoss SD, Cui TJ, Kneppers J, Jinek M, et al. DNA-guided DNA
476 cleavage at moderate temperatures by Clostridium butyricum Argonaute. Nucleic Acids Res.
477 2019;47:5809–21.
478 21. Tutkus M, Sasnauskas G, Rutkauskas D. Probing the dynamics of restriction endonuclease
479 NgoMIV-DNA interaction by single-molecule FRET. Biopolymers. 2017;107:1–9.
480 22. Lushnikov AY, Potaman VN, Oussatcheva EA, Sinden RR, Lyubchenko YL. DNA strand
481 arrangement within the SfiI-DNA complex: Atomic force microscopy analysis. Biochemistry.
482 2006;45:152–8.
483 23. Zaremba M, Sasnauskas G, Siksnys V. The link between restriction endonuclease fidelity and
484 oligomeric state: A study with Bse634I. FEBS Lett. 2012;586:3324–9.
485 doi:10.1016/j.febslet.2012.07.009.
486 24. Pathania S, Jayaram M, Harshey RM. Path of DNA within the Mu transpososome: Transposase
487 interactions bridging two Mu ends and the enhancer trap five DNA supercoils. Cell. 2002;109:425–36.
488 25. Peterson G, Reznikoff W. Tn5 transposase active site mutations suggest position of donor
489 backbone DNA in synaptic complex. J Biol Chem. 2003;278:1904–9.
490 26. Hickman AB, Kailasan S, Genzor P, Haase AD, Dyda F. Casposase structure and the mechanistic
491 link between DNA transposition and spacer acquisition by CRISPR-Cas. Elife. 2020;9:1–26.
492 27. Pouget N, Turlan C, Destainville N, Salomé L, Chandler M. IS911 transpososome assembly as
493 analysed by tethered particle motion. Nucleic Acids Res. 2006;34:4313–23.
494 28. Rutkauskas D, Petkelyte M, Naujalis P, Sasnauskas G, Tamulaitis G, Zaremba M, et al.
495 Restriction enzyme Ecl18kI-induced DNA looping dynamics by single-molecule FRET. J Phys Chem
496 B. 2014;118:8575–82.
497 29. Bellamy SRW, Mina P, Retter SE, Halford SE. Fidelity of DNA Sequence Recognition by the SfiI
498 Restriction Endonuclease Is Determined by Communications between Its Two DNA-Binding Sites. J
499 Mol Biol. 2008;384:557–63. doi:10.1016/j.jmb.2008.09.057.
500 30. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, et al. DNA-guided DNA

501 interference by a prokaryotic Argonaute. Nature. 2014;507:258–61.
502 31. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. Bacterial Argonaute Samples
503 the Transcriptome to Identify Foreign DNA. Mol Cell. 2013;51:594–605.
504 doi:10.1016/j.molcel.2013.08.014.
505 32. Wright A V, Liu J-J, Knott GJ, Doxzen KW, Nogales E, Doudna JA. Structures of the CRISPR
506 genome integration complex. Science (80-). 2017;357:1113 LP – 1118.
507 doi:10.1126/science.aao0679.
508 33. Xiao Y, Ng S, Hyun Nam K, Ke A. How type II CRISPR-Cas establish immunity through Cas1-
509 Cas2-mediated spacer integration. Nature. 2017;550:137–41. doi:10.1038/nature24020.
510 34. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and
511 inhibition of DNA strand transfer. Nature. 2010;464:232–6.
512 35. Grawenhoff J, Engelman AN. Retroviral integrase protein and intasome nucleoprotein complex
513 structures. World J Biol Chem. 2017;8:32.
514 36. Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, et al. Versatile
515 sample environments and automation for biological solution X-ray scattering experiments at the P12
516 beamline (PETRA III, DESY). J Appl Crystallogr. 2015;48:431–43.
517 37. Panjkovich A, Svergun DI. CHROMIXS: Automatic and interactive analysis of chromatography-
518 coupled small-angle X-ray scattering data. Bioinformatics. 2018;34:1944–6.
519 38. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle
520 scattering. J Appl Crystallogr. 2009;42:342–6.
521 39. Svergun DI. Determination of the regularization parameter in indirect-transform methods using
522 perceptual criteria. J Appl Crystallogr. 1992;25:495–503.
523 40. Petoukhov M V., Franke D, Shkumatov A V., Tria G, Kikhney AG, Gajda M, et al. New
524 developments in the ATSAS program package for small-angle scattering data analysis. J Appl
525 Crystallogr. 2012;45:342–50.
526 41. Volkov V V., Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J
527 Appl Crystallogr. 2003;36:860–4. doi:10.1107/S0021889803000268.
528 42. Svergun DI. Restoring low resolution structure of biological macromolecules from solution
529 scattering using simulated annealing. Biophys J. 1999;76 June:2879–86.
530 http://linkinghub.elsevier.com/retrieve/pii/S0006349599774436.

531 43. Svergun DI, Petoukhov M V., Koch MHJ. Determination of domain structure of proteins from x-ray
532 solution scattering. Biophys J. 2001;80:2946–53.
533 44. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. J Appl
534 Crystallogr. 2001;34:33–41.
535 45. Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas Č, Nwokeoji AO, Dickman MJ, et al.
536 Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus
537 thermophilus. Mol Cell. 2014;56:506–17.
538 46. Svergun D, Barberato C, Koch MHJ. CRYSOL - a Program to Evaluate X-ray Solution Scattering
539 of Biological Macromolecules from Atomic Coordinates. J Appl Crystallogr. 1995;28:768–73.
540 47. Lyubchenko YL, Shlyakhtenko LS, Ando T. Imaging of nucleic acids with atomic force microscopy.
541 Methods. 2011;54:274–83. doi:10.1016/j.ymeth.2011.02.001.
542 48. Kapanidis AN, Lee NK, Laurence TA, Doose S, Margeat E, Weiss S. Fluorescence-aided
543 molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single
544 molecules. Proc Natl Acad Sci U S A. 2004;101:8936–41. doi:10.1073/pnas.0401690101.
545 49. Cordes T, Vogelsang J, Tinnefeld P. On the mechanism of trolox as antiblinking and antibleaching
546 reagent. J Am Chem Soc. 2009;131 iv:5018–9.
547 50. Laskowski RA, Jabłonska J, Pravda L, Vareková RS, Thornton JM. PDBsum: Structural
548 summaries of PDB entries. Protein Sci. 2017;27:129–34.
549 51. Ingargiola A, Lerner E, Chung SY, Weiss S, Michalet X. FRETBursts: An open source toolkit for
550 analysis of freely-diffusing Single-molecule FRET. PLoS One. 2016;11:1–27.
551 52. Holden SJ, Uphoff S, Hohlbein J, Yadin D, Le Reste L, Britton OJ, et al. Defining the Limits of
552 Single-Molecule FRET Resolution in TIRF Microscopy. Biophys J. 2010;99:3102–11.
553 53. Durand D, Vivès C, Cannella D, Pérez J, Pebay-Peyroula E, Vachette P, et al. NADPH oxidase
554 activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. J Struct
555 Biol. 2010;169:45–53. doi:10.1016/j.jsb.2009.08.009.
556 54. Chandran S. Rapid Assembly of DNA via Ligase Cycling Reaction (LCR). In: Hughes RA, editor.
557 Synthetic DNA: Methods and Protocols. New York, NY: Springer New York; 2017. p. 105–10.
558 doi:10.1007/978-1-4939-6343-0_8.
559 55. Konarev P V., Volkov V V., Sokolova A V., Koch MHJ, Svergun DI. PRIMUS - a Windows-PC
560 based system for small-angle scattering data analysis. J Appl Crystallogr. 2003;36:1277–82.

561 56. Fischer H, De Oliveira Neto M, Napolitano HB, Polikarpov I, Craievich AF. Determination of the
562 molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a
563 relative scale. J Appl Crystallogr. 2010;43:101–9.

564

565 **TABLES AND FIGURES**

566 **Table 1. AfAgo-DNA complexes observed by AFM.**

Protein	**DNA loops,** %	**Linear**		**Total number of analyzed complexes**
		Bound to one end, %	*Bound to both ends,* %	
WT AfAgo	55	29	16	208
AfAgoΔ	12	29	59	319

567

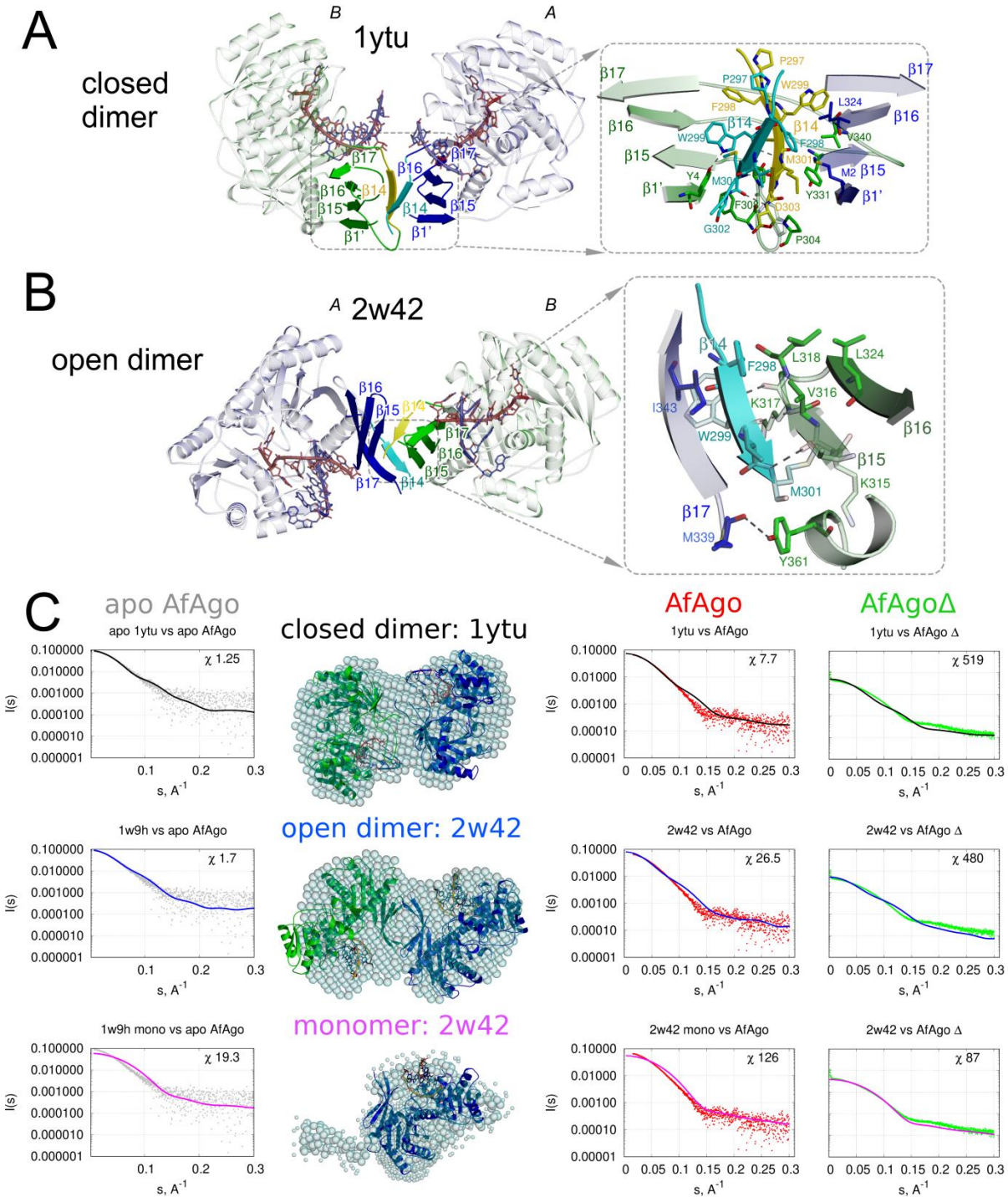
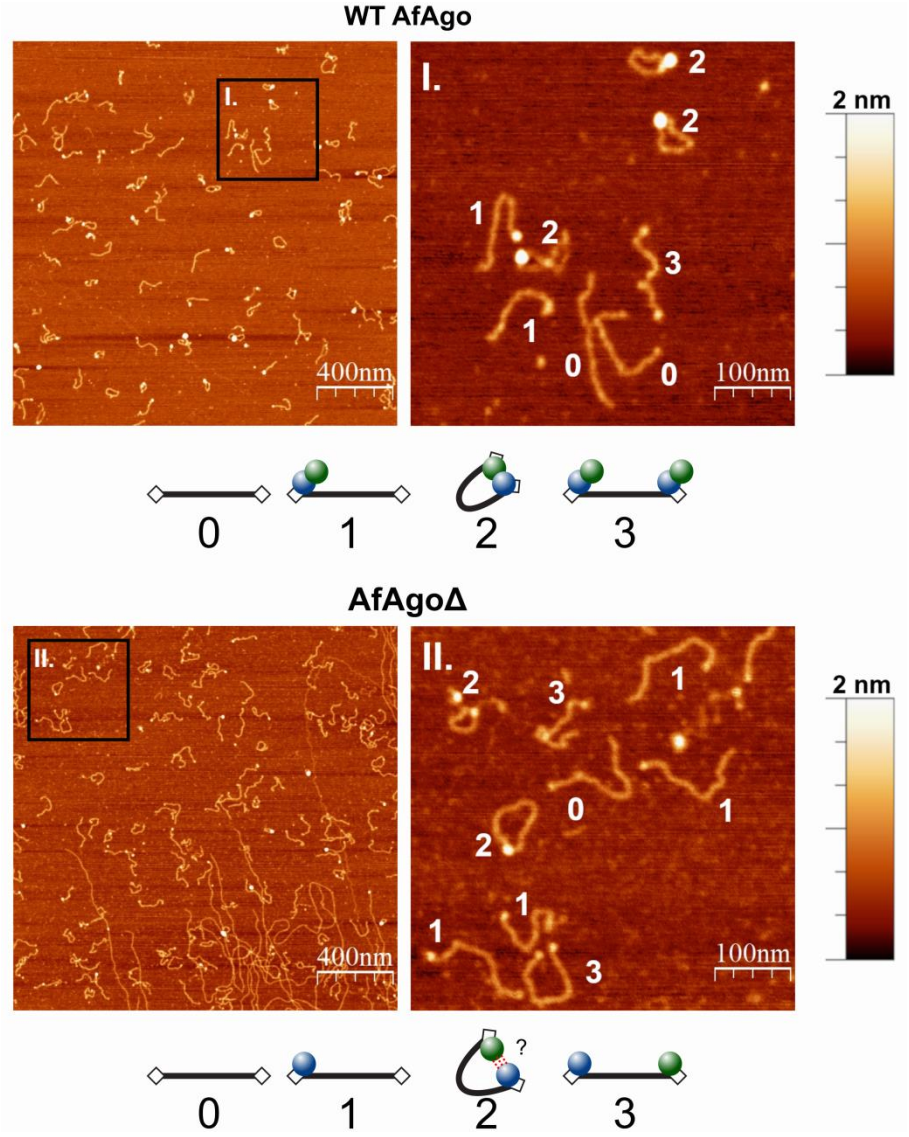
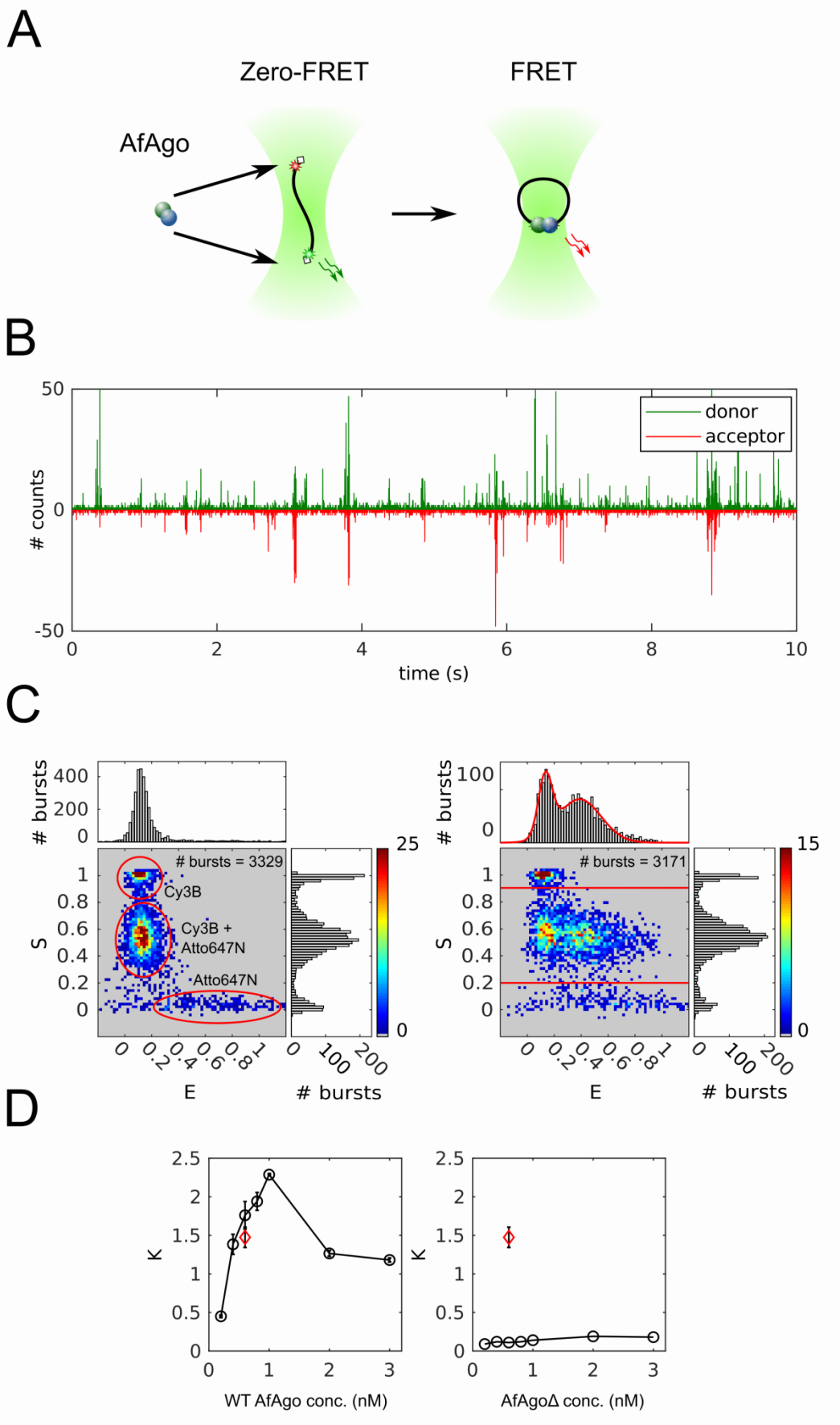


Figure 1. Dimerization of AfAgo. (A-B). Protein subunits are colored blue (protein chain *A*) and green (protein chain *B*). The interface-forming secondary structure elements are highlighted and numbered according to the PDB ID 2w42 assignment made by PDBsum [50]. The 'guide' DNA/RNA strands bound by AfAgo are colored red, 'target' strands — blue. Residues 296-303 deleted in AfAgoΔ are colored cyan and yellow. Hydrogen bonds are shown as dashed lines. (A) AfAgo complex with dsRNA (PDB ID 1ytu, both protein chains as present in ASU), the 'closed' dimer [14]. (B), AfAgo

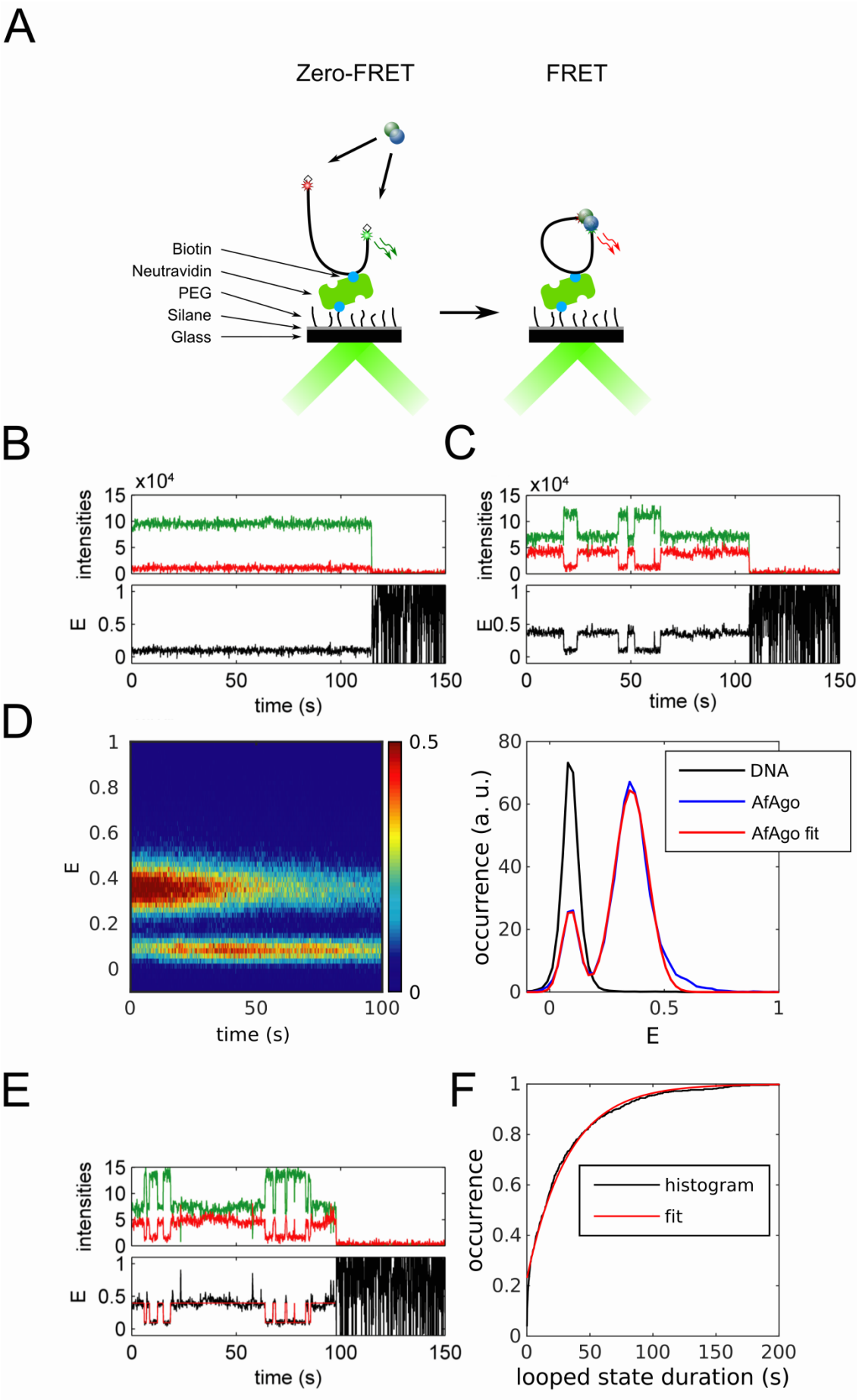
complex with dsDNA (PDB ID 2w42, protein chain B is produced by operator (X,Y,Z+(-1 2 2)&{0 -1 -1}) [16]) - the 'open' dimer. β-strands from both subunits assemble into a closed β-barrel structure, with intersubunit interface formed by β14 and β15 strands of neighboring subunits. (C) SAXS data of WT AfAgo apo and complex with MZ-1289 DNA (grey and red points, respectively), and the dimerization mutant AfAgoΔ with MZ-1289 DNA (green points, right column) are compared with the scattering curves generated from the 'closed' dimer with or without dsRNA (PDB ID: 1ytu, black curves), 'open' dimer (PDB ID: 2w42 or 1w9h for apo AfAgo, blue curves) and monomeric apo and AfAgo-DNA complex (PDB ID: 2w42 or 1w9h, 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 AfAgo complex with MZ-1289 oligoduplex.



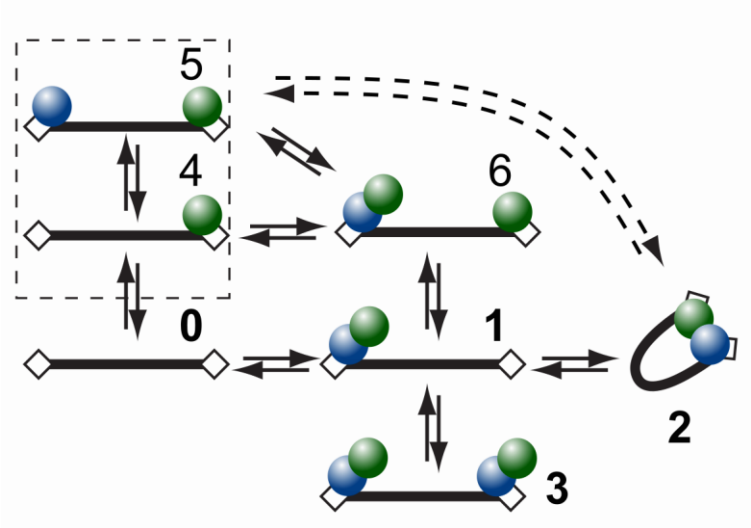
586 **Figure 2. Visualization of AfAgo-induced DNA loops by AFM.** Representative AFM images and
587 enlarged views of DNA-protein complexes adsorbed to APS-mica acquired in the air are shown.
588 DNA+WT AfAgo (top) and DNA + AfAgoΔ (bottom), area of each left column image is 4 μm^2, scale bar
589 is 400 nm. Right column shows regions marked by squares in the respective images on the left,
590 enlarged 4-fold, scale bar 100 nm. Various observed species are depicted below the respective
591 images and are marked as follows: '0' - naked DNA; '1' - protein bound to one DNA end; '3' - protein
592 bound to both DNA ends; '2' - ring-shaped (looped) DNA. In the case of AfAgoΔ, species '2' is
593 presumably formed due to glutaraldehyde crosslinking. The Z scale bar is 2 nm.



595 **Figure 3. Single-molecule studies of AfAgo-DNA interactions in solution.** (A) A schematic
596 overview of the single molecule assay. Left, free DNA, right, DNA-WT AfAgo (blue and green circles)
597 complex in a looped state. (B) Fluorescence intensity trace with 1 ms time bin of 25 pM DNA with
598 1 nM AfAgo. Donor fluorescence upon donor excitation is presented in the upper part of the graph.
599 Inverted acceptor fluorescence upon donor excitation is presented in the lower part of the graph. (C)
600 Left - E-S histogram of DNA alone. The top and side axes contain, respectively, one-dimensional E
601 (proximity ratio) and S (donor/acceptor stoichiometry) histograms of all bursts. Denoted are areas
602 corresponding donor-only DNA, acceptor-only DNA, and dual-labeled DNA. Right – E-S histogram of
603 DNA with 1 nM AfAgo. The one-dimensional E histogram on top is derived from bursts with S = 0.2-
604 0.9, designated by horizontal lines in the E-S histogram. The red curve is a two-Gaussian fit to the
605 data that gave the positions of the Gaussian maxima on the E-axis (0.13±0.01 and 0.39±0.02). (D)
606 Left - dependence of the ratio of looped and unlooped DNA molecules (parameter K) on WT AfAgo
607 concentration (open circles). Right - the dependence of K on the AfAgoΔ concentration (open circles).
608 The red diamonds in both graphs represent the competition experiment performed with 0.6 nM WT
609 AfAgo dimer and 0.6 nM AfAgoΔ monomer. All data points are average values of three measurements
610 ±1 standard deviation.



612 **Figure 4. Dynamics of WT AfAgo-induced DNA loops.** (A) A schematic overview of the single-
613 molecule assay using TIRF microscopy. (B, C) Trajectories of donor (green) and acceptor (red)
614 intensity and corresponding proximity ratio, E, of individual DNA fragments without (B) and with 10 nM
615 AfAgo (C). (D) Left - an image of 287 pooled time traces of the proximity ratio, E, from the
616 measurement with 10 nM of AfAgo. The image is normalized to the maximum image intensity. Right –
617 a section of the image in the left integrated over the first 10 s shown with the two-Gaussian fit. The
618 positions of the Gaussian maxima are 0.09±0.01 and 0.36±0.01. For comparison a trace- and time-
619 averaged section from the measurement of 227 traces on bare DNA is shown. (E) An example of
620 trajectories of donor (green) and acceptor (red) intensity and corresponding proximity ratio, E, with
621 HMM idealization of an individual DNA fragment with 10 nM AfAgo. F – cumulative histogram of the
622 looped state durations from 287 E traces with a single-exponential fit with the exponential factor of
623 33±1 s.



624

625 **Figure 5. Kinetic schemes depicting possible reaction pathways between two-target site DNA**
626 **fragment and AfAgo.** Black bars represent DNA, rectangular boxes – AfAgo-binding targets (DNA
627 ends), green and blue circles – AfAgo monomers. The 0-1-2-3 pathway represents the case of a
628 stable AfAgo dimer and is supported by our experimental data. Pathway 0-4-5-2 represents the
629 reactions that would occur if AfAgo was a monomer assembling into a dimer upon DNA binding.
630 Dashed box represents species limited to AfAgoΔ.

631

Prokaryotic Argonaute from *Archaeoglobus fulgidus* interacts with DNA as a homodimer

Edvardas Golovinas[1], Danielis Rutkauskas[1,2], Elena Manakova[1], Marija Jankunec[1,3], Arunas Silanskas[1], Giedrius Sasnauskas[1], Mindaugas Zaremba[1,*]

[1] Institute of Biotechnology, Life Sciences Center, Vilnius University, Sauletekio av. 7, LT-10257, Vilnius, Lithuania.
[2] Institute of Physics, Center for Physical Sciences and Technology, Savanoriu 231, LT-02300, Vilnius, Lithuania
[3] Institute of Biochemistry, Life Sciences Center, Vilnius University, Sauletekio av. 7, LT-10257, Vilnius, Lithuania.

* To whom correspondence should be addressed. Tel: +370-5-2234357; Fax: +370-5-2234367; Email: zare@ibt.lt.

Supplementary Information

SUPPLEMENTARY METHODS

Single molecule setup

We used a custom single-molecule fluorescence microscopy setup built on a commercial inverted microscope Nikon Eclipse Ti-U equipped with 60× 1.2 WI Plan Apo VC objective (Nikon) used for the excitation and signal collection, two avalanche photodiode (APD)-based single photon counting modules (Tau-SPAD-50, PicoQuant) and 25 mW 532 and 635 nm diode-pumped solid state and diode lasers (Crystalaser), respectively. The laser excitation was reflected off a dichroic mirror (zt532/635rpc-XT, Chroma), and the fluorescence signal filtered off the excitation light with a quadruple-band interference filter (FF01-446/510/581/703, Semrock) and split into two spectral channels with a dichroic mirror (645dcxr, Chroma). ALEX was implemented by directly TTL-modulating the intensity of the 635 nm laser and synchronously modulating the intensity of the 532 nm laser with a mechanical chopper (MC2000B, Thorlabs). The half period of ALEX was 50 µs. Fluorescence photon arrival times were recorded and ALEX was implemented using an FPGA module (PCIe-7851R, National Instruments) and custom Labview (National Instruments) program.

The excitation was focused 50 µm above the sample chamber glass surface. 532 nm excitation intensity was 30 µW, 635 nm - 20 µW. The size of the confocal pinhole was 75 µm. Each measurement was 10 min long.

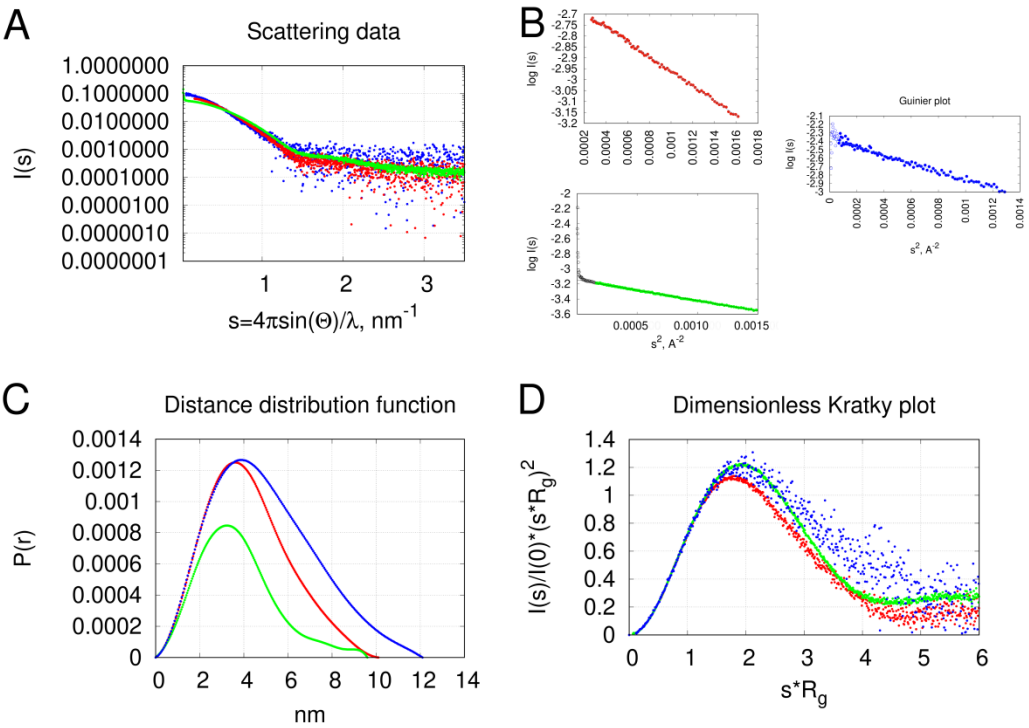
664 **Single Molecule Data analysis**

665 Fluorescence burst analysis was performed using the freely available FRETBursts software [51]. The
666 initial bursts search parameters were m = 10 photons, and F = 6 times the fluorescence background.
667 The total intensity of a burst from both channels and excitation wavelengths was thresholded to be
668 larger than 40 counts, and this yielded ~3000 bursts from a 10 min measurement. Each burst was
669 calculated a proximity ratio, E, according to $E = I_d^a/(I_d^a + I_d^d)$, here I_d^a and I_d^d are acceptor and donor
670 intensities upon donor excitation, respectively, and stoichiometry parameter, S, according to
671 $S = I_d/(I_d + I_a^a)$, here I_d is the total donor and acceptor intensity upon donor excitation, and I_a^a is
672 acceptor intensity upon acceptor excitation. Then we built 2D E-S histograms of bursts. Subsequently,
673 bursts with stoichiometry parameter ranging from 0.2-0.9 were selected to build distributions of the
674 proximity ratio, E, of bursts of DNA molecules labelled with both fluorophores only. E histograms were
675 fit with the sum of two Gaussian functions using unconstrained optimization. Then the ratio of the
676 number of looped and unlooped DNA molecules in the ensemble was calculated as the ratio of the
677 area of the Gaussian of high E with that of low E.

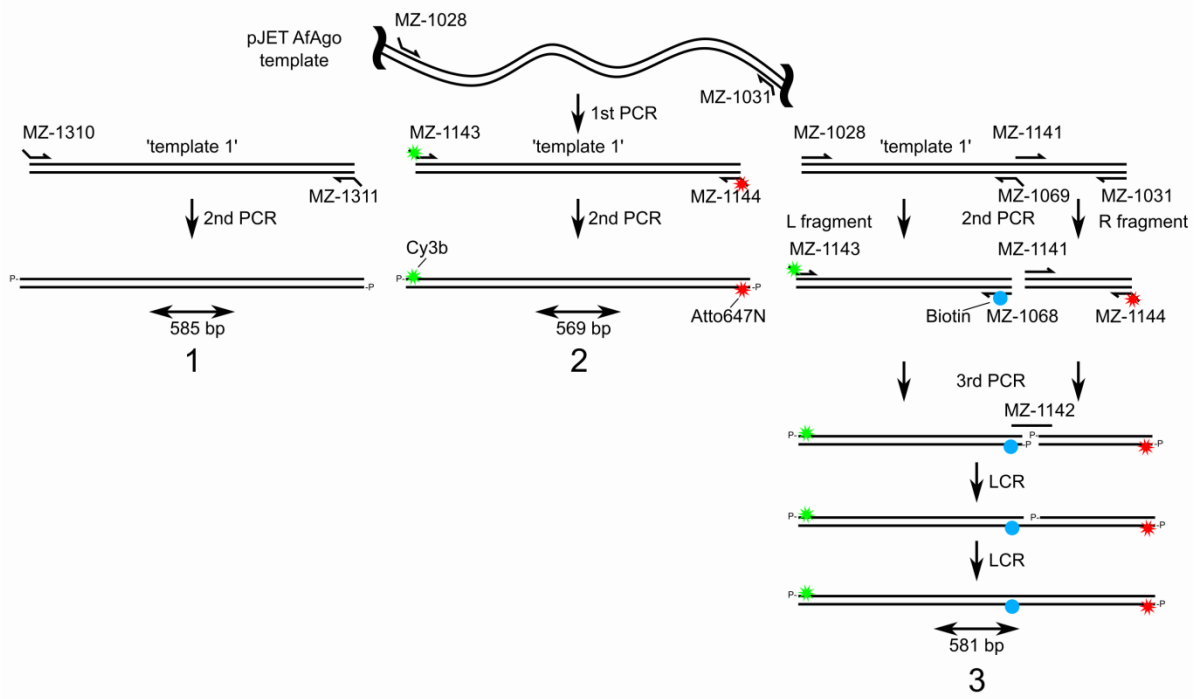
678 The experiment of surface-immobilized DNA fragments was done by first recording a short movie with
679 635 nm excitation to obtain a reference for fluorescent spot identification since the acceptor channel
680 exhibits significantly less fluorescence background than the donor channel. Then a longer actual
681 movie was recorded with the 532 nm excitation. The analysis of the two-spectral channel fluorescence
682 movies was performed using custom software written in Matlab. Briefly, to identify the fluorescent
683 spots, the first 20 frames of the reference and the actual fluorescence movies were averaged, the
684 obtained average images were filtered with the 2D low-pass Gaussian filter 5 pixels large and with the
685 standard deviation of 1 pixel and subtracted the same image filtered with the averaging filter 7 pixels
686 large. The resulting acceptor channel reference image was thresholded with 20 and the donor
687 channel actual image - with 40 counts/pixel. The obtained images were binarized for particle
688 identification. Particles in both binary images were identified and filtered according to the following
689 criteria: 5x5 pixel ROIs (regions of interest) centered on particles' centers of mass had to non-overlap,
690 particle area had to be within 5-100 pixels range, particle eccentricity not larger than 0.8. The
691 coordinates of a particle in the donor channel corresponding to a particle identified in the acceptor
692 channel of the reference movie were calculated using the spatial transformation structure calculated
693 from an image of surface-immobilized 200 nm fluorescent polystyrene beads (F8806, Invitrogen). For

694 trace extraction were considered only those particles in the actual movie whose donor coordinates
695 coincided with the transformed coordinates of the acceptor particles in the reference movie within 1.5
696 pixels. The donor and acceptor intensity traces were extracted using aperture photometry [52] with the
697 background calculated as an average intensity from a 1 pixel-wide annulus around particle's ROI. The
698 proximity ratio, E, was calculated according to the same formula as for the fluorescence bursts.

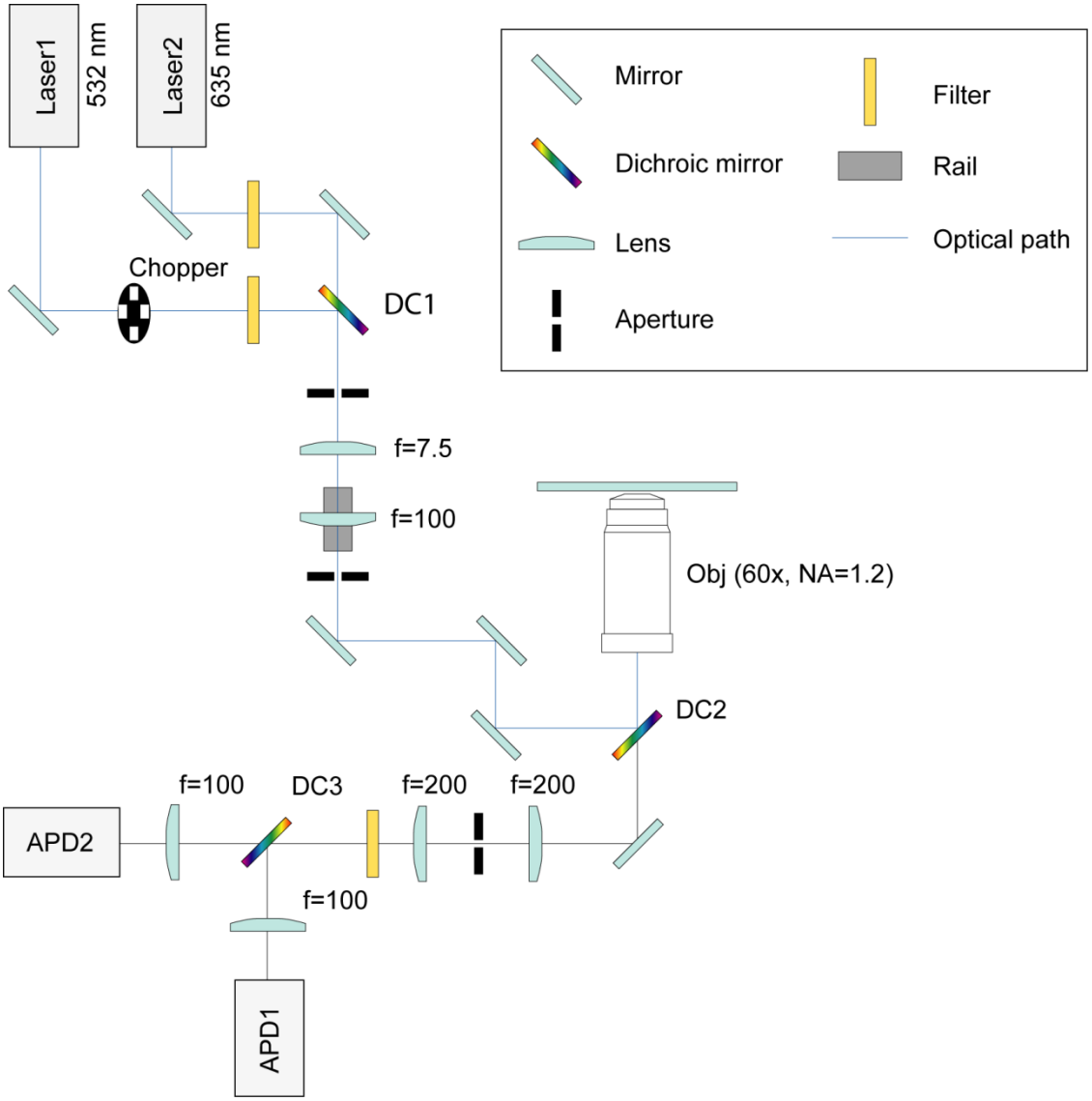
699



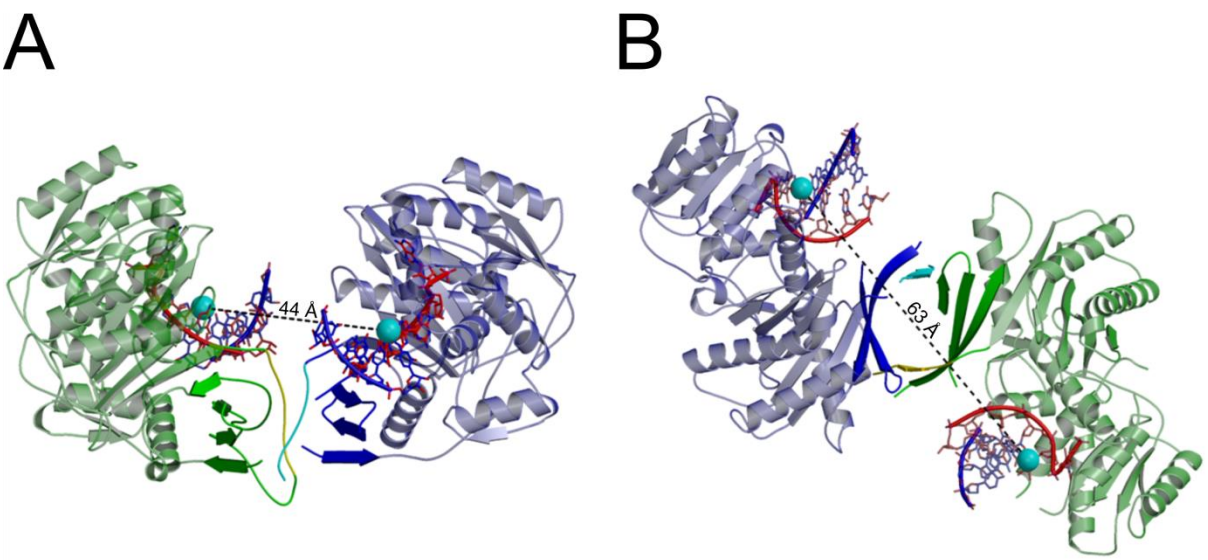
Supplementary figure S1. SAXS data of AfAgo+MZ-1289 (red curves), monomeric mutant AfAgoΔ+MZ-1289 (green curves) complexes and apo AfAgo (blue curves). (A), Scattering curves. (B), Guinier plots log I(s) vs. s^2 of the data at small s values. (C), Pair distance distribution functions. (D), Dimensionless Kratky representation of scattering data $(I(s)/I(0)*(s*Rg)^2$ vs. s*Rg). All curves have similar shape typical for folded proteins [53].



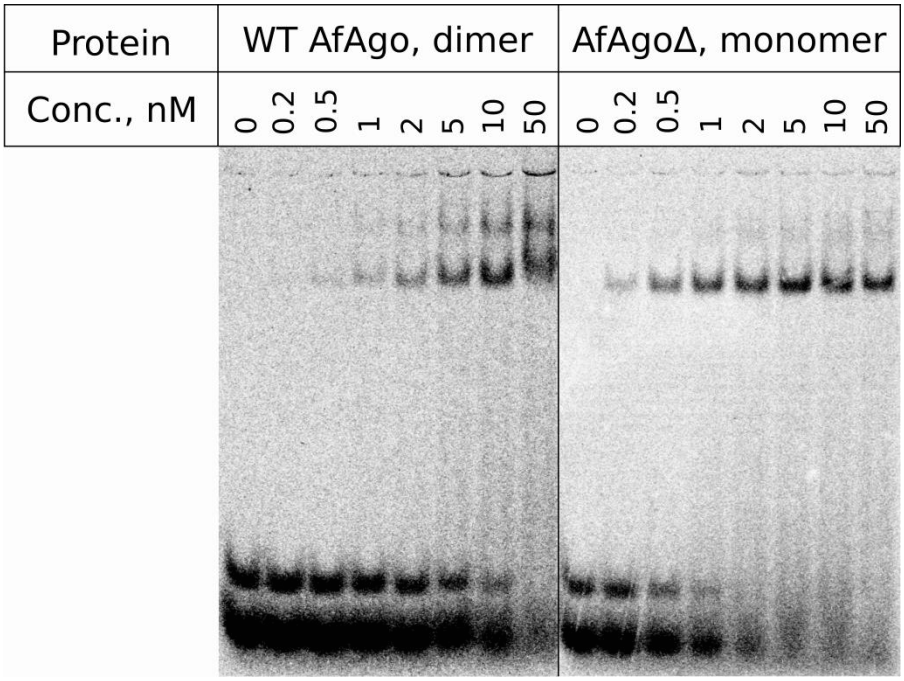
Supplementary figure S2. Synthesis scheme for the DNA fragments. First, a DNA fragment was amplified from a pJET plasmid template containing an AfAgo gene fragment using oligonucleotides MZ-1028 and MZ-1031. The PCR product was then used as a template (dubbed 'template 1') in subsequent reactions. Fragment '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 were 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 b 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 nM. The phosphorylation mix was subsequently ligated by Ampligase® (Epicentre, USA) at 3 nM total DNA and 30 nM bridging oligonucleotide MZ-1142 according to Chandran, 2017 [54]. 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% ethanol and resuspended in water.



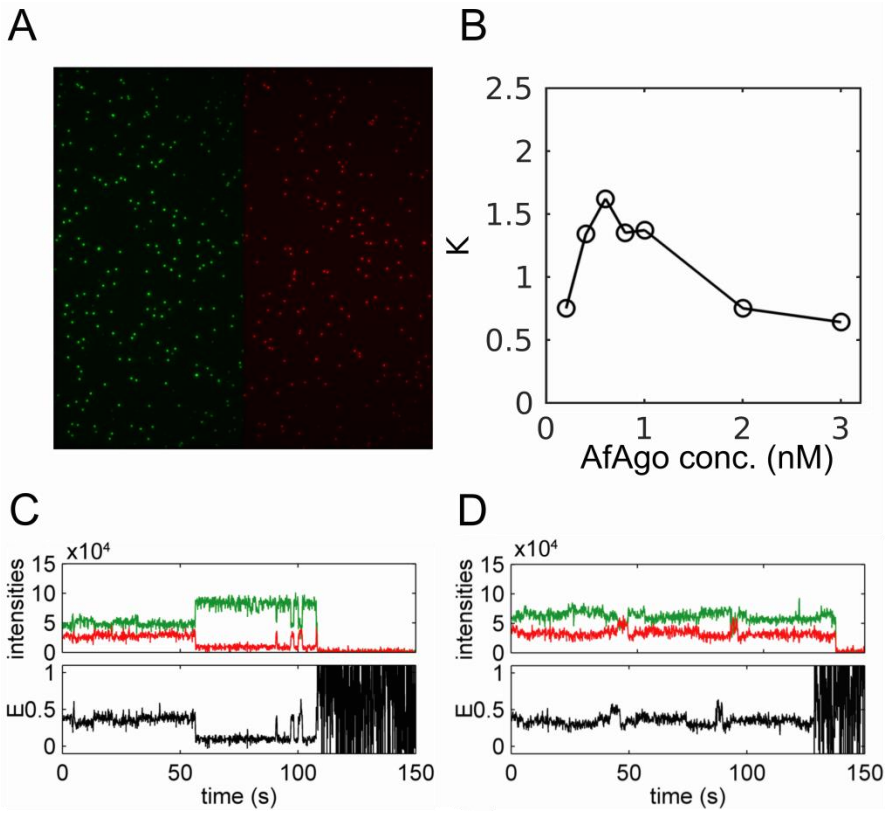
Supplementary figure S3. Optical scheme of custom single-molecule fluorescence microscopy setup used to record fluorescence bursts of single diffusing molecules in this study. APD – avalanche photodiode; f – focal distance; NA – numerical aperture.



Supplementary figure S4. 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, DNA guide and target strands are red and blue respectively.



Supplementary figure S5. DNA binding by AfAgo. DNA binding was verified using electrophoretic mobility shift assay. Self-complementary 5'P^{32} DNA MZ-952 was used as a substrate. Final concentration of DNA duplex in binding reaction was 1 nM, final protein concentrations are shown above each lane. Binding buffer was 40 mM Tris-acetate (pH 8.4 at 25 °C) with 1 mM EDTA (TAE, B49, Thermo Scientific), supplemented with 5 mM $MgCl_2$, 0.1 mg/ml BSA, 1 mM DTT, 10% glycerol. Running buffer – TAE (Thermo Scientific) supplemented with 5 mM $MgCl_2$.



745

746 **Supplementary figure S6. Single molecule experiments.** (A) A fluorescence image of surface-immobilized
747 DNA fragments. It is an average of 20 frames in a fluorescence movie. The left part (green) is the donor image
748 upon donor excitation, and the right part (red) is the acceptor image upon acceptor excitation. (B) The
749 dependence of the ratio, K, of the number of looped and unlooped DNA molecules depending on the
750 concentration of the AfAgo for the biotinylated DNA fragment in solution. (C, D) Examples of different dynamics of
751 DNA looping by AfAgo in TIRF experiments.

752

753 **Supplementary table S1. List of oligonucleotides used in this study.**

Oligonucleotide	Sequence, 5‘->3‘	Modifications
MZ-383	TGATTCTGCAGTTATAGGAACCACGGATTCGTTTGTAATGAGC	
MZ-385	TGATTGGATCCGATGATGGAATATAAAATAGTTGAAAATGGTTTGAC	
MZ-875	GCTATACTTCACTTAAATGAAACTCCTAACAATAGATTTCATCCGTAT G	
MZ-876	CCTTCATACGGATGAAATCTATTGTTAGGAGTTTCATTTAAGTGAAGTATAGC	
MZ-952	ATCGTGGCCACGAT	
MZ-1028	GTGCTGTACCTTGACCTTGATGAACTGGCGCAACACGTATTG	
MZ-1031	ATACTGGCTGCATCTAGCATACGATCTCAACACTTAATGGTTT	
MZ-1068	ATTCTGGTCTCGGACTCCCATTACCCAAAATGGATGAG	**Biotin on T22**
MZ-1069	ATTCTGGTCTCGGACTCCCATTACCCAAAATGGATGAG	
MZ-1141	CCTAACAATAGATTTCATCCG	
MZ-1142	GGGTAATGGGAGTCCGAGACCAGAATCCTAACAATAGATTTCATCC GTATGAAGG	
MZ-1143	ATTATAATTATGTATGTGCTGTACCTTGACCTTGAT	**Cy3b on T3, 5‘P**
MZ-1144	ATTATAATAGGATACTGGCTGCATCTAGCAT	**Atto647N on T3, 5‘P**
MZ-1310	ATTGCTCTACTGTATAATGCTGTGCTGTACCTTGACCTTGAT	
MZ-1311	ATTGCTCTACTGTATAATGCTATACTGGCTGCATCTAGCAT	
MZ-1289	ATTGTACGTACAAT	**5‘P**

754

755

Supplementary table S2. SAXS data collection and main structural parameters

Instrument, Detector	P12, pilatus6m		P12, pilatus2m
Detector-to-sample distance, m	3.0		3.0
Wavelength, nm	0.123981		0.124
Measured s range, nm^{-1}	0.0224526-7.3176000		0.02492870-5.064020
Number of buffer exposure frames averaged (measured) / frame exposure time	101 (101) / 0.995 sec	76 (80) / 0.195 sec	40 (40) / 0.045
Number of sample exposure frames averaged (measured) / frame exposure time	24 (24) / 0.995 sec	30 (40) / 0.195 sec	20 (20) / 0.045
Capillary temperature/ Sample changer temperature	20 °C / Room temperature	20 °C / 10 °C	20 °C / 10 °C
Data reduction and on-line characterization	radaver (r11095), databsolute v0.1 (r11095)		radaver (v. 9729)
Structural parameters			
Sample	WT AfAgo+MZ-1289, SEC peak	AfAgoΔ+MZ-1289, 4 mg/ml	Apo AfAgo, 0.13 mg/ml
Guinier points (AUTORG)	1-87	39-132	20-114
s range, nm^{-1} (points) used in GNOM	0.0640-3.3457 (1-1200)	0.1860-3.3457 (60-1200)	0.0249-2.8342 (20-1000)
Rg, nm (AUTORG/ GNOM)	3.18 ± 0.016/ 3.233 ± 0.005202	2.84 ± 0.03/ 2.879 ± 0.002440	3.83 ± 0.12 / 3.837 ± 0.02216
I(0) (AUTORG/ GNOM)	0.0725 ± 0.00011/ 0.07301 ± 0.00008771	0.0428 ± 3.7e-05/ 0.04289 ± 0.00002499	0.0935 ± 0.00043 / 0.09310 ± 0.0005914
Dmax, nm (DATCLASS/ SHANUM/ GNOM)	11.3/ 10.2/ 10.1	10.9/ 10.5/ 9.6	13.9/ 11.5/ 12.1
Porod volume, nm^3 (DATPOROD)	158.03	108.67	174.1
SASBDB ID	SASDH39	SASDH49	-

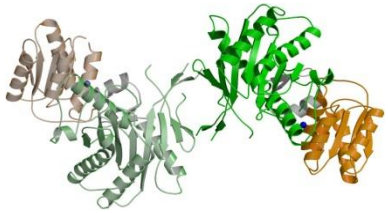
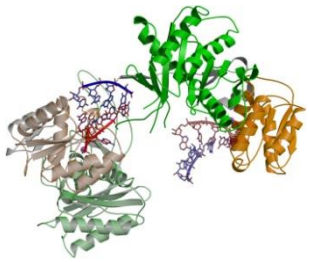
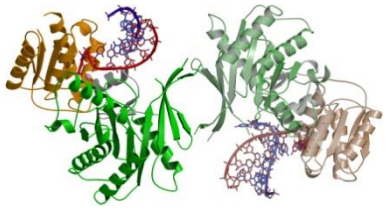
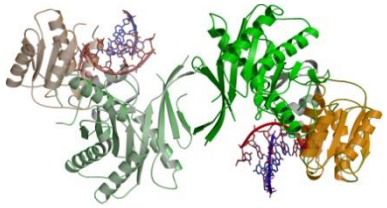
759 **Supplementary table S3. Molecular mass determination from SAXS data using various methods. All**
760 **molecular masses are given in kDa**

Sample			WT AfAgo+MZ-1289	AfAgoΔ+MZ-1289	Apo AfAgo
Expected MW (protein + DNA), kDa			110.5	56.6	101.7
Method	Reference	Sofware	MWcalc		
Absolute scale	[55]	PRIMUS 2.8.4 (r10552)	99.7	55.4	130.2
Qp			102.7	58.5	128.8
Bayes			94.2	56.9	124.5
Size&Shape			100.0	67.9	133.1
Porod volume/1.6	[40]	DATPOROD, ATSAS 2.8.4 (r10552)	98.8	67.9	108.8
SAXSMoW	[56]	SAXSMoW v2.1 http://saxs.ifsc.usp.br/	106.9 (integrated to I0/I(qmax)=$10^{2.25}$)	67.4 (integrated to I0/I(qmax)=$10^{2.25}$)	125.0 (integrated to 8/Rg)
SEC MW		CHROMIXS ATSAS 2.8.4 (r10552)	103.8	n.a.	n.a.

761

762

763 **Supplementary table S4. AfAgo dimerization interfaces as analyzed by PISA (PDBe PISA v1.52**
764 **[20/10/2014])**

PDB ID	Dimer: open/ closed	Image	CSS Complex Formation Significance Score	ΔiG P-values	PISA: dimerization surface, Å^2 (buried in interface)
1w9h	open		0.108	0.004	731
1ytu	closed		1	0	908
2bgg	open		1	0.001	601
2w42	open		1	0.002	748

765