A Random Walk Towards the Golden Fleece
Single-molecule Investigations of Argonaute Target Search

Cui, Thijs

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

A Random Walk Towards the Golden Fleece

Single-Molecule Investigations of Argonaute Target Search

A Random Walk Towards the Golden Fleece

Single-Molecule Investigations of Argonaute Target Search

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 12 december 2019 om 10:00 uur

door

Tao Ju Cui

Master of Science,
Delft University of Technology, Delft, Nederland,
geboren te Leiden, Nederland.

CONTENTS

1

GENERAL INTRODUCTION

In this chapter, we will introduce the fundamental functions of the Argonaute protein and its role inside a post-transcriptional gene regulatory process called RNA interference. Furthermore, we will elaborate upon the structural characteristics of Argonaute and the differences between prokaryotes and eukaryotes.

1.1. GENERAL INTRODUCTION

1.1.1. CELLS AS A BUILDING BLOCK FOR LIFE

Life as we know it today is a very complex phenomenon. Whether it's animate or inanimate matter, we all originate from stardust, but within this framework, life takes a special place on the stage. Living "things" are most of the time highly structured and highly organized. Furthermore, the molecules that make up the building blocks are often very complex. And one of the defining characteristics of life would be that it takes energy from the environment to organize itself, grow itself and replicate itself. It was only until the 17th century that the discovery was made that all organisms are composed of similar building blocks of organisation. Robert Hooke observed plant cells through a microscope and since these walls were reminiscent of the small rooms that the monks inhabit (Figure 1.1), he coined the term *cells* to describe these compartmentalized units. In order for these cells to function and replicate as building blocks of life, they require not only nutrients and a suitable environment for them to thrive, but also a large degree of organization within the cell itself. One could say that these cells are in a way similar to a city, in that it needs energy to keep processes running and therefore it also needs fuel. This is generated by cellular compartments called mitochondria, which convert a proton gradient into the production of ATP. Infrastructure is also needed to transport certain machinery (the proteins) to their place of destination. Numerous other components are needed for damage repair, signalling, regulation of chemical pathways. They serve as building blocks for growth, maintenance of cellular structures, replication and also have many other functionalities that we are now only starting to discover.
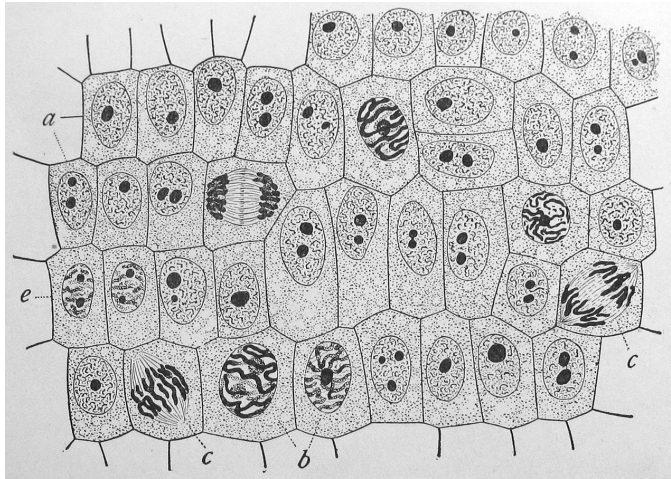
Figure 1.1: Onion root cells in different phases of the cell cycle (drawn by E.B. Wilson)

1.1.2. DNA & RNA

The blueprints of all these cellular components are stored as deoxyribonucleic acid (DNA) inside the cells. The DNA is usually found in its double stranded form, where
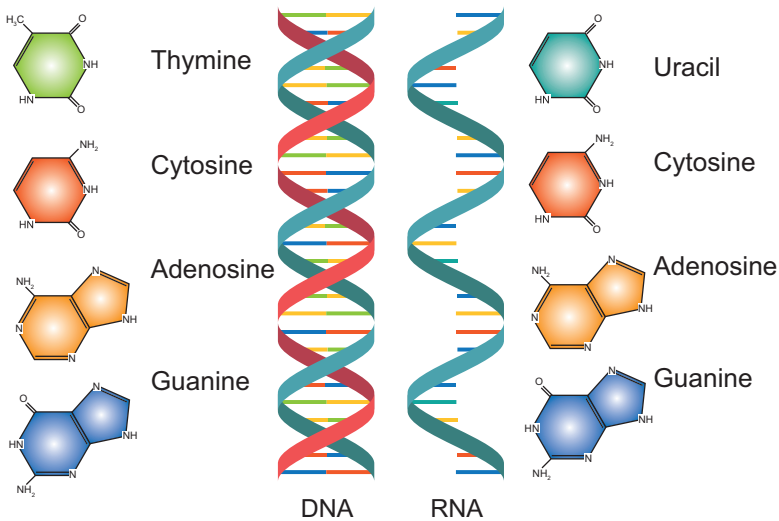
Figure 1.2: Molecular structure of DNA and RNA bases. Thymine can form two hydrogen bonds with Adenosine and guanine can form three hydrogen bonds with cytosine. For RNA, uracil replaces thymine and basepairs with adenosine.

one long polymer chain is paired up with a complementary chain, bound to each other through hydrogen bonds. In the absence of any external forces, this double polymer is found to be in a double helix configuration (Figure 1.2). The DNA consists of smaller subunits called nucleotides. These consists by themselves of a nucleobase and a phosphate backbone. The backbone for all of these are chemically identical. However, the nucleobases can exist in four different varieties: the purines adenine (A) and guanine (G), and the pyrimidines cytosine (C) and thymine (T). Through hydrogen bonds between A and T and between C and G, bases can be paired, which stabilizes the double helix conformation of the DNA. In essence, the bases here encode life's information and the DNA segments that contain the blueprints for proteins are called the genes. The genetic information here is read-out and transcribed into ribonucleic acid (RNA), which is single-stranded and has a chemically similar form to DNA. Thymine (T) is here replaced by uracil (U) but the sequence identity is furthermore identical to the DNA sequence from which it was read out from.

1.1.3. PROTEINS AND GENE REGULATION

The RNAs that are transcribed from DNA strands are used by certain proteins called ribosomes as a template to generate long amino acid chains. These long amino acids will fold in a specific configuration, becoming functional proteins. These are essential for the cell to sustain and replicate itself. Since the discovery that DNA is transcribed into RNA and translated into protein, the term "central dogma" became popular to describe a somewhat simplistic picture of the flow of genetic information inside the cell. In practice, there are many different proteins and some proteins are needed in higher quantities than others at different times. All cells therefore need some type

of regulation in order to control the expression of genetic information. Since DNA is transcribed into RNA, and RNA translated into protein, said regulation of gene expression could happen on all these levels. And indeed, proteins were found to repress or promote expression other proteins or directly act upon the DNA or mRNA to make the transcription site accessible or inaccessible. And while RNAs were mostly seen as an intermediate information carrier after Crick coined the term "central dogma" in 1958, it was discovered later that their role in gene regulation is also much more diverse than initially thought.

1.2. RNA interference

1.2.1. Discovery of RNAi

In the 1980s, small "antisense RNA" were found to repress a complementary RNA target in prokaryotes and eukaryotes. While the mechanism of repression was not understood, the possibility of selectively repressing genes for genetic screening proved to be an attractive option for researchers to identify the function of genes. In antisense studies, sense RNAs have been used as a negative control. Surprisingly perhaps then that Guo and Kemphues found that both sense and anti-sense strands gave the same result in C. Elegans, a nematode [1]. Craig Mello and colleagues coined this effect "RNA interference" to distinguish it from the antisense RNA silencing. By this point, interest in the mechanism of silencing was rekindled. Andrew Fire uncovered one aspect of the mystery by observing that double stranded RNA rather than single-stranded antisense RNA was setting in motion the degradation of target mRNA. Another interesting aspect of this RNAi that distinguished it from the antisense inhibition was the widespread effect of dsRNA: one molecule could trigger potentially the degradation of dozens of mRNAs. Craig Mello and Andrew Fire went to win the 2006 Nobel Prize in Physiology or Medicine for their ground-breaking work on RNAi.

1.2.2. Argonaute as a core protein in the RNAi pathway

RNA interference relies on the usage of small (~22 nt) non coding RNAs called microRNAs (miRNAs) or small-interfering RNAs (siRNAs) (Figure 1.3) [3]. These small RNAs are found in eukaryotic cells and direct post-transcriptional repression of mRNA through complementarity between guide and mRNA target. For miRNAs, these are first transcribed by RNA polymerase II as structured primary miRNA (pri-miRNA)s. These are subsequently processed by the microprocessor complex, consisting of Drosha and DGCR-8, into hairpin loops. Drosha recognizes the stem of the dsRNA region and DGCR-8 ensures cleavage at the correct position. The RNA that is processed is called pre-miRNA and consists of a hairpin of ~70 nts, a 5' phosphate and a 2-3 nt overhang at the 3' end. Exportin-5 recognizes this overhang and exports the pre-miRNA outside the nucleus into the cytoplasm. Dicer, an RNAse III enzyme, will bind the pre-miRNA through recognition of the 5' phosphate, 3' overhang and the loop structure. Subsequently, Dicer acts as a molecular ruler and dices the RNA construct at a pre-defined length [4], which also yields a miRNA duplex with a 2-nt 3' overhang on both sides. The duplex is then transferred to the RNA induced silencing complex (RISC), of which the Argonaute is the core protein. One of the strands is discarded (the
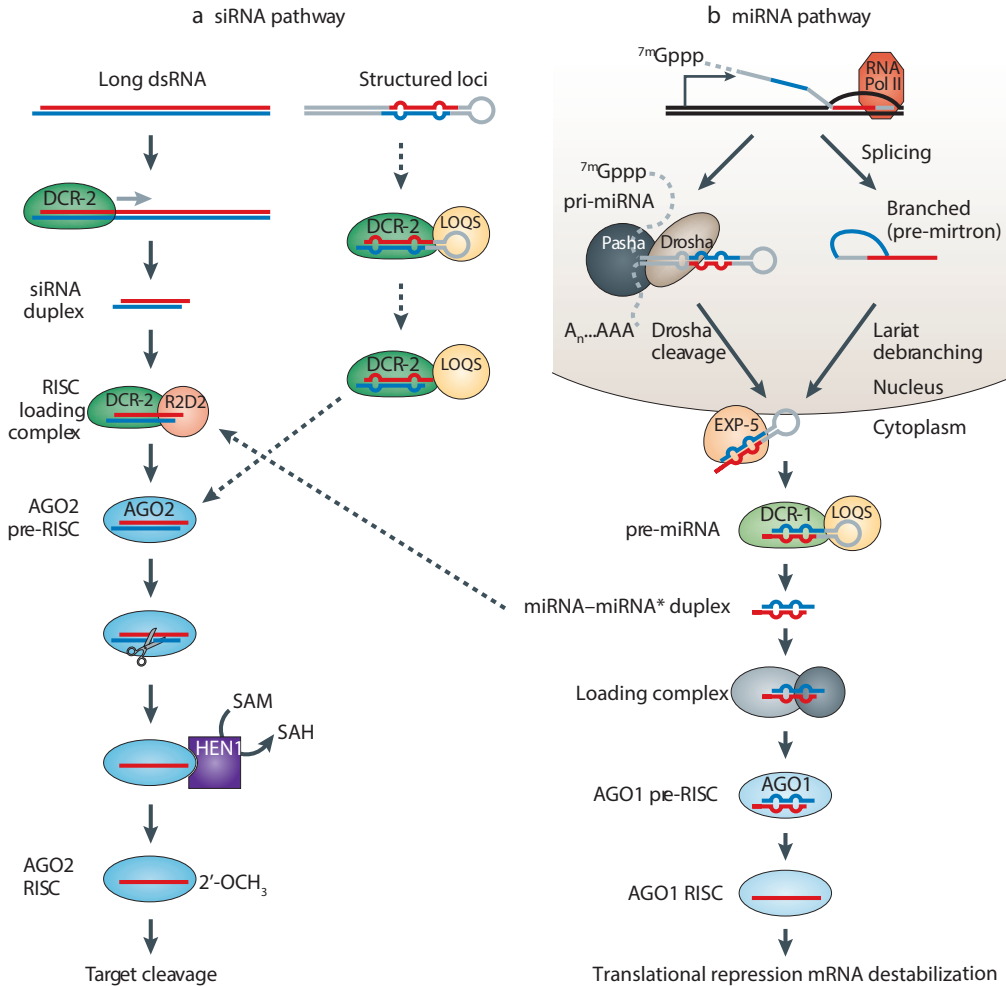
Figure 1.3: RNA interference pathway of siRNA and miRNA in *Drosophila*. (a) dsRNA precursors are processed by Dicer-2 (DCR-2) to generate siRNA duplexes containing guide and passenger strands. These are loaded by Dicer-2 and R2D2 into Argonaute2 (AGO2). The passenger strand is cleaved en ejected, leaving the guide strand inside AGO2 to find a complementary target on mRNA and cleave it. (b) miRNAs are encoded in the genome and are transcribed to yield a primary-miRNA (pri-miRNA) transcript, which is cleaved by Drosha to yield a short precursor miRNA (pre-miRNA). Pre-miRNAs are exported from the nucleus to the cytoplasm, where they are further processed by DCR-1 to generate a duplex containing a passenger strand and guide strand. Again, the passenger strand is ejected and binding of the RISC to the target through the guide strand induces translational repression of target mRNAs. Edited from Ghildiyal et al 2009, Nature [2]

passenger strand), whereas the other strand acts as the guide strand. The preference of guide is determined by the thermodynamic stability of the 5' end [4]. While the above is described as a canonical pathway of pre-processing of miRNA, there are also

1

non-canonical pathways which do not require a Dicer protein. An example of this would be miR-451: after pre-processing by Drosha, the looped structure directly loads in to hAgo2. The passenger strand is cleaved and the overhang of the guide strand is trimmed back to the correct size [5, 6]. In eukaryotes, mRNA silencing by miRNA occurs therefore by binding with the seed region [3], although slicing may also play a role: out of four Argonautes that occur in human cells, only human Argonaute 2 (hAgo2) is able to carry out endonuclease activity in the case of extensive base pairing [7]. Argonaute only requires a small part of the guide (the seed region) for canonical base pairing with the target [3]. This enables the miRNA not to only target the complementary sequence, but also many other sequences that contain a few internal mismatches. In vivo data has shown that miRNAs generally target a whole range of mRNA targets, often resulting in significant effects in gene expression for a multitude of genes [8]. Small-interfering RNAs are similar in size as miRNAs (20-25 nt) and fulfil a similar function as miRNA as post-transcriptional gene regulation. However, the origin of siRNAs is different from miRNAs as the precursor RNA is processed from long dsRNA. This can come from RNA-dependent RNA polymerases, but also invasive nucleic acids such as viruses, transposons and foreign genes. Processing occurs hereafter by Dicer, which acts again as a molecular ruler, dicing the dsRNA construct into shorter (20-25 bp) molecules. Downstream processing is same as for miRNA: short double stranded RNA is loaded into the Argonaute protein, one of the strands is ejected and the other strand is used to find a complementary mRNA target. Contrary to the binding of miRNA mediated Argonaute, the mRNA is now cut and degraded.

Small interfering RNA (siRNA) are similar to miRNA. However, a difference here is that the source of dsRNA is exogenous. A Dicer enzyme processes the dsRNA to the right length (20-24 bp) before one strand is cut and ejected, much like the miRNA biogenesis process. But where miRNA relies on stable base pairing with the target mRNA, siRNAs here direct the RISC to their target to cleave it. The resulting fragments are then further degraded by cellular exonucleases.

1.2.3. Guide architecture
The Argonaute has a bilobal structure and the guide strand is embedded inside a cleft between these lobes. Structural, biochemical and computational data suggest that within this cleft, the guide miRNA is divided into five functional domains: the 5' anchor, the seed, the mid region, the 3' supplementary region and the 3' end region (Figure 1.4A) [10]. Functionally speaking, the seed is arranged such that nucleotides 2-8 (from the 5' end) are arranged in a A-form like geometry and this region is responsible for stable target binding [3]. The pre-arranged helix allows Ago to bind without entropic costs to the target strand, providing stable binding once the complementary sequence is recognized. Computational analysis has shown that many functional/canonical miRNA targets only require seed pairing and matches in the supplementary region appear to as well for mRNA silencing [11, 12]. Additionally, once binding to the seed has occurred, this is followed by propagation of the base pairing to the 3' end of the guide [13].
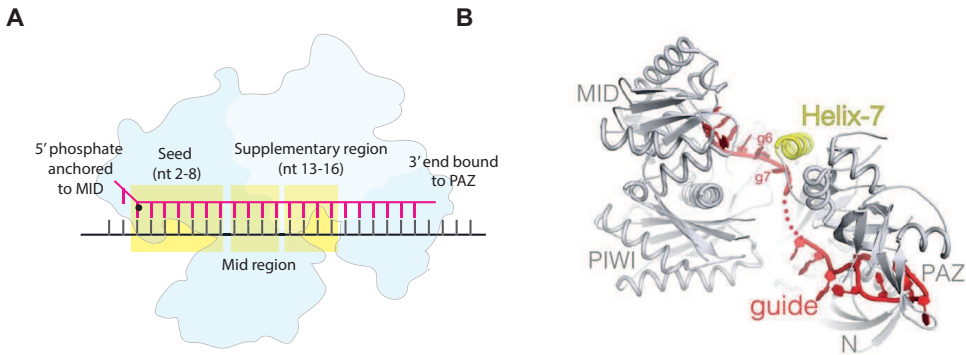
Figure 1.4: Left: Schematic drawing of miRNA, which is divided into five regions: the 5' end anchored in a MID binding pocket, the Seed region (nt 2-8), the Mid region where cleavage occurs on the opposing strand between nucleotide 10 and 11, the Supplementary region which supplements the binding of the seed domain. Lastly, the PAZ domain anchors the 3' end of the guide Right: Cartoon representation of hAgo2-miRNA complex based on the wild-type hAgo2 structure (PDB ID: 4OLA). The different domains of hAgo2 (grey) are indicated. The miRNA guide is indicated in red, while the helix-7 fragment is indicated in yellow. Panel B is taken from [9]

1.2.4. STRUCTURE OF ARGONAUTE PROTEINS

Argonaute proteins belong to the PIWI protein superfamily, which is characterized by a PIWI (P element-induced wimpy testis) domain. Human cells contain 8 PIWI-proteins, of which 4 belong to the Argonaute-clade and only human Argonaute 2 contains slicing activity. The first Argonautes were crystallized without a guide and it was revealed that the protein itself is a bilobed structure with a positively charged cleft in the middle for the guide strand to load [14–16] . One lobe contained the N, L1 and PAZ domain whereas the PIWI lobe contained the MID and PIWI domain. The first crystal structures that contained a guide showed that the guide contacted all the domains of the Argonaute [17, 18]. The MID domain contains a pocket that anchors the 5' end of the guide and some Agos recognize specifically the identity of the first nucleotide [19]. The PIWI domain contains the RNAse H domain for slicing activity. In the cleavage competent conformation the catalytic site consists of a DEDX motif. Lastly the PAZ domain that interacts with the 3' end of the guide. The 3' end of the guide is however not anchored [20–22]. Furthermore, the Argonaute was found to pre-arrange the guide nucleotides 2-10 into a helical arrangement, with the bases 2-6 pointing outward. This suggested that target recognition happened in a stepwise manner, firstly through the "seed" segment where the guide is pre-arranged in a helical manner. Ternary structures of Argonaute that contained both guide and target strand came after. TtAgo, crystallized with DNA guide and RNA target provided key insights in the catalytic activity. In case of extensive base pairing beyond the seed, the Argonaute would undergo a conformational change and this was most apparent in hAgo2. In case of partial base pairing of the seed, a helix-7 of the L2 domain would position itself such that a kink arises due to steric hindrance (Figure 1.4B). If base pairing extended beyond the seed region (2-8), helix-7 would reposition itself to avoid steric clashes. The kink then also relaxes and this in turn allows the seed of the guide to adopt an A-form helix when pairing the target. Furthermore the movements in the helix-7 domain could possibly induce changes in conformation for

the PAZ domain as well, the latter possibly facilitating easier target interactions in the supplemental domain.

In case of extensive base pairing, Ago can cleave the target strand between the 10th and the 11th nucleotide. A glutamic acid residue is directed away from the catalytic pocket if the Argonaute is in a cleavage-incompatible conformation, so that the catalytic site inside the PIWI domain only consist of a DDX triad (here X is often aspartic acid or histidine). Once the conformational change occurs, this glutamate finger is inserted into the catalytic pocket, turning the DDX triad into a DEDX motif.

1.2.5. PROKARYOTIC ARGONAUTES

Historically, the first Argonautes to be crystalized were prokaryotic Argonautes from *Pyrococcus furiosus* [14] and *Aquifex aeolicus* [18]. Similar to their eukaryotic counterparts, these contained also the PIWI, PAZ, MID and N-term domain. Initial attempts to crystallize the protein with RNA guides were not successful and only in 2008 was it possible to crystalize *Thermus thermophilus* Argonaute with a 21 nt DNA guide [23]. While the role of Argonaute in eukaryotes was well characterized, the role of prokaryotic Argonautes (pAgos) remained elusive for years. In recent years, it was implied that pAgos are involved in host defence rather than gene regulation, as some pAgos utilize DNA or RNA guides to target ssDNA instead of RNA [19, 24]. Furthermore, for *Thermus thermophilus* (TtAgo), *Pyrococcus furiosus* (PfAgo) and *Methanocaldococcus jannaschii* (MjAgo), reduced plasmid transformation efficiency and intracellular plasmid concentrations have been observed [25–27]. In vitro, some pAgos were found to target ssDNA in vitro, not dsDNA and currently it is not known how this process takes place in vivo as ssDNA is rarely present in vivo. In the case of thermophilic prokaryotes (such as TtAgo), it is thought that local melting of AT-rich regions would contribute to accessibility for effective cleavage of the target [19]. For TtAgo, a chopping mechanism was observed where new siDNA fragments are generated through endonuclease activity of apo-TtAgo [28]. Through this targeting of ssDNA, one may expect that potentially new genome editing tools are hidden inside a vast pool of uncharacterized pAgos [29]. Recently, a pAgo that uses hydroxylated guides rather than 5' phosphorylated guides have been uncovered [30] and also a modified version [31] would potentially enable one to target RNA instead of DNA, illustrating the diverse species and possibilities of Agos. An increasingly better understanding of how aforementioned Agos are able to find targets in a biological context would allow one to better predict the efficacy of targeting as well.

In 2016, an Argonaute from the species *Natronobacterium gregoryi* was proposed to use 24 nt long DNA to cut double stranded breaks [32]. Since it did not require a PAM sequence in contrast to CRISPR systems, it would make an attractive alternative for genome editing. However, many labs have tried to reproduce the result and none had succeeded, resulting in the end in an retraction of the article. Nevertheless, the controversy sparked the interest to search in the direction of Argonautes that could be successful in gene-editing at moderate temperatures and the search continues today.

1.3. THESIS OUTLINE

This thesis is a culmination of several studies on Argonaute proteins, ranging from a target search study of pAgo to a more practical application of it for super-resolution microscopy. In order to get one acquainted with target search and its relevance, **Chapter 2** gives a short overview of single-molecule target search studies and how they enable us to understand the strategies that proteins utilize to localize their cognate targets.

Before we continue with the pAgo target search study, we first have to understand the biological function of said protein in its physiological context. Therefore, in **Chapter 3**, we focus on the biological characterization of the Argonaute of the mesophilic bacterium *Clostridium butyricum* (also in the context as a potential system for gene-editing). The Argonaute protein is found to mainly associate with DNA guides in vivo, suggesting a host defence role similar to other DNA targeting prokaryotic Agos.

Subsequently, in **Chapter 4**, we experimentally attempt to characterize the target search mechanism of a Argonaute from the mesophilic bacterium *Clostridium butyricum*. We investigated the ability of the protein to deal with junctions and blockades that may occur *in vivo* and find that it utilizes lateral diffusion. However, we found that its mode lateral diffusion is different from what is encountered so far in literature. While the experimental model gives some answers that we have, it also brings many more new questions.

In **Chapter 5**, a theoretical framework is developed to characterize the biophysics of the experimental findings of Chapter 4 and we compare these results with a similar study for hAgo2. From kinetic modelling and simulations, meaningful physical parameters can be derived, such as the mean sliding length and the partitioning of different search modes. These answers may provide some insight in the search strategies of Ago and other nucleic acid guided searchers *in vivo*.

In **Chapter 6** we turn towards possible applications of Argonaute. Here, we aim to use Ago as a tool to speed up current super-resolution imaging techniques. By making use of the property of pre-arranging the guide strand, we can combine the currently existing DNA-PAINT technique with Argonaute, coined AGO-PAINT. We demonstrate that this technique is 10x faster than conventional DNA-PAINT in resolving nanostructures.

REFERENCES

[1] S. Guo and K. J. Kemphues, *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 (1995).

[2] M. Ghildiyal and P. D. Zamore, *Small silencing RNAs: an expanding universe,* Nature Reviews Genetics **10**, 94 (2009).

[3] D. P. Bartel, *MicroRNAs: Target Recognition and Regulatory Functions,* Cell **136**, 215 (2009), arXiv:0208024 [gr-qc] .

[4] L. F. R. Gebert and I. J. MacRae, *Regulation of microRNA function in animals,* Nature Reviews Molecular Cell Biology (2018), 10.1038/s41580-018-0045-7.

[5] S. Cheloufi, C. O. Dos Santos, M. M. W. Chong, and G. J. Hannon, *A dicer-independent miRNA biogenesis pathway that requires Ago catalysis.* Nature **465**, 584 (2010).

[6] G. R. Chen, H. Sive,   and D. P. Bartel, *A Seed Mismatch Enhances Argonaute2-Catalyzed Cleavage and Partially Rescues Severely Impaired Cleavage Found in Fish,* Molecular Cell **68**, 1134 (2017).

[7] J. Liu, M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J.-J. Song, S. M. Hammond, L. Joshua-Tor, and G. J. Hannon, *Argonaute2 Is the Catalytic Engine of Mammalian RNAi,* Science **305**, 1437 LP (2004).

[8] M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, *Widespread changes in protein synthesis induced by microRNAs,* Nature **455**, 58 (2008).

[9] S. M. Klum, S. D. Chandradoss, N. T. Schirle, C. Joo, and I. J. MacRae, *Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition,* The EMBO Journal , e201796474 (2017).

[10] L. M. Wee, C. F. Flores-Jasso, W. E. Salomon, and P. D. Zamore, *Argonaute divides Its RNA guide into domains with distinct functions and RNA-binding properties,* Cell **151**, 1055 (2012), arXiv:NIHMS150003 .

[11] V. Agarwal, G. W. Bell, J. W. Nam, and D. P. Bartel, *Predicting effective microRNA target sites in mammalian mRNAs,* eLife **4**, 1 (2015).

[12] A. Grimson, K. K. H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel, *MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing,* Molecular Cell **27**, 91 (2007).

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

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

[15] U. J. Rashid, D. Paterok, A. Koglin, H. Gohlke, J. Piehler, and J. C.-H. Chen, *Structure*

of Aquifex aeolicus argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. The Journal of biological chemistry **282**, 13824 (2007).

[16] J. S. Parker, S. M. Roe, and D. Barford, *Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity.* The EMBO journal **23**, 4727 (2004).

[17] J.-B. Ma, Y.-R. Yuan, G. Meister, Y. Pei, T. Tuschl, and D. J. Patel, *Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein.* Nature **434**, 666 (2005).

[18] Y.-R. Yuan, Y. Pei, J.-B. Ma, V. Kuryavyi, M. Zhadina, G. Meister, H.-Y. Chen, Z. Dauter, T. Tuschl, and D. J. Patel, *Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage,* Molecular Cell **19**, 405 (2005).

[19] D. C. Swarts, M. M. Jore, E. R. Westra, Y. Zhu, J. H. Janssen, A. P. Snijders, Y. Wang, D. J. Patel, J. Berenguer, S. J. J. Brouns, and J. van der Oost, *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature **507**, 258 (2014), arXiv:15334406 .

[20] Y. Wang, S. Juranek, H. Li, G. Sheng, T. Tuschl, and D. J. Patel, *Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex.* Nature **456**, 921 (2008), arXiv:NIHMS150003 .

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

[22] N. T. Schirle and I. J. MacRae, *The Crystal Structure of Human Argonaute2,* Science **336**, 1037 (2012).

[23] Y. Wang, G. Sheng, S. Juranek, T. Tuschl, and D. J. Patel, *Structure of the guide-strand-containing argonaute silencing complex.* Nature **456**, 209 (2008), arXiv:15334406 .

[24] I. Olovnikov, K. Chan, R. Sachidanandam, D. K. D. Newman, and A. A. A. Aravin, *Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA,* Molecular Cell **51**, 594 (2013), arXiv:NIHMS150003 .

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

[26] S. Willkomm, C. A. Oellig, A. Zander, T. Restle, R. Keegan, D. Grohmann, and S. Schneider, *Structural and mechanistic insights into the DNA-guided DNA endonuclease activity of an archaeal Argonaute,* Nature Microbiology **17035**, 1 (2017).

[27] A. Zander, S. Willkomm, S. Ofer, M. van Wolferen, L. Egert, S. Buchmeier, S. Stöckl, P. Tinnefeld, S. Schneider, A. Klingl, S.-V. Albers, F. Werner, and D. Grohmann, *Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus*

1

jannaschii, Nature Microbiology **2**, 17034 (2017).

[28] D. C. Swarts, J. van der Oost, and M. Jinek, *Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a,* Molecular Cell **66**, 221 (2017).

[29] J. W. Hegge, D. C. Swarts, and J. Van Der Oost, *Prokaryotic argonaute proteins: Novel genome-editing tools?* Nature Reviews Microbiology **16**, 5 (2018).

[30] E. Kaya, K. W. Doxzen, K. R. Knoll, R. C. Wilson, S. C. Strutt, P. J. Kranzusch, and J. A. Doudna, *A bacterial Argonaute with noncanonical guide RNA specificity,* Proceedings of the National Academy of Sciences **113**, 4057 (2016), arXiv:arXiv:1408.1149 .

[31] A. Lapinaite, J. A. Doudna, and J. H. D. Cate, *Programmable RNA recognition using a CRISPR-associated Argonaute,* Proceedings of the National Academy of Sciences **115**, 3368 (2018).

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

2

FACILITATED DIFFUSION OF ARGONAUTE-MEDIATED TARGET SEARCH

Argonaute (Ago) proteins are of key importance in many cellular processes. In eukaryotes, Ago suppresses translation of mRNA molecules through base pairing of microRNAs (miRNAs) with a complementary target on a mRNA sequence. In bacteria, Ago eliminates foreign DNA through base pairing of siDNA (small interfering DNA) with a target on a DNA sequence. Effective targeting activities of Ago require fast recognition of the cognate target sequence among numerous off-target sites. Other target search proteins such as transcription factors (TFs) are known to rely on facilitated diffusion for this goal, but it is undetermined to what extent these small nucleic acid-guided proteins utilize this mechanism. Here, we provide a brief introduction on the value of single molecule measurements to shed light on these questions. Furthermore we review recent single-molecule studies on Ago target search. We discuss the consequences of the recent findings on the search mechanism. Furthermore, we discuss the open standing research questions that need to be addressed for a complete picture of facilitated target search by small nucleic acids.

2.1. INTRODUCTION

Sequence-specific recognition of nucleic acids by proteins is of great importance in cellular development and gene regulation [1]. Since the discovery of regulatory proteins that target specific DNA sequences, it was questioned how these proteins are able to recognize DNA targets among numerous other sequences in a fast yet specific manner. The most intensively studied target search protein is the E. coli lac repressor. For this protein, an extraordinary high binding rate has been observed of 10^10 M^{-1} s^{-1}. This binding rate is a factor of 100 faster than what theory would predict for collisions driven by three dimensional (3D) diffusion (Einstein-Smoluchowski limit) [2, 3]. While this result puzzled many for years, Berg et al. devised a theoretical framework that introduced the facilitated diffusion mechanism: the protein diffuses in 3D before binding non-specifically to a DNA strand, after which the protein diffuses laterally in one dimension (1D) along the strand to find its target [4]. As the dimensionality of the problem has been partially reduced from three dimensions to one dimension, a higher association rate is expected through this mechanism. Since their seminal work, a new field surrounding target search has been developed. Interesting theoretical predictions have been obtained, such as the speed-stability paradox and the optimal partitioning of the different diffusional modes [5],[6]. Experimentally, restriction enzymes such as EcoRV have been studied through a biochemical assay, where the binding rate of the protein was measured as the function of the overall length of the DNA [7]. Non-specific sequences around a target site act here as an antenna, by providing an initial binding site for the protein from solution. After binding to the non-specific sequence, the protein would move laterally to the target site. Thereby, observation of an increased binding rate for a longer DNA construct corresponds well with the facilitated diffusion model. Biochemical studies rely heavily on theoretical assumptions as they can only measure average binding kinetics. Single-molecule methods provide an elegant solution as they visualize the different kinetic steps directly [8]. In recent years, single-molecule experimental groups have been able to observe with improving spatiotemporal resolution the target search mechanism of transcription factors [9], DNA repair proteins [10], zinc finger nucleases [11], TALEN [12] and the homologous recombination protein RecA [13, 14]. Interestingly, some proteins have been found to make use of facilitated diffusion, while others seem to make use of 3D diffusion only. What determines whether a protein moves laterally is not known at this point. Certain considerations such as the type of substrate, biological function and necessity of timely regulation are likely factors that have to be taken into account. More recently, target recognition for certain proteins was found to be mediated by small nucleic acid molecules. Small RNA molecules, which are loaded into proteins such as Argonaute (Ago) [15–17] and Clustered Regularly Interspaced Palindromic Repeats (CRISPR)-associated proteins [18, 19], are utilized as a guide for the recognition of complementary sequences. These nucleic acid-guided proteins are expected to use a different search mechanism than transcription factors or restriction enzymes, the latter of which rely on interrogation of existing features of the DNA grooves. This is because there are no constraints on the sequence identity of the guide and therefore any sequence can be targeted. As a result of that, the amount of possible targets is greatly expanded, and the task of finding the right target in a timely manner becomes

2
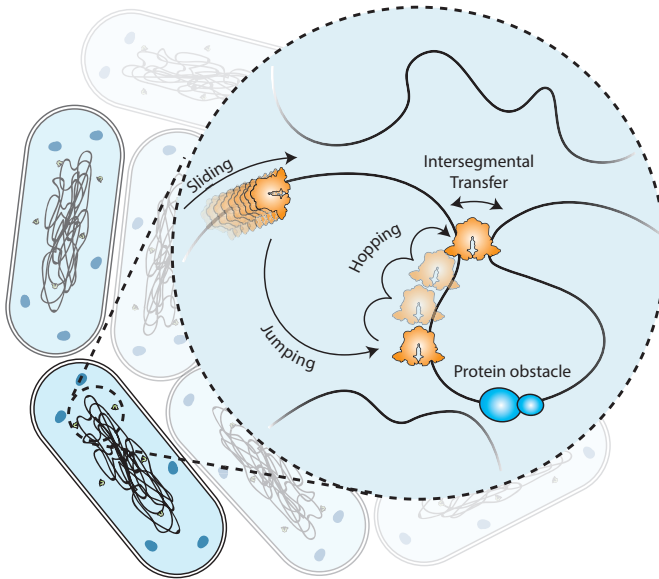
Figure 2.1: Facilitated diffusion consists of cycles of non-specific binding to the DNA strand after three dimensional (3D) search. This is followed by lateral diffusive motions along the DNA strand: (1) sliding, characterized by tight interactions with the strand. (2) Hopping consists of short dissociations away from the strand, however the movement is correlated along the strand. (3) Intersegmental transfer allows a protein with multiple binding sites to momentarily bind first to one than the other strand in a hand-over process. Lastly (4) jumping, or 3D-diffusion allows the protein to diffuse in an uncorrelated manner to new DNA sites

more complex. Here we summarize the findings of recent single-molecule studies on Ago-mediated target search in the context of the theoretical framework. We provide a perspective of facilitated diffusion in target search with respect to its biological function.

2.2. MECHANISMS OF FACILITATED DIFFUSION

In the context of facilitated diffusion, Berg et al have proposed several mechanisms which are conducive for target search [4]. Fundamentally, all modes of target search are driven by thermal energy, and therefore random motion is key in the searching process. The interactions between protein and non-specific nucleic acid sequences are governed through electrostatics, and the ions that screen the negatively charged nucleic acids (counterions) play an important role. For clarity's sake, here are the definitions that we employ in our review (Table 1):

2.3. SINGLE-MOLECULE TECHNIQUES

How are we able to distinguish between these processes that happen on a nanometre length scale at millisecond timescales? In single-molecule target search studies, DNA-protein interactions are visualized through DNA curtains [20], flow stretch assays, Förster Resonance Energy Transfer (FRET) [21], single cell imaging [22] and force spectroscopy methods such as magnetic tweezers, optical tweezers, and atomic force

Name	Description
Jumps / 3D-search	The protein explores the space in the cytosol or solution through 3D Brownian motion. Non-specific binding is followed by the protein dissociating. Subsequent binding to other sites occurs in an uncorrelated manner.
Sliding	The protein binds non-specifically to DNA/RNA and undergoes a movement which is characterized by a tight interaction with the nucleic acid molecule. The protein stays associated to the substrate at all times. In the words of Berg, no net displacement of counterions takes place here and therefore the time spent on non-specific substrate is not affected by a change in the ionic concentration in the surroundings.
Hopping	This mode is characterized by micro-dissociations from the DNA/RNA strand. While the method of diffusion is similar to 3D search, the difference is that the movements are correlated along the contour of the strand. Contrary to sliding, not every base is scanned in the effectively covered distance. Since the protein dissociates momentarily from the strand, condensation of counterions is allowed to occur. Hence the time spent on non-specific DNA is expected to decrease by increased ionic strength.
Intersegmental transfer	The protein with multiple binding sites is bound to one strand. In a hand-over process, the protein can be momentarily bound to two strands through interaction with its binding site, after which it moves to the other strand.

Table 2.1: Description of different mechanisms of facilitated diffusion

spectroscopy [23, 24]. These techniques provide great information on the kinetics of individual molecules. However, technical limitations could withhold one from observing the full dynamics. For example, many camera-based single molecule fluorescence studies rely on the transmission of a fluorescence signal through the imaging system onto the pixels of the camera. The point-like fluorescent signals are mapped on the pixels of the camera as a spread-out image called the point spread function (PSF) of ~200 nm width [25], giving rise to an uncertainty in position. Since the shape of the PSF is known, one can still estimate the true position of the particle. But the finite amount of photons emitted from the fluorescent object - which is limited by photobleaching - determines the accuracy in position: the more photons are collected from a static source, the more accurate the estimation of its true position will be. As the protein of interest will not be static during target search but will undergo many fast movements, the camera needs to be fast enough to capture these dynamics. Most camera-based approaches collect fluorescent light at an acquisition frequency of 10-100 ms time resolution. Therefore this frame rate and the number of photons collected during each time bin defines the time resolution at which one can probe protein DNA interactions. For a more detailed introduction on this we refer the reader to the many reviews available in the field [21, 25]. The method of flow stretch assays, namely DNA curtains, provides

the most intuitive visualization of target search. By labelling the protein of interest with a fluorescent probe, one is able to track the position of the protein on the DNA strand in presence of roadblocks and co-factors [10, 20] [10] (Figure 2.2A).
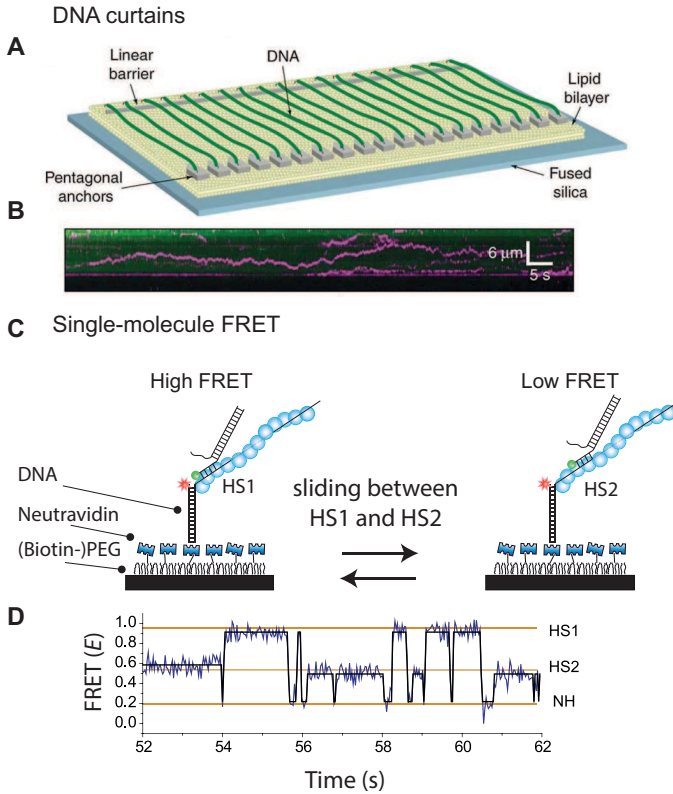
Figure 2.2: (A) Diagram of the nanofabricated DNA curtain device that contains a barrier against lipid diffusion for stretching the DNA. Here, flow is used to stretch the DNA strand. On the other side a pentagonal structure is used to anchor the other end of the strand [10]. As a result of this, hundreds of DNA strands are aligned in parallel and can be imaged simultaneously. (B) The DNA curtains are visualized through YOYO-1 dye staining. A fluorescent probe (pink) is attached to the protein of interest and the position (vertical axis) of said protein is tracked in time (horizontal axis). By imaging both the DNA strand and the fluorescent probe, one can visualize how it travels along the DNA strand [10]. (C) Single-molecule FRET assay showing a RecA filament (blue) containing two homology sites. Recognition of homology site 1 (HS1) or homology site 2 (HS2) results in the appearance of a high FRET state and an intermediate FRET state respectively [14]. (D) Single-molecule time trace showing FRET for an immobilized ssDNA with two identical homology sequences HS1 and HS2. Docking of double stranded DNA at a location outside a FRET sensitive regime results in a low FRET (NH) state [14].

From the position of the protein at different timepoints, one can directly observe whether lateral diffusion takes place (Figure 2.2B), and if so, derive the effective diffusion coefficient. The distinction between sliding and hopping is made by changing the strength of the ionic solution. As stated in Table 1, during sliding no net displacement of counterions takes place, and therefore a change in diffusion coefficient is not expected. In contrast, during hopping, proteins are expected to

2

diffuse faster along the strand at a higher salt concentration, since less time is spent on each non-specific binding site on a strand. This technique has been able to uncover the nature of target search for a wide variety of proteins [10, 14, 26]. It provides for a great in-depth characterization of target search on a large length scale interrogation of sequences. However, due to the large size of the PSF and the thermal fluctuations affecting the position of the DNA strands, observation of proteins on DNA is generally limited to a resolution of ~250 bp. As it cannot be ruled out that lateral movements take place within their observations [14, 27], other high resolution techniques should be used complementarily. Single-molecule Förster Resonance Energy Transfer (smFRET) provides high spatiotemporal resolution. FRET is an energy transfer process between two fluorophores, where due to dipole-dipole interactions, energy from a donor fluorophore is transferred to an acceptor fluorophore if they are within a few nanometres. The FRET efficiency E, (the ratio IA/(IA+ID), where IA is the intensity of acceptor signals and ID is that of donor signals) is given by $E = 1/(1 + (R/R_0)^6$, where R_0 is the characteristic distance of the dye pair (the Förster radius) and is typically a few nanometres. A change of dye-pair distance results in a measurable change in the ratio of intensities of donor and acceptor fluorophore. A seminal smFRET study that investigated target search was on the RecA protein [13], where two identical homology sites were placed on a DNA construct (Figure 2.2C). The design was such that binding to one homology site resulted in a higher FRET efficiency compared to binding the other homology site. The rationale behind it was that while FRET provides high spatiotemporal resolution (~nm at 0.1 s timescale), the dynamics of both 3D and lateral diffusion were expected to occur on a much faster timescale (millisecond timescale). Through the use of local energy traps, one could momentarily trap the RecA nucleofilament at the sites and characterize the nature of their interactions (Figure 2.2C bottom). While most in vitro techniques allow one to probe the mechanisms, it is important to know what the effect would be of the presence of cellular proteins and the crowding environment in physiological conditions [28]. Live cell imaging allows one to study single-molecule facilitated diffusion inside a living cell. The first study of single-molecule live cell imaging was performed on a transcription factor, the lac repressor, LacI, which acts on the operator of lac genes [9]. Binding of LacI to the operator site prevents expression of the lac operon genes that metabolize lactose. However, by adding Isopropyl-β-D-1-thiogalactopyranoside (IPTG), a molecular reagent that binds to the lac repressor, one can prevent LacI from binding to the operator site. Removing IPTG from the solution allows LacI to bind once more to the operator site. The authors of this study used fluorescently labelled LacI to study target search. In the presence of IPTG, the LacI repressor cannot bind to the target site (Figure 2.2D). By measuring the association rate after removing the IPTG inducer, they could measure the average time of a single LacI molecule to find its target. The unbound molecules diffuse too fast to be recorded while bound molecules a stable signal in time and space (Figure 2.2E). When two targets are placed close to each other, the two targets will appear as one target if the distance between two targets is smaller than the mean sliding distance (Figure 2.2F). At distances longer than ~50 bp, the targets were perceived as independent targets, but at distances shorter than ~50 bp, the association rate was comparable to single target association rate kinetics (Figure 2.2G). Additionally, to find

A Live cell methods

Repressed (-IPTG)   Induced (+IPTG)

B

C

203 bp

65 bp

25 bp

D
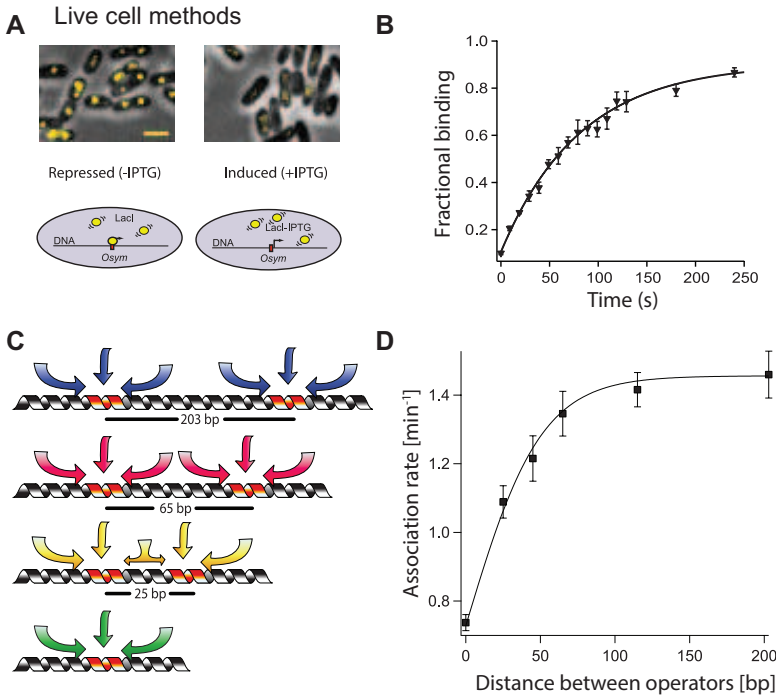
Figure 2.3: (E) Single operator binding assay used by the Elf lab [9]. (Top) Overlays of E. coli cells in phase contrast and with fluorescently labeled LacI (yellow). In absence of IPTG (left), LacI is able to bind to the single operator site LacOsym, resulting in a diffraction limited spot. In presence of IPTG (300 μM) (right), LacI is unable to bind due to the competition with IPTG, and diffuses too rapidly resulting in diffusional smears [9]. (F) Graph plotting the fraction of stable LacI binding vs the time after removal of the IPTG [9]. (G) The mean sliding length is determined by placing two identical targets in varying distances. If the mean sliding length of said targets overlaps, the LacI protein will effectively only sense one target, resulting in a decreased association rate [9]. (H) Rate of binding plotted against the inter-target distance [9].

out whether LacI was able to bypass protein roadblocks, a TetR protein was bound next to one of the targets. In presence of TetR, the association rate was significantly affected indicating that LacI is not able to bypass roadblocks through sliding only. As a whole, the aforementioned studies show that single-molecule methods provide understanding in the molecular processes that govern facilitated diffusion. Similar methods may provide key insights for Ago-mediated target search.

2.4. TWO-MODE SEARCH OF AGO

In order to find a target in a timely manner, a protein needs to bind non-specifically to nucleic acids, search rapidly for the associated target, and bind strongly to a target site. The search of targets needs to happen in both a fast and specific manner, yet this cannot happen at the same time, since specificity imposes as a rule that the energy barriers become too high for lateral diffusion [5]. The paradox is solved by assuming that the protein has two states of binding: a search mode, in which the energy barriers

2

it encounters are minimal, enabling smooth lateral diffusion, and a recognition mode, which is characterized by high affinity and slow diffusion (Figure 2.4A). The encountered energy landscape in the recognition mode ($E_{recognition}$ in Figure 2.4B) is on average higher than the mean energy level landscape in the search mode (E_{search}), so that the protein spends more time in the search mode than in the recognition mode. A key idea here is that the energy landscape the protein encounters during the search mode is well-correlated with the energy landscape in the recognition mode and that an energy gap exists between the two modes. So the deep minimum in the recognition mode would correspond to a more shallow minimum in the search mode. When the protein is trapped inside one of these energy levels during the search mode, it will likely transition into a recognition mode. Effectively, this results in a pre-selection at the minima of the binding landscape, since here it's more likely that a transition will happen from the search mode to the recognition mode. Even if the conformational transition rate from the search mode to the recognition mode is low [29], a gain in search speed is still predicted [6].

2.4.1. EXPERIMENTAL EVIDENCE OF TWO-STATE SEARCH

Theoretically, it was posited that Ago target search may be mediated by such binding modes [31] and there is structural evidence to support this. Since the seed of the guide strand is pre-arranged in a helical manner (Figure 2.3A), this suggests that initial target recognition and perhaps also initial target search commences at this region. From the crystal structure of hAgo2, it has been posited that guide nucleotide 2-5 of hAgo2 is used for initial recognition of target sites [32]. A kink introduced between g6-g7 breaks the A-form structure of the helix, and this is caused by insertion of the residues Ile-365 and Met-364 of α-helix-7 between the bases of 6 and 7 (Figure 2.4C) [32]. Base pairing beyond nucleotide 7 requires a conformational change of helix-7 to accommodate. At the same time, this conformational change stabilizes the base pairing of nucleotides 6 and 7 of the guide (Figure 2.4D) after which base pairing of additional nucleotides can also take place [30]. Fluorescence single-molecule in vitro studies have given further proof of this two-state binding mode. In all the single molecule studies mentioned here, the target strand with fluorophore is immobilized on the surface. A second fluorophore is then attached to the miRNA guide which is loaded into the core-RISC. Through single-molecule FRET or colocalization of both dyes, the binding and unbinding rate can be studied for various base pairing sequences (Figure 2.5). In the case of smFRET, either a donor fluorophore or an acceptor fluorophore can be immobilized on the surface [33, 34] [67,68] (Figure 2.5A & 2.5D). The guide strand that contains the complementary dye required for FRET is loaded into Ago. High FRET indicates binding of the Ago-guide complex to the target site, since the dyes must be in close proximity for energy transfer to occur. From the length of the high FRET signal one can obtain the dissociation rate (Figure 2.5B). Likewise, by measuring the time between introducing Ago-guide complex to the chamber and first binding to a target, one can obtain the binding rate (Figure 2.5C). Additionally, in the case of Salomon's assay (Figure 2.5G), an RNA target was tagged with 17 dyes attached to the 3' end, so that cleavage events can be readily distinguished from photobleaching [35]. From single molecule fluorescence assays, it was found that Ago accelerated the binding rate greatly, compared to binding with guide RNA only, for
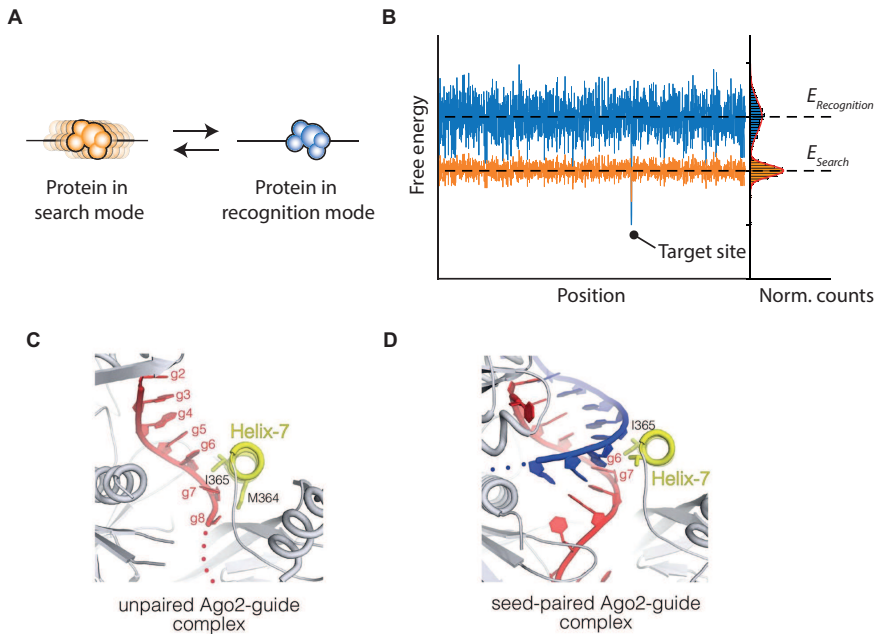
Figure 2.4: (A) The speed stability paradox posited by Slutsky et al. In the search mode (orange), the protein is able to diffuse laterally without encountering significant energy barriers. Once it encounters a potential target site (indicated by the deeper energy level in the binding energy landscape right), it may switch to a recognition mode (blue). In this mode the specificity of the protein is increased and the protein will not diffuse. (B) The energy landscape as encountered by a protein in the search mode (orange) and a protein in the recognition mode (blue). In the search mode, the landscape that the protein encounters is shallow and the variance in energy levels is small. The deeper energy levels in the recognition mode prevent the protein from diffusing laterally. At a potential target site, it's more likely for the protein to switch from a search mode to a recognition mode, since the energy level of the former is higher than the latter. (C) Close-up view of the seed region shows the pre-formed helix of nt 2-6. Helix-7 disrupts base stacking by intercalating itself between g6 and g7 [30]. (PDB ID: 4OLA) (D) Close-up view of the seed region in the event of fully base-paired seed. Helix-7 undergoes a conformational change here, docking into the minor groove of the seed-paired complex [66]. (PDB ID: 4W5O) Permission has been obtained for the above figures. Copyright 2017 John Wiley and Sons.

both hAgo2 and mouse Ago2 (Figure 2.5E and 2.5H) . The rationale is that pre-arranging the seed would result in a higher probability of successful binding to the target strand, hence effectively increasing the binding rate of the complex. Dinucleotide mismatches in the seed were found to be detrimental to the binding rate for both mouse Ago2 and hAgo2 (Figure 2.5F and 2.5I) [34–36]. Disruption of base pairing in the seed would also often result in quick dissociation of mouse RISC [35], showing that seed-pairing is essential for target recognition. Furthermore, it was shown that hAgo2 utilizes a part of the seed (nt 2-4) for initial target search, since shrinking the seed pairing from 2-8 to 2-4 did not change the binding rate (Figure 2.5C) [33]. Varying the seed pairing from 2-4 to 5-7 did reduce the binding rate significantly [33, 35], indicating that the 2-4 seed motif is essential for initial recognition.

2

2.4.2. Transition to recognition mode

Beyond initial recognition, the transition from an initial search mode to a recognition mode of hAgo2 was hinted at through single-molecule FRET. As stated before, the crystal structure of hAgo2 suggests that a conformational change of helix-7 is required for stable base pairing beyond N=6. In single-molecule FRET, through extending complementarity in base pairing from nucleotide 2-4 to 2-19, a significant increase in binding time was observed between N=6 and N=7 (Figure 2.5B). This suggests that a conformational change took place that the strengthened seed-target interactions, as is required for the speed-stability paradox for fast and specific targeting [33]. Furthermore, mutants, in which either the helix-7 is lacking or where two helix-7 residues are mutated, have shown to respectively decrease both the on-rate and the off-rate, indicating that it fulfils additional functionality by rapidly dismissing off-targets while the search itself can be accelerated by pre paying the entropic costs of arranging the guide in a helical manner [30]. In short, the first single-molecule fluorescence studies provided key insights in recognition through visualization of transient kinetics that bulk and static methods could not provide.

2.5. Target search of Ago

2.5.1. Hidden rapid dynamics in target search

How does one envision target search of Ago to take place? The minimal RISC complex would be expected to bind non-specifically to a random position on the target strand, before moving to the target site. This would result in a gradual change in the FRET value. However, the single molecule data of the aforementioned in vitro assays contained only stable traces. The absence of such signature indicates that either the complex binds directly to target from solution, or more likely, that the dynamics take place on a timescale that occur much faster than the acquisition time of 100 ms. In order to characterize the dynamics, one would need to resort to stronger energy traps. The first study that investigated the nature of Ago target search was inspired by the RecA target search assay [13], where two identical strong binding sites were placed on an RNA construct (Figure 2.6A)[33]. Similarly to the RecA assay, binding to one sub-seed target site was designed to result in a higher FRET efficiency compared to binding to the other site. If only one target site was present on the RNA strand, only one FRET state could be observed. If two targets are present on one strand, one did not only observe the addition of binding signatures with a lower FRET state, but also a shuttling signature: the transition from one binding site to the other without interruption (Figure 2.6B) [33].

2.6. Is Ago search time optimal?

The first fluorescence studies have uncovered the mechanism of interactions of the Ago-guide complex with a target strand. However, quantitative understanding is still lacking at this point. One of the outstanding questions is how the gliding and intersegmental jumps are temporally divided for Ago to maximize its target search speed. While theoretical studies have predicted that the optimal search time for a target consists of equal time spent in 1D and 3D diffusion [5], this has proven to be not always the case. Some proteins have been found to have different distributions in vivo [37–39], where a

tenfold or hundredfold more time is spent being bound to nucleic acid strands rather than being diffusing in solution. It will be of interest to see why this is the case for some proteins and whether Ago is one of them. We speculate that, when the tandem target assay is used, the mean first passage time between two targets could be used to infer to what extend Ago partitions its search process into lateral diffusion and intersegmental jumps. Next, it is not always well understood how redundancy and efficiency in target recognition are coupled to each other. Intrinsically, lateral target search is redundant by its very diffusive nature. As an example, the human oxoguanine DNA glycosylase 1 was found to have an effective diffusion coefficient of 5×10^6 bp^2/s [26]. The barrier in energy landscape along the DNA sequence was on the order of k_BT, indicating that lateral diffusion is not limited by the roughness [5]. Persistent contact with DNA indicates that DNA segments are scanned multiple times resulting in an inefficient mechanism. However, the redundancy in target search may be to compensate for the inefficiency of target recognition (i.e. multiple attempts are necessary to recognize the target, but with sliding/hopping the target is bypassed multiple times due to 1D diffusion.). It has been observed that the loosely interacting mode of Ago-guide complex with the nucleic acid strand could potentially allow the protein to skip over bases, implying that multiple scanning attempts might be needed for the protein to recognize a cognate target site. How physiologically relevant is the loosely associated searching mode of Ago? The ability of Ago proteins to bypass secondary structures without any impedance suggests that the search behaviour itself is robust. Secondary structures in mRNA occur frequently in vivo [40–43], providing many functional elements essential for regulation of various post-transcriptional mechanisms [44–47]. In the 3' UTR, where many miRNA target sites are located [48], the RNA structure is more structured than in coding regions [41]. Unimpeded target search allows Ago to efficiently scan for target sites without it being trapped between dsRNA segments. Future measurements in vivo should point out to what extend these weak interactions help Ago speed up its searching process.

2.7. CRISPR ASSOCIATED PROTEINS USE A DIFFERENT STRATEGY

To what extent would the target search mechanism of Ago be conserved among other small RNA-guided systems? The most widely known class of nucleic acid-guided endonucleases are derived from the CRISPR immunity system. As an immunity system against genetic elements from bacteriophages and plasmids, prokaryotes insert short fragments of the foreign DNA into their own genome, the so-called CRISPR array [49, 50]. Spacers from this array are transcribed and processed into short RNA fragments, termed CRISPR RNAs (crRNAs). CRISPR-associated (Cas) proteins utilize the crRNAs to target foreign complementary DNA targets, called protospacers, after which cleavage occurs, either by recruitment of other Cas proteins or by direct slicing by the targeting proteins themselves [50]. The protospacer sequences targeted by Cas proteins are flanked by a short sequence motif, referred to as the protospacer adjacent motif (PAM) [50]. Cas proteins use these motifs to distinguish foreign DNA from endogenous DNA.
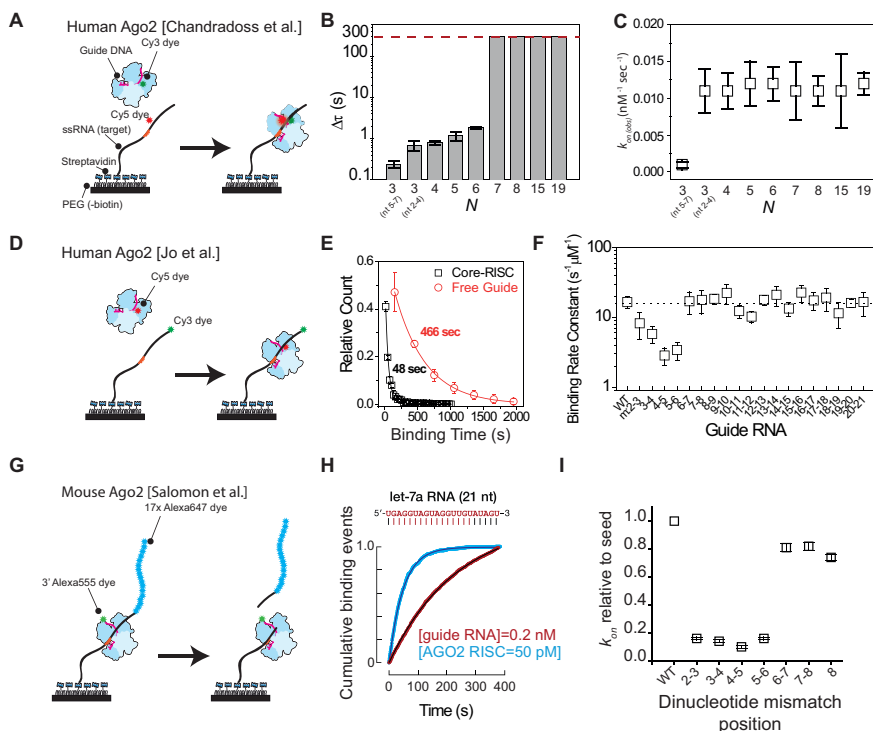
Figure 2.5: (A) Schematic drawing of the experimental assay used by Chandradoss et al [33]. Target RNA is immobilized on the surface through biotin-streptavidin conjugation. The target RNA is labeled with an acceptor-dye (Cy5) while the donor dye (Cy3) is located on the miRNA guide. In absence of hAgo2-RISC binding to the strand, no signal is observed. Once it binds to the minimal target motif, the proximity of donor and acceptor induces Forster resonance energy transfer (FRET), resulting in a high acceptor signal. The duration of the high acceptor signal can be used to estimate the dwell time Δt. (B) The dwelltime Δt plotted versus the number N of seed-paired bases. The dashed line indicates the upper limit of dwell time estimation [33]. (C) The binding rate plotted for various values of N of base pairs [33]. (D) Schematic drawing of the single-molecule assay by Jo et al. where through FRET of the Cy5 and the Cy3 dye the binding can be ascertained [34]. (E) Dwell time distribution of core-RISC (black squares) and free let7 miRNA (red circles), fitted with a single-exponential decay[34]. (F) Binding rate plotted versus dinucleotide mismatched guide RNAs [34]. (G) Schematic assay for Salomon et al [35]. 17 Alexa647 dyes are attached to the 5' end of the target RNA to distinguish cleavage from photobleaching. (H) Comparison of target binding rates (k_{on}) for 21 nt sequences for let-7a RNA and miR-21 RNA. Inset shows a representative intensity trace [69]. (I) Comparison of target binding rate for let-7a sequences with complete seed-matched pairing or seed-matched pairing bearing dinucleotide mismatches [69].

2.7.1. DIFFERENCES BETWEEN CAS PROTEINS AND AGO

The first single-molecule studies on Cas proteins, such as the E. coli Cascade complex [51] and Streptococcus pyogenes Cas9 (spCas9) [52], suggest that the contribution of lateral diffusion of Cas proteins is much smaller than what has been observed of restriction enzymes, repair enzymes and other proteins that interact with double stranded DNA [51–53]. This difference is not unexpected since the price to be paid for the flexibility of programming the guide, is that the double stranded region of DNA needs to be opened up by the Cas proteins for interrogation, which is energetically costly. Eukaryotic Ago proteins do not suffer from this energetic cost as they target single stranded RNA. In this sense, prokaryotic Agos seem to have more in common with CRISPR proteins, as they are thought to be involved as well in host defence and require high fidelity recognition of dsDNA [54]. However, the prokaryotic Agos studied so far seem to interact only with single stranded DNA, and it is currently unknown how they access double stranded binding sites in vivo. In some prokaryotes, the genes of pAgos seem to cluster with genes encoding for nucleases and helicases [55] and it has been posited that pAgo-associated helicases could potentially assist the Ago to unwind the double stranded segments, thereby allowing the endonuclease to access the single stranded DNA molecule for interrogation. At the same time, recent findings indicate that the presence of ssDNA viruses are abundant in certain environments. Here, targeting ssDNA through facilitated diffusion would be highly beneficial for pAgo.

2.7.2. REDUCTION IN SEARCH COMPLEXITY HELPS FOR 3D TARGET SEARCH

In contrast to Ago target search, the target search of Cas protein such as Cas9 requires melting of DNA. It would be unfeasible for Cas proteins to randomly melt DNA sites all over the genome for interrogation. The aforementioned Cas-proteins have been found to interact longer with PAM sequences compared to non-specific sequences, which allows the Cas protein to have enough time to interrogate the sequence [51, 52]. Thus, the PAM sequence recognition, which is critically important for the distinction between self and non-self targeting, also serves as an extra pre-selection step for target recognition. Dividing the recognition process in two or multiple step manner not only circumvents the speed-stability problem in facilitated diffusion, but also reduces the time needed to find the target through 3D diffusion only. Here the term search complexity, a concept, which was introduced for RecA homology search [14], may account for the observed dominance of 3D diffusion for some small RNA-guided proteins [51, 52, 56]. The rationale is that short sequences tend to occur more often inside the genome while longer sequences will have fewer exact matches. A protein relying on initial recognition of a short sequence will only have to search through a small part of the genome, and the rest would be ignored. For Cas proteins, the PAM sequence acts as a pre-selection mechanism filtering out the other sequences that are of no interest. For example, for a three-nucleotide sequence, such as a PAM site, roughly 10% of the genome would need to be interrogated, spending virtually no time on the remaining 90% sequences. In contrast, in the extreme case of a one-step recognition process with the full target of 22 nt, the target search would be inefficient. The probability of 22-nucleotide target occurring randomly would be extremely small:

only once every $1.7 * 10^{13}$ base pairs. This is obviously advantageous in terms of uniqueness. However, the protein has to reject almost every site after opening up the strand for interrogation, and the time it takes to unbind from intermediate base pairing would make this strategy unfeasible [57]. This is again reminiscent of the speed-stability paradox posited by Mirny. The more specific the search will be, the more time the protein will spend on an off-target site. So, the key assumption of reduction in search complexity is that search complexity is proportional to the time it takes for the target to be localized. This assumes that differences in search complexity are coupled to differences in kinetics between short and long sequence base pairing between guide and target. In order for this concept to work, dissociation rates of short pairing sequences have to be substantially higher than those of long pairing sequences, which has been observed for Ago and Cas9 proteins [33, 57, 58].

2.8. STRUCTURAL CONSIDERATIONS OF DIFFERENT TARGET SEARCH PROTEINS

What are the common structural features that determine whether a protein will slide or hop during lateral diffusion? It is assumed that some proteins maintain tight contact with the DNA/RNA substrate relying on their structural features whereas other proteins use only transient interactions. For example, the MutS repair enzyme consists of a clamp-like structure [59]. Once it is in a closed clamp-like conformation, this allows the protein to slide with high processivity along the dsDNA strand. Similarly, the Lac repressor is able to interrogate the grooves of the dsDNA thoroughly, through its structural form that facilitates interaction with the dsDNA groove [60]. It may be therefore unsurprising that some nucleic acid guided endonucleases like Ago or Cas do not use sliding, since their structure does not facilitate these interactions with the DNA substrate in the first place. Interestingly, the degree of lateral diffusion seems to vary from organism to organism even though they are orthologs. Cascade from T. fusca shows lateral diffusive behaviour [61] whereas E. Coli Cascade does not [51] since the latter complex lacks a positive patch in the Cse1 subunit. The question remains why this would be conserved in T. fusca but not in E. coli if they are functionally similar and require the same rapid response in case of a foreign genomic invasion. Is lateral diffusion always required for a protein to find its target in a timely manner? Remarkably, Cas12a shows lateral diffusion over microns at a remarkable coefficient of a micron per second, while it does not enclose the double stranded DNA [62]. Crystal structures may give us hints towards understanding the nature of lateral diffusion. Single molecule kinetics in conjunction with molecular dynamics simulations may give us an idea on which common features are required.

2.9. FUTURE PERSPECTIVES ON AGO TARGET SEARCH

While much has been uncovered in the last decades about small RNA/DNA-mediated target search, there are still many long-standing questions of interest. The molecular nature of targeting through small RNAs has been uncovered to some extent, but the mechanistic picture is still incomplete. For example, while it is now known that Ago
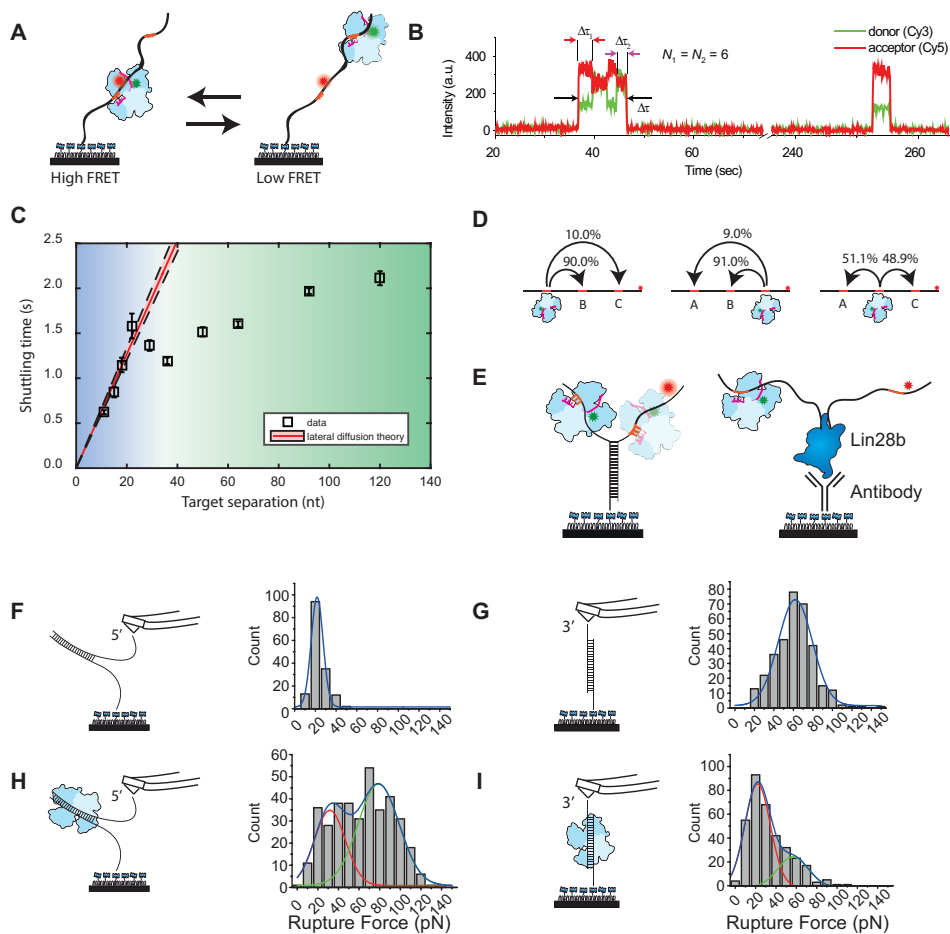
Figure 2.6: Figure 6: Single-molecule experiments on Ago target search (A) Tandem target assay used by Chandradoss et al [33]. The RNA target contains two identical targets, where the Cy5 dye is placed closer to the bottom target. During the experiments, hAgo2 shuttles between the two target sites. (B) A typical shuttling trace for N=6 (nt 2-7). Binding to one target brings the donor dye of the guide closer to the acceptor dye, resulting in a higher FRET value. This way, binding to one target can be readily distinguished from binding to the distal target [33]. (C) Shuttling time of CbAgo plotted versus the distance between two identical targets. The red line indicates a lateral diffusion fit derived from a minimal kinetic model. The data diverges from the kinetic theory beyond the blue region and follows a different trend in the green region [72]. (D) Triple target assay, used to visualize whether CbAgo skips over the middle target when translocating from site A to site C. Percentages indicate the relative amount of transitions from an initial state to a final state versus the total number of transitions [72]. (E) Left: a Y-fork construct that prevents CbAgo to slide from the target on one strand to the other. Right: Lin28 is immobilized as a protein blockade [72]. (F) DNA unzipping with AFM in absence of TtAgo. The AFM was retracted with a velocity of 200 nm/s. The rupture force is plotted with a Gaussian probability distribution fitted on top of it (blue) centered at 21.7 pN [63]. (G) DNA shearing with AFM in absence of TtAgo, with a fit centered at 62.0 pN [63]. (H) DNA unzipping with AFM in presence of TtAgo, fitted with a peak centered at 31.7 pN and 78.9 pN [63]. (I) DNA shearing with AFM in presence of TtAgo fitted with a peak centered at 21.9 pN and 56.8 pN [63]. Permission has been obtained for the above figures. Copyright 2018 American Chemical Society.

2

and CRISPR proteins use lateral diffusion, it is not known what the effect of sequence on target search is. That is, to what degree is the facilitated diffusion reliant on the interaction between guide–target base paring as opposed to an interaction between only the protein and target substrate? For some Cas proteins, which movement is driven by 3D diffusion, the prime determinant in target search would be the PAM recognition step. For Ago proteins, which rely on seed recognition, potential sub-seed-matching sequences could slow down the search process while the absence of such sequences could speed it up. It is known that circular RNA (circRNA) can act as a sponge for miRNA, strongly suppressing miR-7 activity [64]. Cognate seed target sites were found to be responsible for this, but it is not known if and how shorter sub-seed sequences would influence the search dynamics. Perhaps other sequences have evolved over time to ensure temporal control of post-transcriptional gene-regulation by subtly tuning the search time. A high spatio-temporal single molecule technique is required in order to visualize the effect of sequence on diffusion at the length scale of dozens of base pairs. So far, single-molecule Ago target search studies have not focused on the physiological environment of the cells such as the difference between eukaryotic cells and prokaryotic cells. For example, if a prokaryotic Ago is indeed involved in host defence against foreign genetic elements, is the protein then at all times present for surveillance? That seems unlikely: the probability for pAgo to encounter and target its own genome for chopping or slicing would be much higher than necessary and this would have fatal consequences. Besides the biological function, there is also the size difference between eukaryotic cells (∼10-100 μm) and prokaryotic cells (∼ μm). The cytoplasmic volume that Ago has to search through is orders of magnitude different, and this may affect the temporal distribution of 1D target search and 3D target search. Live cell studies, similar to the LacI studies of target search, are needed to validate and verify the questions here. For eukaryotic Agos, the effect of secondary and tertiary structures of mRNA on target search remains largely unexplored. These structures often have physiological regulatory functions [47, 65, 66] and reliable prediction of RNA structures remains a challenge as in vivo and in silico analyses often differ [42, 43]. It has been shown previously that RNA structures that base pair with the target completely abolish silencing [67], suggesting that RISC cannot overcome target inaccessibility by itself. However, miRNA binding sites have been uncovered in some viral sequences [68, 69]. In addition, miR-159 target sites have also been shown to be reliant on the presence of stem loops, as disruption of adjacent stem loops seemed to attenuate miR159 targeting efficacy [70]. The functional role of secondary structures in the context of miRNA silencing is therefore far from understood. Lastly, most studies have focussed on a minimal RISC, consisting of Ago and guide. However in vivo, mRNA silencing happens through translational repression and mRNA decay [71]. Several components in this silencing machinery have been uncovered [108,109], of which the scaffold protein GW182 (TNRC6 in mammals) family is the key component. This protein has been elusive in structural studies due to its intrinsically unstructured nature and therefore insight through biophysical studies has been limited. Recently, a link between the novel field of biological phase separations and RNAi has been uncovered through the molecular interactions of AGO and TNCR6. Promoted by multivalent interactions between glycine/tryptophan domain and the tryptophan binding pockets of hAGO2, condensation of miRISC was found to form in vitro and in

vivo [72]. The finding suggests that phase separation of miRISC could contribute to repression of mRNA targets through sequestration of the targets inside the droplets. Currently, the biophysical properties of RISC condensates remain unexplored and the nature of how these individual components of the condensate interact with each other is still an elusive question. Furthermore, until now Ago target search has been studied in an in vitro environment. How well would the results from the published studies translate to a phase-separated environment, where RISC exists in a high local concentration? In conclusion, we have noted the recent leaps in knowledge of the kinetics of target search and target recognition of individual regulatory complexes. However, the biophysical mechanisms that govern the interplay of these slicing/silencing proteins and their degradation machinery have been largely unexplored. Future studies will allow us to paint a complete picture of targeting by small RNAs in a cellular context.

2

REFERENCES

[1] A. Pauli, J. L. Rinn, and A. F. Schier, *Non-coding RNAs as regulators of embryogenesis,* Nature Reviews Genetics **12**, 136 (2011), arXiv:NIHMS150003 .

[2] A. D. Riggs, S. Bourgeois, and M. Cohn, *The lac represser-operator interaction: III. Kinetic studies,* Journal of Molecular Biology **53**, 401 (1970).

[3] A. D. Riggs, H. Suzuki, and S. Bourgeois, *lac repressor-operator interaction: I. Equilibrium studies,* Journal of Molecular Biology **48**, 67 (1970).

[4] O. G. Berg, R. B. Winter, and P. H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory.* Biochemistry **20**, 6929 (1981).

[5] M. Slutsky and L. A. Mirny, *Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential,* Biophysical Journal **87**, 4021 (2004), arXiv:0402005 [q-bio] .

[6] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, *How a protein searches for its site on DNA: the mechanism of facilitated diffusion,* Journal of Physics A: Mathematical and Theoretical **42**, 434013 (2009).

[7] N. P. Stanford, M. D. Szczelkun, J. F. Marko, and S. E. Halford, *One- and three-dimensional pathways for proteins to reach specific DNA sites.* The EMBO journal **19**, 6546 (2000).

[8] N. Shimamoto, *One-dimensional Diffusion of Proteins along DNA,* Journal of Biological Chemistry **274**, 15293 (1999).

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

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

[11] L. Zandarashvili, A. Esadze, D. Vuzman, C. A. Kemme, Y. Levy, and J. Iwahara, *Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble,* Proceedings of the National Academy of Sciences **112**, E5142 (2015).

[12] L. Cuculis, Z. Abil, H. Zhao, and C. M. Schroeder, *Direct observation of TALE protein dynamics reveals a two-state search mechanism,* Nature Communications **6**, 7277 (2015).

[13] K. Ragunathan, C. Liu, and T. Ha, *RecA filament sliding on DNA facilitates homology search,* eLife **2012**, 1 (2012).

[14] Z. Qi, S. Redding, J. Y. Lee, B. Gibb, Y. Kwon, H. Niu, W. A. Gaines, P. Sung, and E. C. Greene, *DNA sequence alignment by microhomology sampling during homologous recombination,* Cell **160**, 856 (2015), arXiv:NIHMS150003 .

[15] N. C. Lau, L. P. Lim, E. G. Weinstein, and D. P. Bartel, *An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans*, Science **294**, 858 LP (2001).

[16] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, and T. Tuschl, *Identification of Novel Genes Coding for Small Expressed RNAs*, Science **294**, 853 LP (2001).

[17] R. C. Lee and V. Ambros, *An Extensive Class of Small RNAs in Caenorhabditis elegans*, Science **294**, 862 LP (2001).

[18] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*, Science **337**, 816 LP (2012).

[19] S. J. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. van der Oost, *Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes*, Science **321**, 960 LP (2008), arXiv:20 .

[20] E. C. Greene, S. Wind, T. Fazio, J. Gorman, and M.-L. Visnapuu, *DNA Curtains for High-Throughput Single-Molecule Optical Imaging*, Methods in Enzymology **472**, 293 (2010).

[21] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, *Advances in Single-Molecule Fluorescence Methods for Molecular Biology*, Annual Review of Biochemistry **77**, 51 (2008).

[22] J. Yu, *Single-Molecule Studies in Live Cells*, Annual Review of Physical Chemistry **67**, 565 (2016).

[23] K. C. Neuman and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy*, Nature Methods **5**, 491 (2008).

[24] D. Dulin, J. Lipfert, M. C. Moolman, and N. H. Dekker, *Studying genomic processes at the single-molecule level: introducing the tools and applications*, Nature Reviews Genetics **14**, 9 (2012).

[25] M. Fernández-Suárez and A. Y. Ting, *Fluorescent probes for super-resolution imaging in living cells*, Nature Reviews Molecular Cell Biology **9**, 929 (2008).

[26] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie, *A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA*, Proceedings of the National Academy of Sciences **103**, 5752 (2006).

[27] S. H. Sternberg, B. LaFrance, M. Kaplan, and J. A. Doudna, *Conformational control of DNA target cleavage by CRISPR–Cas9*, Nature **527**, 110 (2015).

[28] G.-W. Li, O. G. Berg, and J. Elf, *Effects of macromolecular crowding and DNA looping on gene regulation kinetics*, Nature Physics **5**, 294 (2009).

[29] C. G. Kalodimos, R. Boelens, and R. Kaptein, *A residue-specific view of the association and dissociation pathway in protein-DNA recognition*, Nature Structural

2

Biology **9**, 193 (2002).

[30] S. M. Klum, S. D. Chandradoss, N. T. Schirle, C. Joo, and I. J. MacRae, *Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition,* The EMBO Journal , e201796474 (2017).

[31] M. Klein, S. D. Chandradoss, M. Depken, and C. Joo, *Why Argonaute is needed to make microRNA target search fast and reliable,* Seminars in Cell & Developmental Biology **65**, 20 (2017).

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

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

[34] M. H. Jo, S. Shin, S. R. Jung, E. Kim, J. J. Song, and S. Hohng, *Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs,* Molecular Cell **59**, 117 (2015).

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

[36] C. Yao, H. M. Sasaki, T. Ueda, Y. Tomari, and H. Tadakuma, *Single-Molecule Analysis of the Target Cleavage Reaction by the Drosophila RNAi Enzyme Complex,* Molecular Cell **59**, 125 (2015).

[37] Y. M. Wang, R. H. Austin, and E. C. Cox, *Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA,* Physical Review Letters **97**, 48302 (2006).

[38] P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi, and X. S. Xie, *Nonspecifically bound proteins spin while diffusing along DNA.* Nature structural & molecular biology **16**, 1224 (2009).

[39] J. Elf, G.-W. Li, and X. S. Xie, *Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell,* Science **316**, 1191 LP (2007).

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

[41] J.-D. Beaudoin, E. M. Novoa, C. E. Vejnar, V. Yartseva, C. M. Takacs, M. Kellis, and A. J. Giraldez, *Analyses of mRNA structure dynamics identify embryonic gene regulatory programs,* Nature Structural & Molecular Biology , 1 (2018).

[42] P. C. Bevilacqua, L. E. Ritchey, Z. Su, and S. M. Assmann, *Genome-Wide Analysis of RNA Secondary Structure,* Annual Review of Genetics **50**, 235 (2016).

[43] L. E. Vandivier, S. J. Anderson, S. W. Foley, and B. D. Gregory, *The Conservation and Function of RNA Secondary Structure in Plants,* Annual Review of Plant Biology **67**, 463 (2016), arXiv:1510.01420 .

[44] M. B. Warf and J. A. Berglund, *Role of RNA structure in regulating pre-mRNA splicing,*

Trends in Biochemical Sciences **35**, 169 (2010).

[45] K. C. Martin and A. Ephrussi, *mRNA Localization: Gene Expression in the Spatial Dimension,* Cell **136**, 719 (2009).

[46] P. S. Ray, J. Jia, P. Yao, M. Majumder, M. Hatzoglou, and P. L. Fox, *A stress-responsive RNA switch regulates VEGFA expression,* Nature **457**, 915 (2008).

[47] S. Weingarten-Gabbay, S. Elias-Kirma, R. Nir, A. A. Gritsenko, N. Stern-Ginossar, Z. Yakhini, A. Weinberger, and E. Segal, *Systematic discovery of cap-independent translation sequences in human and viral genomes,* Science **351** (2016).

[48] D. P. Bartel, *MicroRNAs: Target Recognition and Regulatory Functions,* Cell **136**, 215 (2009), arXiv:0208024 [gr-qc] .

[49] P. Horvath and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science (New York, N.Y.) **327**, 167 (2010), arXiv:arXiv:1011.1669v3 .

[50] R. Sorek, C. M. Lawrence, and B. Wiedenheft, *CRISPR-mediated adaptive immune systems in bacteria and archaea,* Annu Rev Biochem **82**, 237 (2013).

[51] S. Redding, S. H. Sternberg, M. Marshall, B. Gibb, P. Bhat, C. K. Guegler, B. Wiedenheft, J. A. Doudna, and E. C. Greene, *Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System,* Cell **163**, 854 (2015).

[52] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, and J. A. Doudna, *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9,* Nature **507**, 62 (2014), arXiv:NIHMS150003 .

[53] V. Globyte, S. H. Lee, T. Bae, J.-S. Kim, and C. Joo, *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion,* The EMBO Journal , e99466 (2018).

[54] D. C. Swarts, M. M. Jore, E. R. Westra, Y. Zhu, J. H. Janssen, A. P. Snijders, Y. Wang, D. J. Patel, J. Berenguer, S. J. J. Brouns, and J. van der Oost, *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature **507**, 258 (2014), arXiv:15334406 .

[55] D. C. Swarts, K. Makarova, Y. Wang, K. Nakanishi, R. F. Ketting, E. V. Koonin, D. J. Patel, and J. van der Oost, *The evolutionary journey of Argonaute proteins.* Nature structural & molecular biology **21**, 743 (2014).

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

[57] D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, and T. Ha, *Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9,* Nature Communications **7**, 1 (2016).

[58] S. R. Jung, E. Kim, S. Shin, J. J. Song, and S. Hohng, *Mechanisms of the Binding/Dissociation Acceleration of the Target–Guide Interaction by Thermus thermophilus Argonaute,* Bulletin of the Korean Chemical Society **39**, 167 (2018).

[59] G. Obmolova, C. Ban, P. Hsieh, and W. Yang, *Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA,* Nature **407**, 703 (2000).

[60] M. A. Schumacher, K. Y. Choi, H. Zalkin, and R. G. Brennan, *Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices,* Science **266**, 763 LP (1994).

[61] K. E. Dillard, M. W. Brown, N. V. Johnson, Y. Xiao, A. Dolan, E. Hernandez, S. D. Dahlhauser, Y. Kim, L. R. Myler, E. V. Anslyn, A. Ke, and I. J. Finkelstein, *Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex,* Cell **175**, 934 (2018).

[62] Y. Jeon, Y. H. Choi, Y. Jang, J. Yu, J. Goo, G. Lee, Y. K. Jeong, S. H. Lee, I. S. Kim, J. S. Kim, C. Jeong, S. Lee, and S. Bae, *Direct observation of DNA target searching and cleavage by CRISPR-Cas12a,* Nature Communications **9** (2018), 10.1038/s41467-018-05245-x, arXiv:15334406 .

[63] G. Song, H. Chen, G. Sheng, Y. Wang, and J. Lou, *Argonaute Facilitates the Lateral Diffusion of the Guide along Its Target and Prevents the Guide from Being Pushed Away by the Ribosome,* Biochemistry **57**, 2179 (2018).

[64] T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard, and J. Kjems, *Natural RNA circles function as efficient microRNA sponges,* Nature **495**, 384 (2013).

[65] M. Mandal and R. R. Breaker, *Gene regulation by riboswitches,* Nature Reviews Molecular Cell Biology **5**, 451 (2004).

[66] M. Kedde, M. Van Kouwenhove, W. Zwart, J. A. Oude Vrielink, R. Elkon, and R. Agami, *A Pumilio-induced RNA structure switch in p27-3'2 UTR controls miR-221 and miR-222 accessibility,* Nature Cell Biology **12**, 1014 (2010).

[67] S. L. Ameres, J. Martinez, and R. Schroeder, *Molecular Basis for Target RNA Recognition and Cleavage by Human RISC,* Cell **130**, 101 (2007).

[68] T. K. Scheel, J. M. Luna, M. Liniger, E. Nishiuchi, K. Rozen-Gagnon, A. Shlomai, G. Auray, M. Gerber, J. Fak, I. Keller, R. Bruggmann, R. B. Darnell, N. Ruggli, and C. M. Rice, *A Broad RNA Virus Survey Reveals Both miRNA Dependence and Functional Sequestration,* Cell Host & Microbe **19**, 409 (2016).

[69] C. L. Jopling, S. Schütz, and P. Sarnow, *Position-Dependent Function for a Tandem MicroRNA miR-122-Binding Site Located in the Hepatitis C Virus RNA Genome,* Cell Host & Microbe **4**, 77 (2008).

[70] Z. Zheng, M. Reichel, I. Deveson, G. Wong, J. Li, and A. A. Millar, *Target RNA Secondary Structure Is a Major Determinant of miR159 Efficacy,* Plant Physiology **174**, 1764 LP (2017).

[71] L. F. R. Gebert and I. J. MacRae, *Regulation of microRNA function in animals,* Nature Reviews Molecular Cell Biology (2018), 10.1038/s41580-018-0045-7.

[72] J. Sheu-Gruttadauria and I. J. MacRae, *Phase Transitions in the Assembly and Function of Human miRISC,* Cell **173**, 946 (2018)

3

DNA-GUIDED DNA CLEAVAGE AT MODERATE TEMPERATURES

Prokaryotic Argonaute proteins (pAgos) constitute a diverse group of endonucleases of which some mediate host defense by utilizing small interfering DNA guides (siDNA) to cleave complementary invading DNA. This activity has been repurposed for programmable DNA cleavage in vitro. However, currently characterized DNA-cleaving pAgos require elevated temperatures (≥ 65°C) for their activity, making them less suitable for in vivo applications at moderate temperatures. Here, using biochemistry, X-ray crystallography, and single-molecule fluorescence methods, we report the functional and structural characterization of the pAgo from the mesophilic bacterium Clostridium butyricum (CbAgo). CbAgo can be reprogrammed with siDNAs to cleave complementary DNA, but not RNA. CbAgo displays a preference for siDNAs that have a deoxyadenosine at the 5'-end and thymidines in the sub-seed segment (siDNA nucleotides 2-4). Furthermore, CbAgo mediates DNA-guided DNA cleavage of AT-rich double stranded DNA at moderate temperatures (37°C). Taken together, this study provides an important step towards the development of pAgos for genetic engineering applications..

3.1. INTRODUCTION

Eukaryotic Argonaute proteins (eAgos) play a key role in RNA interference (RNAi) processes [1–3]. As the core of the multiprotein RNA-induced silencing complex (RISC), eAgos bind small non-coding RNA molecules as guides to direct the RISC complex towards complementary RNA targets [3–5]. Reflecting their physiological function, variation among eAgos is observed with respect to the presence or absence of a catalytic site, and to their potential to interact with other proteins [6]. Depending on the eAgo and on the sequence complementarity between guide and target RNA, eAgo-guide complexes either catalyze endonucleolytic cleavage of the target RNA [7] or indirectly silence the target RNA by repressing its translation and promoting its degradation through recruitment of additional silencing factors [8]. Independent of the mechanism, eAgo-mediated RNA binding generally results in sequence-specific silencing of gene expression. As such, eAgos can coordinate various cellular processes by regulating intracellular RNA levels.

Prokaryotes also encode Argonaute proteins (pAgos) [9, 10]. Various pAgos share a high degree of structural homology with canonical eAgos as both pAgos and eAgos adopt the same four domain (N-PAZ-MID-PIWI) architecture [9–12]. Despite their structural homology, several recently characterized pAgos have distinct functional roles and preferences for their guide and/or target type compared to eAgos; certain pAgos have been implicated in host defense by directly targeting DNA instead of RNA [13–16]. One of the best characterized mechanisms that pAgos utilize is DNA-guided DNA interference, which is mediated by pAgos from Thermus thermophilus (TtAgo), Pyrococcus furiosus (PfAgo), and Methanocaldococcus jannaschii (MjAgo) [14–20]. These pAgos use 5'-end phosphorylated small interfering DNAs (siDNAs) for recognition and successive cleavage of complementary DNA targets. This mechanism enables both TtAgo and PfAgo to mediate host defence against invading nucleic acids. Prokaryotes lack homologs of eukaryotic enzymes that are involved in guide biogenesis [21]. Instead, both TtAgo and MjAgo - besides the canonical siDNA-dependent target cleavage termed 'slicing' - exhibit an alternative nuclease activity termed 'chopping' [16, 18]. Chopping facilitates autonomous generation of small DNA fragments from dsDNA substrates. Subsequently, these DNA fragments can serve as siDNAs for canonical slicing [16, 18].

TtAgo and PfAgo can be programmed with short synthetic siDNA which allows them to target and cleave dsDNA sequences of choice in vitro [14, 15]. This activity has enabled the repurposing of PfAgo as an universal restriction endonuclease for in vitro molecular cloning [22]. In analogy with the now commonly used CRISPR-Cas9 and CRISPR-Cas12a enzymes [23–25], it has been suggested that pAgos could also be repurposed as next-generation genome editing tools [26]. However, due to the thermophilic nature (optimum activity temperature ≥65°C) and low levels of endonuclease activity at the relevant temperatures (20-37°C), it is unlikely that TtAgo, PfAgo and MjAgo are suitable for genome editing. The quest for a pAgo that can cleave dsDNA at moderate temperatures has resulted in the characterization of the Argonaute protein from Natronobacterium gregory (NgAgo). NgAgo was claimed to be the first pAgo suitable for genome editing purposes [27], but the study reporting its characterization was retracted after a series of reproducibility issues [27–29]. Instead, it has been shown that NgAgo might target RNA rather than DNA [30].

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

3.2. RESULTS

3.2.1. CBAGO MEDIATES SIDNA-GUIDED SSDNA CLEAVAGE

CbAgo was successfully expressed in E. coli from a codon-optimized gene using a T7-based pET expression system and purified (Supplementary Fig. 1a). To determine the guide and target binding characteristics of CbAgo, we performed single-molecule experiments using Förster resonance energy transfer (FRET). We immobilized either Cy5-labeled single stranded RNA or DNA targets (FRET acceptor) on a polymer-coated quartz surface (Fig. 3.1a). Next we introduced CbAgo in complex with either a Cy3-labeled siRNA or siDNA guide (FRET donor) and recorded the interactions. Strikingly, CbAgo could utilize both siRNAs and siDNAs to bind DNA or RNA targets (Fig. 3.1b). To test which guide is preferentially bound by CbAgo we performed a competition assay in which CbAgo was immobilized into the microfluidic chamber, and an equimolar mixture of siDNA and siRNAs was introduced. While only short-lived interactions were observed for siRNA, siDNA was strongly bound by CbAgo (Fig. 1c). This results suggests that CbAgo utilizes siDNA rather than siRNA as a guide. CbAgo is phylogenetically closest related to the clade of halobacterial pAgos, among which also pAgo from Natronobacterium gregoryi (NgAgo) can be found (Fig. 3.1d and Supplementary Fig. 2). A multiple sequence alignment of CbAgo with other pAgos (Supplementary Fig. 1b) suggests that CbAgo contains the conserved DEDX catalytic residues (where X can be an D, H or N) which are essential for nuclease activity in 'slicing' Agos [31](Nakanishi et al., 2012). In the case of CbAgo, this concerns residues D541, E577, D611 and D727. To confirm whether CbAgo indeed is an active nuclease, we performed in vitro activity assays in which CbAgo was loaded with either synthetic siDNAs or siRNAs (21 nucleotides in length). Next the complexes were incubated at 37°C with 45-nucleotide complementary single stranded RNA or DNA target oligonucleotides. While no activity was found in any of the combinations in which siRNAs or target RNAs were used, CbAgo was able to cleave target DNAs in a siDNA-dependent manner (Fig. 3.1e). In agreement with the predicted DEDD catalytic site (Supplementary Fig. 1b), alanine substitutions of two of aspartic acids (D541A, D611A) in the expected catalytic tetrad abolished the nuclease activity, demonstrating that the observed siDNA-guided ssDNA endonucleolytic activity was indeed catalyzed by the DEDD catalytic site. To further

investigate the full temperature range at which CbAgo is active, we performed additional cleavage assays at temperatures ranging from 10-95°C. While CbAgo displayed the highest activity at its physiologically relevant temperature (37°C), CbAgo also catalyzed siDNA-guided target DNA cleavage at temperatures as low as 10°C and as high as 50°C (Fig. 3.1f).

When CbAgo-siDNA complexes and target ssDNA substrates (45nt) were mixed in equimolar amounts, cleavage of the target DNA was not complete after 1 hour incubation (Fig. 3.1e). Therefore, we investigated the substrate turnover kinetics of CbAgo by monitoring the cleavage assays in a time course using variable CbAgo:siDNA:target DNA ratios (Fig. 3.1g). A rapid burst of activity was observed during the first minute, likely indicating the first target binding and cleavage event. This stage was followed by a slow steady state, suggesting that under these conditions the CbAgo-siDNA complex slowly dissociates from the cleaved target DNA product before being able to bind and cleave a new target DNA strand. The cleavage kinetics were confirmed using single-molecule assays which demonstrated that the CbAgo-siDNA complex remains bound to the DNA target (N=21) for several minutes (Fig. 3.1h), which prevents CbAgo-siDNA complexes from binding and limits substrate turnover. Thus, while CbAgo functions as a multi-turnover nuclease enzyme, its steady-state rate is limited by product release.
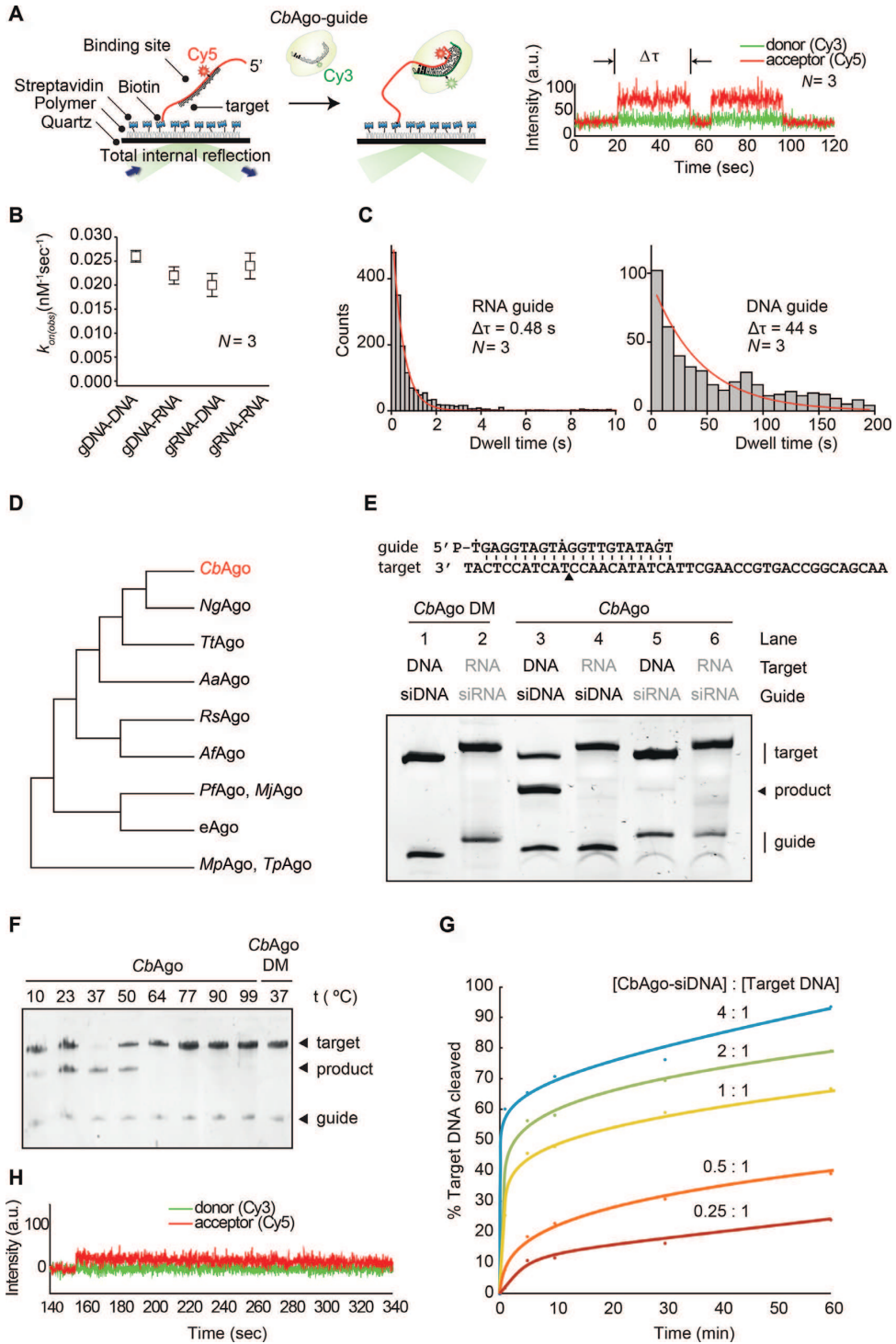
Figure 3.1: CbAgo exhibits DNA-guided DNA endonuclease activity at 37°C.

Figure 3.1 *(previous page)*: **CbAgo exhibits DNA-guided DNA endonuclease activity at 37°C.** (A) Left: Overview of the single molecule assay to determine the binding characteristics of CbAgo. Right: FRET diagram of a CbAgo–siDNA complex that has three complementary base pairs (2–4nt) to the DNA target. Indicated is the dwell time ($\Delta\tau$). (B) Comparison of the binding rates (kon) of CbAgo in complex with siDNA or siRNA to bind DNA or RNA targets. The rates are similar for each nucleic acid type guide and target. N is the number of base paired nucleotides. (C) Dwell time histograms showing CbAgo preferentially binds siDNAs in siDNA-siRNA competition experiments. (D) Schematic phylogenetic tree of characterized pAgos. (E) CbAgo exhibits DNA-guided DNA endonuclease activity. Upper panel: Sequence of the synthetic let7 miRNA-based siDNA guide and target DNA sequences that were used for the in vitro cleavage assays. Lower panel: CbAgo, guides and targets were mixed in a 1:1:1 molar ratio and incubated for 1 h at 37°C. Catalytic mutant CbAgoDM was used as a control. Cleavage products were analysed by denaturing polyacrylamide electrophoresis. (F) CbAgo displays highest activity at 37°C. CbAgo and siDNA were mixed and pre-incubated at various temperatures for 10 min. Next, target DNA was added and the sample was incubated for 1 h at the same temperature. CbAgoDM was used as a control. Cleavage products were analysed by denaturing polyacrylamide electrophoresis. (G) Quantified data of a CbAgo-mediated siDNA-guided ssDNA cleavage turnover experiment using 5 pmol target DNA and increasing concentrations of CbAgo-siDNA (1.25–20 pmol). (H) FRET diagram showing that a cleavage compatible CbAgo-siDNA remains bound to a fully complementary target DNA (N = 21) during the entire the measurement (340 seconds).

3.2.2. Structure of CbAgo in the cleavage-competent conformation

To investigate the molecular architecture of CbAgo in light of its biochemical activity, we crystalized CbAgoDM in complex with both a 21-nt siDNA and a 19-nt DNA target, and solved the structure of the complex at 3.54 Å resolution (Fig. 2 and Table 1). Like other Agos, CbAgo adopts a bilobed conformation in which one of the lobes comprises the N-terminal, linker L1, and PAZ domains, which are linked by linker L2 to the other lobe comprising the MID and PIWI domains. Nucleotides 2-16 of the siDNA constitute a 15 base-pairs, A-form-like duplex with the target DNA (Fig. 3.2a). The 5'-terminal nucleotide of the siDNA is anchored in the MID domain pocket, where the 5'-phosphate group of the siDNA makes numerous interactions with MID domain residues and the C-terminal carboxyl group of CbAgo (Supplementary Fig. 3). However, there are no base-specific interactions with the 5'-terminal cytosine base. To test whether these interactions are important for CbAgo activity, we performed target DNA cleavage assays in which we used siDNAs with a 5' phosphate or a 5' hydroxyl group (Supplementary Fig. 4). As for other pAgos, CbAgo was able to utilize both siDNAs for target DNA cleavage, but it cleaved target DNA much more efficiently when the siDNA contained a 5' phosphate group, in agreement with the observed mode of siDNA binding in the structure. Furthermore, the backbone phosphates of the siDNA seed segment form hydrogen-bonding and ionic interactions with specific residues in the MID, PIWI and L1 domains (Supplementary Fig. 3). At the distal end of the siDNA-target DNA duplex, the N-domain residue His35 caps the duplex by stacking onto the last base pair. After this point, the remaining 3'-terminal nucleotides of the siDNA are unordered, while the target DNA bends away from the duplex and enters the cleft between the N-terminal and PAZ domains. In agreement with other ternary pAgo complexes [20, 32, 33], the PAZ domain pocket, which normally binds the 3' end of the guide in a binary Ago-guide complex, is empty. CbAgo is phylogenetically closely related to TtAgo (Fig. 3.2d). However, CbAgo is 63 amino acids (9.2%) longer than TtAgo (748 amino acids vs. 685

amino acids) and CbAgo and TtAgo share only 23% sequence identity. Superposition of the CbAgo complex structure with the structure of TtAgo bound to a siDNA and DNA target (PDB: 4NCB) (Fig. 3.2c) reveals that the macromolecular architecture and conformation of these TtAgo and CbAgo structures are highly similar (Core root mean square deviation of 3.0 Å over 563 residues), with differences found mostly in the loop regions. This agrees with the fact that loops of thermostable proteins are generally more compact and shorter [34, 35]. In the TtAgo structure, which is thought to represent a catalytically competent state, a 'glutamate finger' side chain (Glu512TtAgo) is inserted into the catalytic site completing the catalytic DDED tetrad [32]. Similarly, the corresponding residue in CbAgo (Glu577) is located within a flexible loop and is positioned near the other catalytic residues (Fig. 3.2d; Asp541, Asp611, and Asp727). This observation implies that this CbAgo complex structure also represents the cleavage competent conformation. Since only 15 siDNA-target DNA base pairs are formed in the complex, this suggests that additional siDNA-target DNA binding is not essential for target DNA cleavage. To determine the minimum siDNA length that CbAgo requires for target binding, we performed single-molecule fluorescence assays. First, CbAgo was immobilized on a surface and next it was incubated with 5'-phosphorylated Cy3-labelled siDNAs (Fig. 3.2e). These assays demonstrate that CbAgo can bind siDNAs with a minimal length of 12 nucleotides. Next, we determined the minimum siDNA length for CbAgo-siDNA mediated target DNA cleavage (Fig. 3.2f). In line with the observation that the CbAgo complex structure adopts a catalytically active confirmation, CbAgo was active with siDNAs as short as 14 nt under the tested conditions. This resembles PfAgo, MjAgo, and MpAgo, all of which require siDNAs with a minimal length of 15 nt to catalyze target DNA cleavage [1, 15, 16]. Only TtAgo has been reported to mediate target DNA cleavage with siDNAs as short as 9 nt [12].

3.2.3. CBAGO ASSOCIATES WITH PLASMID-DERIVED SIDNAS IN VIVO

It has previously been demonstrated that certain pAgos co-purify with their guides and/or targets during heterologous expression in Escherichia coli [13, 14]. To determine whether CbAgo also acquires siDNAs during expression, we isolated and analyzed the nucleic acid fraction that co-purified with CbAgo. Denaturing polyacrylamide gel electrophoresis revealed that CbAgo co-purified with small nucleotides with a length of 12-19 nucleotides (Fig. 3.3a). These nucleic acids were susceptible to DNase I but not to RNase A treatment, indicating that CbAgo acquires 12-19 nucleotide long siDNAs in vivo, which fits with its observed binding and cleavage activities in vitro (Fig. 3.1 and 3.2). We cloned and sequenced the siDNAs that co-purified with CbAgo to determine their exact length and sequence. The majority of the siDNAs had a length of 16 nucleotides and are complementary to the plasmid used for expression of CbAgo (Fig. 3.3b and 3.3c). Likewise the siRNAs and siDNAs that co-purify with respectively Rhodobacter sphaeroides (RsAgo) and TtAgo are also mostly complementary to their expression plasmids [13, 14]. As both TtAgo and RsAgo have been demonstrated to interfere with plasmid DNA, this suggests that also CbAgo might play a role in protecting its host against invading DNA. However, no significant reduction of plasmid content could be detected during or upon expression of CbAgo in *E. coli* (Supplementary Fig. 6). We also investigated whether CbAgo co-purified with nucleic acids that were enriched
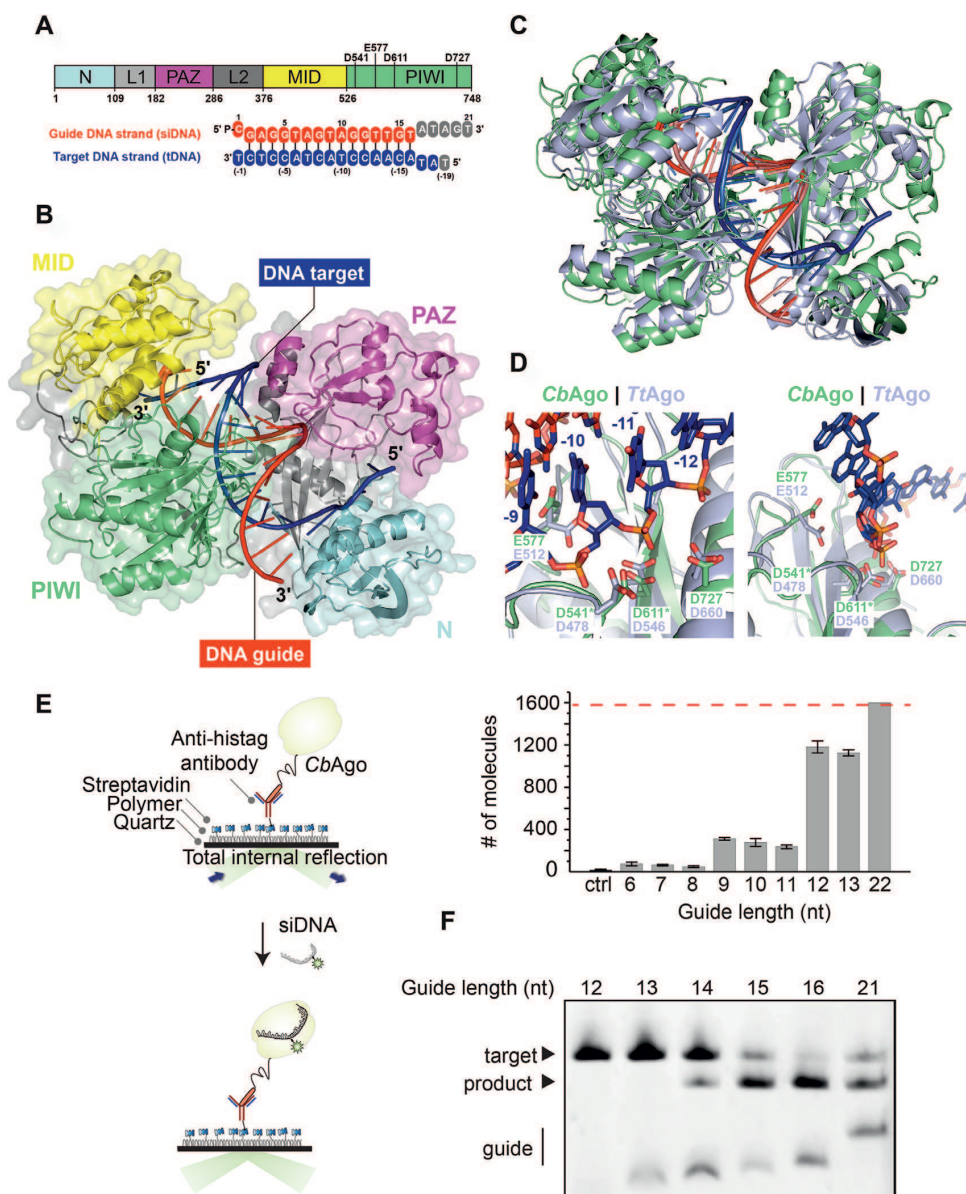
Figure 3.2: Structure of CbAgo in complex with a siDNA and a DNA target. (A) Upper panel: Schematic diagram of the domain organization of CbAgo. L1 and L2 are linker domains. Lower panel: Sequences of the siDNA (red) and target DNA (blue). Nucleotides that are unordered in the structure are coloured grey. See also Supplementary Table S1. (B) Overall structure of the CbAgo-siDNA-target DNA complex. Domains are coloured according to the colour scheme in panel A. (C) Structural alignment of CbAgo (green) and TtAgo (light purple; PDB: 4NCB). Core Root Mean Square Deviation of 3.0 Å over 563 residues.

Figure 3.2 *(previous page)*: (D) Close-up views of the aligned DDED catalytic sites of CbAgo (green) and TtAgo (light purple; PDB: 4NCB). Modelled side chains of D541 and D611 in CbAgo are indicated with green asterisks. The glutamate finger of both pAgos (E512 in TtAgo or E577 in CbAgo) are inserted into the catalytic site. The scissile phosphate between nucleotide -10 and -11 of the target DNA strand (blue) is indicated with a black asterisk in the left panel. (E) Total internal reflection microscopy (TIRM) was used to determine the minimal length for siDNA to be bound by CbAgo. Left panel: Graphical overview of the TIRM method. Right panel: Histogram with TIRM results demonstrated that synthetic siDNAs of at least 12 nt in length are efficiently bound by CbAgo. The red line indicates the total number of countable molecules within the microscope image. The raw microscope images are given in Supplementary Figure S5. (F) CbAgo mediates target DNA cleavage with siDNAs as short as 14 nucleotides. CbAgo was incubated with siDNA and target DNA in a 1:1:1 ratio. Cleavage products were analyzed by denaturing polyacrylamide electrophoresis.

for certain motifs. Sequence analysis revealed that most siDNAs co-purified with CbAgo contain a deoxyadenosine at their 5' ends (Fig. 3.3d). In addition, we observed an enrichment of thymidine nucleotides in the three positions directly downstream of the siDNA 5' end (nt 2-4) (Fig. 3.3d).

3.2.4. THE SEQUENCE OF THE SIDNA AFFECTS CBAGO ACTIVITY

To investigate if the 5'-terminal nucleotide of the siDNA affects the activity of CbAgo, we performed cleavage assays. CbAgo was loaded with siDNA guides with varied nucleotides at position 1 (g1N) and incubated with complementary target DNAs (Fig. 3.4a). Surprisingly, the highest cleavage rates were observed with CbAgo loaded with siDNAs containing a 5'-T, followed by siDNAs containing 5'-A. CbAgo bound 5'-G or 5'-C siDNAs displayed slightly lower initial cleavage rates. Also for other pAgos the g1N preference is not reflected in the in vitro activities; TtAgo (which preferentially co-purifies with g1C siDNAs) as well as PfAgo and MpAgo (of which the in vivo g1N preferences are unknown) demonstrate no clear preference for a specific g1N during in vitro cleavage reactions [1, 14, 18]. Instead, the preference of TtAgo for 5'-C siDNAs is determined by specific recognition of a guanosine nucleotide in the corresponding position (t1) in the target DNA [18]. Indeed, TtAgo structures and models have revealed base-specific interactions with target strand guanine, while base-specific interactions with the 5'-terminal cytidine in the siDNA are less obvious [18]. In the structure of the CbAgo-siDNA-target DNA complex, we observe no base-specific interactions with the 5'-terminal (Supplementary Fig. 7). Strikingly, the thymine base is not placed in the t1 binding pocket as has been observed in TtAgo, RsAgo and hAGO2 [18, 36, 37]. Instead, the thymine bases is flipped and stacks on Phe557 that also caps the siDNA-target DNA duplex (Supplementary Fig. 7). At present, we are unable to rationalize the preferentially co-purification of 5'-A siDNAs with CbAgo. In order to characterize the seed segment of CbAgo, and test whether the seed length changes depending on the nature of the guide and the target (i.e. DNA vs. RNA), we performed additional single-molecule binding assays. The length of seed was determined based on the minimal number of complementary nucleotide pairs between guide and target that were required to achieve a stable binding event. We first tested the sub-seed (nt 2-4), a 3-nt motif involved in initial target recognition in hAgo2 [38, 39]. When only the sub-seed segment of the siDNA was complementary to the DNA and RNA targets, CbAgo-siDNA complexes bound to the DNA target with an average dwell time 58-fold longer than for the RNA target (Fig.
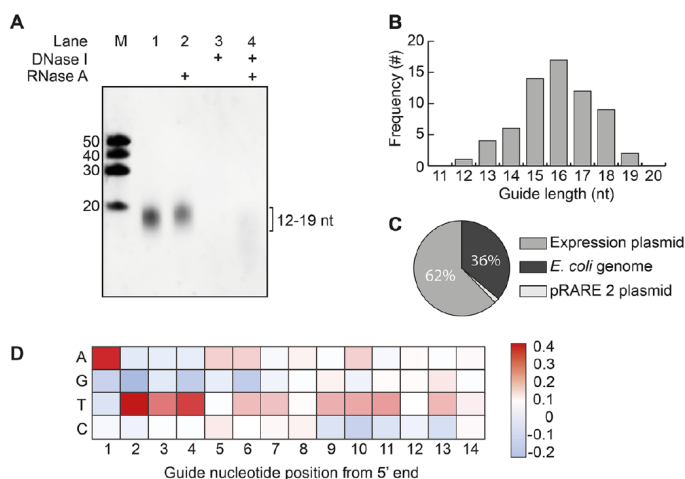
Figure 3.3: CbAgo associates with small plasmid derived siDNA in vivo. (A) Nucleic acids that co-purified with CbAgo were treated with either RNAse A, DNAse I or both, and were analyzed by denaturing polyacrylamide gel electrophoresis. (B) Histogram displaying the length of DNA co-purified with CbAgo as determined by sequencing. (C) Sequenced nucleic acids that co-purified with CbAgo are mostly complementary to the CbAgo expression plasmid. (D) Heat map showing the base preference of the co-purified nucleic acids at each position. The red squares indicate bases that were more often found compared to a random distribution (25%); blue squares indicate bases that were less frequently found.

3.4b). When nt 2-7 of the guide were complementary to the target, the CbAgo-siDNA complex stably bound to both to target DNA and RNA beyond our observation time of 300 s. This suggests CbAgo prefers DNA above RNA targets and that the seed segment of the siDNAs bound by CbAgo comprises nucleotides 2-7. Next, we set out to investigate whether CbAgo displays a preference for siDNAs with a TTT sub-seed (nt 2-4) in vitro, similar to the observed sequence preference for siDNAs that co-purified with CbAgo in vivo. CbAgo was incubated with siDNAs in which the sub-seed was varied and complementary target DNAs were added. In agreement with the in vivo preference, CbAgo displays the highest cleavage rates with TTT sub-seed siDNAs (Fig. 3.4c). To confirm these findings, we performed single-molecule assays in which we compared the target binding properties of CbAgo-siDNA complexes containing siDNAs with either a TTT or an AAA sub-seed segment. These assays demonstrate that the dwell time of CbAgo loaded with an TTT sub-seed siDNA on a target was 18-fold longer compared to CbAgo loaded with siDNA containing a AAA sub-seed (Fig. 3.4d). Combined, these data indicate that CbAgo displays a preference for siDNAs containing a TTT sub-seed segment.

3.2.5. A pair of CbAgo-siDNA complexes can cleave double stranded DNA

Thermophilic pAgos have successfully been used to generate double stranded DNA breaks in plasmid DNA [14, 15]. As each pAgo-siDNA complex targets and cleaves a single strand of DNA only, two individual pAgo-siDNA complexes are required for
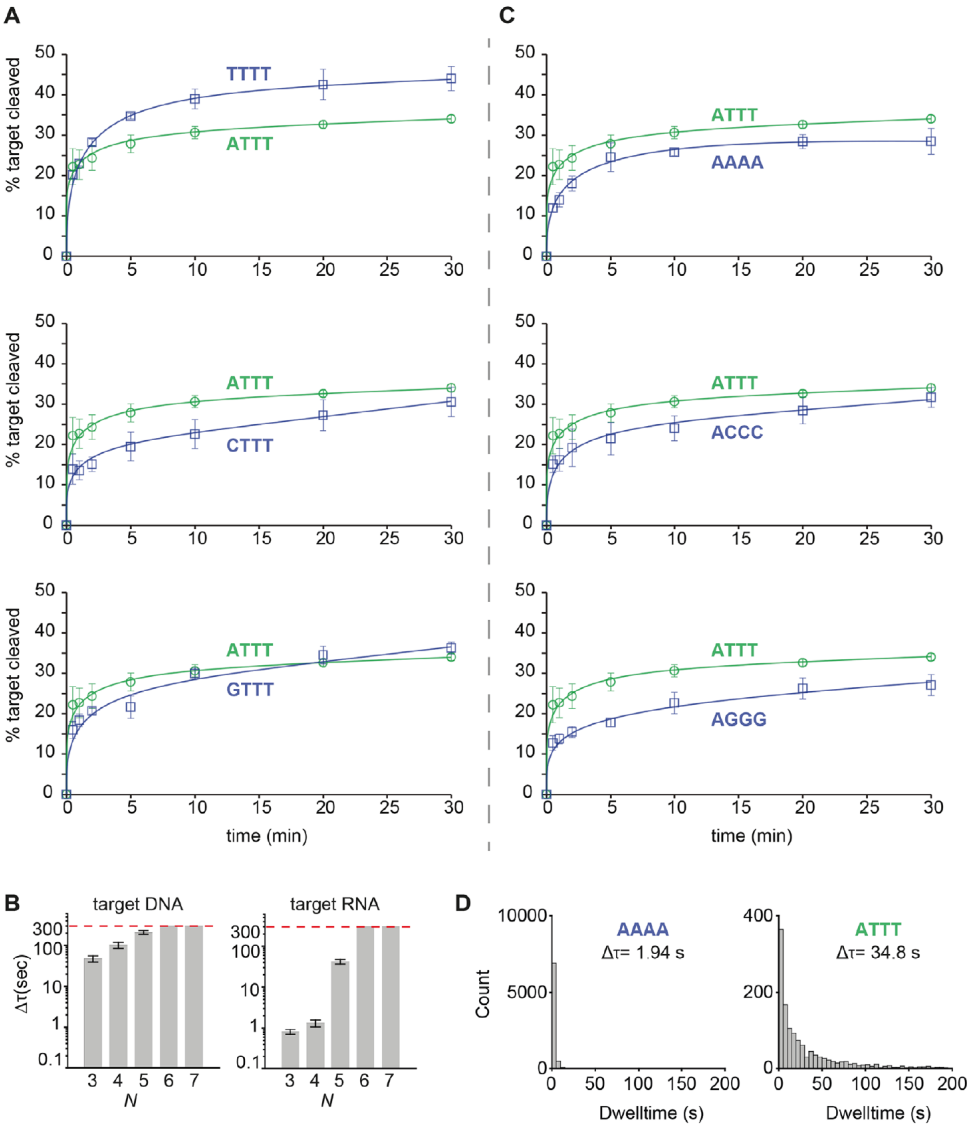
Figure 3.4: The siDNA sequence affects CbAgo activity. (A) CbAgo has no strong 5'-end nucleotide preference. CbAgo was incubated with siDNA with varied 5'-end and incubated with complementary DNA targets. Cleavage products were analysed by denaturing polyacrylamide electrophoresis and quantified. Graphs display the amount of target DNA cleaved. Error bars indicate the standard variation of three independent experiments. (B) Histograms displaying dwell time of CbAgoDM-siDNA complexes binding either DNA or RNA targets with a varied sequence complementarity (N = number of complementary nucleotides between the siDNAand the target, starting at nt 2. Thus N 3 = nt 2–4) The photobleaching limit is reachedwhere the signal is deactivated (300s). (C) CbAgo preferentially utilizes siDNAs with thymidines at position 2–4. CbAgo-siDNA complexes with siDNA in which nt 2–4 varied were incubated with complementary DNA targets. Cleavage products were analysed by denaturing polyacrylamide electrophoresis and quantified. Graphs display the amount of target DNA cleaved. Error bars indicate the standard variation of three independent experiments. (D) Histograms displaying dwell time of CbAgoDM in complex with a 5'-ATTT siDNA or 5'-AAAA siDNA binding to a target DNA. Interactions of CbAgo are on average 18-fold longer with the siDNA containing a 5-ATTT motif compared to interactions with siDNAs containing a 5'-AAAA motif.

dsDNA cleavage, each targeting another strand of the target dsDNA. Although all pAgos characterized so far appear to lack the ability to actively unwind or displace a dsDNA duplex substrate, it has been proposed that, at least in vitro, thermophilic pAgos rely on elevated temperatures (≥65 °C) to facilitate local melting of the dsDNA targets to target each strand of the DNA individually. However, CbAgo is derived from a mesophilic organism and we therefore hypothesize that is also able to mediate protection against invading DNA at moderate temperatures (37°C). To test if CbAgo can indeed cleave dsDNA targets at 37°C, we incubated apo-CbAgo and pre-assembled CbAgo-siDNA complexes with a target plasmid. Previous studies showed that the 'chopping' activity of siDNA-free apo-TtAgo and apo-MjAgo can result in plasmid linearization or degradation, respectively [16, 18]. We observe that apo-CbAgo converted the plasmid substrate from a supercoiled to open-circular state, possibly by nicking one of the strands, but did not observe significant linearization or degradation of the plasmid DNA (Fig. 3.5a). When the plasmid was targeted by CbAgo loaded with a single siDNA, we also observed loss of supercoiling (Fig. 3.5a). As this activity was not observed with nuclease-deficient CbAgoDM, we conclude that apo-CbAgo and CbAgo-siDNA complexes are able to generate nicks in dsDNA plasmid targets. When using two CbAgo-siDNA complexes, each targeting one strand of the plasmid, we observed that a large fraction of the plasmid DNA was linearized (Fig. 3.5a). This implies that CbAgo-siDNA complex-mediated nicking of each of the plasmid DNA strands generated a double stranded DNA break. Next, we investigated if the spacing between the two siDNAs affects the ability of CbAgo to cleave the plasmid. The most efficient plasmid linearization was achieved when the siDNAs were orientated exactly opposite each other (Fig. 3.5a). Finally, we investigated whether the GC-content of the target DNA plays a role during DNA targeting by CbAgo. For TtAgo, it was previously observed that AT-rich DNA is cleaved more efficiently than GC-rich DNA [18]. To test this for CbAgo, we designed a target plasmid containing 16 gene fragments of 100 base pairs complementary to sequences from the human genome, with an increasing GC content (Fig. 3.5b). CbAgo-siDNA complexes were only able to generate dsDNA in gene fragments with a GC-content of up to 31% (Fig. 3.5c). This indicates that, at least in vitro, the GC-content is an important factor that determines target DNA binding by CbAgo.

3.3. DISCUSSION

Several prokaryotic Argonaute proteins have been demonstrated to protect their host against invading nucleic acids, such as plasmid DNA [13–15]. Similar to TtAgo and RsAgo, CbAgo co-purifies with guides which are preferentially acquired from the plasmid used for its heterologous expression in E. coli. In addition, CbAgo mediates programmable DNA-guided DNA cleavage in vitro. This suggests that, similar to the phylogenetically related TtAgo, also CbAgo provides host defense via DNA-guided DNA interference. Sequencing of the nucleic acids that co-purified with CbAgo revealed that CbAgo preferentially associates with DNA with a 5'-ATTT-3' sequence at the 5' end. It was previously shown that the guide RNA utilized by eAgos can be divided into functional segments. These segments are (from 5' to 3') the anchor nucleotide (nt 1), the seed (nt 2-8) and sub-seed segments (nt 2-4), and the central (nt 9-12), 3' supplementary (nt 13-16) and tail (nt 17-21) segments [39, 40]. Extending this to the
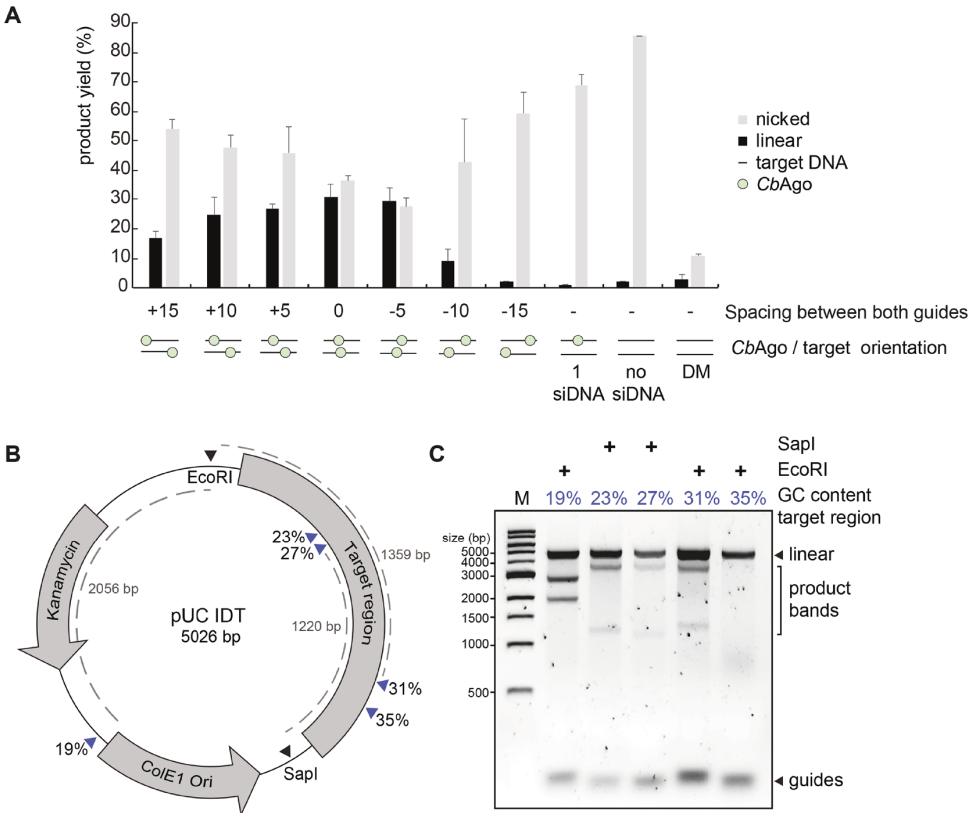
Figure 3.5: Double stranded plasmid DNA cleavage by CbAgo. (A) Two CbAgo–siDNA complexes can generate double stranded DNA breaks in plasmid DNA. CbAgo-siDNA complexes were pre-assembled and incubated with target plasmid DNA. Cleavage products were analysed by agarose gel electrophoresis (Supplementary Figure S8B) and quantified. The spacing between both CbAgo-siDNA target sites affects the linearization efficiency (nucleotide spacing between the predicted cleavage sites: +15 nt, +10 nt, +5 nt, 0 nt, -5 nt, -10 nt, -15 nt, a single siDNA, no siDNA). With 0 nt spacing, both CbAgo-siDNA complexes are exactly on top of each other. (B) Schematic overview of the pUC IDT target plasmid. Blue arrows indicate target sites while percentages indicate the GC-content of the 100 bp segments in which these target sites are located. (C) Pre-assembled CbAgo-siDNA complexes targeting various pUC IDT segments were incubated with pUC IDT. Cleavage products were incubated with EcoRI or SapI and were further analysed by agarose gel electrophoresis. The GC-content of the segments in which the target sites were located are indicated by the percentage (in blue)

3

siDNAs that co-purified with CbAgo, CbAgo preferentially associates with siDNAs that have a 5'-terminal A anchor and a T-rich sub-seed. In RNAi pathways, the preference for a specific 5'-terminal nucleotide is important for guide RNA loading into a subset of eAgos [41–43]. Similarly, several pAgos including RsAgo, TtAgo, and now CbAgo also preferentially associate with specific 5'-terminal nucleotides in vivo [13, 14]. However, for both CbAgo and TtAgo, there is no clear preference for the same 5'-base during cleavage assays in vitro. Rather than having a functional importance, the 5' end preference of pAgos might be a consequence of the way siDNAs are being generated and/or loaded, as has been demonstrated for TtAgo [18]. Several studies on human Ago2 have described the importance of the sub-seed segment (nt 2-4) in its RNA guides [38, 39, 44]. For hAgo2, a complete match between the guide RNA sub-seed segment and the target RNA triggers a conformational change that first exposes the remainder of the seed (nt 5-8), and eventually the rest of the guide. This facilitates progressive base paring between the guide RNA and the target [?]. However, a specific nucleotide preference in the sub-seed segment, as we have observed for CbAgo, has not been described for any other Argonaute protein. The preference for the T-rich sub-seed is not only observed in the in vivo acquired siDNAs, but also plays a clear role during target binding and cleavage assays in vitro. This may reflect a structural preference for these thymidines in the cleft of the PIWI domain. We have not been able to obtain diffracting crystals of CbAgo in complex with siDNAs that have a 5'-ATTT-3' sequence at the 5'-end. Future research will thus be necessary to determine the structural basis the apparent preference for these nucleotides at these positions. We hypothesize that this bias might reflect the mesophilic nature of CbAgo, which might have better access to AT-rich dsDNA fragments, both for siDNA acquisition and for target cleavage. The potential of pAgos for genome editing applications has previously been discussed [26]. However, the previously characterized DNA-cleaving pAgos originate from thermophilic prokaryotes and therefore require temperatures for their activity that are too high for genome editing at moderate temperatures. The biochemical characterization of CbAgo reporter herein is the first example of a pAgo that catalyzes efficient DNA-guided DNA cleavage at 37°C. Yet, we were unable to demonstrate CbAgo-mediated genome editing in human HEK293 cells (unpublished work). It is conceivable that our experiments were not performed under the optimal conditions for efficient editing. If CbAgo or other pAgos could be harnessed for genome editing, they will have certain advantages over the currently well-established genome editing tools CRISPR-Cas9 and CRISPR-Cas12a. While CRISPR-based genome editing tools can be programmed with a guide RNA to target DNA sequences of choice, target DNA cleavage additionally requires the presence of a protospacer adjacent motif (PAM) next to the targeted sequence (5'-NGG-3' for Cas9 and 5'-TTTV-3' for Cas12a) [45]. This limits the possible target sites of Cas9 and Cas12a. In contrast, pAgos do not require a PAM for DNA targeting, which would make them much more versatile tools compared to CRISPR-associated nucleases. However, PAM binding by Cas9 and Cas12a also promotes unwinding of dsDNA targets [46–48] which subsequently facilitates strand displacement by the RNA guide, and eventually R-Loop formation. The absence of such mechanism in pAgos might explain their limited nuclease activity on dsDNA targets. Here, we have demonstrated that CbAgo does not strictly rely on other proteins when targeting AT-rich DNA sequences in vitro. As

such, this study provides a foundation for future efforts to improve double stranded DNA target accessibility of pAgos and to facilitate the further development of pAgos as genome editing tools.

3.4. METHODS

3.4.1. PLASMID CONSTRUCTION

The CbAgo gene was codon harmonized for E.coli Bl21 (DE3) and inserted into a pET-His6 MBP TEV cloning vector (obtained from the UC Berkeley MacroLab, Addgene #29656) using ligation independent cloning (LIC) using oligonucleotides oDS067 and oDS068 (Table S1) to generate a protein expression construct that encodes the CbAgo polypeptide sequence fused to an N-terminal tag comprising a hexahistidine sequence, a maltose binding protein (MBP) and a Tobacco Etch Virus (TEV) protease cleavage site.

3.4.2. GENERATION OF THE DOUBLE MUTANT

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

3.4.3. CBAGO EXPRESSION AND PURIFICATION

The CbAgo WT and DM proteins were expressed in E.coli Bl21(DE3) Rosetta™ 2 (Novagen). Cultures were grown at 37°C in LB medium containing 50 μg ml-1 kanamycin and 34 μg ml-1 chloramphenicol until an OD600nm of 0.7 was reached. CbAgo expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. During the expression cells were incubated at 18°C for 16 hours with continues shaking. Cells were harvested by centrifugation and lysed by sonication (Bandelin, Sonopuls. 30% power, 1s on/2s off for 5min) in lysis buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM imidazole, supplemented with a EDTA free protease inhibitor cocktail tablet (Roche). The soluble fraction of the lysate was loaded on a nickel column (HisTrap Hp, GE healthcare). The column was extensively washed with wash buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl and 30 mM imidazole. Bound protein was eluted by increasing the concentration of imidazole in the wash buffer to 250 mM. The eluted protein was dialysed at 4°C overnight against 20 mM HEPES pH 7.5, 250 mM KCl, and 1mM dithiothreitol (DTT) in the presence of 1mg TEV protease (expressed and purified according to [49]Tropea et al. 2009) to cleave of the His6-MBP tag. Next the cleaved protein was diluted in 20mM HEPES pH 7.5 to lower the final salt concentration to 125 mM KCl. The diluted protein was applied to a heparin column (HiTrap Heparin HP, GE Healthcare), washed with 20 mM HEPES pH 7.5, 125 mM KCl and eluted with a linear gradient of 0.125-2 M KCl. Next, the eluted protein was loaded onto a size exclusion column (Superdex 200 16/600 column, GE Healthcare) and eluted with 20 mM HEPES pH 7.5, 500mM KCl and 1 mM DTT. Purified CbAgo protein was diluted in size exclusion buffer to a final concentration of 5 μM. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

3.4.4. Co-purification nucleic acids

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

3.4.5. Single stranded Activity assays

Unless stated otherwise 5 pmoles of each CbAgo, siDNA and target were mixed in a ratio of 1:1:1, in 2x reaction buffer containing 20 mM Tris-HCl (pH 7.5) supplemented with 500 µM MnCl2+. The target was added after the CbAgo and siDNA had been incubation for 15 min at 37°C. Then the complete reaction mixture was incubated for 1 hour at 37°C. The reaction was terminated by adding 2x RNA loading dye (95% Formamide, 0.025% bromophenol blue, 5 mM ETDA) and heating it for 5 minutes at 95°C. After this the samples were resolved on a 20% denaturing (7 M Urea) polyacrylamide gel. The gel was stained with SYBR gold nucleic acid stain (Invitrogen) and imaged using a G:BOX Chemi imager (Syngene).

3.4.6. Double stranded Activity assay

In two half reactions 12.5 pmoles of CbAgo was loaded with either 12.5pmoles of forward or reverse siDNA in reaction buffer containing 10 mM Tris-HCl, 10 µg/ml BSA, 250 µM MnCl2. The half reactions were incubated for 15 min at 37°C. Next, both half reactions were mixed together and 120 ng target plasmid was added after which the mixture was incubated for 1 hour of 37°C. After the incubation the target plasmid was purified from the mixture using a DNA clean and concentrate kit (DNA Clean & Concentrator™-5, Zymogen) via the supplied protocol. The purified plasmid was subsequently cut using either EcoRI-HF (NEB) or SapI-HF (NEB) in Cutsmart buffer (NEB) for 30 min at 37°C. A 6x DNA loading dye (NEB) was added to the plasmid sample prior to resolving it on a 0.7% agarose gel stained with SYBR gold (invitrogen).

3.4.7. Crystallization

To reconstitute the CbAgo DM-siDNA-target DNA complex, siDNA and target DNA were pre-mixed at a 1:1 ratio, heated to 95°C, and slowly cooled to room temperature. The formed dsDNA duplex (0.5M) was mixed with CbAgo DM in SEC buffer at a 1:1:4 ratio (CbAgo DM:duplex DNA), and MgCl2 was added to a final concentration of 5 mM. The sample was incubated for 15 minutes at 20°C to allow complex formation. The complex was crystallized at 20°C using the hanging drop vapour diffusion method by mixing equal volumes of complex and reservoir solution. Initial crystals were obtained at a CbAgo DM concentration of 5 mg/ml with a reservoir solution consisting of 4 M Sodium Formate.

Data was collected from crystals grown obtained using a complex concentration of 4.3 mg/ml and reservoir solution containing 3.8 M Sodium Formate and 5 mM NiCl2 at 20°C. For cryoprotection, crystals were transferred to a drop of reservoir solution and flash-cooled in liquid nitrogen. X-ray diffraction data were measured at beamline X06DA (PXIII) of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). Data were indexed, integrated, and scaled using AutoPROC (Vonrhein et al (2011)). Crystals of the CbAgo-siDNA-target DNA complex diffracted to a resolution of 3.55 Å and belonged to space group P63 2 2, with one copy of the complex in the asymmetric unit. The structure was solved by molecular replacement using Phaser-MR (McCoy et al., 2007). As search model, the structure of TtAgo in complex with guide and target DNA strands (PDB: 5GQ9) was used after removing loops and truncating amino acid side chains. Phases obtained using the initial molecular replacement solution were improved by density modification using phenix.resolve (Terwilliger 2003) and phenix.morph_model (Terwilliger 2013). The atomic model was built manually in Coot (Emsley et al., 2010) and refined using phenix.refine (Afonine et al., 2012). The final binary complex model contains CbAgo residues 1-463 and 466-748, guide DNA residues 1–16, and target DNA residues (-18)–(-1).

3.4.8. STRUCTURE ANALYSIS

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

3.4.9. SINGLE-MOLECULE EXPERIMENTAL SET-UP

Single-molecule fluorescence FRET measurements were performed with a prism-type total internal reflection fluorescence microscope. Cy3 and Cy5 molecules were excited with 532 nm and 637 nm wavelength, respectively. Resulting Cy3 and Cy5 fluorescence signal was collected through a 60X water immersion objective (UplanSApo, Olympus) with an inverted microscope (IX73, Olympus) and split by a dichroic mirror (635dcxr, Chroma). Scattered laser light was blocked out by a triple notch filter (NF01-488/532/635, Semrock). The Cy3 and Cy5 signals were recorded using a EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology) with exposure time 0.1 s. All single-molecule experiments were done at room temperature (22 ± 2C). Fluorescent DNA and RNA preparation The RNAs with amine-modification (amino-modifier C6-U phosphoramidite, 10-3039, Glen Research) were purchased from STPharm (South Korea) and DNAs with amine-modification (internal amino modifier iAmMC6T) Ella biotech (Germany). The guide and target strands were labeled with donor (Cy3) and acceptor (Cy5), respectively, using the NHS-ester form of Cy dyes (GE Healthcare). 2012).1 µL of 1 mM of DNA/RNA dissolved in MilliQ H20 is added to 5 µL labeling buffer of (freshly prepared) sodiumtetraborate (380 mg/10mL, pH 8.5). 1 µL of 20 mM dye (1 mg in 56 µL DMSO) is added and incubated overnight at room temperature in the dark, followed by washing and ethanol precipitation. The labeling efficiency was 100%.

3.4.10. Single-molecule sample preparation

A microfluidic chamber was incubated with 20 μL Streptavidin (0.1 mg/mL, Sigma) for 30 sec. Unbound Streptavidin was washed with 100 μL of buffer T50 (10 mM Tris-HCl [pH8.0], 50 mM NaCl buffer). The fifty microliters of 50 pM acceptor-labelled target construct were introduced into the chamber and incubated for 1 min. Unbound labeled constructs were washed with 100 μL of buffer T50. The CbAgo binary complex was formed by incubating 10 nM purified CbAgo with 1 nM of donor-labeled guide in a buffer containing 50 mM Tris-HCl [pH 8.0] (Ambion), 1mM MnCl2, and 100 mM NaCl (Ambion) at 37°C for 20 min. For binding rate (kon) measurements, the binary complex was introduced into the fluidics chamber using syringe during the measurement. The experiments were performed at the room temperature (23 ± 1°C). For fluorescence Guide Loading Experiments before immobilizing CbAgo on the single-molecule surface, 1 μL of 5 μM His-tagged apo-CbAgo was incubated with 1 μL of 1 μg/ml biotinylated anti-6x His antibody (Abcam) for 10 min. Afterward, the mixture was diluted 500x in T50 and 50 μL were loaded in the microfluidic channel for 30 s incubation, followed by washing with 100 μL of T50 buffer. Cy3-labeled ssDNA (0.1) was applied to the microfluidic chamber in imaging buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM MnCl2, 1 mM Trolox ((±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid), supplemented with an oxygen-scavenging system (0.5 mg/mL glucose oxidase (Sigma), 85 mg/mL catalase (Merck), and 0.8% (v/v) glucose (Sigma)).

3.4.11. Single-molecule data acquisition and analysis

CCD images of time resolution 0.1 or 0.3 sec were recorded, and time traces were extracted from the CCD image series using IDL (ITT Visual Information Solution). Co-localization between Cy3 and Cy5 signals was carried out with a custom-made mapping algorithm written in IDL. The extracted time traces were processed using Matlab (MathWorks) and Origin (Origin Lab). The binding rate (kon) was determined by first measuring the time between when CbAgo binary complex was introduced to a microfluidic chamber and when the first CbAgo- guide docked to a target; and then fitting the time distribution with a single-exponential growth curve, $A(1 - e^{-k_o nt})$. The dissociation rate was estimated by measuring the dwell time of a binding event. A dwell time distribution was fitted by single-exponential decay curve ($Ae^{-t/\Delta\tau}$).

3.4.12. Fluorescence competition experiments

MBP-tagged CbAgo was immobilized on the quartz surface using an anti-MBP antibody. An equimolar mixture of let7 DNA guide (Cy3 labeled) and let7 RNA guide (Cy5 labeled) in imaging buffer was introduced to the microfluidic chamber. After 5 minutes, 10 snapshots of independent fields of view with simultaneous illumination were collected to estimate the amount of guide molecules bound to protein. Movies were taken for 200 s (2000 frames) at continuous illumination of Cy3 and Cy5 molecules to determine the dwell times of the binding events. Dwell times were binned in a histogram and fitted with a single exponential decay curve. FRET targeting experiments of ATTT and AAAA guide target combinations 100 pM of target construct annealed with biotin handle were flushed in the microfluidic chamber. After incubation of 1 min, the microfluidic chamber was rinsed with 100 uL T50 buffer. 10 nM of apo-CbAgo was loaded with 1 nM

of ATTT seed DNA guide or with AAAA seed DNA guide at 37°C for 30 minutes in imaging buffer after which the mixture is introduced inside the microfluidic chamber. Movies of 200 s were taken at continuous illumination of the Cy3 signal. Site specific protein target interactions were identified as FRET signals and were further analysed.

3

REFERENCES

[1] Emine Kaya, Kevin W. Doxzen, Kilian R. Knoll, Ross C. Wilson, Steven C. Strutt, Philip J. Kranzusch, and Jennifer A. Doudna. A bacterial Argonaute with noncanonical guide RNA specificity. *Proceedings of the National Academy of Sciences*, 113(15):4057–4062, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1524385113. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1524385113.

[2] René F. Ketting. MicroRNA biogenesis and function: An overview. *Advances in Experimental Medicine and Biology*, 700:1–14, 2010. ISSN 00652598. doi: 10.1007/978-1-4419-7823-3_1.

[3] Gunter Meister. Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*, 14(7):447–59, 2013. ISSN 1471-0064. doi: 10.1038/nrg3462.

[4] David P. Bartel. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136 (2):215–233, 2009. ISSN 00928674. doi: 10.1016/j.cell.2009.01.002.

[5] Ashley J. Pratt and Ian J. MacRae. The RNA-induced silencing complex: A versatile gene-silencing machine. *Journal of Biological Chemistry*, 284(27):17897–17901, 2009. ISSN 00219258. doi: 10.1074/jbc.R900012200.

[6] Claus D. Kuhn and Leemor Joshua-Tor. Eukaryotic Argonautes come into focus. *Trends in Biochemical Sciences*, 38(5):263–271, 2013. ISSN 09680004. doi: 10.1016/j.tibs.2013.02.008.

[7] Gyorgy Hutvagner and Martin J. Simard. Argonaute proteins: Key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, 9(1):22–32, 2008. ISSN 14710072. doi: 10.1038/nrm2321.

[8] René F. Ketting. The Many Faces of RNAi. *Developmental Cell*, 20(2):148–161, 2011. ISSN 15345807. doi: 10.1016/j.devcel.2011.01.012.

[9] Kira S S Makarova, Yuri I I Wolf, John van der Oost, and Eugene V V Koonin. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology direct*, 4(1):29, 2009. ISSN 1745-6150. doi: 10.1186/1745-6150-4-29. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2743648{&}tool=pmcentrez{&}rendertype=abstracthttp://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-4-29.

[10] Daan C Swarts, Kira Makarova, Yanli Wang, Kotaro Nakanishi, René F Ketting, Eugene V Koonin, Dinshaw J Patel, and John van der Oost. The evolutionary journey of Argonaute proteins. *Nature structural & molecular biology*, 21(9):743–53, 2014. ISSN 1545-9985. doi: 10.1038/nsmb.2879. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4691850{&}tool=pmcentrez{&}rendertype=abstract.

[11] Ji-joon Song, Stephanie K Smith, Gregory J Hannon, and Leemor Joshua-Tor. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science*, 305(September):1434–1437, 2004. ISSN 0036-8075. doi: 10.1126/science.1102514.

[12] Yanli Wang, Stefan Juranek, Haitao Li, Gang Sheng, Thomas Tuschl, and Dinshaw J Patel. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, 456(7224):921–926, 2008. ISSN 0028-0836. doi: 10.1038/nature07666. URL http://dx.doi.org/10.1038/nature07666.

[13] Ivan Olovnikov, Ken Chan, Ravi Sachidanandam, Dianne K. DianneK Newman, and Alexei A. AlexeiA Aravin. Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. *Molecular Cell*, 51(5):594–605, sep 2013. ISSN 10972765. doi: 10.1016/j.molcel.2013.08.014. URL http://dx.doi.org/10.1016/j.molcel. 2013.08.014https://www.sciencedirect.com/science/article/pii/ S109727651300587X?via{%}3Dihub.

[14] Daan C Swarts, Matthijs M Jore, Edze R Westra, Yifan Zhu, Jorijn H Janssen, Ambrosius P Snijders, Yanli Wang, Dinshaw J Patel, José Berenguer, Stan J J Brouns, and John van der Oost. DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*, 507(7491):258–61, 2014. ISSN 1476-4687. doi: 10.1038/nature12971. URL http://www.ncbi.nlm.nih.gov/pubmed/24531762.

[15] Daan C. Swarts, Jorrit W. Hegge, Ismael Hinojo, Masami Shiimori, Michael A. Ellis, Justin Dumrongkulraksa, Rebecca M. Terns, Michael P. Terns, and John Van Der Oost. Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Research*, 43(10):5120–5129, 2015. ISSN 13624962. doi: 10.1093/nar/gkv415.

[16] Adrian Zander, Sarah Willkomm, Sapir Ofer, Marleen van Wolferen, Luisa Egert, Sabine Buchmeier, Sarah Stöckl, Philip Tinnefeld, Sabine Schneider, Andreas Klingl, Sonja-Verena Albers, Finn Werner, and Dina Grohmann. Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nature Microbiology*, 2(March):17034, 2017. ISSN 2058-5276. doi: 10.1038/nmicrobiol. 2017.34. URL http://www.nature.com/articles/nmicrobiol201734.

[17] Daan C. Swarts, Jasper J. Koehorst, Edze R. Westra, Peter J. Schaap, and John Van Der Oost. Effects of argonaute on gene expression in Thermus thermophilus. *PLoS ONE*, 10(4):1–13, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0124880.

[18] Daan C C. Swarts, Malwina Szczepaniak, Gang Sheng, Stanley D D. Chandradoss, Yifan Zhu, Elizabeth M. M. Timmers, Yong Zhang, Hongtu Zhao, Jizhong Lou, Yanli Wang, Chirlmin Joo, and John van der Oost. Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Molecular Cell*, 65(6):985–998.e6, 2017. ISSN 10974164. doi: 10.1016/j.molcel.2017.01.033.

[19] S. Willkomm, C. A. Oellig, A. Zander, T. Restle, R. Keegan, D. Grohmann, and S. Schneider. Structural and mechanistic insights into the DNA-guided DNA endonuclease activity of an archaeal Argonaute. *Nature Microbiology*, 17035 (March):1–7, 2017. ISSN 2058-5276. doi: 10.1038/nmicrobiol.2017.35.

[20] Adrian Zander, Phil Holzmeister, Daniel Klose, Philip Tinnefeld, and Dina Grohmann. Single-molecule FRET supports the two-state model of Argonaute action. *RNA biology*, 11(1):45–56, 2014. ISSN 1555-8584. doi: 10.4161/

rna.27446. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3929424{&}tool=pmcentrez{&}rendertype=abstract`.

[21] Svetlana A. Shabalina and Eugene V. Koonin. Origins and evolution of eukaryotic RNA interference. *Trends in Ecology and Evolution*, 23(10):578–587, 2008. ISSN 01695347. doi: 10.1016/j.tree.2008.06.005.

[22] Behnam Enghiad and Huimin Zhao. Programmable DNA-Guided Artificial Restriction Enzymes. *ACS Synthetic Biology*, (February), 2017. ISSN 2161-5063. doi: 10.1021/acssynbio.6b00324.

[23] Christof Fellmann, Benjamin G. Gowen, Pei Chun Lin, Jennifer A. Doudna, and Jacob E. Corn. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nature Reviews Drug Discovery*, 16(2):89–100, 2017. ISSN 14741784. doi: 10.1038/nrd.2016.238.

[24] Gavin J. Knott and Jennifer A. Doudna. CRISPR-Cas guides the future of genetic engineering. *Science*, 361(6405):866–869, 2018. ISSN 0036-8075. doi: 10.1126/science.aat5011.

[25] Nataša Savić and Gerald Schwank. Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research*, 168:15–21, feb 2016. ISSN 1931-5244. doi: 10.1016/J.TRSL.2015.09.008.

[26] Jorrit W. Hegge, Daan C. Swarts, and John Van Der Oost. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nature Reviews Microbiology*, 16(1):5–11, 2018. ISSN 17401534. doi: 10.1038/nrmicro.2017.73. URL `http://dx.doi.org/10.1038/nrmicro.2017.73`.

[27] Feng Gao, Xiao Z Shen, Feng Jiang, Yongqiang Wu, and Chunyu Han. DNA-guided genome editing using the Natronobacterium gregoryi Argonaute. *Nat Biotech*, 34 (7):768–772, may 2016. ISSN 1546-1696.

[28] David Cyranoski. Replications, ridicule and a recluse: the controversy over NgAgo gene-editing intensifies. *Nature*, 536:136–137, 2016.

[29] Seung Hwan Lee, Giandomenico Turchiano, Hirotaka Ata, Somaira Nowsheen, Marianna Romito, Zhenkun Lou, Seuk-Min Ryu, Stephen C Ekker, Toni Cathomen, and Jin-Soo Kim. Failure to detect DNA-guided genome editing using Natronobacterium gregoryi Argonaute. *Nature Biotechnology*, pages 3–4, 2016. ISSN 1087-0156. doi: 10.1038/nbt.3753.

[30] Sunghyeok Ye, Taegeun Bae, Kyoungmi Kim, Omer Habib, Yoon Seung Hwan Lee, Young Kim, Kang-In Lee, Seokjoong Kim, Kim, and Jin-Soo. DNA-dependent RNA cleavage by the Natronobacterium gregoryi. *preprint at bioRxiv*, pages 1–9, 2017.

[31] Kotaro Nakanishi, David E Weinberg, David P Bartel, and Dinshaw J Patel. Structure of yeast Argonaute with guide RNA. *Nature*, 486:368, may 2012. URL `http://dx.doi.org/10.1038/nature11211http://10.0.4.14/nature11211https://www.nature.com/articles/nature11211{#}supplementary-information`.

[32] Gang Sheng, Hongtu Zhao, Jiuyu Wang, Yu Rao, Wenwen Tian, Daan C Swarts, John van der Oost, Dinshaw J Patel, and Yanli Wang. Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 111(2):652–7, 2014. ISSN 1091-6490. doi: 10.1073/pnas.1321032111. URL http://www.pnas.org/content/111/2/652.full.

[33] Yanli Wang, Stefan Juranek, Haitao Li, Gang Sheng, Greg S. Wardle, Thomas Tuschl, and Dinshaw J. Patel. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, 461(7265):754–761, 2009. ISSN 0028-0836. doi: 10. 1038/nature08434. URL http://www.ncbi.nlm.nih.gov/pubmed/19812667.

[34] Rupert J.M. Russell, Jacqueline M.C. Ferguson, David W. Hough, Michael J. Danson, and Garry L. Taylor. The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiosus at 1.9 Å resolution. *Biochemistry*, 36(33):9983–9994, 1997. ISSN 00062960. doi: 10.1021/bi9705321.

[35] Michael J. Thompson and David Eisenberg. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *Journal of Molecular Biology*, 290(2):595–604, 1999. ISSN 00222836. doi: 10.1006/jmbi.1999. 2889.

[36] Yiwei Liu, Daria Esyunina, Ivan Olovnikov, Marianna Teplova, Andrey Kulbachinskiy, Alexei A. Aravin, and Dinshaw J. Patel. Accommodation of Helical Imperfections in Rhodobacter sphaeroides Argonaute Ternary Complexes with Guide RNA and Target DNA. *Cell Reports*, 24(2):453–462, 2018. ISSN 22111247. doi: 10.1016/j.celrep.2018.06.021.

[37] Nicole T. Schirle, Jessica Sheu-Gruttadauria, Stanley D. Chandradoss, Chirlmin Joo, and Ian J. MacRae. Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. *eLife*, 4(September):1–16, 2015. ISSN 2050084X. doi: 10.7554/eLife.07646.

[38] Stanley D. Chandradoss, Nicole T. Schirle, Malwina Szczepaniak, Ian J. Macrae, and Chirlmin Joo. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell*, 162 (1):96–107, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.06.032. URL http://dx. doi.org/10.1016/j.cell.2015.06.032.

[39] N. T. Schirle, J. Sheu-Gruttadauria, and I. J. MacRae. Structural basis for microRNA targeting. *Science*, 346(6209):608–613, 2014. ISSN 0036-0075. doi: 10.1126/science. 1258040. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi? artid=4313529{&}tool=pmcentrez{&}rendertype=abstract.

[40] Liang Meng Wee, C. Fabián Flores-Jasso, William E. Salomon, and Phillip D. Zamore. Argonaute divides Its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*, 151(5):1055–1067, 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.10.036.

[41] Alexei Aravin, Dimos Gaidatzis, Sébastien Pfeffer, Mariana Lagos-Quintana, Pablo Landgraf, Nicola Iovino, Patricia Morris, Michael J. Brownstein, Satomi

Kuramochi-Miyagawa, Toru Nakano, Minchen Chien, James J. Russo, Jingyue Ju, Robert Sheridan, Chris Sander, Mihaela Zavolan, and Thomas Tuschl. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442(7099):203–207, 2006. ISSN 14764687. doi: 10.1038/nature04916.

[42] Filipp Frank, Nahum Sonenberg, and Bhushan Nagar. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 465 (7299):818–822, 2010. ISSN 0028-0836. doi: 10.1038/nature09039.

[43] Filipp Frank, Jesse Hauver, Nahum Sonenberg, and Bhushan Nagar. Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *The EMBO journal*, 31(17):3588–95, 2012. ISSN 1460-2075. doi: 10.1038/emboj. 2012.204.

[44] William E. Salomon, Samson M. Jolly, Melissa J. Moore, Phillip D. Zamore, and Victor Serebrov. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell*, 162(1):84–95, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.06.029. URL http://dx.doi.org/10.1016/j.cell.2015.06.029.

[45] Daan C. Swarts and Martin Jinek. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdisciplinary Reviews: RNA*, 9(5):1–19, 2018. ISSN 17577012. doi: 10.1002/wrna.1481.

[46] Carolin Anders, Ole Niewoehner, Alessia Duerst, and Martin Jinek. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513 (7519):569–573, 2014. ISSN 14764687. doi: 10.1038/nature13579.

[47] Daan C. Swarts, John van der Oost, and Martin Jinek. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular Cell*, 66(2):221–233.e4, apr 2017. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2017.03.016.

[48] Takashi Yamano, Bernd Zetsche, Ryuichiro Ishitani, Feng Zhang, Hiroshi Nishimasu, and Osamu Nureki. Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. *Molecular Cell*, 67(4): 633–645.e3, aug 2017. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2017.06.035.

[49] Joseph E Tropea, Scott Cherry, and David S Waugh. Expression and Purification of Soluble His6-Tagged TEV Protease. In Sharon A Doyle, editor, *High Throughput Protein Expression and Purification: Methods and Protocols*, pages 297–307. Humana Press, Totowa, 2009. ISBN 978-1-59745-196-3. doi: 10.1007/978-1-59745-196-3_19.

4

ARGONAUTE BYPASSES CELLULAR OBSTACLES WITHOUT HINDRANCE DURING TARGET SEARCH

Argonaute (Ago) proteins are key players in both gene regulation (eukaryotes) and host defense (prokaryotes). Acting on single-stranded nucleic-acid substrates, Ago relies on base pairing between a small nucleic-acid guide and its complementary target sequences for specificity. To efficiently scan nucleic-acid chains for targets, Ago diffuses laterally along the substrate and must bypass secondary structures as well as protein barriers. Using single-molecule FRET in conjunction with kinetic modelling, we reveal that target scanning is mediated through loose protein-nucleic acid interactions, allowing Ago to slide short distances over secondary structures, as well as to bypass protein barriers via intersegmental jumps. Our combined single-molecule experiment and kinetic modelling approach may serve as a novel platform to dissect search process and study the effect of sequence on search kinetics for other nucleic acid-guided proteins.

4.1. INTRODUCTION

Target recognition by oligonucleotide guides is essential in cellular development, differentiation and immunity [1, 2]. Argonaute (Ago) proteins are key mediators of the target interference process, utilizing short oligo-nucleotides (20-30 nt) as guides for finding complementary target sequences [3, 4]. The guide-target interaction initiates at the 5' end of the guide, and progresses through Watson-Crick base pairing at the "seed" segment, which propagates along the guide, resulting in target interference upon completion [5]. While eukaryotic Argonautes use RNA guides to target RNA, prokaryotic Agos (pAgo) have been demonstrated to use a variety of guides and targets [6–8]. Depending on the pAgo type, it uses either DNA or RNA guides to target single-stranded (ss) DNA, ssRNA or both2. The ability of pAgos to cleave ssDNA but not double stranded DNA (dsDNA) suggests a physiological role as a host defense system against ss mobile genetic elements6–8. Recently, a new family of CRISPR-Cas systems that targets ssDNA—not dsDNA—have been discovered in archaea, suggesting that these defense systems may be more widespread than previously thought [9]. The number of potential targets encoded in cellular DNA/RNA is vast [5, 10, 11] and Ago needs to search long stretches of polymer before finding a canonical target. Single-molecule studies have shown that a mixture of excursions into solution and one-dimensional movements results in a search that is orders of magnitude more efficient than is possible without lateral diffusion [12, 13]. In a previous biophysical study we suggested that human Argonaute 2 (hAGO2) uses lateral diffusion along RNA for target search [14]. Yet, the degree of lateral diffusion remains unclear, as excessive usage of 1D diffusion would lead to redundant re-sampling of potential target sites and to problems at various roadblocks present on the target nucleic acids [15, 16]. In addition to complete dissociation into solution, intersegmental jumping, in which a protein transfers between two spatially close-by segments, has been shown to occur for DNA binding proteins such as restriction enzyme EcoRV [17]. After binding to DNA non-specifically from solution, the protein diffusively scans only a limited section [12, 18–20], and dissociates into solution before rebinding to a new section. Use of such a mechanism would lead to reduced sampling redundancy, and the possibility to circumvent obstructions when proteins search for their targets.

Previous studies have shown that certain DNA/RNA-guided proteins interact with DNA through non-specific electrostatic interactions [21–23], but the strength of these interactions and their behaviour on roadblocks and secondary structures is not understood. Since these interactions are typically short-ranged [24–26] and short-lived [14, 21, 23, 24, 27–29], a method offering high spatio-temporal resolution is required to study these interactions. Here we make use of single molecule Förster Resonance Energy Transfer (FRET) to elucidate the mechanism of ssDNA target search by a mesophilic Ago from the bacterium *Clostridium butyricum* (CbAgo). We show that CbAgo does not remain in tight contact with the DNA backbone, enabling it to bypass secondary structures along the nucleic-acid chain—all while retaining the ability to recognize its target. After sliding locally, the protein is able to reach distant sites (>100 nt) along the DNA through intersegmental jumps and then resumes sliding. These different modes of facilitated diffusion allow Ago to rapidly search through nucleic acid segments, as well as to bypass substantial obstacles during target scanning.

4.2. RESULTS

4.2.1. SINGLE-MOLECULE KINETICS OF CBAGO BINDING

To elucidate the complexity of the target search mechanism, we made use of the high spatial sensitivity of single-molecule FRET. We studied a minimal Argonaute complex that consists of CbAgo, loaded with a 22-nt DNA guide (small interfering DNA, siDNA) [30]. By using total internal reflection fluorescence (TIRF) microscopy, we recorded the interactions of CbAgo-siDNA with target DNA. Target DNA was immobilized on a PEG-coated quartz surface in a microfluidic chamber through biotin-streptavidin conjugation. Guide-loaded CbAgo was introduced to the microfluidic chamber by flow. The target was embedded within a poly-thymine sequence and labelled with an acceptor dye (Cy5) (Figure 4.1a). The guide construct was labelled at nt 9 from the 5'-end with a donor dye (Cy3) (Figure 4.1b). A 532-nm laser excitation resulted in donor excitation when the protein loaded with the guide DNA interacted with the target DNA. Once the CbAgo-siDNA complex became bound to the target, the proximity of the donor dye to the acceptor dye on the target resulted in high FRET efficiency. This was followed by a sudden disappearance of the signal, indicating that the complex dissociated from the target and diffused into the free solution. Freely diffusing molecules move too rapidly (~ μs) in and out of the evanescent field for the current time resolution of the experimental setup (100 ms) and were therefore not recorded. We found that CbAgo is not able to target dsDNA directly (Supplementary Fig. 1a-b). Likewise, when a ssDNA target with one base pair complementarity to the seed motif of the guide was used, only transient interactions (~0.45 s) were detected (Figure 4.1c-d), and no accurate binding profile could be extracted from the FRET histogram (Figure 4.1e).

To observe target search that involves intrinsically transient interactions, we determined the optimal target motif for recording binding events. The optimal motif should provide binding events longer than our detection limit of 100 ms, but still lead to dissociation events within the time of our measurement (200 s). To determine the optimal motif, the complementarity between guide and target was incrementally extended from nt 2 to 8 of the guide, showing a gradually increasing dwell time of the Ago-siDNA complex. We found that increasing the number of complementary base pairs above 6 resulted in stable binding beyond the photobleaching time (Supplementary Fig. 1c). To maintain weak interactions, we continued our experiments using a siDNA with three-base complementarity (N=3) with the target (nt 2-4) (Figure 4.1f). This gives a well-defined FRET population in the FRET histogram (Figure 4.1h), unlike one base-pair complementarity. Our estimation of the photobleaching rate (1.4 x 10^{-3} s^{-1}) (Supplementary Fig. 1d) was an order of magnitude lower than the dissociation rate (2.7 x 10^{-2} s^{-1}) (Figure 4.1g), indicating that photobleaching does not affect our estimation of the dissociation rate.

4.2.2. LATERAL DIFFUSION OF CBAGO

It was previously shown that an Ago-guide complex does not directly bind a specific target site from solution, but rather binds non-specifically to random positions along a surfaced-immobilized nucleic acid construct [14]. Such non-specific interactions of CbAgo-siDNA along target DNA are too short-lived to resolve in the absence of a

Figure 4.1: Single molecule imaging of target binding by siDNA:CbAgo complex. a, Immobilization scheme of the Argonaute-guide DNA complex. ssDNA is immobilized on a PEGylated quartz slide surface. Presence of the Ago-siDNA complex is detected by specific binding to target site (light yellow) resulting in high FRET. b, Sequences of guide (green) and target DNA. Guide is labelled on the 9th nucleotide position from the 5' side. c, Representative FRET trace of a single molecule experiment at 100 mM NaCl showing a transient interaction between CbAgo and a poly-T strand. Time resolution is 100 ms. d, Dwell time distribution of the Argonaute in absence of target motif. e, FRET values of the transient interactions of (d). f, Representative FRET trace of a single molecule experiment showing the interaction between CbAgo and a 2-4 nt (N=3) motif. g, Dwell time distribution Dwell time distribution of N=3 binding events with the mean dwell time of 37 s. h, FRET histogram of binding events, showing a single FRET population for N=3 at E=0.78.

canonical target motif (Figure 4.1c), and in the presence of such a motif there was still no lateral diffusion visible (Figure 4.1f). As we were unable to resolve lateral diffusion by CbAgo from non-specifically bound regions to the target, we questioned whether the observed stable signal for three complementary base pairs is due to stable binding to the target or contains lateral excursions away from the target but below our time resolution. In case of the latter, measured apparent dwell times (Figure 4.1g) would consist of the combined dwell times of many target escapes through lateral diffusion, each followed by rapid recapture below the detection limit, before CbAgo eventually unbinds from the DNA (Supplementary Fig. 1g). We show that such a process of repeated recapture would result in an exponential distribution of apparent dwell times, in accordance with Figure 4.1g (see Supplementary Note: Methods Kinetic Modelling). To overcome the temporal resolution limit, we adopted a tandem target assay [14, 31]. While lateral diffusive excursions from a trap are too short-lived to be resolved in the presence of only a single target, a second target can trap an excursion for long enough to be observed. We placed two identical optimal targets (site 1 and site 2) separated by 22 nt (Figure 4.2a) along the DNA construct. Both targets base pair only with the first three nucleotides (nt 2-4) of the

guide bound by CbAgo. As the second target is located further away from the acceptor dye, binding the second target results in a lower FRET efficiency than binding the first target. This difference in FRET values allows us to determine which of the two targets CbAgo-siDNA is bound to (Figure 4.2b). The respective distance and FRET efficiency between the first binding site (site 1) and the acceptor dye (Cy5) remained the same as for the single target assay (E 0.78), while an additional peak appeared at a lower FRET efficiency for the second target (E 0.43, Figure 4.2c). After binding to one of the target sites, a majority of the binding events (87.8%) resulted in CbAgo-siDNA shuttling to the other target without loss of FRET signal. Under our standard experimental condition (100 mM NaCl), an average of 13.5 shuttling events occur per binding event (Figure 4.2d). When the experiment was repeated for guides and targets with complementary increased to N=6 (nt 2-7), only 15.1% of the traces showed the shuttling signature within our time window (Supplementary Fig. 2f). This shows that the shuttling signature is controlled by interactions between CbAgo-ssDNA and the target motif. With a 6-nt match, the target is strongly bound, and we are less likely to observe a shuttling event within our observation window.

Interestingly, the average dwell time of the first target (Figure 4.1g) decreased from 37 s to 1.7 and 1.8 s after adding a second target in its vicinity (Figure 4.2e). This observation is in agreement with our lateral diffusion model, since with close-by targets, each sub-resolution diffusive excursion has some probability to be caught at the opposing target. To further test our claim that the transition between targets occur through lateral diffusion, we use single-molecule analysis software [32] to extract the average time between shuttling events ($\Delta\tau_{\text{shuttle}}$) from traces (Supplementary Fig. 3).

4.2.3. Kinetic modelling of lateral diffusion

To determine how lateral diffusion contributes to the shuttling, we kinetically model how $\Delta\tau_{\text{shuttle}}$ depends on the distance between traps. The DNA construct is modelled as a series of binding sites along which CbAgo will perform an unbiased random walk by stepping to neighboring nucleotides. The rate of stepping away from the target is k_{esc} in both directions, while at non-specific sites (poly-T), stepping is assumed to be near instantaneous—an approximation justified by the fact that lateral excursions are never resolved in the experiments. The time needed for FRET transitions to occur (named "shuttling time", $\Delta\tau_{\text{shuttle}}$) is equivalent to the apparent dwell time at a single FRET state. In the Supplementary Note: Kinetic Modelling we construct a diffusive model capturing the effect of Ago's repeated retrapping before shuttling to the other trap. The model shows that the shuttling time from the target grows linearly with the separation x_{target} between the targets

$$\Delta\tau_{\text{shuttle}}(x_{\text{target}}) = \frac{x_{\text{target}}}{k_{\text{esc}}} \qquad \text{(Supplementary Equation 1)}$$

The linear dependence of the shuttling time with trap separation might seem puzzling at first, given that diffusive timescales usually show a quadratic dependence on distances. Here though, it is not the diffusive steps themselves that directly contributes to the shuttling time, but rather the changing probability to getting retrapped before shuttling. In support of this model, we observed that the apparent shuttling time
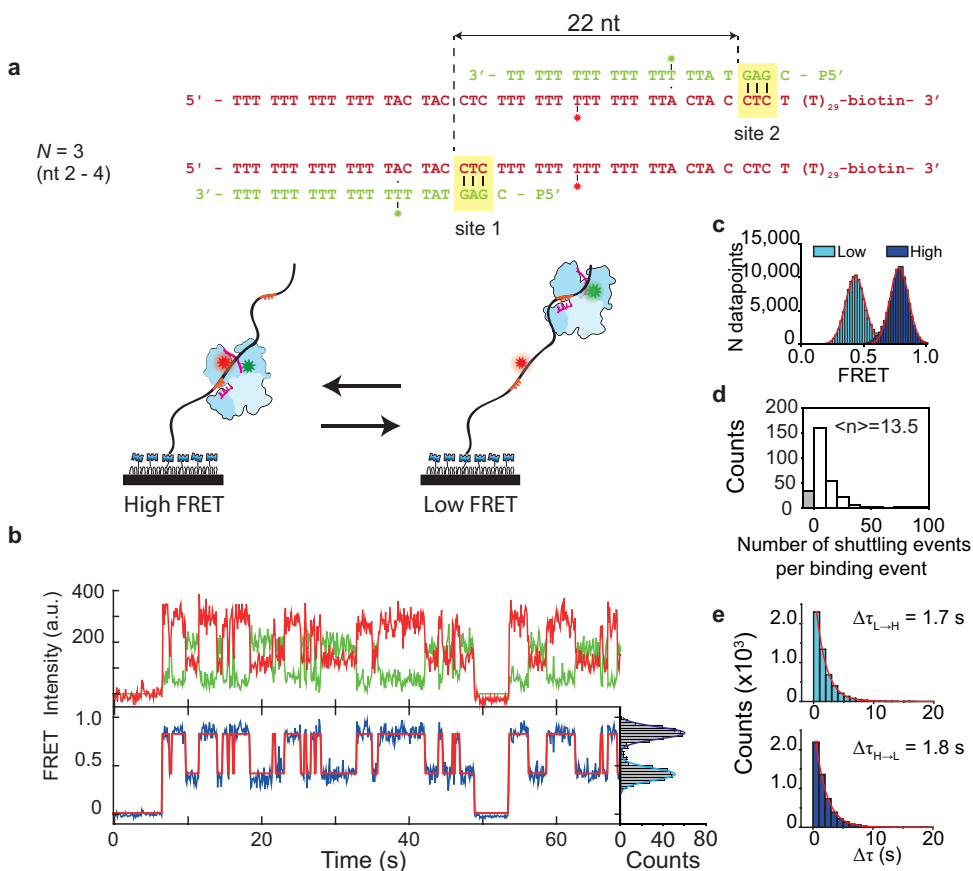
Figure 4.2: Shuttling signature of CbAgo appears in presence of two targets. a, In the top right corner the DNA sequence of guide and target for 22 nt separation between targets. Here the distance is defined as the distance from beginning of a target to the beginning of the next target. The placement of the second target (site 2) results in the appearance of an additional FRET signal, with lower FRET efficiency. b, (Top) Representative shuttling trace of a 22 nt separation tandem target at 100 mM NaCl for N=3. (Bottom) The corresponding FRET states (blue) with the fitted HMM trace on top (red). (Right) FRET histogram of the respective time trace. Time resolution is 100 ms. c, FRET histograms of respective states, with peaks at 0.43 and 0.78. d, Shuttling event distribution for the same conditions (n=309). Bin size = 10. On average 13.5 shuttling events take place before dissociation. The grey bar (n=33) marks binding events followed by dissociation (no shuttling). e, Dwell time distributions of respectively the transitions from low FRET state to high FRET state (top) and vice versa (bottom).

$\Delta\tau_{\text{shuttle}}(x_{\text{target}})$ increases approximately linearly when the distance between the targets increases through 11, 15, 18 and 22 nt (**Figure 4.3**). A fit to Equation 1 reveals that CbAgo-siDNA complexes escape the target site at a rate of 15.8 times per second ($k_{\text{esc}} = 15.8 s^{-1}$) in either direction.
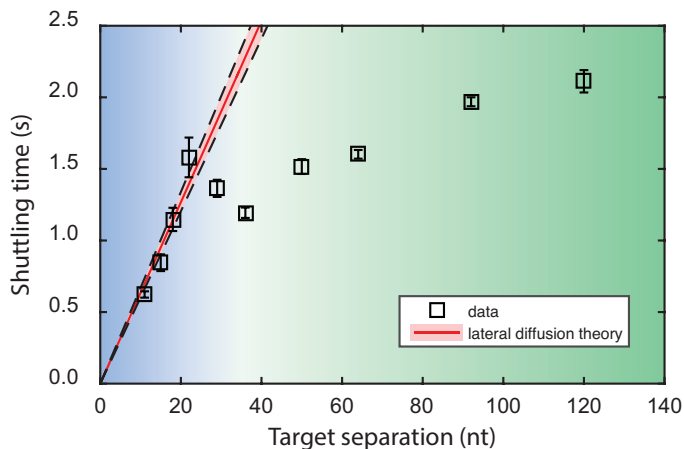
Figure 4.3: CbAgo shuttling behaviour differs across short and large distances Shuttling time is plotted versus distance between targets. Squares indicate the mean shuttling time for each DNA construct. The plotted error bars indicate the 95% confidence interval of 10^5 bootstrapped dwell times. The red line indicates the lateral diffusion model where the first four data points are fitted with $k_{esc} = 15.8 s^{-1}$. The shaded red region indicates its 95% confidence interval. The blue region indicates where the shuttling time follows lateral diffusion theory. This theory breaks down for larger distances (green).

4.2.4. AGO PROBES FOR TARGETS DURING LATERAL DIFFUSION

Next, we placed a third target on the tandem construct (Figure 4.4a), keeping the distance between each set of neighboring targets well within the regime for which we find good agreement to Equation 1 using the assay discussed above (i.e. at 11 nt trap separation, see Figure 4.3). We observed three different FRET levels, corresponding to CbAgo getting trapped at the three different targets (Figure 4.4b). Using Hidden Markov Modelling (HMM), states can be assigned (Figure 4.4b) and transition probabilities can be extracted (Figure 4.4c). If CbAgo returns to solution between binding targets, transitions between any pair of targets will be equally probable, resulting in equal effective rates between all targets. However, if lateral diffusion dominates, transitions between adjacent sites will be favored. The transition probabilities (Figure 4.4c) indicate that over 90% of the transitions between the two outer targets (from state A to C, or from C to A) proceed through the intermediate target site (state B). The rate to transfer from B to C and B to A is twice as much as that of the opposite path (A to B or C to B). Using the fitted escape rate from above, $k_{esc} = 15.8 s^{-1}$, we predict similar shuttling times based on our theoretical model for lateral diffusion (Figure 4.4d, Supplementary Note: Theoretical Modelling). With no more free-parameters remaining for this prediction, we take this experimental agreement with our prediction as further evidence of lateral diffusion. It is noteworthy that there are about 10% direct transitions from A to C and C to A without any intervening dissociation. The exponential distribution of the dwell times (Supplementary Fig. 4b) suggests that at our current time resolution this 10% may be either due to missed events or due to the existence of an additional translocation mode through which Ago is able to bypass the intermediate target.
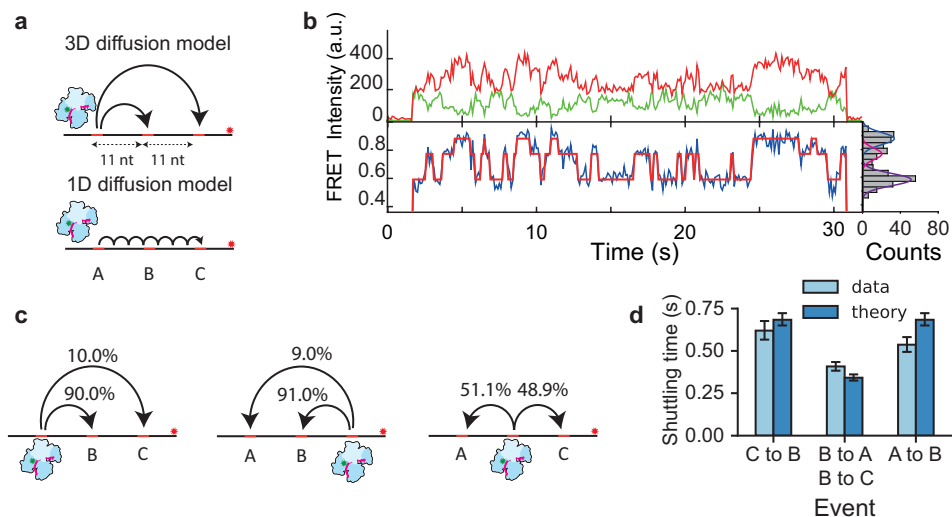
Figure 4.4: CbAgo undergoes short range diffusion through correlated steps.   a, Models for target translocation at short range. In the 3D diffusion model, target dissociation occurs from A followed by random 3D diffusion through solution. In effect, the neighboring two targets (B and C) will compete for binding. In the lateral diffusion model, the CbAgo complex will have to bypass the adjacent target B before binding to target C. b, Representative FRET trace showing the shuttling behavior between three targets. Top: donor (green) and acceptor (red) intensities. Bottom: FRET trace (blue) and HMM assigned states (red). Right: The fitted states from this data trace with dark blue: state C, pink: state B and purple state A. c, Transition probabilities from state A to B,C (left), from state C to A and B (middle) and from state B to A or C (right). d, Experimental values of the shuttling time of the three target construct were compared against the parameter-free theoretical model that only uses the $k_{esc} = 15.8s^{-1}$ from Figure 3. Error bars indicate the 95% confidence interval acquired from 10^5 bootstraps.

4.2.5. AGO TARGET SEARCH IS UNHINDERED BY STRUCTURAL AND PROTEIN BARRIERS

Secondary structures are commonly found in mRNA and are also predicted to exist in single stranded viruses [33, 34]. It is not known whether CbAgo is able to bypass the numerous junctions it encounters upon scanning a DNA segment. To examine this, a Y-fork structure (DNA junction) was introduced as a road block between two targets (Figure 4.5a), while keeping their separation (11 nt) the same as in the tandem target variant (Supplementary Fig. 4f). The construct was designed such that the labelled target was partially annealed at the stem with a biotinylated target, thus only annealed constructs were observable on the surface of the microfluidic device. When CbAgo binds to either of the two targets, it can reach the other target only by crossing the junction. Our measurement showed that there was no significant difference in shuttling time between the standard tandem-target construct and the Y-fork construct (Figure 4.5b-c), indicating that the Y-fork does not impede any of the lateral diffusion modes present. We have previously observed that the CbAgo-siDNA complex is not able to stably bind to dsDNA31, demonstrating that the protein cannot simply track the backbone of dsDNA (Supplementary Fig. 1a-b). Thus, our result suggests that the Ago-siDNA complex does not maintain tight contact with DNA during lateral diffusion.

Maintaining a weak interaction with the DNA molecule allows CbAgo-siDNA to move past the junction. Next, we questioned whether CbAgo is also able to overcome larger barriers, such as proteins which cannot reasonably be traversable through sliding alone. Lin28, a sequence-specific inhibitor of let-7 miRNA biogenesis, has been found to associate sequence specifically to RNA and DNA [35]. His-tagged Lin28 was immobilized on the surface of the microfluidic chamber (Supplementary Fig. 4d) after which a fluorescent ssDNA fragment was added containing a central Lin28 binding motif and an Ago target motif on either side (Figure 4.5d & Supplementary Fig. 4g). The presence of the protein blockade did not preclude Ago from reaching the distal site (Figure 4.5e) but noticeably broadened the FRET peak (Figure 4.5f), possibly due to protein-protein interactions. Although the shuttling rate was lowered from $0.60 s^{-1}$ to $0.27 s^{-1}$ (Figure 4.5g & Supplementary Fig. 4e), Ago is able to bypass the obstacle. Since short-range lateral movement is now blocked by the protein barrier, Ago's ability to move between targets demonstrates that the target search process also allows for intersegmental jumps, in accordance with our observation that the middle target is sometimes skipped when transitioning between the outer targets in Figure 4.4c.

4.2.6. AGO RELIES ON FLEXIBILITY OF DNA SEGMENTS OF BYPASSING BLOCKADES

Since Ago was observed to be able to bypass junctions and proteins, we questioned whether Ago could bypass other large-profile barriers. Previously, we observed that Ago only interacts transiently with dsDNA (Supplementary Fig. 1a-b) and thus we repurposed dsDNA as an extended blockade. We made a construct analogous to the tandem target construct used in Figure 4.2a, but the targets were separated by 36 nt and complementary strands of 17 nt, 21 nt, and 25 nt were annealed to the region in between the targets (Figure 4.5h-i). For the construct with a 17-nt blockade we observed a large number of shuttling events (shuttling probability 65.3% upon binding) indicating that a dsDNA blockade does not prohibit CbAgo from reaching the other site (Figure 4.5j and Figure 4.5l black squares). Upon extending the length of the dsDNA blockade, to 21 nt and 25 nt, we noticed a drop in the percentage of shuttling events (63.1% and 40.4% respectively) although shuttling still persisted (Supplementary Fig 5). Since the stiff segment of dsDNA decreases the shuttling probability, we conclude that Ago relies on the flexibility of segments for lateral diffusion. To further investigate the contribution of DNA flexibility, we used another construct which was shortened (by 15 nt from 19 nt) from the 5' side (Figure 4.5h bottom sequence). Here, ssDNA coiling was no longer possible from the 5' side of the DNA construct (Figure 4.5k). We measured a significant decrease (50%) in shuttling probability for all three blockades compared to the untruncated construct (Figure 4.5l), which supports that Ago relies on the flexibility of DNA segments when transferring between them.

4.2.7. AGO USES HOPS TO ACCESS DISTANT DNA SEGMENTS

Sliding is not expected to dominate across large distances, as the linear increase in shuttling time (Equation Supplementary Equation 1) would render the search process prohibitively slow. However, when CbAgo was studied with tandem targets that were separated 36 nt or more, we observed that the shuttling still persisted across larger
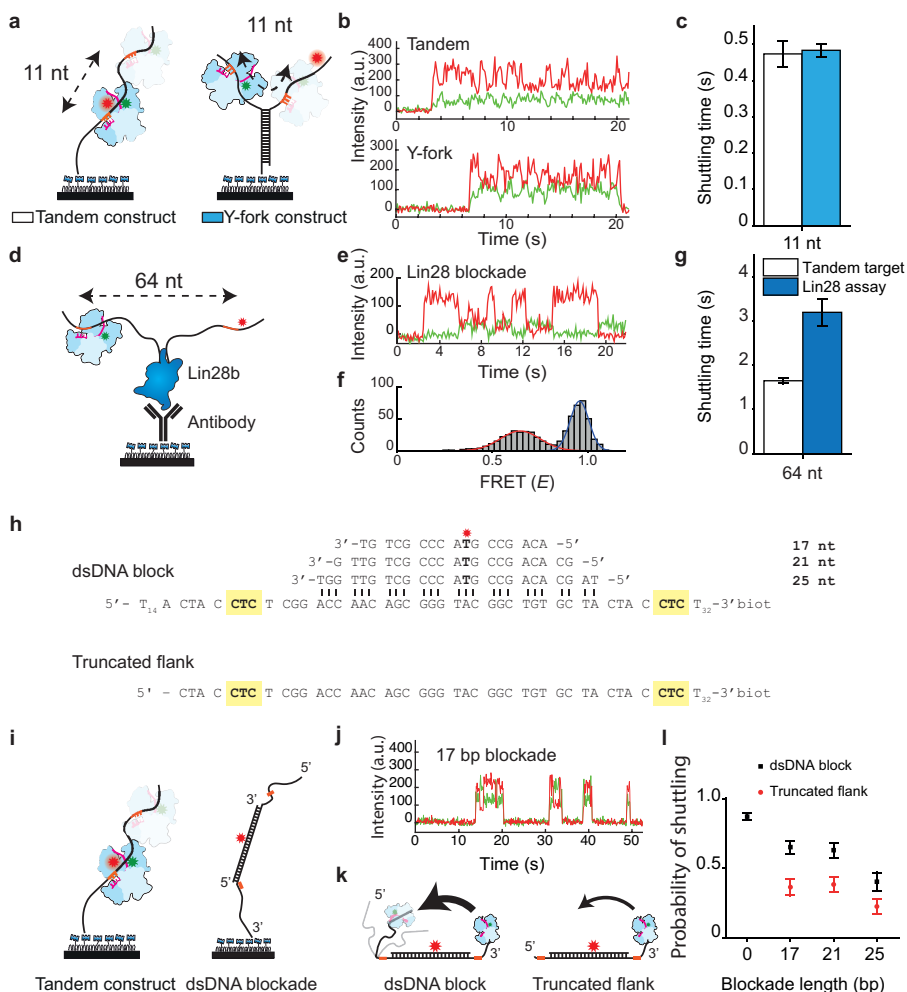
Figure 4.5: Argonaute can overcome structural and protein barriers. a, Schematic drawing tandem target assay (left) and the Y-fork assay (right) with 11 nt separation between targets. b, Representative shuttling traces of the tandem target assay (top) and Y-fork assay. c, The shuttling time of the Y-fork junction (blue bar) compared with the tandem assay (white bar). The experimental data of both sets were taken on the same days. Error bars indicate the 95% confidence interval acquired from 105 bootstraps d, Schematic drawing of the His-Lin28b blockade assay, where targets are separated by 64 nt. Immobilization happens through a biotin-anti-His antibody. e, Example of a shuttling trace with Lin28b located in between two targets. Exposure time is 100 ms. f, FRET histogram (molecules n = 46) fit with two Gaussian functions (E=0.64 for red fit and E=0.95 for dark blue fit). g, The shuttling time of the Lin28 assay compared with the tandem target assay for 64 nt separation between targets. h, Sequences used for the dsDNA block assay, indicating the base pairing between a 17 nt, 21 nt and a 25 nt long blockade and the target strand. The dsDNA block construct has a 19 nt flank on the 5' side, whereas the "truncated flank" has a 4 nt flank. i, Schematic of a dsDNA block assay, where the CTC targets are highlighted with orange. j, Representative trace of binding and shuttling of CbAgo on a 17 bp blockade DNA construct. k, (left) Schematic of dsDNA block construct with full length flanks. (right) schematic of the truncated version where the flank on the 5' side is removed. The thickness of the arrows indicate the observed shuttling probability. l, The probability of shuttling upon binding to a CTC target plotted versus the blockade length (none, 17 nt, 21 nt and 25 nt) for full length flanks (black squares) and for the truncated flanks (red circles). Error bars are given by the 95% confidence interval acquired from 10^5 bootstraps

distances (Figure 4.3, green region, Supplementary Table 1 and Supplementary Fig. 6). Together with the evidence of intersegmental jumping above, and the fact that the ssDNA can easily be coiled back to bring the second target close to the Ago protein [36], we speculate that there is a second mechanism of lateral diffusion: after local scanning for the target through sliding, the CbAgo complex jumps to a different part of the segment that has looped back into proximity of the complex. From this point on, we refer to these hops as intersegmental transfers in accordance with the current literature (Supplementary Fig. 7) [17, 37]. This intersegmental jumping mechanism would enable CbAgo to travel to new sites without fully dissociating, and rescanning of the same sections would be minimized [16, 18]. Based on the dependence of the single-target off-rate on the ionic strength (Supplementary Fig. 1f), we expect the rate of the intersegmental jumps to also be dependent on salt concentration, while sliding should only be moderately effected since it has no net effect on the ion condensation along the substrate. In order to test the hypothesis that short-ranged lateral diffusion is governed by sliding and long-range diffusion is governed by intersegmental jumps, we altered the ionic strength of the buffer solution from 10 mM NaCl to 200 mM NaCl. Here, we expect the degree of DNA coiling not to be significantly affected by the change in salt concentration, since the persistence length is only expected to vary between 20 Å and 14 Å when exchanging the buffers, and in both buffers it is smaller than the contour length of the constructs [36]. We used dual-target constructs with 15-nt separation and 64-nt separation (Figure 4.6), taken from the two different regions in Figure 4.3 (indicated by blue and green shading). At a separation of 64 nt, we observed a 13-fold increase of the shuttling rate when increasing the salt concentration from 10 mM NaCl to 200 mM NaCl. In contrast, we observed that for the dual-target construct with 15-nt separation, the shuttling time changed roughly only two-fold for the same change in ionic strength (Figure 4.6)—a modest change compared to 13-fold of the dual-target constructs with 64-nt separation. We take the relative ionic-strength insensitivity of shuttling times for 15-nt trap separation as evidence of translocation being dominate by sliding over short distances. In contrast, given the relative ionic-strength sensitivity for the 64-nt construct, the Ago complex is here unlikely to first reach the distal site through sliding only, and requires partial dissociation from the DNA strand. In conclusion, lateral diffusion during CbAgo target search is governed by two distinct modes. For short distances, lateral diffusion takes place through a sliding process characterized by loose contact with the DNA strand. This allows the protein to "glide" past secondary structures. To traverse larger distances, CbAgo is able to take advantage of the fact that the softness of the substrate allows it to bend back on itself to enable frequent intersegmental jumps between nearby segments (Supplementary Fig. 8).

4.3. Discussion

Within a vast number of potential targets, Ago-guide complexes have to minimize the time spent unproductively diffusing through solution or redundantly checking off-targets, as timely regulation is crucial for both cell development and host defense [38]. Our single-molecule study shows that Argonaute from C. butryicum (CbAgo) uses a loose sliding mode to bypass junctions and relies on intersegmental jumps to cover larger distances and to bypass substantial barriers.
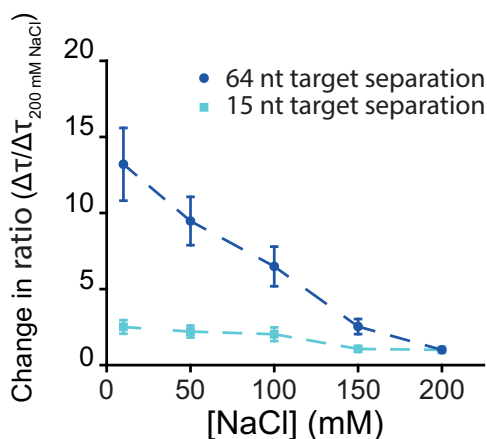
Figure 4.6: The relative change in shuttling time of two constructs from Figure 4.3, 64 nt separation (dark blue circles) and 15 nt separation (light blue squares), normalized against $\Delta\tau_{shuttle}$ at 200 mM NaCl. Errors of the ratio were determined through bootstrapping 10^5 times the ratio of $\Delta\tau/\Delta\tau_{200\text{ mM NaCl}}$

We have shown that bacterial Ago binds DNA loosely and slides along the DNA to locally scan for complementary targets. While such sliding mechanism has been characterized for several proteins [12, 20, 39, 40], little was previously known for DNA/RNA-guided target searchers like Ago. Proteins searching along nucleic acids with secondary structures may be blocked from sliding further. However, this does not seem to be true for Ago. Instead, the loose interaction with the substrate allows the protein to slide past junctions while still probing potential target sequence through base pairing. To the best of our knowledge, this mode of loose-contact sliding has not been reported for any nucleic-acid guided proteins. In addition, we show that the loose binding further allows Ago to move to a new segment via intersegmental jumps, reducing redundant scanning of the same segment and allowing Ago to bypass large-profile roadblocks.

The ability of CbAgo to target specifically ssDNA but not dsDNA [30] (Supplementary Fig. 1a-b) suggests a role as host defense against mobile genetic elements and ssDNA viruses. In environments where ssDNA viruses can be abundant, such as in sea water, fresh water, sediment, terrestrial, extreme, metazoan-associated and marine microbial mats [41–43], pAgo's targeting ssDNA would be very beneficial for the host. Upon entry in the infected cell, ssDNA binding and recombination proteins may associate with the invading nucleic acid, and DNA polymerase will start to generate the second strand. In addition, it is anticipated that secondary structures will be formed in the ssDNA viral genome [33]. This will generate road blocks that may affect scanning by defense systems such as restriction enzymes but—as shown here—not Argonaute. Likewise, insertion of transposons in prokaryotes often proceeds via a ssDNA-intermediate state [44–46], and pAgos may here encounter the same type of obstacles. In case of ssRNA, both in prokaryotes and in eukaryotes, it is well known that complex secondary structures can be formed by base pairing different anti-parallel RNA segments [47–50]. The presence of secondary structures suggests that it is necessary for Agos to "glide"—the type of loosely bound sliding we report—past such roadblocks to enable search along ssRNA. Based

on the functional and structural similarities of prokaryotic Agos and eukaryotic Agos [2, 14], we expect eAgo to also slide past RNA secondary structures, minimizing time spent trapped at such structures.

The effect of lateral diffusion on the total target search time is dependent on the roughness of the energy landscape that the DNA binding protein encounters once it binds non-specifically. We have shown how to determine the escape time for a 3-nt complementary target. This can be extended to estimate the escape time for any complementarity and consequently the diffusion constant on DNA with any base composition [51]. Here we have inferred a 15.8 s^{-1} escape rate from the 3-nt CTC guide sequence (Figure 4.3), indicating that if a target strand were to consists only of GA in repeating order, the effective diffusion $D = \frac{1}{2}\frac{dx^2}{dt} = \frac{\text{nt}^2}{2(2 \cdot k_{\text{esc}}^{-1})} = \text{nt}^2 k_{\text{esc}} = 15.8\frac{\text{nt}^2}{s}$. Changing the number of base-paring nucleotides as well as the identity of nucleotides in the guide/target could provide insights into how sequence variation would affect the rate of diffusion for other nucleic acid proteins. Since the guide strand only provides the specificity needed for accurate targeting, lateral diffusion could be reliant on the non-specific surface interactions with the protein. We envision that the positive surface charge distribution inside the Ago cleft could orient Ago with the guide towards the negatively charged nucleic acid strand (Supplementary Fig. 9), thereby promoting target interrogation while traveling along the target strand. It is unknown whether Ago is able to scan each base during this process or whether it skips over nucleotides. For our triple-target construct, we have observed that 90% of the time the middle target traps Ago. It will be of interest to investigate whether this level of effective target trapping is achieved by a low trapping efficiency offset by repeated passes over the target.

For a longer range target search, we have observed that at distances >100 nt separation, the shuttling time remains well below what would be expected for sliding (Figure 4.3). We show that coiling of the ssDNA (persistence length ∼ 1 nm) may bring distant segments in close proximity, allowing intersegmental jumps over longer distances (beyond 30 nt target separation), and so speeding up lateral diffusion. Interestingly, Ago cannot use intersegmental jumps to cover shorter distances, as implied by the sudden increase in shuttling time when the trap separation goes below 30 nt (Figure 4.3). Experimentally, one could further investigate the nature of intersegmental jumps through a combined tweezer-fluorescence single-molecule assay, where forces strong enough to pull on entropically coiled ssDNA can be applied [17, 37]. Furthermore, theoretical modelling and additional experiments are required in order to establish to what extent partitioning the search modes on different length scales will allow nucleic acid guided proteins to optimize the search process [52–54] since the absence of cooperative binding was recentley reported for another Ago system [29].

We hypothesize that similar target search strategies may be used by Agos from different families, which are structurally and functionally similar [2]. For example, in RNA induced transcriptional silencing (RITS), guide-loaded AGO1 binds to a transcript after which other proteins are recruited for heterochromatin assembly [55, 56]. Similarly, in the piRNA pathway PIWI proteins associate with piRNA in germline cells to bind and cleave transposon transcripts in the cytoplasm [57–59] or to nascent RNA in the nucleus in order to induce heterochromatin formation [60]. In each of these functions, the reliance on guide-complementary for sequential target search likely necessitates the

usage of facilitated diffusion strategies to optimize the search time for proper regulation of cell development or gene stability.

4.4. METHODS

4.4.1. PURIFICATION OF CBAGO

The CbAgo gene was codon harmonized for E.coli Bl21 (DE3) and inserted into a pET-His6 MBP TEV cloning vector (Addgene plasmid # 29656) using ligation independent cloning. The CbAgo protein was expressed in E.coli Bl21(DE3) Rosetta^{TM} 2 (Novagen). Cultures were grown at 37 °C in LB medium containing 50µg ml-1 kanamycin and 34µg ml-1 chloramphenicol till an OD600nm of 0.7 was reached. CbAgo expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1mM. During the expression cells were incubated at 18∘C for 16 hours with continues shaking. Cells were harvested by centrifugation and lysed, through sonication (Bandelin, Sonopuls. 30% power, 1s on/2s off for 5min) in lysis buffer containing 20mM Tris-HCl pH 7.5, 250mM NaCl, 5mM imidazole, supplemented with a EDTA free protease inhibitor cocktail tablet (Roche). The soluble fraction of the lysate was loaded on a nickel column (HisTrap Hp, GE healthcare). The column was extensively washed with wash buffer containing 20mM Tris-HCl pH 7.5, 250mM NaCl and 30mM imidazole. Bound protein was eluted by increasing the concentration of imidazole in the wash buffer to 250mM. The eluted protein was dialysed at 4oC overnight against 20mM HEPES pH 7.5, 250mM KCl, and 1mM dithiothreitol (DTT) in the presence of 1mg TEV protease (expressed and purified according to Tropea et al.63) to cleave of the His6-MBP tag. Next the cleaved protein was diluted in 20mM HEPES pH 7.5 to lower the final salt concentration to 125mM KCl. The diluted protein was applied to a heparin column (HiTrap Heparin HP, GE Healthcare), washed with 20mM HEPES pH 7.5, 125mM KCl and eluted with a linear gradient of 0.125-2M KCl. Next, the eluted protein was loaded onto a size exclusion column (Superdex 200 16/600 column, GE Healthcare) and eluted with 20mM HEPES pH 7.5, 500mM KCl and 1mM DTT. Purified CbAgo protein was diluted in size exclusion buffer to a final concentration of 5uM. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

4.4.2. PURIFICATION OF HIS-TAGGED LIN28B

The protein was prepared following the protocol of Yeom et al. [61].Briefly, recombinant Lin28b was prepared by subcloning cDNA with BamHI and XhoI into pET28-a vector (Novagen). Subsequently, the strain was transformed to E. coli BL21-RIL strain. The expression and purification of recombinant Lin28b was performed according to the manufacturer's protocol.

4.4.3. SINGLE MOLECULE EXPERIMENTAL SETUP

Single molecule FRET experiments were performed with an inverted microscope (IX73, Olympus) with prism-based total internal reflection. Excitation of the donor dye Cy3 is done by illuminating with a 532nm diode laser (Compass 215M/50mW, Coherent). A 60X water immersion objective (UPLSAPO60XW, Olympus) was used for collection of photons from the Cy3 and Cy5 dyes on the surface, after which a 532 nm long

pass filter (LDP01-532RU-25, Semrock) blocks the excitation light. A dichroic mirror (635 dcxr, Chroma) separates the fluorescence signal which is then projected onto an EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology). All experiments were performed at an exposure time of 0.1 s at room temperature (22 ± 0.1 °C)

4.4.4. FLUORESCENT DYE LABELING OF NUCLEIC ACID CONSTRUCTS
All DNA constructs were ordered from ELLA Biotech. Nucleic acid constructs that have an internal amino modification were labeled with fluorescent dyes based on the CSHL protocol 65.1 uL of 1 mM of DNA/RNA dissolved in MilliQ H20 is added to 5 uL labeling buffer of (freshly prepared) sodiumbicarbonate (84 mg/10mL, pH 8.5). 1 uL of 20 mM dye (1 mg in 56 uL DMSO) is added and incubated overnight at 4°C in the dark, followed by washing and ethanol precipitation. Concentration of nucleic acid and labeling efficiency was determined with a Nanodrop spectrophotometer.

4.4.5. SINGLE MOLECULE CHAMBER PREPARATION
Quartz slides were coated with a polyethylene-glycol through the use of amino-silane chemistry. This is followed by assembly of microfluidic chambers with the use of double sided scotchtape. For a detailed protocol, we refer to 66. Further improvement of surface quality occurs through 15 min incubation of T50 and 5% Tween20 67 after which the channel is rinsed with 100 μL T50 buffer. Streptavidin (5 mg/mL) was diluted in T50 to 0.1 mg/mL. 50 μL of this solution is then flowed inside the chamber. This is followed by incubation for 1 min followed by rinsing with approximately 10-fold the volume of the chamber with T50 (10 mM Tris-HCl [pH 8.0], 50 mM NaCl). 100 pM of DNA/RNA target with biotin construct is then flushed in the chamber, followed by 1 min incubation. This is followed subsequently by rinsing with T50. The chamber is subsequently flushed with CbAgo buffer, containing 50 mM Tris-HCl [pH 8.0], 1 mM Trolox, 1 mM MnCl2, 100 mM NaCl. Guide-loading of apo-CbAGO occurs by incubation of the protein (10 nM) with 1 nM guide construct in a buffer containing 50 mM Tris-HCl [pH 8.0], 1 mM Trolox, 1 mM MnCl2, 100 mM NaCl, 0.8% glucose at 37°C for 30 min. Following incubation, glucose oxidase and catalase is added (0.1 mg/mL glucose oxidase) after which the sample is flushed in the microfluidic chamber containing the DNA targets.

4.4.6. LIN28 ASSAY
Immobilization of Lin28b occurred in the following way: 50 μl of streptavidin (0.1 mg/mL) in T50 is flowed inside the chamber and incubated for 1 minute. After this, the chamber is rinsed with approximately 100 μL of T50. 1 μl of Anti-6X His tag® antibody (Biotin) diluted 100-fold in T50 and subsequently flowed inside the chamber. After 5 minutes, the chamber is rinsed with 100 μL of T50. Stock of Lin28b (100 μM) is diluted to 100 nM and incubated with the target DNA (10 nM) and 10 mM MgCl2 for 5 minutes, after which the solution is flushed inside the chamber, followed by incubation of 5 minutes. Lastly, the CbAgo buffer is flushed inside the chamber. Guide-loading of apo-CbAgo occurs in the same way as described above (Single molecule chamber preparation) after which the CbAgo:siDNA complex is also flushed inside the chamber.

4.4.7. QUANTIFICATION AND STATISTICAL ANALYSIS

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

4

4.5. SUPPLEMENTARY INFORMATION

4.5.1. SUPPLEMENTARY FIGURES
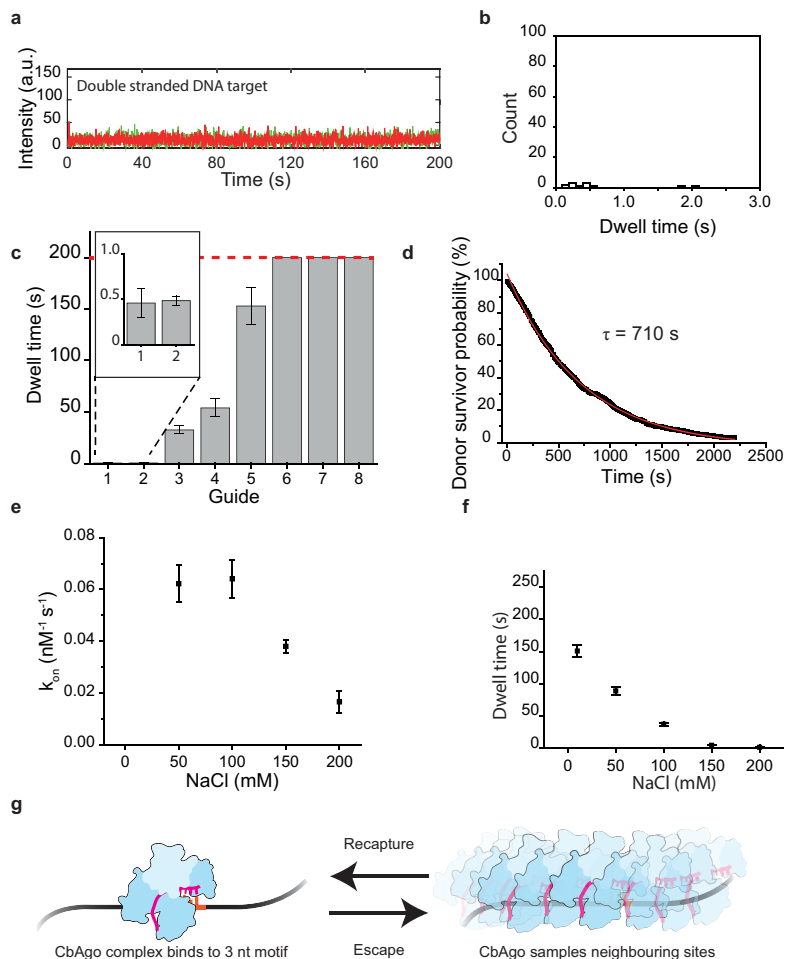
Supplementary Figure 1: Single molecule interactions of CbAgo:siDNA (2-4 nt) at different conditions. (a) Representative trace single-molecule interaction of CbAgo-siDNA (let7) with full target dsDNA target immobilized on the surface (300 per FoV). Exposure time is 100 ms. (b) Dwelltime distribution of CbAgo-guide 3-dsDNA target interactions. Number of molecules recorded n = 540. Number of datapoints n = 12 (c) Average dwell time of protein bound to target versus guide length for N=1 to N=8. The error bars are taken from the 95% confidence interval of bootstrapped dwelltimes (20,000 empirical bootstraps). The striped red line indicates the observation time, limited by photobleaching. (d) Survival plot of donor only (Cy3) constructs in standard experimental conditions (100 mM NaCl, 50 mM Tris-HCl pH 8.0). Mean donor bleaching time was obtained by a single exponential fit to the survival probability plot. (e) Binding rate for different salt concentrations for N=3 (nt 2-4) between guide and single target. (f) Dwell time of CbAgo and a single-stranded single target DNA construct (N=3) at 10, 50, 100, 150 and 200 mM NaCl concentration. Total measurement time = 250 s. Error bars are indicating the 95% percentile of 20,000 empirical bootstraps of the mean dwell time. (G) Schematic image indicating the dynamic escape and recapture events of CbAgo.
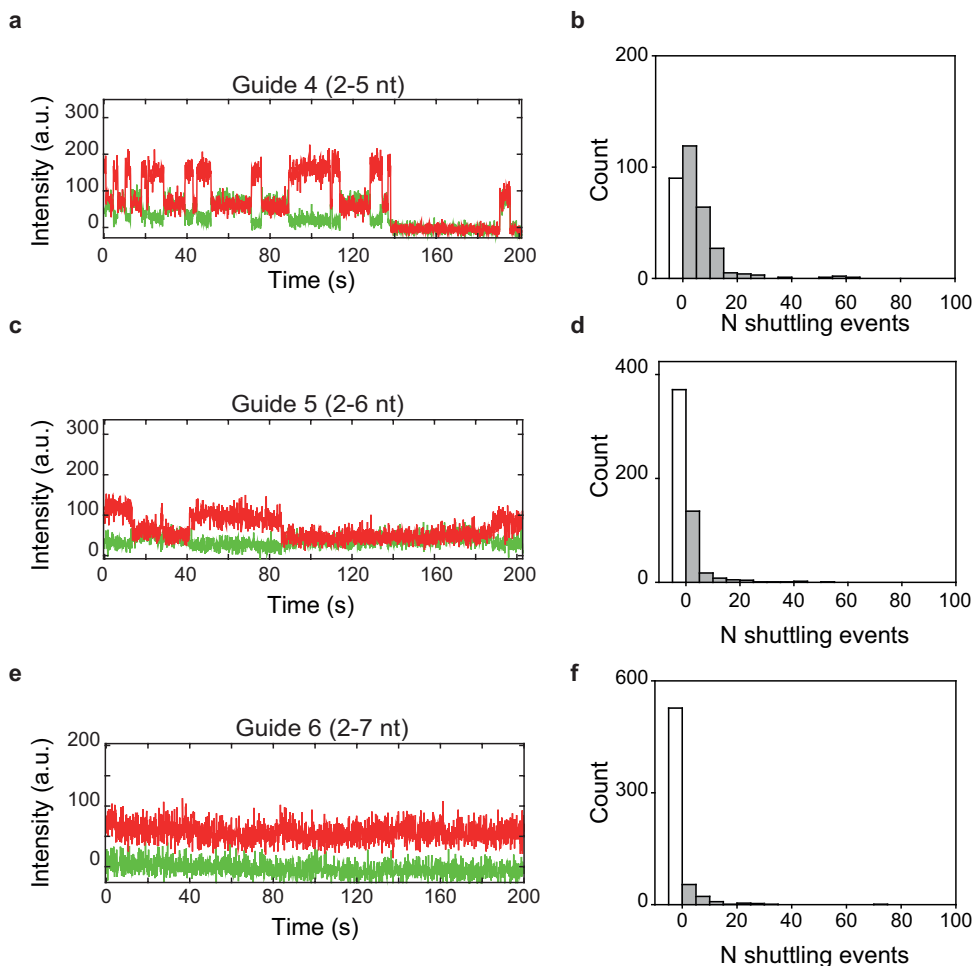
Supplementary Figure 2: Single-molecule interactions of CbAgo with guide 4, 5, 6 and tandem target (22 nt separation). (a) Representative trace of binding events by CbAgo with guide 4 (nt 2-5). Duration of observation 200 s. (b) Shuttling event distribution for guide 4 (nt 2-5). Bin size = 5. The white bar represents binding (no shuttling) events followed by dissociation. N = 317. (c) Representative trace of binding events by CbAgo with guide 5 (2-6). (d) Shuttling event distribution for guide 5 (2-6 nt). Bin size = 10. The white bar represents events that consists of single molecule binding followed by dissociation. n = 550. (e) Representative trace of guide 6 (2-7 nt) interaction.(f) Shuttling event distribution for guide 6. The white bar represents events that consists of single molecule binding followed by dissociation. n = 621.
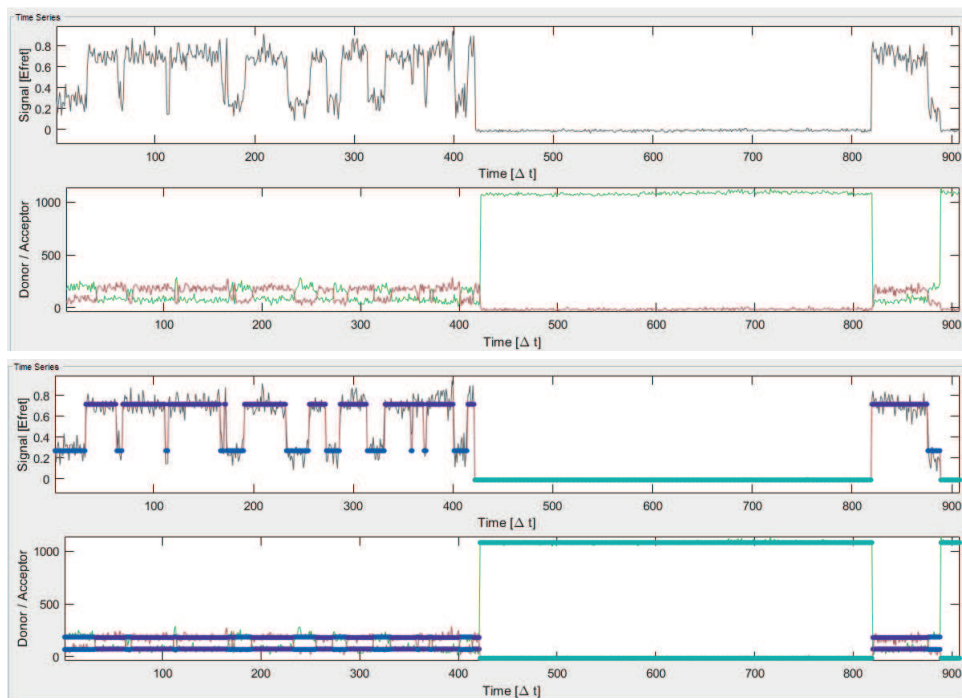
Supplementary Figure 3: Example of HMM software applied to data trace. (Top) An example shuttling trace of CbAgo in the user interface of ebFRET. The donor and acceptor intensities plotted versus time. The donor intensity is enhanced artificially in absence of any signal, resulting in an extra zero FRET state (upper subfigure). (Bottom) The donor, acceptor and FRET intensities overlaid with states resulting from the Hidden Markov Modeling. The HMM analysis program recognizes the unbound state as an extra state (light blue), while low FRET and high FRET are respectively assigned dark blue and purple.

a

b

c

d

e

f

g
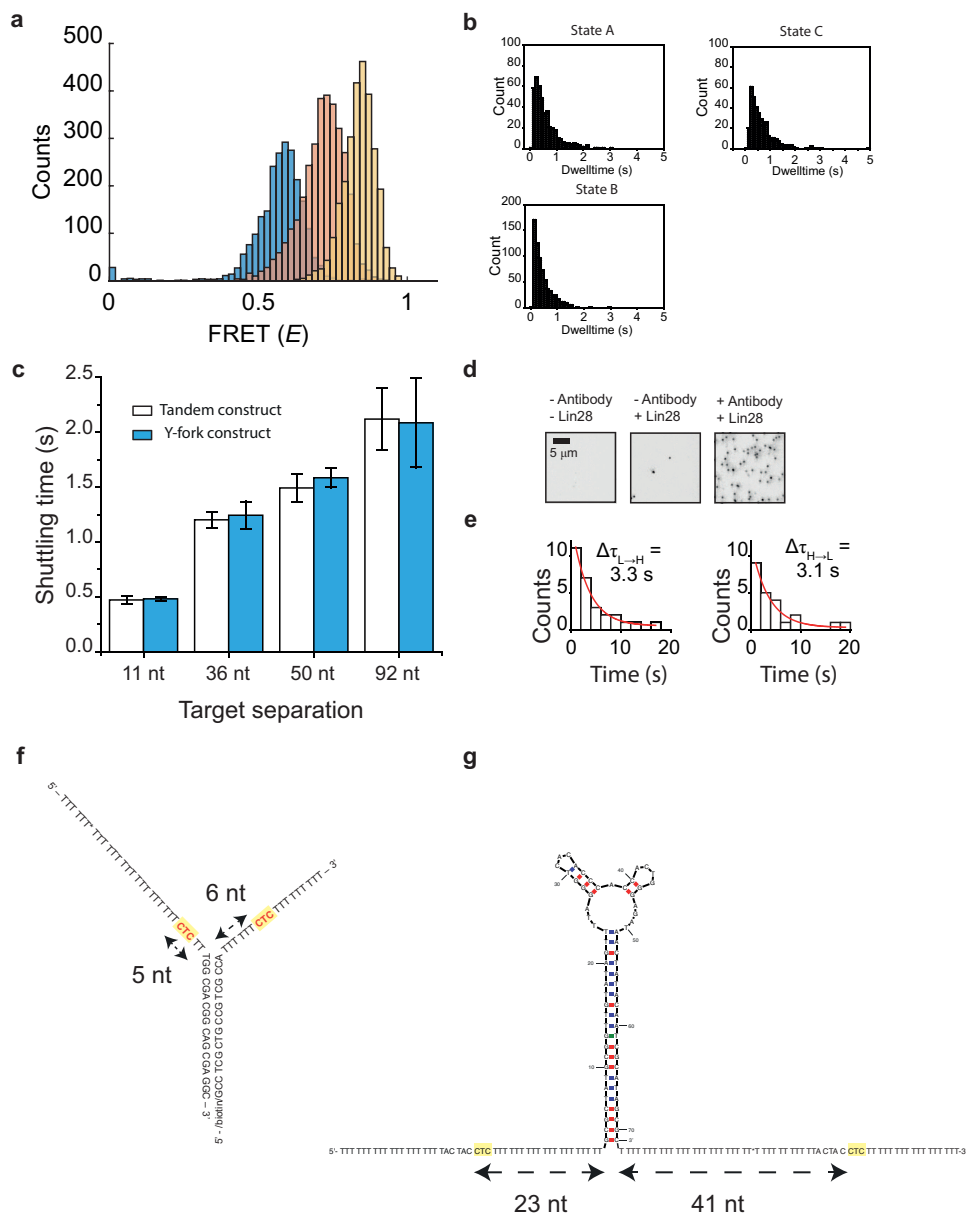
Supplementary Figure 4: Triple target assay, Y-fork assay and Lin28 assay. (a) FRET histogram of three-target assay. n = 168 molecules (b) Dwell time histograms for respectively the low FRET, mid FRET and high FRET state of the three target assay. (c) Shuttling rate of Y-fork constructs (blue) compared to tandem target assay (white) for 11 nt, 36 nt, 50 nt and 92 nt target separation. The error bars indicate the 95% percentile of 20,000 bootstrapped mean dwell times. (d) An EMCCD image of the acceptor channel. (Left) In absence of Lin28 protein and antibody with Cy5 labeled DNA. (Middle) In absence of antibody, but in presence of Lin28 protein and Cy5 labeled DNA. (Right) In presence of antibody, Lin28 protein and Cy5 labeled DNA. (e) Individual dwell times from low FRET state to high FRET state (left) and vice versa (right). (f) Sequence schematic for the Y-fork 11 nt, indicating the target sites and their respective distances to the junction. (g) Sequence schematic for the Lin28 blockade assay, indicating the target sites and their respective distances to the junction/protein.
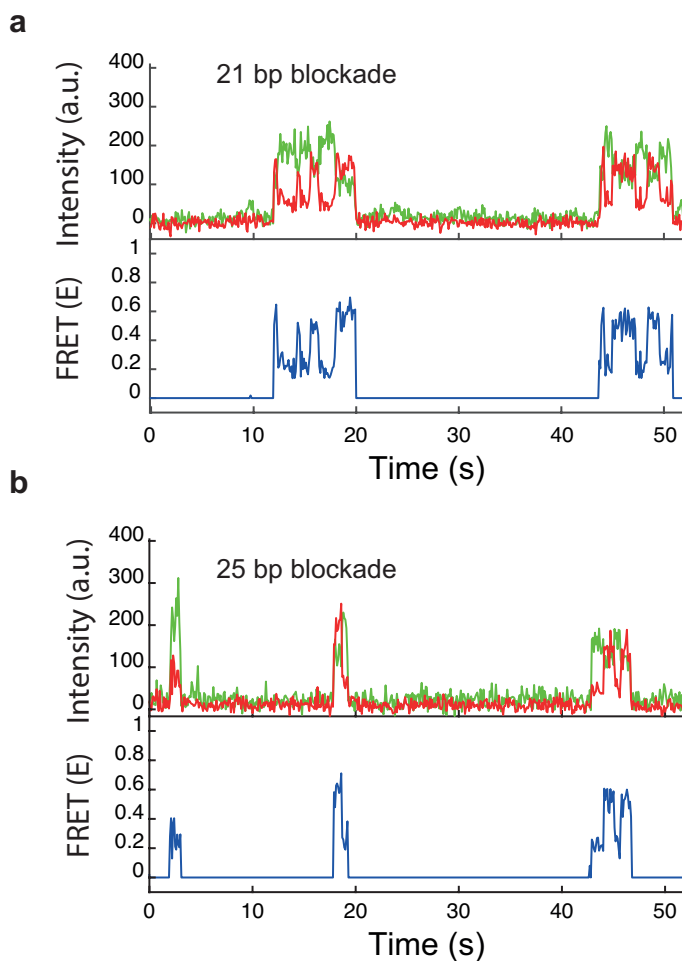
Supplementary Figure 5: Interactions of CbAgo with the dsDNA block construct. (a) Representative trace of CbAgo interacting with a 21 bp DNA blockade construct. (b) Representative trace of CbAgo interacting with a 25 bp DNA blockade construct.
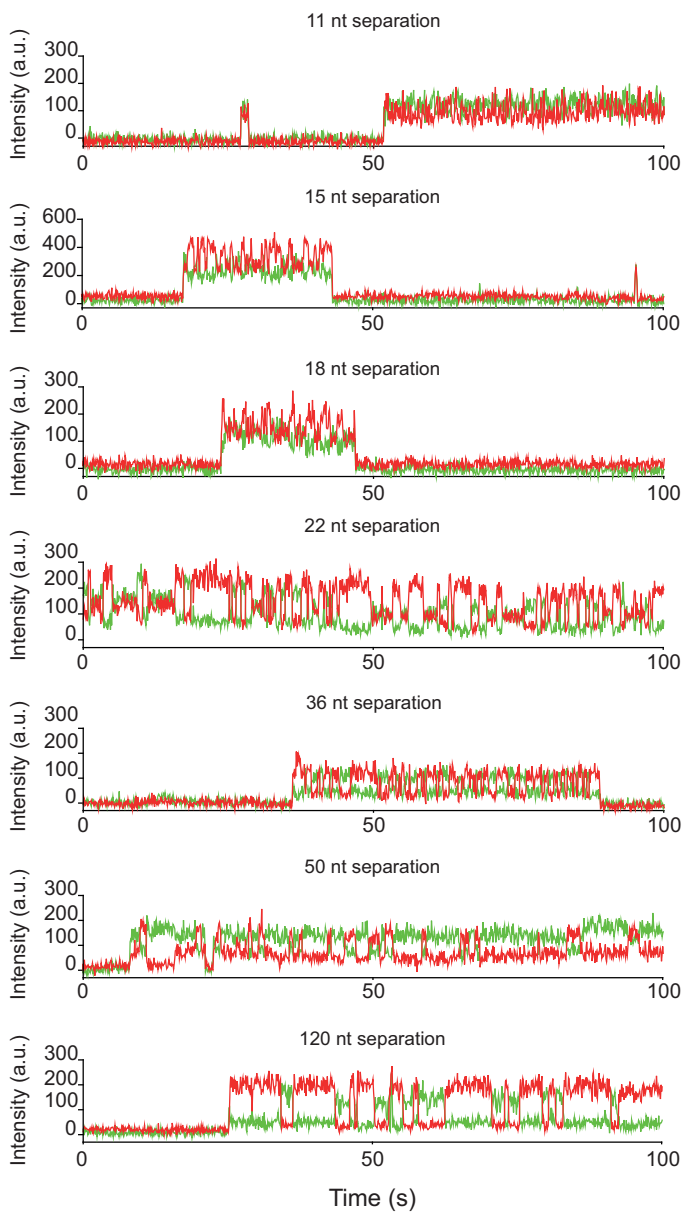
Supplementary Figure 6: Example shuttling traces for 11 nt, 15 nt, 18 nt, 22 nt, 29 nt, 36 nt, 50 nt and 120 nt target separation.

Sliding
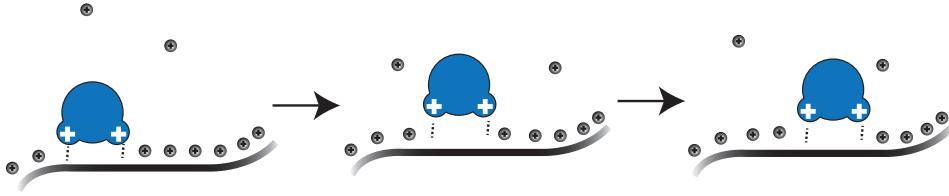
Hopping
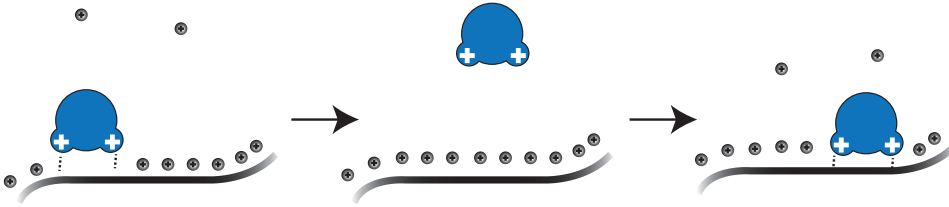
Intersegmental transfer via hopping

Neighbouring
DNA segment
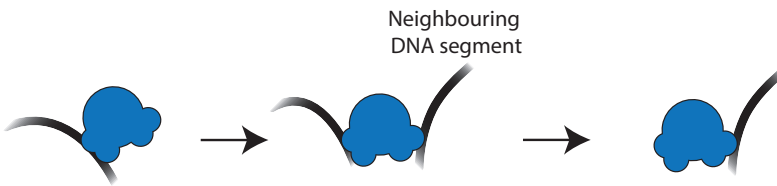
Supplementary Figure 7: Cartoon representation of target search mechanisms. Sliding: Proteins that undergo sliding make a well-correlated movement along the contour of the nucleic acid substrate. There is no net displacement of counterions (grey circles). Hopping: Proteins alternate quickly between a bound and unbound state with respect to DNA and there is counterion condensation upon dissociation of the protein. The method of diffusion is similar to 3D search, but its movements are correlated along the contour of the strand. Intersegmental transfer: This mechanism is a specialized form of hopping where segments appear transiently close by allow the protein to transfer to this new segment.

a

Argonaute is weakly associated
with the DNA strand

Local scanning

Intersegmental transfer allow Ago
to bypass protein barriers
and access distant DNA/RNA segments

RNA binding protein

hAGO2

RNA

4

b

DNA binding protein

CbAgo

DNA

Subseed match
Target

cleavage of
target strand
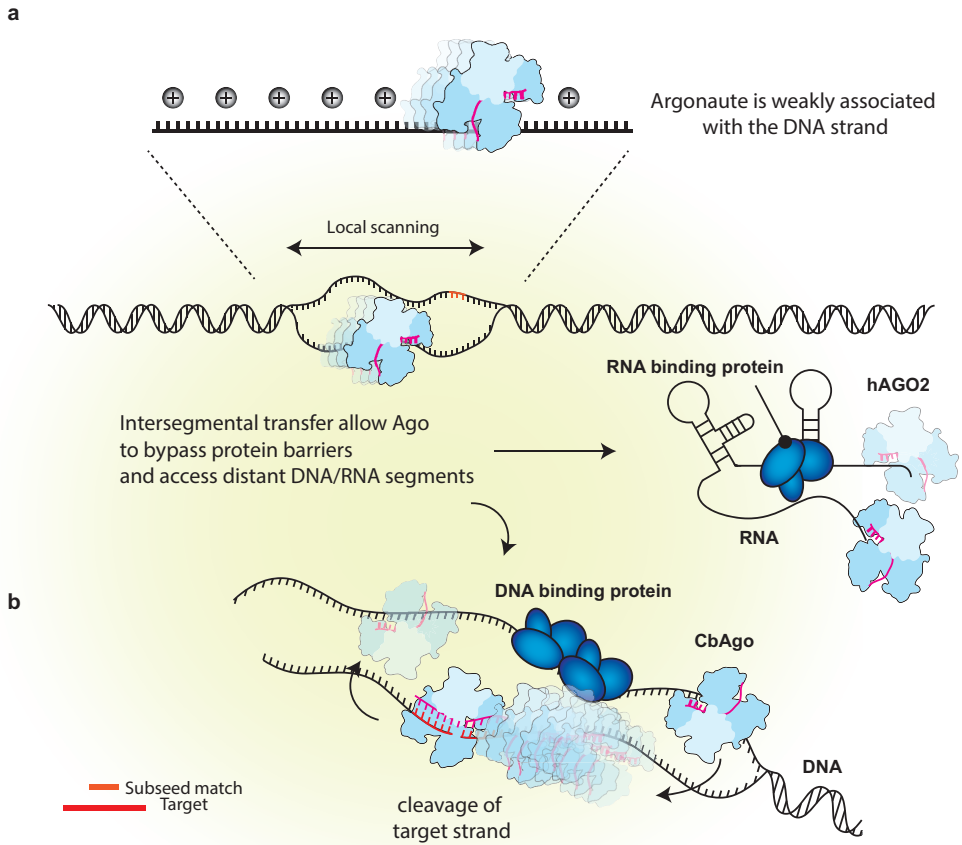
Supplementary Figure 8: Cartoon representation of Ago search model. The Ago complex utilizes short transient interactions with nucleic acid strands to rapidly sample the adjacent (tens of nucleotides away) sites for possible targets. Loose interaction with the nucleic acid strand persists. Obstacles can be overcome through intersegmental jumps.
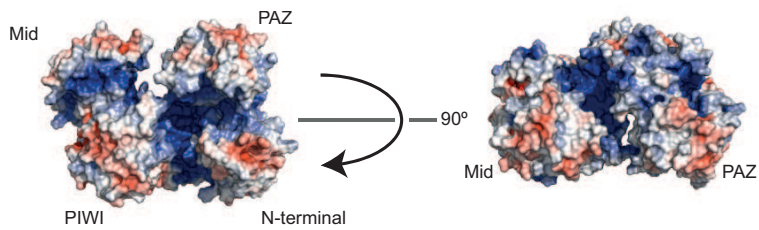
Supplementary Figure 9: Coulombic surface coloring of *Clostridium butyricum Argonaute* (CbAgo). The crystal structure of CbAgo (PDB 6qzk) (3.23 Å resolution) reveals the charge distribution. The cleft that contains the guide DNA and the target DNA is highly positively charged (blue).

4

4.5.2. SUPPLEMENTARY NOTE 1: BINDING TIMES SINGLE-TARGET INCLUDING RECAPTURE EVENTS FOLLOW SINGLE-EXPONENTIAL DISTRIBUTION

We here build a kinetic model for the lateral diffusion by CbAgo. Since Argonaute can in principle bind to any sequence along the DNA, we imagine the binding sites to be located a nucleotide apart. Further, we shall here only explicitly take sliding into account, which is represented as an unbiased random walk with unit step length. Assuming sliding should be a good approximation when considering only short distances traveled. If the protein is bound at the designed 3-nt sub-seed 'target' it can move to either of its neighbors at a rate of k_{esc} or unbind from the ssDNA at a rate of k_{ub}. When bound elsewhere movement and dissociation are assumed to happen instantaneously. To establish the manner in which these undetectable movements contribute to the observed dwell time distribution ($p_{bound}(\Delta t)$) we count all possible paths that the protein can take to dissociate following initial association to the sub-seed. In Laplace space the unbinding-time distribution, $P_{ub}(s) = \mathscr{L}\{p_{bound}(\Delta t)\}$, can be calculated as a product of the distributions of individual transitions (rather than their convolutions), summed over the possible paths towards unbinding. With an exponential distribution of stepping/escape times from the sub-seed trap,

$$p_{esc}(s) = \frac{2k_{esc}}{s + 2k_{esc} + k_{ub}}$$

(Supplementary Equation 1)

, an unbinding time distribution from the trap

$$p_{ub}(s) = \frac{k_{ub}}{s + 2k_{esc} + k_{ub}}$$

(Supplementary Equation 2)

and a probability to return, get recaptured at the trap, from either flank without unbinding P_{retrap} we can write

$$P_{ub}(s) = \sum_{m=0}^{\infty} \left(p_{esc}(s)P_{retrap}\right)^m \left[p_{ub}(s) + p_{esc}(s)(1 - P_{retrap})\right]$$
$$= \frac{k_{ub} + 2k_{esc}(1 - P_{retrap})}{s + k_{ub} + 2k_{esc}(1 - P_{retrap})}$$

(Supplementary Equation 3)

The sum on the left hand side of **Supplementary Equation 3** therefore accounts for the protein escaping from, and getting recaptured at the target an arbitrary amount of times (see **additional Supplementary Figure** 1 below). The two terms outside the sum represent the probability distributions to unbind from either the target directly or after having escaped one final time respectively (**additional Supplementary Figure** 1 below). Taking the inverse Laplace transform, we derive the observed dwell time distribution.

$$p_{bound}(\Delta t) = \mathscr{L}^{-1}\left\{\frac{k_{ub} + 2k_{esc}(1 - P_{retrap})}{s + k_{ub} + 2k_{esc}(1 - P_{retrap})}\right\}$$
$$= (k_{ub} + 2k_{esc}(1 - P_{retrap}))e^{-(k_{ub} + 2k_{esc}(1 - P_{retrap}))\Delta t}$$

(Supplementary Equation 4)
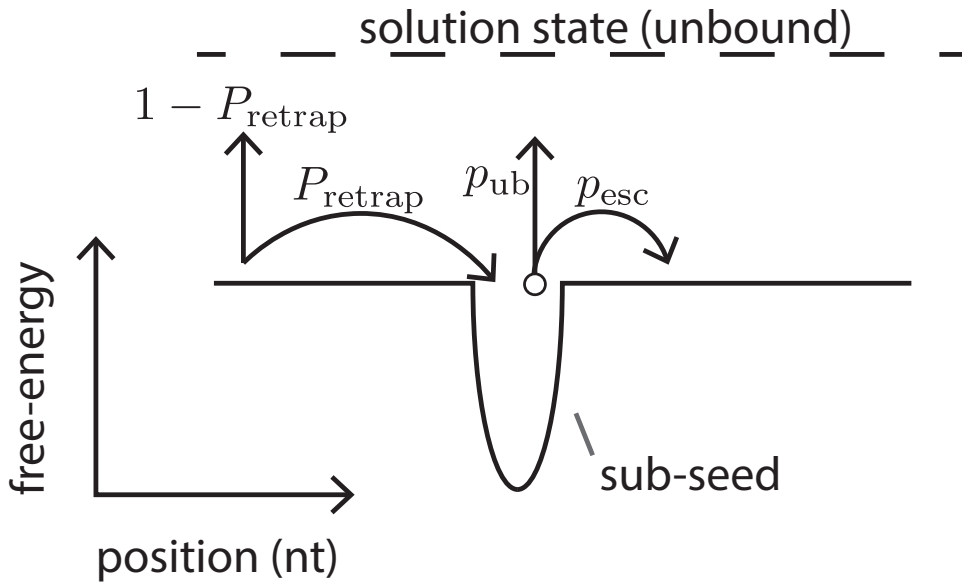
additional Supplementary Figure 1: This figure illustrates how to construct Supplementary Equation 3. Starting from the sub-seed, Ago can either unbind directly (probability p_{ub}) or slide onto the non-specific binding sites flanking the trap (probability p_{esc}). When non-specifically bound, Ago can either laterally diffuse back into the sub-seed (probability P_{retrap}), or unbind (probability $1 - P_{retrap}$)

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

$$k_{ub}^{observed} = k_{ub} + 2k_{esc}(1 - P_{retrap}) \qquad \text{(Supplementary Equation 5)}$$

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

4.5.3. Supplementary Note 2: Shuttling rate due to sliding alone

We seek to explain to what extend sliding contributes to the observed shuttling rate from the tandem-target assay. Given under the current experimental conditions about 13 shuttle events occur prior to unbinding, we shall ignore unbinding in the following analysis ($k_{ub} \ll k_{esc}$). To get the distribution of shuttle times ($p(\Delta t_{shuttle})$) we count all possible paths that lead the protein from one sub-seed to the other. If the two 3-nt nucleotide long sub-seeds are separated by x_{poly-T} thymine nucleotides, the shuttle times are distributed as (setting $x_{target} = x_{poly-T} + 3 \geq 3$) (see **additional Supplementary Figure** 2 below).)

$$P_{shuttle}(s, x_{target}) = \sum_{m=0}^{\infty} \left(p_{esc}(s) \left(\frac{1}{2} \times 1 + \frac{1}{2} \times P_R(x_{target}) \right) \right)^m p_{esc}(s) P_S(x_{target}) = \frac{k_{esc} P_S(x_{target})}{s + k_{esc} P_S(x_{target})}$$

$$\text{(Supplementary Equation 6)}$$

The two terms within the sum shown above represent recapture events at the initial trap via either the the flanking sequence (from which it always returns) or the poly-T stretch in between the traps (from which it returns with a probability $P_R(x_{target})$ without shuttling) (**additional Supplementary Figure** 2 shown below). Finally, the term outside the sum accounts for successful shuttling events (which occurs with probability $P_S(x_{target}) = 1 - P_R(x_{target})$). Once the protein has left the initial trap $P_R(x)$ and $P_S(x)$ denote the distributions for either returning back to the initial trap or shuttling/making it across to the other, if the two traps are x nucleotides apart (see **additional Supplementary Figure** 3 below)). Inverting the Laplace transformation of Supplementary Equation 6 we obtain

$$p(\Delta t_{shuttle}) = \mathscr{L}^{-1}\left\{\frac{k_{esc}P_S(x_{target})}{s + k_{esc}P_S(x_{target})}\right\}$$

(Supplementary Equation 7)

$$= k_{esc}P_S(x_{target})e^{-(k_{esc}P_S(x_{target})\Delta t_{shuttle})}$$

Hence, the observed dwell time distributions are indeed single exponential. In terms of the microscopic model the average time is set by the escape rate from the trap modified by the probability to make it across once outside of it ($P_S(x_{target})$).

The probabilities P_R and P_S, for a given inter-trap distance x_{target} follow (see **additional Supplementary Figure** 3 below)

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

(Supplementary Equation 8)

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

(Supplementary Equation 9)

- from which we can write the recurrence relation

$$P_S(x_{target}) = P_R(x_{target})P_S(x_{target}-1)$$

(Supplementary Equation 10)

Using ($P_S(x_{target}) = 1 - P_R(x_{target})$) the above can be re-written as

$$P_S(x_{target}) = \frac{P_S(x_{target}-1)}{P_S(x_{target}-1)+1}$$

(Supplementary Equation 11)

which subjected to the boundary condition $P_S(1) = 1$ - signifying that if the traps are placed adjacent to each other, the shuttle is complete once the protein escaped the initial trap - has the simple solution

$$P_S(x_{target}) = \frac{1}{x_{target}}$$

(Supplementary Equation 12)

Taken together, the observed shuttling time equals

$$\Delta\tau_{shuttle} = \frac{1}{k_{esc}P_S(x_{target})} = \frac{x_{target}}{k_{esc}}$$

(Supplementary Equation 13)

Note that $x_{target} \geq 3$, as the two sub-seeds cannot overlap. A fit of Supplementary Equation 13 to the experimental data for x_{target} of 11nt, 15nt, 18nt and 22nt in Figure 4.3 of the main manuscript were used to estimate the value of k_{esc} for CbAgo.

4.5.4. SUPPLEMENTARY NOTE 3: SHUTTLING RATE TRIPLE-TARGET CONSTRUCT

For the assay using three sub-seed targets, we can now predict both the time needed to slide from any of the outer ones to the inner ($C \to B$) and the average time needed to slide along the opposite path ($B \to C$). The former is equal to the time measured on the tandem target construct, denoted above as $\Delta t_{\text{shuttle}}$ (Supplementary Equation 13, $\Delta \tau_{\text{CB}} = \Delta t_{\text{shuttle}}$). We obtain $\Delta \tau_{\text{BC}}$, via the distribution of lifetimes in the middle trap

$$P(\text{leave } B|\text{arrive at } C)(t) = \frac{P(\text{leave } B)(t)}{P(\text{arrive at } C(\text{and not } A))} \quad \text{(Supplementary Equation 14)}$$

Using that the distance between A and B is equal to that in between B and C, in Laplace space, the time spent at target B is distributed as ($P_{\text{B}}(t) \equiv P(\text{leave } B)(t)$)

$$P_{\text{B}}(s, x_{\text{target}}, k_{\text{esc}}) = \sum_{m=0}^{\infty} \left(\frac{1}{2} p_{\text{esc}}(s) \times 2 \times P_{\text{R}}(x_{\text{target}}) \right)^{m} \frac{1}{2} p_{\text{esc}}(s) P_{\text{S}}(x_{\text{target}})$$

$$\text{(Supplementary Equation 15)}$$

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

$$\tau_{\text{B}}(x_{\text{target}}) = \frac{x_{\text{target}}}{4 k_{\text{esc}}} \quad \text{(Supplementary Equation 16)}$$

Using that half of the times the protein arrives at A, rather than C, results in the average dwelltime/shuttling time conditioned on moving from B to C (using eq. Supplementary Equation 14):

$$\Delta \tau_{\text{BC}}(x_{\text{target}}) = 2\tau_{\text{B}}(x_{\text{target}}) = \frac{x_{\text{target}}}{2 k_{\text{esc}}} \quad \text{(Supplementary Equation 17)}$$

4.5.5. SUPPLEMENTARY NOTE 4: ERROR ESTIMATES USING BOOTSTRAPPING

Fitting the data from the tandem target assay to Supplementary Equation 13 provides the estimate of k_{esc}. We bootstrapped the dwell time distributions acquired using the original tandem target assay (distances of 11nt, 15nt, 18nt and 22nt). For each of the 10^5 bootstrap samples we calculated new values for the associated $\Delta t_{\text{shuttle}}$'s and repeated the fit to Supplementary Equation 13 to obtain an error estimate in the fitted value of the escape rate.

After using the data from the tandem target assay to estimate k_{esc} there are no more free parameters remaining when predicting the data for the triple-target assay. Performing the bootstrap procedure for k_{esc}, and using Supplementary Equation 13 and Supplementary Equation 17 results in the 95% confidence intervals shown in figure 4D in the main manuscript.

An error estimate for the experimental values of $\Delta \tau_{\text{BC}}$ and $\Delta \tau_{\text{CB}}$ were obtained using 10^5 bootstrap samples of the dwell time distributions measured using the triple-target assay. All analysis was performed with a custom code written in Python.
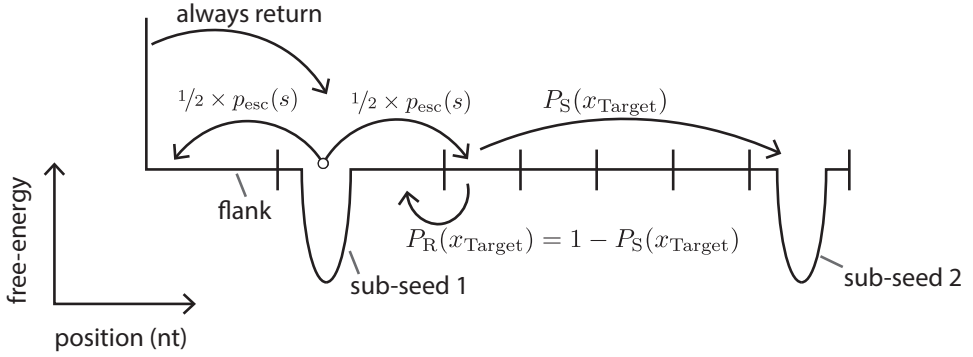
additional Supplementary Figure 2: This figure illustrates how to construct Supplementary Equation 6. Ago slides to either of its neighboring sites with equal probability. Every shuttle event starts with Ago bound to one of the sub-seed sequences. After residing there for a time distributed as $p_{esc}(s)$, half of the times Ago moves onto the flank (from which it always returns by assumption), while the other half of the times the protein slid onto the poly-T sequence in between the two sub-seeds. All movements along these intermediate sites occur too fast to observe, which is why we only take into account to probability $P_S(x_{target})$ of completing the shuttle event when x_{target} sites separate Ago from the second sub-seed.
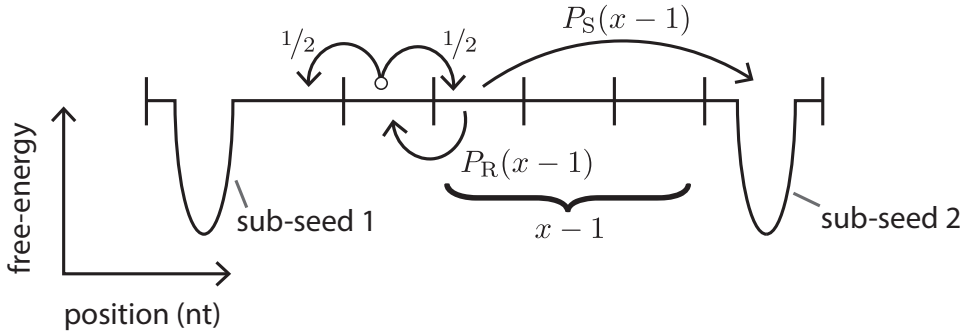
additional Supplementary Figure 3: This figure illustrates how to construct Supplementary Equation 8 and Supplementary Equation 9. Let $P_S(x)$ denote the probability to complete the shuttle when x sites separate Ago from the second sub-seed. Ago walks to either of its neighboring sites with equal probability. Therefore, when situated next to the first sub-seed, Ago gets recaptured half of the times it makes a move, while the other half has a probability of $P_S(x-1)$ to result in a completed shuttle event.

4.5.6. SUPPLEMENTARY TABLES

Table S1: Dwell times of different two target DNA constructs for several distances. The upper bound and lower bound are estimated through 20000 bootstraps of the acquired dwell times.

Target distance (nt)	Lifetime (sec)	Lower bound lifetime (sec)	Upper bound lifetime (sec)	Shuttling rate (sec-1)	Lower bound shuttling rate (sec-1)	Upper bound shuttling rate (sec-1)
11	0.47	0.46	0.49	2.11	2.04	2.19
15	0.83	0.81	0.87	1.19	1.15	1.24
18	1.17	1.11	1.24	0.85	0.81	0.90
22	1.79	1.74	1.86	0.56	0.54	0.57
29	1.36	1.30	1.42	0.73	0.7	0.77
36	1.19	1.16	1.23	0.84	0.81	0.86
50	1.52	1.46	1.57	0.66	0.64	0.68
64	1.65	1.59	1.71	0.61	0.59	0.63
92	1.94	1.85	2.02	0.52	0.49	0.54
120	2.11	2.03	2.19	0.47	0.46	0.49

4

Table S2: Oligonucleotides used for this study

Name Oligo	Sequence 5'->3'	Length (nt)
GUIDE		
Guide 3nt (2-4)	5- /5Phos/CGA GTA TT/iAmMC6T/ TTT TTT TTT TTT T – 3'	22
Guide 4nt (2-5)	5'-/5Phos/CGA GGA TT/iAmMC6T/ TTT TTT TTT TTT T - 3'	22
Guide 5nt (2-6)	5'- /5Phos/CGA GGT TT/iAmMC6T/ TTT TTT TTT TTT T - 3'	22
Guide 6nt (2-7)	5'- /5Phos/CGA GGT AT/iAmMC6T/ TTT TTT TTT TTT T - 3 '	22
Guide 7nt (2-8)	5'- /5Phos/CGA GGT AGA /iAmMC6T/TT TTT TTT TTT T -3'	22
Guide 8nt (2-9)	5'- /5Phos/ CGA GGT AG/iAmMC6T/ TTT TTT TTT TTT T - 3 '	22
TARGET		
8nt tandem target 7nt separation	5' - TTT TTT TTT TTT TTT TTT CTC TTT TCT CT/iAmMC6T/ TTT TTT TTT TTT TTT TTT TTT TTT TTT T/biotin/ -3'	58
8nt tandem target 11nt separation	5' - TTT TTT TTT TTT TTT TTT CTC TTT TTT TT CT CT/iAmMC6T/ TTT TTT TTT TTT TTT TTT TTT TTT TTT T/biotin/ -3'	62
8nt tandem target 15nt separation	5' - TTT TTT TTT TTT TAC TAC CTC TTT TTT TA CTA CCT CT/iAmMC6T/ TTT TTT TTT TTT TTT TTT TTT TTT TTT T/biotin/ -3'	66
8nt tandem target 18nt separation	5' - TTT TTT TTT TTT TAC TAC CTC TTT TTT TTT TA CTA CCT CT/iAmMC6T/ TTT TTT TTT TTT TTT TTT TTT TTT TTT T/biotin/ -3'	69
8nt tandem target 22nt separation	5' - TTT TTT TTT TTT TAC TAC CTC TTT TTT /iAmMC6T/TT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT T/biotin/ -3'	73
8nt tandem target 29nt separation	5' –TTT TTT TTT TTT TA CTA CCT CTT TT TTT TT/iAmMC6T/ TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin/-3'	81
8nt double target 36nt separation	5' –TTT TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TT/iAmMC6T/ TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin dT//Phos/-3'	89

8 nt tandem target 50nt separation	5' –TTT TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TT TTT TTT TT/iAmMC6T/ TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin dT//Phos/-3'	104
8 nt tandem target 64 nt separation	5' –TTT TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT T/iAmMC6T/T TTT TTT TTT TTT ACT ACC TCT TTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin-dT/ /Phos/-3'	117
8 nt tandem target 92 nt separation	5' –TTT TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT T TTT TTT TTT TTT TTT TT TTT TTT TTT TTT TTT TTT TTT T/iAmMC6T/T TTT TTT TTT T ACT ACC TCT TTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin-dT/ /Phos/-3'	145
8nt double target 120nt separation	5' –TTT TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TT/iAmMC6T/ TTT TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin dT//Phos/-3'	171
11nt Y-fork	5' – TTT TTT* TTT TTT TTT TTT TTT TTT CTC TT TGG CGA CGG CAG CGA GGC – 3'	47
11nt Y-fork biotin	5' - /biotin/GCC TCG CTG CCG TCG CCA TTT TTT CTC TTT TTT TTT – 3'	36
50nt Y-fork	5'- TTT TTT TTT TTT* TTT TTT TAC TAC CTC TTT TTT TTT TTT TTT TT TTT TGG CGA CGG CAG CGA GGC – 3'	65
Y-fork stem (not for Y11)	5' – /biotin/GCC TCG CTG CCG TCG CCA TTT TTT TTT TTT TTT TTT TTT TAC TAC CTC TTT TTT TTT – 3'	57
36nt dsDNA target	5' – TTT TTT TTT TTT T TA CTA C CTC T CGG ACC AAC AGC GGG /T-biotin/AC GGC TGT GC TA CTA CCT CTT TTT TTT TTT TTT TTT TTT - 3'	78
36nt dsDNA block v2 3' biotin	5' –TTT TTT TTT TTT T TA CTA C CTC T CGG ACC AAC AGC GGG TAC GGC TGT GC TA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin dT/- 3'	91
36nt dsDNA block 5'end truncated	5' –CTA C CTC T CGG ACC AAC AGC GGG TAC GGC TGT GC TA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TT/biotin dT/- 3'	75
36nt 25nt block	5' - TA GC ACA GCC GT* A CCC GCT GTT GGT- 3'	25
36nt 21nt block	5'- GC ACA GCC GT* A CCC GCT GTT G- 3'	21
36nt 17nt block	5' -ACA GCC GT* A CCC GCT GT- 3'	17

Triple target	5' – T/iAmMC6T/ TTT TTT TTT TAC CTC TTT TTT ACC TCT TTT TTA CCT C TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT/biotin/ -3'	69
No target DNA	5'- TTT TTT TTT TTT TTT TTT TTT TTT TTT /iAmMC6T/TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TGG CGA CGG CAG CGA GGC -3'	90
8nt single target	5' - TTT TTT TTT TTT TTT TTT TTT TTT TTT /iAmMC6T/TT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT TTT TTT TTT TTT T/biotin/-3'	73
3' biotin stem	5' - GCC TCG CTG CCG TCG CCA biotin – 3'	18
Lin28 double target	5'- TTT TTT TTT TTT TTT TTT TAC TAC CTC TTT TTT TTT TTT TTT TTT TTG CGC TAT GCG GTT GTA TAG TTT TAG GGT CAC ACC CAC CAC TGG GAG ATA ACT ATA CAA TCG CAT AGC GCT TTT TTT TTT TTT TTT TTT TTT T/iAmMC6T/T TTT TTT TTA CTA CCT CTT TTT TTT TTT TTT TTT-3'	174

REFERENCES

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

[2] D. C. Swarts, K. Makarova, Y. Wang, K. Nakanishi, R. F. Ketting, E. V. Koonin, D. J. Patel, and J. van der Oost, *The evolutionary journey of Argonaute proteins.* Nature structural & molecular biology **21**, 743 (2014).

[3] L. He and G. J. Hannon, *MicroRNAs: Small RNAs with a big role in gene regulation,* Nature Reviews Genetics **5**, 522 (2004).

[4] F. V. Rivas, N. H. Tolia, J.-J. Song, J. P. Aragon, J. Liu, G. J. Hannon, and L. Joshua-Tor, *Purified Argonaute2 and an siRNA form recombinant human RISC,* Nature structural & molecular biology **12**, 340 (2005).

[5] D. P. Bartel, *MicroRNAs: Target Recognition and Regulatory Functions,* Cell **136**, 215 (2009), arXiv:0208024 [gr-qc] .

[6] I. Olovnikov, K. Chan, R. Sachidanandam, D. K. D. Newman, and A. A. A. Aravin, *Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA,* Molecular Cell **51**, 594 (2013), arXiv:NIHMS150003 .

[7] D. C. Swarts, M. M. Jore, E. R. Westra, Y. Zhu, J. H. Janssen, A. P. Snijders, Y. Wang, D. J. Patel, J. Berenguer, S. J. J. Brouns, and J. van der Oost, *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature **507**, 258 (2014), arXiv:15334406 .

[8] D. C. Swarts, J. van der Oost, and M. Jinek, *Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a,* Molecular Cell **66**, 221 (2017).

[9] L. B. Harrington, L. B. Harrington, D. Burstein, J. S. Chen, D. Paez-espino, E. Ma, I. P. Witte, J. C. Cofsky, N. C. Kyrpides, J. F. Banfield, and J. A. Doudna, *Programmed DNA destruction by miniature CRISPR-Cas14 enzymes,* Science **4294**, 1 (2018).

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

[11] D. Kim, Y. M. Sung, J. Park, S. Kim, J. Kim, J. Park, H. Ha, J. Y. Bae, S. Kim, and D. Baek, *General rules for functional microRNA targeting,* Nature Genetics **48**, 1517 (2016).

[12] O. G. Berg, R. B. Winter, and P. H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory.* Biochemistry **20**, 6929 (1981).

[13] A. D. Riggs, H. Suzuki, and S. Bourgeois, *lac repressor-operator interaction: I. Equilibrium studies,* Journal of Molecular Biology **48**, 67 (1970).

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

[15] U. Gerland, J. D. Moroz, and T. Hwa, *Physical constraints and functional*

characteristics of transcription factor-DNA interaction, Proceedings of the National Academy of Sciences **99**, 12015 (2002), arXiv:0112083 [physics] .

[16] A. B. Kolomeisky and A. Veksler, *How to accelerate protein search on DNA: Location and dissociation,* Journal of Chemical Physics **136** (2012), 10.1063/1.3697763.

[17] B. van den Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler, and G. J. L. Wuite, *How DNA coiling enhances target localization by proteins.* Proceedings of the National Academy of Sciences of the United States of America **105**, 15738 (2008).

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

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

[20] P. H. Von Hippel and O. G. Berg, *Facilitated target location in biological systems,* Journal of Biological Chemistry **264**, 675 (1989).

[21] V. Globyte, S. H. Kim, and C. Joo, *Single-Molecule View of Small RNA–Guided Target Search and Recognition,* Annual Review of Biophysics **47**, 569 (2018).

[22] D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, and T. Ha, *Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9,* Nature Communications **7**, 1 (2016).

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

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

[25] C. Li, V. V. Vagin, S. Lee, J. Xu, S. Ma, H. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. Weng, and P. D. Zamore, *Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies,* Cell **137**, 509 (2009).

[26] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, and J. A. Doudna, *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9,* Nature **507**, 62 (2014), arXiv:NIHMS150003 .

[27] V. Globyte, S. H. Lee, T. Bae, J.-S. Kim, and C. Joo, *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion,* The EMBO Journal , e99466 (2018).

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

[29] C. S. Smith, K. Jouravleva, M. Huisman, S. M. Jolly, P. D. Zamore, and D. Grunwald, *An automated Bayesian pipeline for rapid analysis of single-molecule binding data,* Nature Communications **10**, 1 (2019).

[30] J. W. Hegge, D. C. Swarts, S. D. Chandradoss, T. J. Cui, J. Kneppers, M. Jinek, C. Joo, and J. van der Oost, *DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute,* Nucleic Acids Research **47**, 5809 (2019), arXiv:/doi.org/10.1101/534206 [https:] .

[31] K. Ragunathan, C. Liu, and T. Ha, *RecA filament sliding on DNA facilitates homology search,* eLife **2012**, 1 (2012).

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

[33] B. M. Muhire, M. Golden, B. Murrell, P. Lefeuvre, J.-M. Lett, A. Gray, A. Y. F. Poon, N. K. Ngandu, Y. Semegni, E. P. Tanov, A. L. Monjane, G. W. Harkins, A. Varsani, D. N. Shepherd, and D. P. Martin, *Evidence of Pervasive Biologically Functional Secondary Structures within the Genomes of Eukaryotic Single-Stranded DNA Viruses,* Journal of Virology **88**, 1972 (2014).

[34] E. J. Strobel, A. M. Yu, and J. B. Lucks, *High-throughput determination of RNA structures,* Nature Reviews Genetics **19**, 615 (2018).

[35] Y. Nam, C. Chen, R. I. Gregory, J. J. Chou, and P. Sliz, *Molecular basis for interaction of let-7 MicroRNAs with Lin28,* Cell **147**, 1080 (2011), arXiv:NIHMS150003 .

[36] H. Chen, S. P. Meisburger, S. a. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, *Ionic strength-dependent persistence lengths of single-stranded RNA and DNA,* Proceedings of the National Academy of Sciences **109**, 799 (2012).

[37] M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. L. Wuite, and R. Metzler, *Facilitated diffusion with DNA coiling,* Proceedings of the National Academy of Sciences **106**, 8204 (2009).

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

[39] R. B. Winter and P. H. Von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli lac repressor-operator interaction: equilibrium measurements,* Biochemistry **20**, 6948 (1981).

[40] R. B. Winter, O. G. Berg, and P. H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor–operator interaction: kinetic measurements and conclusions.* Biochemistry **20**, 6961 (1981).

[41] C. Desnues, B. Rodriguez-Brito, S. Rayhawk, S. Kelley, T. Tran, M. Haynes, H. Liu, M. Furlan, L. Wegley, B. Chau, Y. Ruan, D. Hall, F. E. Angly, R. A. Edwards, L. Li, R. V. Thurber, R. P. Reid, J. Siefert, V. Souza, D. L. Valentine, B. K. Swan, M. Breitbart, and F. Rohwer, *Biodiversity and biogeography of phages in modern stromatolites and thrombolites,* Nature **452**, 340 (2008).

[42] J. M. Labonté and C. A. Suttle, *Previously unknown and highly divergent ssDNA viruses populate the oceans,* ISME Journal **7**, 2169 (2013).

[43] M. Yoshida, T. Mochizuki, S. I. Urayama, Y. Yoshida-Takashima, S. Nishi, M. Hirai, H. Nomaki, Y. Takaki, T. Nunoura, and K. Takai, *Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku, Japan,* Frontiers in Microbiology **9**, 1 (2018).

[44] O. Barabas, D. R. Ronning, C. Guynet, A. B. Hickman, B. Ton-Hoang, M. Chandler, and F. Dyda, *Mechanism of IS200/IS605 Family DNA Transposases: Activation and Transposon-Directed Target Site Selection,* Cell **132**, 208 (2008).

[45] M. J. Curcio and K. M. Derbyshire, *The outs and ins of transposition: From MU to kangaroo,* Nature Reviews Molecular Cell Biology **4**, 865 (2003).

[46] B. Ton-Hoang, C. Pasternak, P. Siguier, C. Guynet, A. B. Hickman, F. Dyda, S. Sommer, and M. Chandler, *Single-stranded DNA transposition is coupled to host replication,* Cell **142**, 398 (2010).

[47] J.-D. Beaudoin, E. M. Novoa, C. E. Vejnar, V. Yartseva, C. M. Takacs, M. Kellis, and A. J. Giraldez, *Analyses of mRNA structure dynamics identify embryonic gene regulatory programs,* Nature Structural & Molecular Biology , 1 (2018).

[48] P. C. Bevilacqua, L. E. Ritchey, Z. Su, and S. M. Assmann, *Genome-Wide Analysis of RNA Secondary Structure,* Annual Review of Genetics **50**, 235 (2016).

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

[50] L. E. Vandivier, S. J. Anderson, S. W. Foley, and B. D. Gregory, *The Conservation and Function of RNA Secondary Structure in Plants,* Annual Review of Plant Biology **67**, 463 (2016), arXiv:1510.01420 .

[51] R. Zwanzig, *Diffusion in a rough potential.* Proceedings of the National Academy of Sciences **85**, 2029 (1988).

[52] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, *How a protein searches for its site on DNA: the mechanism of facilitated diffusion,* Journal of Physics A: Mathematical and Theoretical **42**, 434013 (2009).

[53] M. Sheinman and Y. Kafri, *The effects of intersegmental transfers on target location by proteins.* Physical biology **6**, 016003 (2009), arXiv:0807.3639 .

[54] M. Slutsky and L. A. Mirny, *Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential,* Biophysical Journal **87**, 4021 (2004), arXiv:0402005 [q-bio] .

[55] D. Holoch and D. Moazed, *Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex.* Nature structural & molecular biology **22**, 328 (2015).

[56] D. Moazed, *Small RNAs in transcriptional gene silencing and genome defence,* Nature **457**, 413 (2009), arXiv:NIHMS150003 .

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

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

[59] G.-W. Li, O. G. Berg, and J. Elf, *Effects of macromolecular crowding and DNA looping on gene regulation kinetics,* Nature Physics **5**, 294 (2009).

[60] A. A. Aravin, R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor, and G. J. Hannon, *A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice,* Molecular Cell **31**, 785 (2008), arXiv:NIHMS150003 .

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

4

5

OPTIMAL TARGET SEARCH USING FREQUENT SKIP-N-SLIDES

The timed action of proteins at specific DNA or RNA sequences plays a crucial role in the cell. A special class of such target searchers, amongst which Argonaute and CRISPR-Cas9, use small RNA or DNA molecules to define their target site. Through applying the Watson-Crick base pairing rules, these guide RNAs (or DNAs) can readily be designed, enabling the repurposing of the target searching proteins for genome engineering. Here we employ a combination of single-molecule FRET and theoretical modeling to understand the microscopic kinetics underlying the (guided) target search. We show Argonaute (both a prokaryotic and a eukaryotic variant) only sparsely interrogates its ssDNA/mRNA substrates – frequently skipping over about 2 out of 3 bases. Next, we show in general a mixture of sliding (interrogating neighboring sites) and frequent skipping (interrogating distant sites) minimizes the time needed to locate the target, given sufficiently long skips. Hence, we suggest Argonaute seems to operate at near optimal conditions using a mechanism likely applicable to other (guided and non-guided) target searchers.

5.1. INTRODUCTION

A multitude of cellular processes, including gene regulation, DNA repair, and immune response rely on proteins binding to specific DNA or RNA sequences. Even if the protein interacts only with the correct target sequence, the sheer size of the intracellular volume restricts the rate at which it can be found through diffusive collisions alone [1–4]. Still, measured search speeds can exceed the upper limit for diffusive collisions with up to two orders of magnitude [5]. To reach the observed speeds, target searching proteins can reduce the effective size of their search space by spending some fraction of time non-specifically associated and diffusively sliding along the DNA—partially replacing excursions into the solution (3D motion) as a means of reaching new sites to interrogate [1, 3, 4, 6].

Theoretical work [7] showed that an equal split of time spent sliding along the DNA and diffusing through solution would minimize the search time. While experiments have indeed confirmed such facilitated diffusion (a mix of 1D and 3D motion) for a variety of proteins [8–21], in vivo studies suggested the system spends considerably more than half the time associated to DNA [22, 23].

Repeated transfer between 1D sliding motion and 3D diffusion through solution and rebinding at an uncorrelated site is beneficial, as the sliding motion by itself will inevitably double back on itself and wastes time interrogating sites already visited. Early theoretical work recognized that this scanning redundancy could be further reduced if the non-specific interactions allow for intersegmental jumps [24–27], where the protein quickly transfers between close by DNA segments without fully returning to the solution state [1, 28, 29]. If the search process is optimized for time, and the total time spent transferring between segments is assumed negligible, we expect intersegmental jumps to minimize search redundancy (and so search time) by occurring as frequently as the geometry of the substrate allows. It has been shown theoretically that allowing for a (small) fixed amount of rapid intersegmental jumps shifts the optimal partitioning between 1D and 3D diffusion toward spending more time associated with the DNA [27].

However, when such jumps occur frequently the total time spent transferring between segments cannot be neglected. For instance, we expect this to be the case for proteins searching along single-stranded (ss) RNA or DNA with persistence lengths on the order of one nucleotide (nt) [30]. We may expect similar behavior for proteins that bind genomic targets, due to the strongly compacted double-stranded (ds) DNA within the nucleus or bacterial nucleoid. Furthermore, cellular RNA or DNA is typically occupied by various other (non-)specific binding proteins [31, 32], or can form secondary structures (i.e. plectonemes on dsDNA, or hairpins on ssRNA), all forming roadblocks along the target searcher's path. Bypassing such obstacles is often impossible through sliding, thereby necessitating the frequent use of some form of base-skipping, such as intersegmental jumps in case of sufficiently flexible substrates. Irrespective of the particular mechanism used, bases along the substrate are not interrogated, and we will simply refer to this process as 'skipping'. Little is known of the effect the frequent skips have on the search time.

Here we use Argonaute (Ago) as a model system for searches along flexible ss substrates. Ago belongs to a particular class of target searchers that pair with a small non-coding RNA (or DNA) guide, and then targets its complementary sequence [33]. The common usage of the CRISPR-Cas9 and CRISPR-Cas12a as next-generation genome editing tools [34], further highlights the importance of understanding also how such guided target searchers operate [33, 35–39]. A recent study showed the prokaryotic Clostridium butyricum Argonaute (CbAgo) uses a ssDNA guide to cleave ssDNA at moderate temperatures ($\sim$ 37 °C) [40], making it a potential candidate as a genome-editing tool. In a previous study [41] we demonstrated CbAgo can bypass roadblocks while diffusing along its substrate. Here, we start by establishing the generality of this base-skipping behavior by confirming its existence also for the eukaryotic human Argonaute 2 (hAgo2), using single-molecule (sm) Förster resonance energy transfer (FRET).

Next, we ask under what conditions skips can speed-up the target search process. To this end, we draw inspiration from established models [7, 27, 42, 43] and consider the search as consisting of three parts, but crucially allow all parts to take a finite time to complete: (i) interrogation of off-targets through sliding, (ii) base skipping, and (iii) diffusion through solution, followed by rebinding at an uncorrelated site. Through our modeling we discover the existence of two optimal partitioning between the three search modes: one coinciding with the known optimum of an equal time-split between 1D and 3D diffusion through solution when no skipping is allowed [7], and one novel optimum where skipping and sliding coexist during lateral diffusion. We fully characterize the search optima, and show that as a general rule, the system can never spend more time in solution than on the substrate when optimized, in accordance with experimental results [22, 23].
Using the presented smFRET data, we conclude by arguing that Ago operates far from the sliding-only optimum, and that its search characteristics are consistent with the skip-and-slide optimum. Our work suggests that any search involving many skips soon becomes beneficial over using only sliding, and thus raises the question whether skip-and-slide search could also be the preferred search mode for other searchers.

5.2. RESULTS

5.2.1. SINGLE-MOLECULE FRET ASSAY TO PROBE LATERAL DIFFUSION

Diffusive motion is often characterized by measuring the mean square displacement as a function of time [8, 9, 11, 13, 14, 16, 20, 21]. Even in the best of scenarios, when considering a stretched and uncoiled substrate, direct observation of lateral diffusion would require us to track target searchers over several hundreds of nucleotides. Such long trajectories would imply very redundant scanning by Ago, and might therefore not be performed by the protein [39]. In an attempt to capture also short diffusive excursions [15, 18, 35, 36, 41], we utilized the high spatial resolution of smFRET [44].
The experimental procedure has been described in detail elsewhere [35, 41], and we here restate only the core components. To trap any diffusive excursions for long enough to detect it (>100 ms), and have it complete before photobleaching (<700s), we design ss thymine (CbAgo [41]) and uracil (hAgo2, present study) repeats that contain two 3-nt
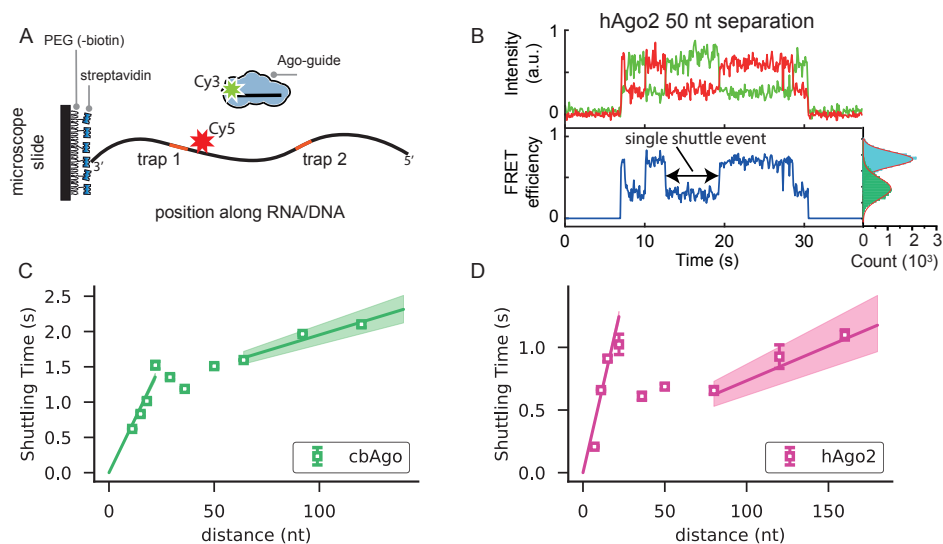
Figure 5.1: A Schematic of assay. DNA/RNA constructs, containing the two trapping sequences (shown in red) are passivated to the microscope slide via a 3' biotin-streptavidin linker and are labelled with the acceptor die. The Ago-guide complex is labelled with the donor die. B Representative trace for hAgo2 at a trap separation of 50nt. Top shows donor (green) and acceptor (red) signals. Bottom shows corresponding FRET efficiency and side panel shows histogram of all FRET efficiency values obtained for the population of molecules. C Shuttling time versus trapping distance (average ± sem) for CbAgo. Solid lines represent linear fits to data points at 11 nt, 15 nt, 18 nt, 22 nt (initial slope) and 64 nt, 92 nt, 120 nt (final slope). Shaded regions represent 95% confidence interval obtained using bootstrapping (see Methods). D Same as C for hAgo2. Data points at 7 nt, 11nt, 15 nt (initial slope) and 80 nt, 120 nt, 160 nt (final slope) are used for linear fits.

targets and two 4-nt targets respectively (Figure **5.1**A and S5.1). In order to accurately determine whether the protein is binding to one target as opposed to the other, one of the traps is labeled with an acceptor fluorophore (Cy5), while the guide is labeled with the donor fluorophore (Cy3) (Figure **5.1**A and S5.1). High FRET efficiency is observed when the protein binds to the site in close proximity of the acceptor dye, whereas lower FRET efficiency is obtained when Ago is trapped at the target far away from the dye (Figure **5.1**B). To reduce the background fluorescence, traces were recorded using total internal reflection (TIRF) microscopy.

5.2.2. AGO SLIDES OVER SHORT DISTANCES

As shown in **Figure 5.1**B, the FRET efficiency shifts almost instantaneously between those corresponding to the two trap locations. Though smFRET solves the problem of spatial resolution, the total time spent diffusing now seems to have fallen below our time resolution (30-100ms). In a recent paper [41] we showed both experimentally and theoretically that for small trap separations, the average shuttling time is directly proportional to the trap separation

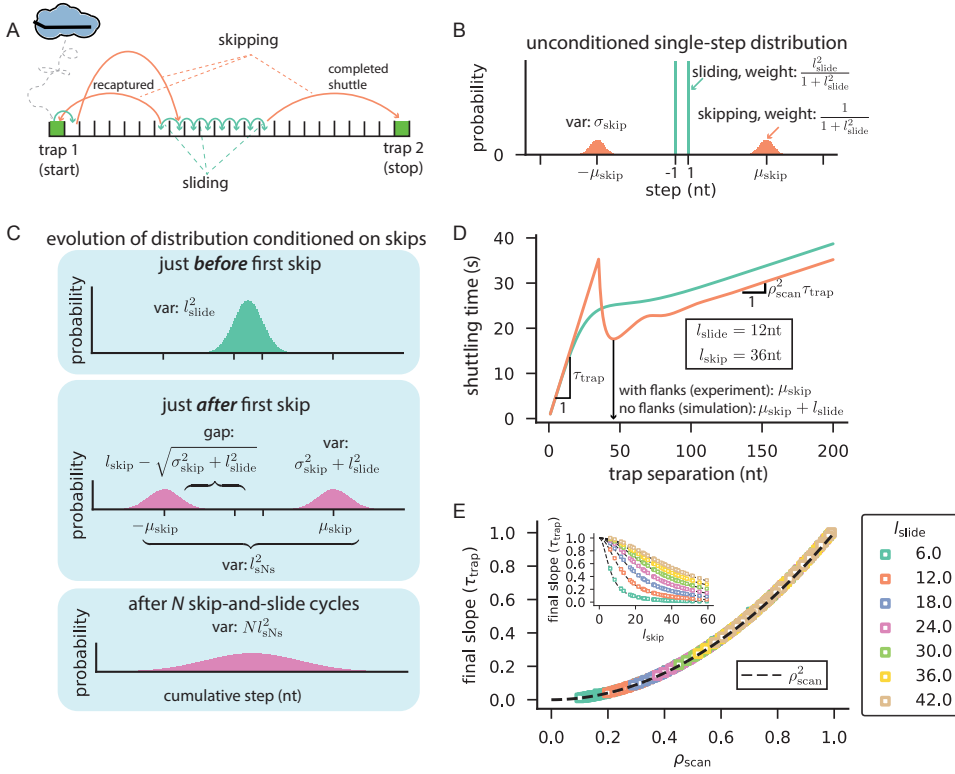$$T_{\text{shuttle}}(d_{\text{trap}} < 25\text{nt}) \approx d_{\text{trap}}\tau_{\text{trap}} \tag{1}$$

Figure 5.2: Modeling skip-and-slide search (shuttling events). A Schematic of shuttling event. Starting from the leftmost trap, the protein uses a combination of single-nucleotide steps (sliding) and larger steps (skipping) to reach the opposite trap after possibly getting recaptured at the initial traps several times. **B** single-step distribution of random walk defining our model. The protein either slides to a neighboring site or skips to sites located at $\pm(\mu_{\text{skip}} \pm \sigma_{\text{skip}})$. **C** Distribution of visited sites conditioned on skips. (top) The protein covers a rms distance l_{slide}^2 between consecutive skips. (middle) The first skip takes the protein μ_{skip} away (in either direction) with an uncertainty of σ_{skip} in the landing site. (bottom) Repeated skip-and-slide (sNs) cycles result in a distribution that resembles a simple random walk (top panel) with an adjusted effective step length of l_{sNs}. **D** Representative numerical solutions (**S.I.**) for shuttling time versus trapping distance. **E** Final slope versus scanning density. Inset shows equivalent versus skipping length (see **S.I.** for values in parameter sweep).

with τ_{trap}^{-1} being the one-sided escape rate from the trapping sequence. The linear increase in shuttling time with trap separation is consistent with Ago performing rapid lateral diffusion (undetected), with numerous escape and re-trapping events before eventually making it across to the other trap (**Figure 5.2**A). In **Figure 5.1**C we show data for CbAgo [41], and in **Figure 5.1**D we confirm that the initial proportionality (**Equation** 1) reported for CbAgo also holds for hAgo2 (new data).

5.2.3. AGO USES A MIXTURE OF SKIPPING AND SLIDING OVER LARGER DISTANCES

As the distance between traps grows beyond the initial linear regime, the shuttling time drops, before it eventually settles into a gentler linear increase over large trap separations (**Figures 5.1**C and D). The drop in shuttling time suggests that a new avenue for traversing the gap between traps has opened up, while the shuttling time's eventual linearity with regard to trap separation suggests that also this avenue is governed by lateral diffusion and repeated re-trapping to the original trap, before reaching the second trap.To explain the linear long-range behavior, we consider the fact that CbAgo has previously been shown to bypass both protein roadblocks and secondary structures [41]. Exactly how such obstacles are traversed is not fully understood, but it is clear that bases would be skipped (i.e. not interrogated) in any process able to bypass roadblocks, and we will therefore simply refer to this process as skipping.

In **Figure 5.2**A we show a schematic of the skip-and-slide dynamics, and in **Figure 5.2**B we show the single step distribution of such a random walker within our model. In **Figure 5.2**C we show the cumulative step distribution conditioned on skipping. Measuring all lengths in nucleotides, Ago has diffused the average root-mean square (rms) distance l_{slide} after taking l_{slide}^2 sliding steps between consecutive skips (see **S.I.** for derivation). After having slid the l_{slide}^2 steps, Ago skips on average μ_{skip} nucleotides away in either direction, with a standard deviation of σ_{skip} nucleotides in the length of every skip (**Figures 5.2**B and C). In the **S.I.** we calculate the average shuttling time for such a system numerically using a master-equation formulation. In **Figure 5.2**D we show the resulting shuttling time for a fixed sliding length $l_{\mathrm{slide}} = 12\mathrm{nt}$, while the average skip distance and its standard deviation is either $\mu_{\mathrm{skip}} = 36\mathrm{nt}$ and $\sigma_{\mathrm{skip}} = 0\mathrm{nt}$ (green curve) or $\sigma_{\mathrm{skip}} = 36\mathrm{nt}$ and $\mu_{\mathrm{skip}} = 0\mathrm{nt}$ (orange curve). Both have the same rms skipping length, $l_{\mathrm{skip}} = \sqrt{\mu_{\mathrm{skip}}^2 + \sigma_{\mathrm{skip}}^2} = 36\mathrm{nt}$, with the $\sigma_{\mathrm{skip}} = 0\mathrm{nt}$ case representing skips of definite length that take the protein to a location not reachable in a single round of sliding ($l_{\mathrm{skip}} \gg l_{\mathrm{slide}}$). Contrarily, the protein may (likely) skip to a site already interrogated when $\mu_{\mathrm{skip}} = 0\mathrm{nt}$ – depleting the 'gap' shown in the middle panel of **Figure 5.2**C causes the distributions shown in the middle panel to overlap with that of the top panel. We note a clear resemblance of our numerical solutions to the empirical curves (**Figures 5.1**C and D), including the possibility of non-monotonic behavior when the skip length distribution is tight enough that there is a central gap in the cumulative step distribution just after the first skip (middle panel **Figure 5.2**C).

From the central-limit theorem it follows that the distribution of Ago positions after repeated skip-and-slide (sNs) cycles will approach that of simple diffusive motion with average mean squared step length $l_{\mathrm{sNs}} = \sqrt{l_{\mathrm{slide}}^2 + l_{\mathrm{skip}}^2}$ between each unbinding cycle (bottom panel **Figure 5.2**C), where $l_{\mathrm{skip}}^2 = \mu_{\mathrm{skip}}^2 + \sigma_{\mathrm{skip}}^2$ is the variance added to the cumulative translocation by one skip. In the **S.I.** we use a description conditioned on skips to construct scaling arguments showing that for large trap separations (**Figure 5.2**D)

$$T_{\mathrm{shuttle}}(d_{\mathrm{trap}} \gg l_{\mathrm{sNs}}) \approx \mathrm{const.} + \rho_{\mathrm{scan}}^2 \tau_{\mathrm{trap}} d_{\mathrm{trap}} \text{ with } \rho_{\mathrm{scan}} = l_{\mathrm{slide}}/l_{\mathrm{sNs}} \qquad (2)$$

Here we have introduced the scanning density ρ_{scan} as the fraction of unique bases interrogated by Ago within a single skip-and-slide cycle. Having used our numerical approach to obtain $T_{\text{shuttle}}(d_{\text{trap}})$ curves for a wide range of l_{slide}, μ_{skip} and σ_{skip} (**S.I.**), the resulting final slopes from those curves indeed coincide with the derivative of **Equation** 2, thereby validating our scaling arguments (**Figure 5.2**E).

5.2.4. AGO SKIPS STRAIGHT INTO THE SECOND TRAP FOR INTERMEDIATE TRAP SEPARATIONS

In between the two linear regimes, the shuttling time varies non-monotonically (**Figures 5.1**C, D and **5.2**D). At short distances, when only sliding, the protein's motion is well described by a simple random walk, with consecutive steps being uncorrelated (**Equation** 1). Using the scaling arguments leading up to **Equation** 2, a similar uncorrelated motion over segments of length l_{sNs} is expected at large trap separations. Although we expect said scaling arguments to fail (i.e. ignoring the constant in **Equation** 2) within the intermediate (non-monotonic) regime, preventing us from estimating the corresponding shuttling times, we can still estimate the trap separation at which we expect a local minimum shuttling time.

If the trap is not the outermost sequence on the construct, as is the case in our experiment (**Figure** S5.1), the initial sliding induces no average shift in position, and it stands to reason that the local minimum in shuttling times appears at a trap separation μ_{skip}, from where Ago typically slides straight into the second trap after the first skip. Below, we shall use this reasoning to estimate l_{slide} and l_{skip} from the data. Note that our numerical calculations have been performed for traps placed as the most outer sequence on the construct. For such a system Ago drifts an approximate distance l_{slide} towards the other trap before skipping, which is why **Figure 5.2**D shows a curve with its minimum around a trap separation of $\mu_{\text{skip}} + l_{\text{slide}} = 48$nt (orange curve).

5.2.5. AGO SKIPS OVER TWO THIRDS OF ALL BASES

Applying the above arguments to our experimental data, we estimate the trapping time τ_{trap} by fitting **Equation** 1 to the initial linear part of the shuttling time dependence on trap distance (left most line in **Figures 5.1**C and D, $\tau_{\text{trap}} = 0.062 \pm 0.003$s for CbAgo and $\tau_{\text{trap}} = 0.057 \pm 0.002$sfor hAgo2)(see **Methods**). Next, we can determine the scanning density ρ_{scan} by fitting **Equation** 2 to the final linear part of the data (right most line in **Figures 5.1**C and D). The resulting scanning densities ($\rho_{\text{scan}} = 0.38 \pm 0.03$ for CbAgo and $\rho_{\text{scan}} = 0.31 \pm 0.04$ for hAgo2) indicate that only approximately one in three bases are checked by Ago while moving along its substrate.

We can further give rough estimates of the sliding distance and skip length as follows. As we see a dip in the shuttling time we know that skipping can only be a viable avenue of translocation above a certain trap separation, and thus there should be a gap in the position distribution of a skip-and-slide cycle just after the first skip (middle panel **Figure 5.2**C). For there to be a substantial gap in this distribution we need a clear separation between the distributions shown in the first two panels of **Figure 5.2**C. In mathematical terms, $\sigma_{\text{skip}}^2 + l_{\text{slide}}^2 \ll \mu_{\text{skip}}^2$, implying that $l_{\text{sNs}} \approx l_{\text{skip}} \approx \mu_{\text{skip}}$, and that the dip visible in the shuttling time (**Figures 5.1**C and D) essentially reports on this

quantity. With a dip for both systems occurring around trap-separations of 30 nt, this implies a skipping distance of around $l_{skip} \approx 30$nt. With a scanning density of a third, this skip distance in turn suggests that both sliding distances are around $l_{slide} \approx 10$nt, or equivalently, Ago takes around 100 sliding steps between skips.
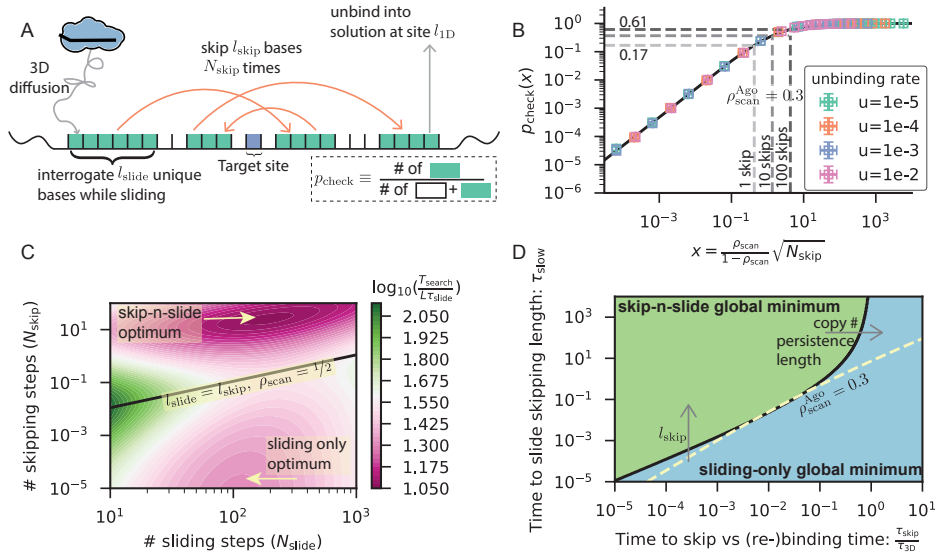

Figure 5.3: Optimal search times. (A) Schematic of a single search round. In search of the unique target, the protein uses a combination of skipping and sliding along the substrate before it unbinds into solution and must perform 3D diffusion before it can return. Only sites slid past (at least once) are interrogated (green), resulting in a probability p_{check} to interrogate a particular site. **(B)** Comparison of $p_{check}(x)$ (solid line, Equation 3) to Monte Carlo simulations (symbols) (details given in Methods). Dashed lines indicate Argonautes ($\rho_{scan} \approx 0.3$) that typically skip once (light grey), 10 (dark grey), and 100 times (black) before unbinding. **(C)** Search time versus N_{slide} and N_{skip}. Region above the solid line represents sparse scanning ($\rho_{scan} < 0.5$), while the region below it represents dense scanning ($\rho_{scan} > 0.5$). **(D)** Phase-diagram showing when $T_{search}^{sNs} < T_{search}^{sliding}$. Dashed line represents the constant scanning density of 0.3 (approx. the value estimated for both Ago). Arrows represent directions of increasing l_{skip}, protein copy number (concentration) and substrate persistence length.

5.2.6. THE TOTAL SEARCH TIME

Having shown that both hAgo2 and CbAgo skip over a significant number of bases—about double the number of bases it actually scans in any skip-and-slide cycle—we now turn to the question why both Argonaute – from different kingdoms of life – behave so similarly. Under what conditions does skipping speed up a protein's search for a single target in the genome or mRNA pool? To answer this question, we now theoretically consider what combinations of the number and length of skipping and sliding steps – and thereby scanning density – lead to minimal overall search times.

We consider a target searcher that after diffusing through solution, binds its substrate randomly and non-specifically to perform a lateral excursion consisting of both

skipping and sliding before unbinding (or finding the target). In a lateral excursion that ends with unbinding, we take the protein to undergo an average of N_{skip} skips, and N_{slide} slides. Note that N_{slide} does not equal the previously defined l_{slide}^2, as the latter is the number of sliding steps between consecutive skips, while the former equals l_{slide}^2 multiplied by the number of skips prior to unbinding (see **S.I.**). To estimate the total time to find the target, we first determine the average number N_{rnd} of search rounds ('rnd') (binding-skip-and-slide-unbinding) needed before the target is found, and then the average time T_{rnd} of each search round [7, 27, 42]. In what follows, we express both N_{rnd} and T_{rnd} for target searchers using a mixture of skipping and sliding corresponding to a scanning density ρ_{scan}, after which we shall proceed to minimize the search time in terms of the frequency of skipping and sliding steps taken. To properly model the skip-and-slide process between unbinding events, we must cover the scenario presented in **Figure 5.3**A: even though the target sits in between the binding and unbinding locations, it might still be skipped over. In the **S.I.** we show that the average fraction of bases checked at least once over the rms lateral diffusion distance $l_{1D} = \sqrt{N_{skip}} l_{sNs}$ between binding and unbinding can be estimated using the scanning density and the typical number of skips prior to unbinding as (see **Figures 5.3**A and B)

$$p_{check}(x) = 1 - \frac{\log(1+2x)}{2x}, \text{ with } x = \frac{\rho_{scan}}{1-\rho_{scan}}\sqrt{N_{skip}} \tag{3}$$

The total number of checked sites at a fixed scanning density increases with increasing number of skips per binding event. The logic being that an increased number of skips allows for repeated rescanning of the same region of DNA sites, with the protein every time interrogating about ρ_{scan} of these sites. **Figure 5.3**B shows that if the Argonaute proteins ($\rho_{scan} \approx 0.3$) are to skip on average 100 times before unbinding, they still interrogate only about 60% of all sites spanned within its lateral excursion (dashed lines). Hence, after correcting for repeated scanning due to skipping, Ago likely still leaves a significant portion of the RNA/DNA unseen. We validated **Equation** 3 (solid line in **Figure 5.3**B) using Monte Carlo simulations (colored data points, **Methods**).
Each lateral diffusion event checks on average $p_{check} l_{1D}$ distinct bases, and with a single target on a substrate of L nucleotides, it will take on average $N_{rnd} = L/p_{check}l_{1D}$ cycles before the target is found.

Each search round can be split between base interrogation through 1D lateral diffusion and 3D diffusion through solution. The 1D lateral diffusion time $\tau_{1D} = T_{slide} + T_{skip}$ can further be split into the total time spent interrogating off-targets after a sliding step $T_{slide} = N_{slide}\tau_{slide}$, and the total time spent completing skips and interrogating the landing site $T_{skip} = N_{skip}\tau_{skip}$. The timescales for interrogating off-targets after a sliding event τ_{slide}, executing skips τ_{skip} (including the time to interrogating the site of arrival), and executing excursions into solution τ_{3D} (including the time to interrogating the site of binding), together with the average number of rounds to find the target, leads us to the total search time

$$T_{search} = T_{rnd}N_{rnd} = (\underbrace{N_{slide}\tau_{slide}}_{T_{slide}} + \underbrace{N_{skip}\tau_{skip}}_{T_{skip}} + \tau_{3D})\frac{L}{l_{1D}p_{check}} \tag{4}$$

We will seek the minima of the search time, but before proceeding we must consider what variables evolution could act upon to create a balance between skipping, sliding, and unbinding.

From the definition of the microscopic timescales we immediately have $\tau_{skip}, \tau_{3D} > \tau_{slide}$ as the sliding motion itself costs negligible time by assumption, and both skipping and excursions into solution are ended by interrogating the base at arrival (τ_{slide}). Further, we only ever expect to find an optimum with a balance between skipping and unbinding when the time to complete a skip is shorter than the time to return from solution. If returning from solution would be faster than completing a skip, skipping would always be eliminated and unbinding favored because it has both lower redundancy and is completed quicker. Further, decreasing any of the microscopic timescales associated with different search modes will clearly speed it up. Therefore, we assume these times to already be reduced as far as possible, and ordered as $\tau_{3D} > \tau_{skip} > \tau_{slide}$.

Apart from the three microscopic timescales, there are three more independent parameters evolution could act upon. These are the total number N_{skip} of skips N_{skip} in one search round, the number N_{slide} of off-targets checked after sliding in one search round, and the rms skip distance l_{skip} (or equivalently N_{skip}, N_{slide} and ρ_{scan}, see **S.I.**). Increasing only the rms skipping distance l_{skip} will always reduce the scanning redundancy, and so will always reduce the search time. Since we observe skips of finite length, we also assume these to be externally limited, and take also l_{skip} to be fixed. We are left with two independent parameters, and in **Figure 5.3**C we plot the search time as a function of N_{skip} and N_{slide} when $l_{skip} = 30$nt and $\tau_{3D} = 10\tau_{skip} = 100\tau_{slide}$.

Minimization of the search time over our remaining two independent variables – the number of skips N_{skip} and the scanning density ρ_{scan} (defined in **Equation** 2) – results in two conditions that need to be satisfied at any optimum (see **S.I.**). We present the general conditions in the **S.I.**, and here present solutions valid in regimes of both high and low scanning densities to determine when skip-and-slide search, of the kind observed for Ago, is favored.

5.2.7. Sliding is optimal for scanning densities above 1/2

One local minimum exists in the densely scanned region ($1 - \rho_{scan} \ll 1/2$) and corresponds to the protein using sliding as its only lateral diffusion mode, eliminating skips entirely. The minimum is defined by, $\rho_{scan}^{sliding} = 1$, and (see **S.I.**)

$$N_{skip}^{sliding} = 0, \; N_{slide}^{sliding} = \frac{\tau_{3D}}{\tau_{slide}} \Rightarrow T_{slide} = \frac{1}{2}T_{rnd} \tag{5}$$

This minimum corresponds to the known minimum when a priori assuming that there are no skips [7, 27]. Namely, the protein spends half its time diffusing through solution and the other half of the time sliding (the rightmost identity in **Equation** 5 is equivalent to $\tau_{1D} = \tau_{3D}$). The search time at this minimum equals (see **S.I.**)

$$T_{search}^{sliding} = 2L\sqrt{\tau_{slide}\tau_{3D}} \tag{6}$$

The non-skip minimum is the only minimum in the densely scanned regime ($\rho_{scan} > 1/2$) (**Figure 5.3**C, minimum coinciding with horizontal axis), suggesting that it might be hard to evolve away from the it by incremental steps.

5.2.8. A MIX OF SKIPPING AND SLIDING IS OPTIMAL FOR SCANNING DENSITIES BELOW $1/2$

For the skipping to be beneficial, skips must be large enough ($l_{skip} \gg l_{slide}$ or equivalently $\rho_{scan} \ll 1/2$) to get the system beyond the barrier visible in **Figure 5.3**C. In the **S.I.** we show that after recognizing

$$\tau_{slow} = \tau_{slide} l_{skip}^2 \tag{7}$$

as the time needed to traverse the length of a skip purely through sliding (diffusion with 1 nt steps) – a measure of the added benefit of using skipping – we obtain the location of the skip-n-slide optimum corresponding to a scanning density of $\rho_{scan}^{sNs} = l_{skip}^{-1} \sqrt{N_{slide}^{sNs}/N_{skip}^{sNs}} < 0.5$, with (see **S.I.**)

$$N_{slide}^{sNs} = \frac{\tau_{3D}}{\tau_{slide}} \left(\frac{\tau_{slow}}{\tau_{3D}} \right)^{\frac{1}{3}}, \; N_{skip}^{sNs} = \frac{\tau_{3D}}{\tau_{skip}} \left(1 + \left(\frac{\tau_{slow}}{\tau_{3D}} \right)^{\frac{1}{3}} \right) \Rightarrow T_{skip} = \frac{1}{2} T_{rnd} \tag{8}$$

Note the final identity shown in **Equation** 8 says that at the skip-and-slide optimum, the protein spends half of its time skipping, and the other half on a combination of sliding and diffusing through solution. In agreement with experimental studies [22, 23], this indicates the protein spends more time diffusing along the DNA then it does through solution ($\tau_{1D} > \tau_{3D}$). The search time at this skip-and-slide optimum equals (see **S.I.**)

$$T_{search}^{sNs} = \frac{2L\sqrt{\tau_{skip}\tau_{3D}}}{l_{skip}} \frac{\sqrt{1 + \left(\frac{\tau_{slow}}{\tau_{3D}} \right)^{\frac{1}{3}}}}{p_{check} \left(\left(\frac{\tau_{slow}}{\tau_{3D}} \right)^{\frac{2}{3}} \right)} \tag{9}$$

5.2.9. GLOBAL OPTIMAL SEARCH STRATEGY

As there are local minima in both the sparsely and densely scanned regions (**Equations** 6 and 9), the global optimal search strategy is defined by which of these two minima have the smallest search time. The condition for the slip-and-slide minimum being the global minimum ($T_{search}^{sNs} < T_{search}^{sliding}$) can be written as (see **S.I.**)

$$\frac{\tau_{skip}}{\tau_{slow}} < \frac{p_{check}^2 \left(\left(\frac{\tau_{slow}}{\tau_{3D}} \right)^{\frac{2}{3}} \right)}{1 + \left(\frac{\tau_{slow}}{\tau_{3D}} \right)^{\frac{1}{3}}} < 1 \tag{10}$$

Figure 5.3D shows the corresponding phase diagram – in $\left\{ \tau_{slow}, \frac{\tau_{skip}}{\tau_{3D}} \right\}$-space – showing when the skip-and-slide minimum is the global minimum. We previously argued that if $\tau_{3D} < \tau_{skip}$ there will be no skip-and-slide minimum. Now we see that for $\tau_{3D} > \tau_{skip}$ we can always find an l_{skip} long enough that the skip-and-slide optimum is also the global optimum (upward arrow in **Figure 5.3**D). Logically, the skip-and-slide optimum is only preferred over the sliding-only one for $\tau_{slow} > \tau_{skip}$, indicating the typical return time of a skip may not exceed the time needed to cover the same distance by just sliding, and **Equation** 10 gives the more stringent condition that must be satisfied.

We conclude by noting both of the Argonaute proteins considered above have $\rho_{\text{scan}} \approx 0.3$ (yellow dashed line in **Figure 5.3**D), putting the system above the line separating the sparse and dense scanning regimes (**Figure 5.3**C). Certainly, hAgo2 and CbAgo operate far from the sliding-only optimum, and, as we shall discuss further below, are working in the regime where the skip-and-slide optimum is found (crossing point **Figure 5.3**D).

5.3. DISCUSSION

Site-specific DNA or RNA binding proteins must find a single sequence amongst megabase (prokaryotes) to gigabase (eukaryotes) pools of off-targets. Here we have shown that facilitated diffusion with a mixture of sliding (single-nucleotide steps) with frequent and large skips (multi-nucleotide steps) is capable of reducing the overall search time beyond using sliding by itself. Interestingly, pure sliding is a possible optimal strategy, and the search time for skips shorter than the sliding length is minimal only after eliminating skips entirely as their temporal cost is no longer accompanied by the benefit of visiting off-targets not encountered before (**Figure 5.3**C). Contrarily, skips greater than the sliding length reduce the probability of redundantly sampling off-targets, and we find another optimum where the search time is minimal if skips are used so frequently that the system spends half of the time skipping. We further showed how single-molecule FRET experiments (**Figure 5.1**) can be used to extract what we termed the scanning density, a measure of the fraction of bases directly interrogated during a skip-and-slide cycle (**Figure 5.2**). Our experiments performed on a prokaryotic (CbAgo) and eukaryotic (hAgo2) Argonaute revealed both to have scanning densities around 0.3 (**Figures 5.1**C and **5.1**D)—well within the sparse scanning regime (**Figure 5.3**C).

As shown in **Figure 5.3**C, the scanning densities of the Argonaute proteins are consistent with having skip-n-slide search as an optimal strategy. However, according to **Figure 5.3**D it appears at this the system just touches the separating line determining the global optimum. One might speculate what other factor, not taken into account in our modeling, could have driven Ago away from the sliding only optimum. As shown in reference [41], skips are needed to surpass roadblocks present on any physiological substrate. We therefore hypothesize that if one limits the sliding length to be less than the typical separation between other (high affinity) binding proteins – i.e. about 30 nt for cellular RNA or 30-100nt for DNA [31, 32] – it to always be beneficial to include skips ($T_{\text{search}}^{\text{sNs}} < T_{\text{search}}^{\text{sliding}}$).

Based on our results, for a low scanning density to be preferred, the binding rate from solution should not exceed the return rate after skipping (**Figure 5.3**D). As binding rates scale linearly with concentrations (before reaching saturating levels), we thus expect binding proteins present at lower copy numbers to be prone to use more frequent skips (arrow in **Figure 5.3**D). For example, E.coli cells express about 1-10 copies of the lac repressor [48] and experiments have indeed seen signatures of a skipping-and-sliding mixture [49].

Instead of increasing (reducing copy number), a reduction in τ_{skip} is to be expected on more flexible substrates, such as single-stranded DNA or RNA. We therefore deem it likely that skip-n-slide search to also be used by sequence specific single-stranded binding proteins other than Argonaute, such as ribosomes searching for the transcription start site. We hope to motivate future experiments utilizing different DNA binding proteins to investigate whether they belong to the "sliding only" ($\rho_{\text{scan}} \gg 1/2$) or the "skipping-and-sliding" ($\rho_{\text{scan}} \ll \frac{1}{2}$) class (**Figure 5.3**C).

Within our analysis of the total search time we have decoupled the return time from a skip (τ_{skip}) from the average length thereof (l_{skip}). Hence, fixing the time, there is no penalty for ever increasing skipping distances. In fact, for large enough skipping distances we can always reach a situation where the skip-and-slide optimum is the global optimum (provided $\tau_{\text{skip}} < \tau_{\text{3D}}$)(**Figure 5.3**D). In our previous work [41] we demonstrated the duration of skips to be limited by the time needed to escape the bound site – rather than the time needed to find the distant location – justifying our assumption for Argonaute. However, skips limited by the rate of rebinding – for instance through diffusion – couple τ_{skip} to l_{skip} and we expect an optimal l_{skip} to exist. As we here focused on the coupling between search time and the experimentally measurable ρ_{scan}, we deem such an analysis beyond the scope of the presented research, but an interesting future direction.

A previous study [7] has pointed out that speeding up the lateral diffusion – by reducing the variation in binding strengths along the genome – comes at the cost of reducing the protein's specificity. The authors proposed that in order to overcome this apparent 'search-stability paradox' the protein must switch between two conformations – one with higher affinity (for specificity) and one with a lower one (for speed) – and detail the tight constrains on the binding energies for such a solution to exist [28, 43]. Selected target searchers – including selected RNA guided nucleases [28, 33, 45–47] – indeed adopt multiple conformations during target interrogation [16, 20]. The necessity for two protein conformations, however, arises from assuming the protein is only capable of sliding, thereby forcing the protein to sample every site along the genome. We hypothesize that using the different skip-and-slide scheme described here could provide a complementary/alternative route to being both fast and specific –allowing for wider spreads in binding energies – especially for proteins that are not known to exhibit multiple conformations.

The experiments performed here – together with our theoretical analysis – are in principle applicable to other DNA binding proteins. Proteins not guided by non-coding DNA/RNA should be labeled with the donor dye directly. Moreover, both Ago proteins examined here bind single-stranded nucleic acids, which have close to nucleotide persistent lengths [30] and thereby offer a clear possible mechanism of introducing frequent skips – Ago can skip to distant sequences as they can come close together in space. Yet, the presented analysis and experiment do not rely on such, and proteins binding double-stranded DNA – persistence lengths ~50 nt– can similarly be investigated for the presence of (presumably larger and less frequent) skips, without

prior knowledge of a possible microscopic mechanism for skipping.

In conclusion, a search strategy combining skipping and sliding can significantly increase the rate of association to the cognate target – which is of critical importance for proper functioning of the cell – and Argonaute proteins adopt scanning densities consistent with their mixture being optimal.

5.4. METHODS

5.4.1. MONTE CARLO SIMULATIONS FOR VALIDATING

To test the validity of **Equation** 3, we set up Monte Carlo simulations (code written in Python). The proteins are assigned a unity step rate to either side, as well as an unbinding rate u. Hence in every move, the protein diffuses to one of its neighboring site with a probability $1/2+u$ and unbinds with a probability $u/2+u$. Before every move, the protein interrogates the site currently located at with a fixed probability of ρ_{scan}. Each of the 1000 runs ends when the protein unbinds. The corresponding value of x is evaluated using the distance between binding and unbinding sites (see definition of x above **Equation** S5.47). We estimate the value of p_{check} as the fraction of sites visited that are interrogated. Error bars in **Figure 5.3**B show 95% confidence intervals for both x and p_{check}. Simulations we repeated for in $[10^{-10}, 10^{-9}, 10^{-8}, ..., 10^{-2}, 0.9, 0.8, ...0.1]$, and u in $[10^{-5}, ..., 10^{-2}]$ as indicated in **Figure 5.3**B.

5.4.2. BOOTSTRAPPING FOR ERROR ESTIMATION AND BASED ON SMFRET DATA

Fitting the data from the tandem target assay to **Equation** 1 provides the estimate of τ_{trap}. We bootstrapped the dwell time distributions acquired using the original tandem target assay (distances of 11 nt, 15 nt, 18 nt and 22 nt (CbAgo) and 7 nt, 11 nt, and 15 nt (hAgo2)). For each of the 10^5 bootstrap samples we calculated new values for the associated $T_{shuttle}$'s and repeated the fit to **Equation** 1 to obtain an error estimate in the fitted value of the escape rate. In similar fashion, we used **Equation** 2, together with the estimate of τ_{trap} from the original dataset, to determine ρ_{scan} (distances of 64 nt, 92 nt and 120 nt (CbAgo) and 80 nt, 120 nt, 160 nt (hAgo2)). All analysis was performed with a custom code written in Python. Shaded areas in **Figures 5.3**C and D represent 95% confidence intervals.

5.4.3. PROTEIN PURIFICATION

CbAgo was purified according to Hegge et al, 2019 [40]. hAgo2 was purified according to Chandradoss et al, 2015 [35].

5.4.4. NUCLEIC ACID PREPARATION

RNA constructs with a single amine-C6-uridine modification were ordered from STPharm. After labelling with Cy5 according to [44], the constructs were precipitated. The RNA constructs were subsequently annealed to a DNA splint (specific for RNA and U40 mer), a second DNA splint (for ligating U40 mers) and a U40 mer (in the ratio 1:2:3:3). After ligation with T4 RNA ligase II (NEB), the ligated constructs were run on

a 10% PAGE. Different ligated populations are created through this process (for example, TGT- U40 or TGT-U40-U40 etc) and these are then excised from the gel and concentrated through ethanol precipitation. The concentrated and ligated RNA constructs were again annealed to a DNA construct and an RNA target with biotin on the 3' end. Ligation was again performed with T4 RNA ligase II. DNA oligos with a single amine-C6-thymine modification were ordered from ELLA Biotech GmbH and labeled in the same way as the RNA.

5.4.5. SAMPLE PREPARATION
Quartz slides were prepared according to [50]. Briefly, quartz slides were cleaned with detergent, sonicated and treated with acetone and subsequently KOH. Coverslips were directly sonicated with KOH. Piranha cleaning was done followed by treatment with methanol and incubation of (3-Aminopropyl)triethoxysilane (APTES) for both coverslips and quartz slides. PEGylation took place overnight and slides and coverslips were stored at -20 °C. Before single-molecule experiments, an extra round of PEGylation took place with MSPEG-4. The quartz slide was then assembled with scotch tape and epoxy glue and the chamber is flushed in T50 and 1% Tween-20 for >10min to further improve the surface quality of the single-molecule chambers [51]. Channels were thoroughly washed with T50 before adding in streptavidin (0.1 mg/mL) for 1 min. Subsequently, DNA or RNA was immobilized on the surface through biotin-streptavidin conjugation. 10 nM CbAgo or hAgo2 was incubated with 1 nM guide in (100 mM NaCl for CbAgo, 50 mM NaCl for hAgo2), 50 mM Tris, 1 mM Trolox, 0.8% glucose for 30 min. Lastly, glucose oxidase (0.1 mg/mL final conc.) and catalase (17 g/mL final conc.) were added and introduced in the chamber.

5.4.6. EXPERIMENTAL SETUP
Single-molecule experiments were performed on a custom built inverted microscope (IX73, Olympus) using prism-TIRF and a 60X water immersion objective (UPLSAPO60XW, Olympus). The Cy3 dye was excited using a 532 nm diode laser Compass 215M/50mW, Coherent) and the Cy5 dye was excited using a 637 nm diode laser (OBIS 637 nm LX 140 mW). The scattered light was blocked by a 532 nm notch filter (NF03-532E-25, Semrock) and a 633 nm notch filter (NF03-633E-25, Semrock) after which the remaining signal from the fluophores was separated into two separate channels. Lastly, the light is projected on a EM-CCD camera (iXon Ultra, DU-897U-CS0-# BV, Andor Technology). Before each experiment, a reference movie was taken with the red laser to excite the Cy5 dyes on the nucleic acid molecules of interest. After that, a movie is taken with the green laser. The single-molecule experiments were taken at room temperature (20 ± 0.1 °C).

5.4.7. ANALYSIS OF RAW DATA
The raw data was analysed using custom written code in IDL, where the reference movie is used to take into account only the regions of interest (i.e. the regions that contain a Cy5). The resulting time traces where further analysed in MATLAB (Mathworks) where the shuttling rates were extracted through the use of Hidden Markov software called ebFRET (http://ebfret.github.io/) and custom written code in Matlab.

5.5. AUTHOR CONTRIBUTIONS

M.K. and M.D. performed all theoretical analysis, T.J.C. and C.J. designed the experiments, and T.J.C. performed the experiments. I.M. provided the hAgo2 protein. M.K., T.J.C., M.D. and C.J. wrote the manuscript.

5.6. ACKNOWLEDGMENTS

5.7. SUPPLEMENTAL INFORMATION

5.7.1. DETERMINING SHUTTLING TIMES USING A MIXTURE OF SKIPPING AND SLIDING

We here build a kinetic model for the lateral diffusion by target searching proteins capable of explaining the experimental data shown in **Figure 5.1**.

MODELING SKIPPING-AND-SLIDING LATERAL DIFFUSION

Given the protein can in principle (attempt to) bind any sequence along the DNA or RNA, we imagine binding sites to be a nucleotide apart. When bound to site i, the protein diffuses away (in either direction) at a rate

$$k_{\text{move}}(i) = \begin{cases} k_{\text{trap}} & \text{at trap} \\ k_{\text{ns}} & \text{at non-specific site} \end{cases} \tag{S5.1}$$

We assume the binding energy at the trap is significantly greater than at any non-specific site, with both still being significantly more stable than the unbound state. As a result, the (average) shuttling time measured in our *in vitro* experiments - the system contains two stronger binding traps and a limited amount of remaining off-targets - is governed by movements from the trap.

$$k_{\text{ns}} \gg k_{\text{trap}} \tag{S5.2}$$

Ignoring any temporal contribution from the non-specific sites reflects the lack of any directly observable FRET signal corresponding to the protein being at these locations (**Figure 5.1**). Furthermore, given the TIRF microscopy assay ensures we are tracking laterally diffusing proteins that did not unbind - proteins diffusing through solution move in and out of the evanescent field too fast to be detected - we shall ignore the protein's intrinsic unbinding rate at all sites for now - an assumption that is further justified by noting that typically more than 10 shuttle events occur prior to unbinding.

In every move, taking an average time of k_{move}^{-1}, the protein can either slide - step to its neighbors - or skip - step further. We let the rate to step away from site i still be set by **Equation** S5.1 and assign a probability that such a step is of definite length $|l|$ (in nucleotides). Letting $\delta_{x,y}$ denote the Kronecker delta,

$$P(l, l_{\text{slide}}, l_{\text{skip}}) = \frac{n_{\text{slide}}(l_{\text{slide}})}{1 + n_{\text{slide}}(l_{\text{slide}})} \delta_{|l|,1} + \frac{1}{1 + n_{\text{slide}}(l_{\text{slide}})} s(|l|, l_{\text{skip}}) \tag{S5.3}$$

, with $\sum_{n>0} P(n) = 1$. The weight of a skip of length $|l|$ as a function of the typical skipping length l_{skip}, is denoted by $s(|l|, l_{\text{skip}})$. Further, n_{slide} is the typical number of sliding steps taken between two consecutive skips. Given a sliding step displaces the protein by a single nucleotide, the stochastic variable Δn_i representing the number of nucleotides moved during one such step follows

$$\Delta n_i = \begin{cases} +1 \text{ nt} & p = 1/2 \\ -1 \text{ nt} & 1 - p = 1/2 \end{cases} \tag{S5.4}$$

Hence, the mean squared displacement after n_{slide} of such steps equals

$$\begin{aligned} (1 \text{ nt})^2 \, l_{\text{slide}}^2 &= \langle \left(\sum_{i=1}^{n_{\text{slide}}} \Delta n_i \right)^2 \rangle \\ &= \sum_{i=1}^{n_{\text{slide}}} \sum_{j=1}^{n_{\text{slide}}} \langle \Delta n_i \Delta n_j \rangle \\ &= \sum_{i=1}^{n_{\text{slide}}} \langle (\Delta n_i)^2 \rangle + \sum_{i \neq j} \langle \Delta n_i \rangle \langle \Delta n_j \rangle \\ &= n_{\text{slide}} \times \langle (\Delta n_1)^2 \rangle \\ &= n_{\text{slide}} (1 \text{ nt})^2 \end{aligned} \tag{S5.5}$$

, where in the third line we have used the independence of individual steps. We define the 'sliding length', $l_{\text{slide}} = \sqrt{n_{\text{slide}}}$, as the typical number of nucleotides covered sliding between two consecutive skips - the rms displacement of a simple random walk with n_{slide} steps. Rewritten in terms of the now defined sliding length l_{slide}, the probability of taking a step of length $|n|$ reads

$$P(n, l_{\text{slide}}, l_{\text{skip}}) = \frac{l_{\text{slide}}^2}{1 + l_{\text{slide}}^2} \delta_{|n|,1} + \frac{1}{1 + l_{\text{slide}}^2} s(|n|, l_{\text{skip}}) \tag{S5.6}$$

The (effective) rate from i to j then equals

$$\kappa(i, j | l_{\text{slide}}, l_{\text{skip}}) = k_{\text{move}}(i) P(|i - j|, l_{\text{slide}}, l_{\text{skip}}) \tag{S5.7}$$

As we will show below, the behavior of the resulting shuttling times both at short and long distances is independent of the choice of the distribution s. Yet, all numerical results are obtained using

$$s(n, \mu_{\text{skip}}, \sigma_{\text{skip}}) = \int_{n-1/2}^{n+1/2} \left[G(n | \mu_{\text{skip}}, \sigma_{\text{skip}}) + G(n | -\mu_{\text{skip}}, \sigma_{\text{skip}}) \right] dn \tag{S5.8}$$

with

$$G(x, \mu_{\text{skip}}, \sigma_{\text{skip}}) = \frac{1}{\sqrt{2\pi\sigma_{\text{skip}}^2}} e^{\frac{-(x-\mu_{\text{skip}})^2}{2\sigma_{\text{skip}}^2}} \tag{S5.9}$$

denoting the Gaussian distribution with average μ_{skip} and standard deviation σ_{skip}. Hence, the length of each skip is normally distributed, with a typical (rms) skipping length of

$$l_{\text{skip}} = \sqrt{\mu_{\text{skip}}^2 + \sigma_{\text{skip}}^2} \tag{S5.10}$$

NUMERICAL METHOD TO SOLVE FOR SHUTTLING TIME

Every shuttling event starts with the protein bound at one of the two trapping sites ($t = 0$) and ends the first time it reaches the other ($t = T_{\text{shuttle}}$), located d_{trap} sites away. Using the transition rates of **Equation** S5.7, letting $P_i(t)$ denote the probability for the protein to reside at site i at time t, and defining the vector

$$\vec{P}(t) = \left[P_1(t), P_2(t), ..., P_{d_{\text{trap}}-1}(t) \right]^T \tag{S5.11}$$

(for ease of notation we omit the sites flanking either trap $i < 1$ and $i > d_{\text{trap}}$, but note the approach mentioned here is applicable also if the traps are not the outermost sites on the construct)

the following set of Master Equations determine the evolution of the occupancies at every site during a shuttling event with the first trap at site 1 and the second at d_{trap}.

$$\frac{\partial \vec{P}}{\partial t} = -K\vec{P}(t) \tag{S5.12}$$

with the elements in rate matrix K given by

$$K_{ij} = \begin{cases} -\kappa(|j - i|, l_{\text{slide}}, l_{\text{skip}}) & \forall i \neq j \\ \sum_{i \neq j} \kappa(|i - j|, l_{\text{slide}}, l_{\text{skip}}) & \forall i = j \end{cases} \tag{S5.13}$$

The shuttle event starts with the protein located at the first trap,

$$P_1(0) = 1, \quad P_i(0) = 0 \ \forall i \neq 1 \tag{S5.14}$$

, and ends when the second trap is reached, whose corresponding outgoing rates are set to zero ($j = d_{\text{trap}}$ in **Equation** S5.13). The probability of completing a shuttle within the time interval $[\tau, \ \tau + \Delta t]$ should be proportional to the change in occupancy at the destination trap ($P_{d_{\text{trap}}}(\tau + \Delta t) - P_{d_{\text{trap}}}(\tau)$). Letting $p_{\text{shuttle}}(\tau)$ denote the probability density of completing the shuttle at time τ, ($p_{\text{shuttle}}(\tau)\Delta t = P_{d_{\text{trap}}}(\tau + \Delta t) - P_{d_{\text{trap}}}(\tau)$, for small enough Δt. Taking $\Delta t \to 0$, we recognize the rate of change of the second trap's occupancy ($\frac{\partial P_{d_{\text{trap}}}(t)}{\partial t}|_{t=\tau}$) as the instantaneous probability that the shuttling time equals τ ($p_{\text{shuttle}}(\tau)$). Denoting the basis vectors $\vec{p}_j$ as $\vec{p}_0 = [1,0,0,.....0]^T$, $\vec{p}_1 = [0,1,0,.....0]^T$, $\vec{p}_2 = [0,0,1,.....0]^T$ and so on, the shuttle times are distributed as

$$\begin{aligned} p_{\text{shuttle}}(\tau) &= \frac{\partial P_{d_{\text{trap}}}(t)}{\partial t}\Big|_{t=\tau} \\ &= -\sum_{j \neq d_{\text{trap}}} \frac{\partial P_j(t)}{\partial t}\Big|_{t=\tau} \\ &\equiv -\sum_{j \neq d_{\text{trap}}} \vec{p}_j^T \frac{\partial \vec{P}(t)}{\partial t}\Big|_{t=\tau} \\ &= +\sum_{j \neq d_{\text{trap}}} \vec{p}_j^T K \vec{P}(\tau) \\ &= +\sum_{j \neq d_{\text{trap}}} \vec{p}_j^T K e^{-K\tau} \vec{P}(0) \end{aligned} \tag{S5.15}$$

In the second line we have used that any additional occupancy at the trap must come from somewhere else on the RNA/DNA ($P_{d_{\text{trap}}}(t) = 1 - \sum_{j \neq d_{\text{trap}}} P_j$). The next lines makes use of **Equation** S5.11 together with the basis vectors to write the elements of $\vec{P}$ as its projections, and the Master Equation, **Equation** S5.12, to work in the rate matrix K and its matrix exponential. The desired average shuttling time (T_{shuttle}) is the first moment of the distribution $p_{\text{shuttle}}(\tau)$,

$$
\begin{aligned}
T_{\text{shuttle}}(d_{\text{trap}}) &= \int_0^\infty \tau \, p_{\text{shuttle}}(\tau) \mathrm{d}\tau \\
&= \int_0^\infty \tau \sum_{j \neq d_{\text{trap}}} \vec{p}_j^{\,T} K e^{-K\tau} \vec{P}(0) \mathrm{d}\tau \\
&= \sum_{j \neq d_{\text{trap}}} \vec{p}_j^{\,T} \left(\int_0^\infty \tau K e^{-K\tau} \mathrm{d}\tau \right) \vec{P}(0) \\
&= \sum_{j \neq d_{\text{trap}}} \vec{p}_j^{\,T} K^{-1} \vec{P}(0)
\end{aligned}
\tag{S5.16}
$$

Using the values of l_{slide}, μ_{skip} and σ_{skip} (thereby knowing l_{skip} via **Equation** S5.10) and the distance between traps d_{trap}, we construct the rates in **Equation** S5.7, build the matrix K, invert it and compute $T_{\text{shuttle}}(d_{\text{trap}})$ as the inner product shown in **Equation** S5.16. Note that if the trap located at d_{trap} is not the outermost binding site on the construct, **Equation** S5.16 is still valid after substituting matrix K for the sub-matrix with its d_{trap}-th row and column removed.

5.7.2. Shuttling times scales with square of scanning density at large trap separations

Given movements along the non-specific parts of the substrate occurred too fast to be observed, T_{shuttle} should be proportional to the time needed to escape the initial trap towards the region in between traps ($\tau_{\text{trap}} = k_{\text{trap}}^{-1}$) multiplied by the number of re-trapping events.

$$
T_{\text{shuttle}}(d_{\text{trap}}) = n_{\text{return}} \tau_{\text{trap}}
\tag{S5.17}
$$

After sufficient rounds of skipping and sliding, the protein's excursion is well described by a random walk with basic step length (**Figure 5.2**, 'sNs':'skip-N-slide'):

$$
l_{\text{sNs}} = \sqrt{l_{\text{slide}}^2 + l_{\text{skip}}^2} = \sqrt{l_{\text{slide}}^2 + \mu_{\text{skip}}^2 + \sigma_{\text{skip}}^2}
\tag{S5.18}
$$

The protein slides - covering l_{slide} nucleotides - before skipping to the next segment of length l_{sNs}. For this coarse-grained system, we once again expect the escaping of the trap to be rate limiting, resulting again in a linear increase of the shuttling time with inter-trap distance, similar to the case of diffusion purely by sliding (**Equation** S5.17),

$$
T_{\text{shuttle}}(d_{\text{trap}}) = \text{const.} + \hat{n}_{\text{return}} \tau_{\text{trap}}
\tag{S5.19}
$$

Here we are concerned only with $T_{\text{shuttle}}(d_{\text{trap}})$'s scaling with d_{trap}, for which it is only the term proportional to τ_{trap} that has to be taken into account. In the coarse-grained

system

$$\hat{n}_{\text{return}} = (\text{\# returns to segment that contains the first trap})$$
$$\times (\text{\# returns to trap when in first segment}) \qquad \text{(S5.20)}$$
$$\equiv \hat{n}_{\text{segment}} \times \hat{n}_{\text{retrap}}$$

To get the average number of re-entries to the first segment we must derive its corresponding probability. First, given a skip translocates the protein to an adjacent segment of l_{sNs} nucleotides, and l_{slide}^2 steps are taken within each segment

$$\rho_{\text{scan}} = \frac{l_{\text{slide}}}{l_{\text{sNs}}} = \frac{l_{\text{slide}}}{\sqrt{l_{\text{slide}}^2 + l_{\text{skip}}^2}} \qquad \text{(S5.21)}$$

denotes the typical fraction of interrogated sites along the substrate, or 'scanning density'. In other words, any particular site within a l_{sNs}-long region of DNA/RNA has a probability of ρ_{scan} to be interrogated prior to the protein moving beyond this segment. Equivalently, the protein visits a segment without checking (all) the sites within it with a probability of $1 - \rho_{\text{scan}}$. Next, let $P_{\text{shuttle}}(\hat{d})$ denote the probability of traversing/shuttling across $\hat{d}$ segments without entering the previous segment. We shall derive $P_{\text{shuttle}}(d)$ below. Having entered the first of the $\hat{d}_{\text{trap}} = d_{\text{trap}}/l_{\text{sNs}}$ segments that lie between the traps, the probability of returning to the segment that contains the initially bound trap equals (**Figure** S5.2).

$$P_{\text{segment}} = \left(1 - P_{\text{shuttle}}(\hat{d}_{\text{trap}})\right)$$
$$+ P_{\text{shuttle}}(\hat{d}_{\text{trap}}) \sum_{m=0}^{\infty} \left(\left(1 - \rho_{\text{scan}}\right)\left(1 - P_{\text{shuttle}}\right)\right)^m \left(1 - \rho_{\text{scan}}\right) P_{\text{shuttle}}(\hat{d}_{\text{trap}}) \qquad \text{(S5.22)}$$

The first term is the probability of immediately going back to the segment the protein started from, while the sum accounts for the probability of all paths that reach the segment that contains the second trap, do not get captured by it, and eventually return back to the first trap (**Figure** S5.2). For instance, the $m = 0$ term ($P_{\text{shuttle}}(1 - \rho_{\text{scan}})P_{\text{shuttle}}$) represents the path that walks to the opposite side of the construct, does not interrogate the final trap and walks back across the construct to arrive back at the segment with the initially bound trap.

Using a similar type of 'path counting', we find the probabilities P_{shuttle} and $P_{\text{no shuttle}} = 1 - P_{\text{shuttle}}$, for a given inter-trap distance $\hat{d}_{\text{trap}}$ to equal (**Figure** S5.3)

$$P_{\text{no shuttle}}(\hat{d}_{\text{trap}}) = \sum_{m=0}^{\infty} \left(\frac{1}{2}\left(1 - P_{\text{shuttle}}(\hat{d}_{\text{trap}} - 1)\right)\right)^m \frac{1}{2} \qquad \text{(S5.23)}$$

$$P_{\text{shuttle}}(\hat{d}_{\text{trap}}) = \sum_{m=0}^{\infty} \left(\frac{1}{2}\left(1 - P_{\text{shuttle}}(\hat{d}_{\text{trap}} - 1)\right)\right)^m \frac{1}{2} P_{\text{shuttle}}(\hat{d}_{\text{trap}} - 1) \qquad \text{(S5.24)}$$

- from which we can write the recurrence relation

$$P_{\text{shuttle}}(\hat{d}_{\text{trap}}) = P_{\text{no shuttle}}(\hat{d}_{\text{trap}}) P_{\text{shuttle}}(\hat{d}_{\text{trap}} - 1) \qquad \text{(S5.25)}$$

The above can be re-written as

$$P_{\text{shuttle}}(\hat{d}_{\text{trap}}) = \frac{P_{\text{shuttle}}(\hat{d}_{\text{trap}} - 1)}{P_{\text{shuttle}}(\hat{d}_{\text{trap}} - 1) + 1} \tag{S5.26}$$

, which subjected to the boundary condition $P_{\text{shuttle}}(1) = 1$ - signifying that if the traps are placed in adjacent segments, the shuttle is complete once the protein escaped the trap for the first time - has the simple solution

$$P_{\text{shuttle}}(\hat{d}_{\text{trap}}) = \frac{1}{\hat{d}_{\text{trap}}} \tag{S5.27}$$

Given the probability of re-entering the first segment, the average number of times this occurs prior to eventually shuttling across equals

$$\hat{n}_{\text{segment}} = \sum_{n=0}^{\infty} n P_{\text{segment}}^n (1 - P_{\text{segment}}) = \frac{P_{\text{segment}}}{1 - P_{\text{segment}}} \tag{S5.28}$$

Using **Equation** S5.22 we find that the protein on average re-enters the segment with the initial trap

$$\hat{n}_{\text{segment}} = \frac{d_{\text{trap}}}{l_{\text{sNs}}} + \frac{l_{\text{sNs}}}{l_{\text{slide}}} - 2 \tag{S5.29}$$

times prior to completing the shuttling event. Once arrived back within the first segment, we must count the (average) number of times the protein gets recaptured by the actual trap ($\hat{n}_{\text{retrap}}$). Assuming sufficient 'skip-and-slide cycles' have taken place, the protein's position is uniformly spread throughout the l_{sNs}-long segment (**Figure 5.2**C). Hence, every step taken within the segment has a probability of $1/l_{\text{sNs}}$ to lead to the trap. Given there are typically $n_{\text{slide}} = l_{\text{slide}}^2$ steps taken prior to a skip (that moves the protein outside of the l_{sNs}-long region),

$$\hat{n}_{\text{retrap}} = \frac{n_{\text{slide}}}{l_{\text{sNs}}} = \frac{l_{\text{slide}}^2}{l_{\text{sNs}}} \tag{S5.30}$$

Taken together, **Equations** S5.29 and S5.30 - by virtue of **Equation** S5.20:

$$\hat{n}_{\text{return}} = \frac{l_{\text{slide}}^2}{l_{\text{sNs}}} \times \left[\frac{d_{\text{trap}}}{l_{\text{sNs}}} + \frac{l_{\text{sNs}}}{l_{\text{slide}}} - 2 \right] \equiv \text{const.} + \frac{l_{\text{slide}}^2}{l_{\text{sNs}}^2} d_{\text{trap}} \tag{S5.31}$$

Hence, when placed sufficiently far apart, the shuttling time (**Equation** S5.19),

$$T_{\text{shuttle}}(d_{\text{trap}}) = \text{const.} + \rho_{\text{scan}}^2 \tau_{\text{trap}} d_{\text{trap}} = \left(\frac{1}{1 + \left(\frac{l_{\text{skip}}}{l_{\text{slide}}} \right)^2} \right) d_{\text{trap}} \tag{S5.32}$$

grows linearly with a slope that scales quadratically with the scanning density (**Equation** S5.21) from which we obtain the ratio between sliding and skipping lengths.

5.7.3. PARAMETER SWEEP AND ESTIMATION OF SLOPES

To construct **Figure 5.2**E, we evaluate **Equation** S5.16 for $l_{slide} \in$ [1 nt, 6 nt, 12 nt, 18 nt, 24 nt, 30 nt, 36 nt, 42 nt], $\mu_{skip} \in$ [0 nt,6 nt, 12 nt, 18 nt, 24 nt, 30 nt, 36 nt, 42 nt] and $\sigma_{skip} \in$ [0.01 nt, 6 nt, 12 nt, 18 nt, 24 nt, 30 nt, 36 nt, 42 nt]. The distance between traps varied from 1-250 nt. The values of l_{slide}, μ_{skip} and σ_{skip} where chosen such that at the largest trap separation of 250 nt the system is always in the regime for which we expect **Equation** S5.32 to hold.

For every $T_{shuttle}$ vs d_{trap} curve, we use the first two points (1 nt , 2 nt) to estimate τ_{trap} (**Equation** 1) and the final two points (249 nt, 250 nt), together with the estimate of τ_{trap}, to estimate ρ_{scan} (**Equation** S5.32).

5.7.4. SEARCH TIME USING SKIPPING AND SLIDING SHOWS TWO OPTIMA

Here we connect the scanning density (ρ_{scan}) that we can extract from experiments to the time needed for a protein to locate a single target embedded within a larger pool of L binding sites. Following [7],

$$T_{search} = N_{rnd} T_{rnd} \tag{S5.33}$$

with T_{rnd} the (average) time each round of facilitated diffusion takes and N_{rnd} the number of such rounds ('rnd') needed to find the target. As mentioned in the main text, we seek to find the minimum search time with respect to the number skips (N_{skip}) and slides (N_{slide}) within every round (binding - lateral diffusion - unbinding).

The length of a skip (l_{skip}), as well as the times to interrogate (slide past) a binding site (τ_{slide}), execute a skip and interrogate the landing site (τ_{skip}), and the time spent on 3D diffusion and interrogating the landing site (τ_{3D}) are all kept constant. The time per round consists of the time spent on the DNA performing lateral diffusion and the time spent in solution performing 3D diffusion.

$$T_{rnd} = \tau_{1D} + \tau_{3D} \tag{S5.34}$$

We further write the time spent on lateral diffusion as the time spent interrogating off-targets either by sliding or skipping,

$$\tau_{1D} = T_{slide} + T_{skip} \tag{S5.35}$$

For ease of calculation, we define the following variables with respect to which we have minimized the search time

$$x = \frac{\rho_{scan}}{1 - \rho_{scan}} \sqrt{N_{skip}} \tag{S5.36}$$

$$y = \frac{\rho_{scan}}{1 - \rho_{scan}} \tag{S5.37}$$

Written in terms of x and y (**Equations** S5.36 and S5.37), the total times spent either on sliding or skipping become

$$T_{slide} = N_{slide}\tau_{slide} = (x\delta l)^2 \tau_{slide} \tag{S5.38}$$

$$T_{skip} = N_{skip}\tau_{skip} = (x/y)^2 \tau_{skip} \tag{S5.39}$$

Here we have introduced the variable $\delta l = l_{\text{sNs}} - l_{\text{slide}} = \frac{l_{\text{skip}}}{\sqrt{1+2y}}$ for ease of notation. To complete **Equation** S5.33 we need the average number of search rounds (binding-lateral diffusion-unbinding) needed to locate a single target amongst L potential binding/target sites,

$$N_{\text{rnd}} = \frac{L}{l_{1D}\, p_{\text{check}}(x)} \tag{S5.40}$$

In here, we set the typical length of a lateral excursion to span l_{1D} sites, out of which a fraction $p_{\text{check}}(x)$ have been interrogated (slid past) at least once prior to unbinding (derivation shown below) (see **Figure 5.3A**). Further, l_{1D} represents the (rms) distance between binding and unbinding sites

$$
\begin{aligned}
l_{1D} &= \sqrt{N_{\text{slide}} + N_{\text{skip}} l_{\text{skip}}^2} \\
&= \sqrt{N_{\text{skip}} n_{\text{slide}} + N_{\text{skip}} l_{\text{skip}}^2} \\
&= \sqrt{N_{\text{skip}} l_{\text{slide}}^2 + N_{\text{skip}} l_{\text{skip}}^2} \\
&= \sqrt{N_{\text{skip}}}\, l_{\text{sNs}} = \left(\frac{y+1}{y}\right) x \delta l
\end{aligned}
\tag{S5.41}
$$

In the second line of **Equation** S5.41 we have rewritten the total number of sites visited through sliding as the product of the number of skip-n-slide cycles (N_{skip}) and the number of sliding steps between two skips (n_{slide}). The latter is related to the sliding length as we have defined it above ($l_{\text{slide}}^2 = n_{\text{slide}}$, **Equation** S5.5). In the last line, we recognize the rms length covered in a skip-n-slide cycle ($l_{\text{sNs}} = \sqrt{l_{\text{slide}}^2 + l_{\text{skip}}^2}$). We note that l_{1D} is what can be determined experimentally as the span of a lateral excursion, which is not equal to the variable l_{slide} - even when the protein only performs sliding. Namely, as we have defined l_{slide} to be the rms between consecutive skips, this quantity becomes much greater than l_{1D} if on average less than a skip occurs per search round ($n_{\text{slide}} \gg 1$ when $N_{\text{skip}} \ll 1$).

Taken together, the search time can be written as

$$T_{\text{search}} = N_{\text{rnd}} T_{\text{rnd}} = N_{\text{rnd}} \left[T_{\text{slide}} + T_{\text{skip}} + \tau_{3D} \right] = L \times y \frac{(x\delta l)^2 \tau_{\text{slide}} + (x/y)^2 \tau_{\text{skip}} + \tau_{3D}}{(1+y) x\, p_{\text{check}}(x)\delta l(y)} \tag{S5.42}$$

In what follows, we shall first derive p_{check}, and proceed to show T_{search} has minima both for large scanning densities (sliding only) and low scanning densities (skip-n-slide).

PROBABILITY TO INTERROGATE ALL SITES WITHIN A GIVEN SECTION OF SEQUENCE SPACE
As discussed in the derivation leading up to **Equation** S5.32, after sufficient 'skip-and-slide cycles' the protein's motion is approximately described by a simple random walk with basic step length l_{sNs} and a probability ρ_{scan} to check all the bases within each segment per visit. Here, we derive an approximate equation for p_{check} for which we used Monte Carlo simulations to show it has the correct scaling with the model parameters (see main text and **Figure 5.3**) - thereby validating our analysis of the

search time done below.

Let the protein bind to the DNA at segment 1 and leave it at $\hat{l}_{1D} = l_{1D}/l_{sNs}$. Towards calculating the probability to check all sites along its path at least once, we first pick a segment $\hat{l}$ between start- and endpoints and determine the probability to visit/interrogate all sites in this segment at least once prior to making it to segment $\hat{l}_{1D}$ for the first time (**Figure** S5.4A). Assuming the protein does not visit any other segments outside the interval $[1, \hat{l}_{1D}]$, the probability to reach $\hat{l}_{1D}$ after having checked the sites within $\hat{l}$ equals the probability of making it from $\hat{l}$ to $\hat{l}_{1D}$,

$$P(1 \to \hat{l}_{1D}|\text{check } \hat{l}) = P(1 \to \hat{l}) \times P(\hat{l} \to \hat{l}_{1D}|\text{check } \hat{l}) = P(\hat{l} \to \hat{l}_{1D}|\text{check } \hat{l}), \qquad \text{(S5.43)}$$

as the protein will always return from the first segment to the intermediate (with or without checking sites in between) ($P(1 \to \hat{l}) = 1$). The probability of making it from 1 to $\hat{l}_{1d}$ *without* checking the intermediate site equals (**Figure** S5.4A)

$$\begin{aligned}
P(\hat{l} \to \hat{l}_{1D}|\text{no check } \hat{l}) &= \frac{1}{2}(1 - \rho_{\text{scan}})P_{\text{no shuttle}}(\hat{l}_{1D} - \hat{l}) \\
&\quad \times \sum_{m=0}^{\infty} \left(\frac{1}{2}(1 - \rho_{\text{scan}}) \left[P_{\text{no shuttle}}(\hat{l}) + P_{\text{shuttle}}(\hat{l}) + P_{\text{no shuttle}}(\hat{l}_{1D} - \hat{l}) \right] \right)^m \\
&= \frac{1}{1 + \frac{2\rho_{\text{scan}}(\hat{l}_{1D} - \hat{l})}{1 - \rho_{\text{scan}}}},
\end{aligned}$$
$$\text{(S5.44)}$$

with $P_{\text{shuttle}}(d)$ given by **Equation** S5.27. The common term in **Equation** S5.44 represents the path that leads directly from segment $\hat{l}$ to the final one at $\hat{l}_{1D}$ without having checked the intermediate site. The first set of terms within the sum are all paths that attempt to reach segment 1, but do not make it (**Figure ??A**). The middle terms within the sum count all paths that do make it to the first segment and return with unit probability. The final term within the sum represents all paths that attempt to walk to the final segment, but do not make it across. From this we derive

$$P(\hat{l} \to \hat{l}_{1D}|\text{check } \hat{l}) = 1 - P(\hat{l} \to \hat{l}_{1D}|\text{no check } \hat{l}) = \frac{(\hat{l}_{1D} - \hat{l})}{(\hat{l}_{1D} - \hat{l}) + \frac{1 - \rho_{\text{scan}}}{2\rho_{\text{scan}}}} \qquad \text{(S5.45)}$$

As this holds for any segment within $[1, \hat{l}_{1D}]$, we get the probability of interrogating all sites/segments by averaging over all positions of $\hat{l}$,

$$p_{\text{check}}(\rho_{\text{scan}}, l_{1D}) \approx \frac{1}{\hat{l}_{1D}} \int_0^{\hat{l}_{1D}} P(\hat{l} \to \hat{l}_{1D}|\text{check } \hat{l}) d\hat{l} = 1 - \frac{1 - \rho_{\text{scan}}}{2\rho_{\text{scan}} l_{1D}/l_{sNs}} \log \left[1 + \frac{2\rho_{\text{scan}} l_{1D}/l_{sNs}}{1 - \rho_{\text{scan}}} \right],$$
$$\text{(S5.46)}$$

for which we assumed large enough distances $\hat{l}_{1D}$ such that $\frac{1}{\hat{l}_{1D}} \sum_{\hat{l}=1}^{\hat{l}_{1D}} P(\hat{l} \to$

$\hat{l}_{1D}|\text{check }) \approx \frac{1}{\hat{l}_{1D}} \int_1^{\hat{l}_{1D}} P(\hat{l} \to \hat{l}_{1D}|\text{check } \hat{l}) d\hat{l} \approx \frac{1}{\hat{l}_{1D}} \int_0^{\hat{l}_{1D}} P(\hat{l} \to \hat{l}_{1D}|\text{check } \hat{l}) d\hat{l}.$

We can rewrite **Equation** S5.46 using $x = \frac{l_{1D}}{l_{sNs}} \frac{\rho_{scan}}{1 - \rho_{scan}}$ (which is equal to **Equation** S5.36, by virtue of **Equation** S5.41),

$$p_{check}(x) = 1 - \frac{\log(1 + 2x)}{2x} \approx \begin{cases} x - \frac{4}{3}x^2 & x \ll 1 \\ 1 & x \gg 1 \end{cases} \tag{S5.47}$$

CONDITIONS FOR OPTIMAL SEARCH TIME

We now proceed to find the optima of **Equation** S5.42 in terms of x and y. Its derivative with respect to x equals

$$\partial_x \log T_{search} = \frac{2}{x} \frac{\tau_{1D}}{\tau_{1D} + \tau_{3D}} - \frac{1}{x} - \partial_x \log p_{check} \tag{S5.48}$$

Setting it equal to zero results in the following condition

$$\frac{2\tau_{1D}}{\tau_{1D} + \tau_{3D}} = 1 + x \partial_x \log p_{check} \tag{S5.49}$$

Similarly, setting $\partial_y \log T_{search}$ equal to zero results in

$$2 \left[\frac{y}{1 + 2y} \frac{T_{slide}}{T_{rnd}} + \frac{T_{skip}}{T_{rnd}} \right] = 1 + \frac{y}{1 + 2y} - \frac{y}{1 + y} \tag{S5.50}$$

In what follows it is our goal to prove the existence of (at least) two minima - sets of coordinates in $\{N_{slide}, N_{skip}\}$-space, or equivalently $\{x, y\}$-space, that simultaneously satisfy **Equations** S5.49 and S5.50.

HIGH SCANNING DENSITIES: SLIDING-ONLY OPTIMUM

Here, we seek a local minimum of **Equation** S5.42 - - satisfying both **Equations** S5.49 and S5.50 - in the 'densely scanned' regime ($\rho_{scan} \gg 0.5$). For sufficiently large scanning densities, $y \gg 1$, for which **Equations** S5.38 and S5.39 make the second term on the left hand side of **Equation** S5.50 vanish, and we are left with

$$T_{slide} = \frac{1}{2} T_{rnd} \tag{S5.51}$$

If we additionally assume (close to) no skipping takes place ($N_{skip} \to 0$), or $y \gg x$ (**Equation** S5.39), this condition simplifies further to

$$\tau_{1D} = \tau_{3D} \tag{S5.52}$$

We see that at (close to) unit scanning density it is most beneficial to spend half of the time searching laterally along the substrate and the other half using excursions through solutions to reach distant sites. This result was obtained by Slutsky and Mirny [Slutsky and Mirny, Biophysical Journal 2004], whose model does not allow for skips to take place. Hence, our model coincides with theirs when shutting down skipping. Using **Equation** S5.52 in S5.49 yields

$$x \partial_x p_{check} = 0 \tag{S5.53}$$

As this equation is satisfied both for $x \gg 1$ (**Equation** S5.47), and for $x = 0$ (using the $x \ll 1$ case in **Equation** S5.47), we identify the sliding-only case,

$$N_{\text{skip}}^{\text{sliding}} \to 0, \;\; N_{\text{slide}}^{\text{sliding}} = \frac{\tau_{\text{3D}}}{\tau_{\text{slide}}}, \;\; l_{\text{slide}}^{\text{sliding}} \gg l_{\text{skip}}^{\text{sliding}}, \; \rho_{\text{scan}}^{\text{sliding}} \to 1, \tag{S5.54}$$

as a (local) optimal search strategy. Recognizing that $l_{\text{1D}} = \sqrt{N_{\text{slide}}}$ for $N_{\text{skip}} = 0$ (**Equation** S5.41), and using **Equations** S5.51, S5.52 and S5.54 results in a search time (**Equation** S5.42) at the "sliding-only" optimum of

$$T_{\text{search}}^{\text{sliding}} = 2L\sqrt{\tau_{\text{slide}}\tau_{\text{3D}}} \tag{S5.55}$$

Hence, the search time can be minimized by eliminating skips altogether and adopting a scanning density of 1 ($l_{\text{slide}} \gg l_{\text{skip}}$).

LOW SCANNING DENSITIES: SKIPPING-AND-SLIDING OPTIMUM
Next, we seek to find an optimal search strategy that involves (frequent) skips. Returning to the y-derivative shown in **Equation** S5.50, we now explore the opposite limit of low scanning densities ($\rho_{\text{scan}} \ll 0.5$, $l_{\text{slide}} \ll l_{\text{skip}}$), $y \ll 1$, for which

$$T_{\text{skip}} = \frac{1}{2}T_{\text{rnd}} \tag{S5.56}$$

We see that at low scanning densities, it is most beneficial for the protein to spent half of its time interrogating sites following skips. Before proceeding, we introduce

$$\tau_{\text{slow}} = \tau_{\text{slide}}l_{\text{skip}}^2 \tag{S5.57}$$

as the time required to travel a full skipping length purely through sliding. That is, $\tau_{\text{skip}}/\tau_{\text{slow}} < 1$ indicates, after having taken into account the temporal cost of performing the skip, it remains beneficial to skip instead of just using sliding to reach the same region of the DNA/RNA. Having defined this variable, **Equation** S5.56 results in

$$(y^{\text{sNs}})^2 = \underbrace{\frac{\tau_{\text{skip}}}{\tau_{\text{slow}}}}_{y_0^2} \frac{(x^{\text{sNs}})^2}{(x^{\text{sNs}})^2 + \underbrace{\frac{\tau_{\text{3D}}}{\tau_{\text{slow}}}}_{x_0^2}} = y_0^2 \frac{\left(x^{\text{sNs}}/x_0\right)^2}{1 + \left(x^{\text{sNs}}/x_0\right)^2}, \tag{S5.58}$$

where we have introduced x_0 and y_0 for notational convenience. Using this y-coordinate reduces **Equation** S5.49 into a condition for the x-coordinate only

$$x\partial_x \log p_{\text{check}}|_{x=x^{\text{sNs}}} = \frac{\left(x^{\text{sNs}}/x_0\right)^2}{1 + \left(x^{\text{sNs}}/x_0\right)^2} \tag{S5.59}$$

Both sides of **Equation** S5.59 are monotonic functions in x (**Figure** S5.4B). Hence, there is an optimal $T_{\text{search}}^{\text{sNs}}$ at $\{x^{\text{sNs}}, y^{\text{sNs}}\}$ corresponding to small scanning densities ($\rho_{\text{scan}} < 0.5$).

To obtain the corresponding value of the search time ($T_{\text{search}}^{\text{sNs}}$), we proceed to solve **Equation** S5.59. Although we are unable to solve **Equation** S5.59 for general x, we can however obtain an approximate solution by assuming $x \ll 1$, for which (using **Equation** S5.47 to simplify the left hand side of **Equation** S5.59)

$$\frac{(x^{\text{sNs}})^3}{2x_0^2} + x^{\text{sNs}} = \frac{3}{8} \tag{S5.60}$$

If we further assume $2x_0^2 \ll 1$, or equivalently, $\tau_{\text{3D}} \ll \tau_{\text{slow}}$,

$$x^{\text{sNs}} \approx \left(\frac{3}{4}\right)^{1/3} x_0^{2/3} \approx x_0^{2/3} = \left(\frac{\tau_{\text{3D}}}{\tau_{\text{slow}}}\right)^{1/3} \tag{S5.61}$$

To demonstrate the validity of this assumption we compared the numerical solution to **Equation** S5.59 to the above approximation thereof (**Equation** S5.61). **Figure S5.4C** shows these to differ less than a factor 3 over a range in $\tau_{\text{3D}}/\tau_{\text{slow}}$ that spans 20 orders of magnitude. We therefore deem **Equation** S5.61 to be valid also outside the $\tau_{\text{3D}}/\tau_{\text{slow}} \ll 1$ taken to obtain it initially (further allowing us to ignore the factor of $(3/4)^{1/3} \approx 0.91$). Using the x-coordinate, we obtain the following y-coordinate (**Equation** S5.58)

$$y^{\text{sNs}} = y_0 \sqrt{\frac{1}{1 + x_0^{2/3}}} \tag{S5.62}$$

Next, using that $\delta l \approx l_{\text{skip}}$ for $y \ll 1$ (the limit already taken), we find the following number of skipping and sliding steps taken in every search round (**Equations** S5.38 and S5.39)

$$N_{\text{slide}}^{\text{sNs}} = x_0^{4/3} l_{\text{skip}}^2 = \frac{\tau_{\text{3D}}}{\tau_{\text{slide}}} \left(\frac{\tau_{\text{slow}}}{\tau_{\text{3D}}}\right)^{1/3} \tag{S5.63}$$

$$N_{\text{skip}}^{\text{sNs}} = \frac{x_0^{4/3}}{y_0^2} \left(1 + x_0^{2/3}\right) = \frac{\tau_{\text{3D}}}{\tau_{\text{skip}}} \left(1 + \left(\frac{\tau_{\text{slow}}}{\tau_{\text{3D}}}\right)^{1/3}\right) \tag{S5.64}$$

Combining **Equations** S5.38 and S5.39 together with the skip-n-slide optimum set by **Equations** S5.61, S5.62, S5.63 and S5.64 , results in a search time (**Equation** S5.42)

$$T_{\text{search}}^{\text{sNs}} = 2L \frac{\sqrt{\tau_{\text{skip}} \tau_{\text{3D}}}}{l_{\text{skip}}} \left(\frac{\sqrt{1 + \left(\frac{\tau_{\text{slow}}}{\tau_{\text{3D}}}\right)^{1/3}}}{p_{\text{check}}(x^{\text{sNs}})}\right) \tag{S5.65}$$

In conclusion, the search time is minimized both at a maximum scanning density of 1 ($\rho_{\text{scan}}^{\text{sliding}} \approx 1$) - with a search time of $T_{\text{search}}^{\text{sliding}}$ (**Equation** S5.55) - and at a lower scanning density ($\rho_{\text{scan}}^{\text{sNs}} = \frac{1}{l_{\text{skip}}} \sqrt{N_{\text{slide}}^{\text{sNs}}/N_{\text{skip}}^{\text{sNs}}} < 0.5$) - with a search time $T_{\text{search}}^{\text{sNs}}$ (**Equation** S5.65).

GLOBAL OPTIMUM

Having found two local optima, the more favorable search strategy is the one corresponding to the lowest search time. Hence, a combination of skipping and sliding is preferred (over just sliding) when $T_{\text{searh}}^{\text{sNs}} < T_{\text{search}}^{\text{sliding}}$. Using **Equations** S5.55 and S5.65

$$\frac{T_{\text{search}}^{\text{sNs}}}{T_{\text{search}}^{\text{sliding}}} = \sqrt{\frac{\tau_{\text{skip}}}{\tau_{\text{slow}}} \left(\frac{\sqrt{1 + \left(\frac{\tau_{\text{slow}}}{\tau_{\text{3D}}}\right)^{1/3}}}{p_{\text{check}}(x^{\text{sNs}})} \right)} < 1 \tag{S5.66}$$

This can be rewritten as

$$\frac{\tau_{\text{skip}}}{\tau_{\text{slow}}} < \frac{p_{\text{check}}^2(x^{\text{sNs}})}{1 + \left(\frac{\tau_{\text{slow}}}{\tau_{\text{3D}}}\right)^{1/3}} < 1 \tag{S5.67}$$

The second inequality ('less than 1') follows from noticing that $p_{\text{check}}(x) \leq 1$ for any x as it is a probability, and $\left(\frac{\tau_{\text{slow}}}{\tau_{\text{3D}}}\right)^{1/3} > 0$ as all τ's are positive, together making the middle identity always less than 1. As expected, $\frac{\tau_{\text{skip}}}{\tau_{\text{slow}}} < 1$, for skipping to be beneficial. However, **Equation** S5.67 refines this statement and gives the exact boundary shown in the phase diagram of **Figure 5.3D**.
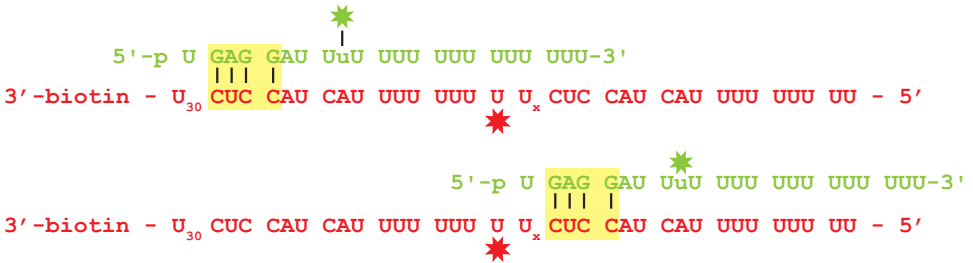

Figure S5.1: related to Figure 5.1. construct design hAgo2. ssRNA constructs (red) are passivated to the microscope slide using a 3'-biotin-streptavadin linkage. The two trapping sequences, 4 nt sequences that are complementary to the corresponding guide nucleotides (green), are highlighted in yellow. Top figure represents the 'high FRET' configuration, while the bottom figure displays Ago bound to the trap resulting in 'low FRET'. The distance between traps is varied by adding Uracil nucleotides (Ux reads: 'x times a U'). To embed the traps within the sequence, as opposed to them being the outermost sites, poly-U sequences flank both traps.
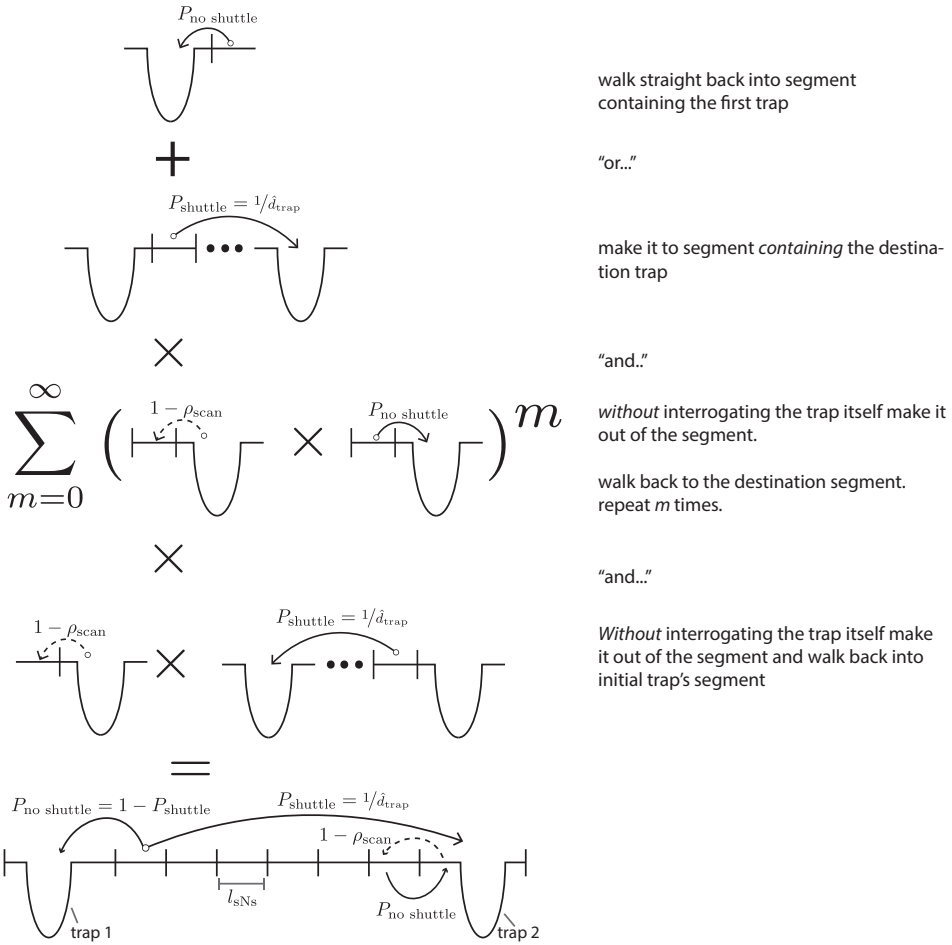
Figure S5.2: related to Figure 5.2. Path counting to derive scaling of shuttling time with distance. A graphical explanation of **Equation** S5.22. Subsequent figures will only show the equivalent of the bottom schematic shown here.
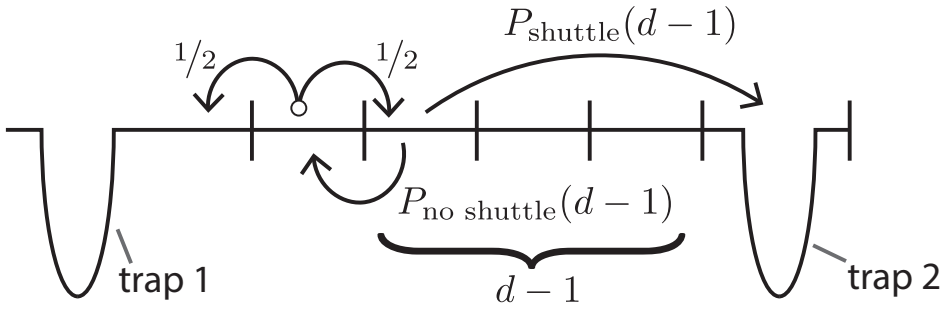
Figure S5.3: related to Figure 5.2. Shuttling time simple diffusion scales linearly with d_{trap}. Illustration of recursion relation dictating probability to shuttle P_{shuttle} (or get recaptured $P_{\text{no shuttle}}$) in terms of number of binding sites separating the two traps. Relates to **Equations** S5.23 and S5.24.

5
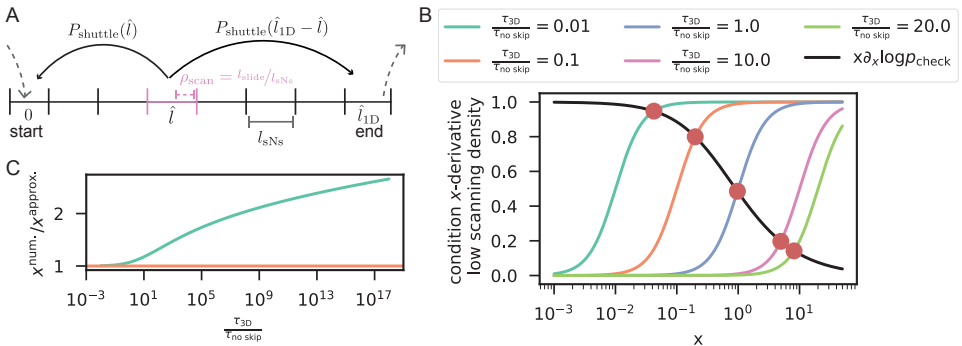
Figure S5.4: related to Figure 5.3.derivation of search time at given scanning density. **(A)** Illustration of paths (and corresponding probabilities) that lead the protein from segment 1 to $\hat{l}_{\text{1D}}$ (size l_{sNs}) without having interrogated all binding sites within segment $\hat{l}$. Relates to **Equation** S5.44. **(B)** At low scanning densities, the search time exhibits a unique minimum. Colored lines show right hand side of **Equation** S5.59 for varying values of $\tau_{\text{3D}}/\tau_{\text{slow}}$ and black line shows the left hand side. Intersections (red dots) our found numerically and –together with **Equation** S5.58 -indicate the location in $\{x, y\}$-space the skip-and-slide optimum can be found at (**Equation** 8). **(C)** Approximate location of skip-and-slide optimum (x-coördinate) from **Equation** S5.61 versus numerical solution to **Equation** S5.59.

REFERENCES

[1] O. G. Berg, R. B. Winter, and P. H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory.* Biochemistry **20**, 6929 (1981).

[2] M. B. Elowitz, M. G. Surette, P.-e. Wolf, J. B. Stock, S. Leibler, M. B. Elowitz, M. G. Surette, P. E. Wolf, J. Stock, and S. Leibler, *Protein Mobility in the Cytoplasm of Escherichia coli,* Journal of Bacteriology **181**, 197 (1999).

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

[4] P. H. Vonhippel and O. G. Berg, *Facilitated Target Location in Biological-Systems,* Journal of Biological Chemistry **264**, 675 (1989).

[5] A. D. Riggs, S. Bourgeois, and M. Cohn, *The lac represser-operator interaction. III. Kinetic studies,* Journal of Molecular Biology **53**, 401 (1970).

[6] P. H. Richter and M. Eigen, *Diffusion controlled reaction rates in spheroidal geometry,* Biophysical Chemistry **2**, 255 (1974).

[7] M. Slutsky and L. A. Mirny, *Kinetics of Protein-DNA Interaction : Facilitated Target Location in Sequence-Dependent Potential,* Biophysical Journal **87**, 4021 (2004).

[8] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie, *A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA,* Proceedings of the National Academy of Sciences **103**, 5752 (2006).

[9] I. Bonnet, A. Biebricher, P. L. Porté, C. Loverdo, O. Bénichou, R. Voituriez, C. Escudé, W. Wende, A. Pingoud, and P. Desbiolles, *Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA,* Nucleic Acids Research **36**, 4118 (2008).

[10] D. M. Gowers, G. G. Wilson, and S. E. Halford, *Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA,* Proceedings of the National Academy of Sciences **102**, 15883 (2005).

[11] A. Graneli, C. C. Yeykal, R. B. Robertson, and E. C. Greene, *Long-distance lateral diffusion of human Rad51 on double-stranded DNA,* Proceedings of the National Academy of Sciences **103**, 1221 (2006).

[12] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, *The lac Repressor Displays Facilitated,* Science **336**, 1595 (2012).

[13] J. H. Kim and R. G. Larson, *Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules,* Nucleic Acids Research **35**, 3848 (2007).

[14] A. B. Kochaniak, S. Habuchi, J. J. Loparo, D. J. Chang, K. A. Cimprich, J. C. Walter, and A. M. van Oijen, *Proliferating cell nuclear antigen uses two distinct modes to move along DNA,* Journal of Biological Chemistry **284**, 17700 (2009).

[15] H. R. Koh, M. A. Kidwell, J. Doudna, and S. Myong, *RNA scanning of a molecular*

machine with a built-in ruler, Journal of the American Chemical Society **139**, 262 (2017).

[16] J. S. Leith, A. Tafvizi, F. Huang, W. E. Uspal, P. S. Doyle, A. R. Fersht, L. A. Mirny, and A. M. van Oijen, *Sequence-dependent sliding kinetics of p53,* Proceedings of the National Academy of Sciences **109**, 16552 (2012).

[17] D. Normanno, L. Boudarène, C. Dugast-Darzacq, J. Chen, C. Richter, F. Proux, O. Bénichou, R. Voituriez, X. Darzacq, and M. Dahan, *Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher,* Nature Communications **6** (2015), 10.1038/ncomms8357.

[18] K. Ragunathan, C. Liu, and T. Ha, *RecA filament sliding on DNA facilitates homology search,* eLife , 1 (2012).

[19] N. P. Stanford, M. D. Szczelkun, J. F. Marko, and S. E. Halford, *One- and three-dimensional pathways for proteins to reach specific DNA sites.* The EMBO journal **19**, 6546 (2000).

[20] A. Tafvizi, F. Huang, A. R. Fersht, L. A. Mirny, and A. M. van Oijen, *A single-molecule characterization of p53 search on DNA,* Proceedings of the National Academy of Sciences **108**, 563 (2011).

[21] Y. M. Wang, R. H. Austin, and E. C. Cox, *Single molecule measurements of repressor protein 1D diffusion on DNA,* Physical Review Letters **97**, 1 (2006).

[22] J. Elf, G.-W. Li, and X. S. Xie, *Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell,* Science **316**, 1191 (2007).

[23] Y. Kao-Huang, A. Revzin, A. P. Butler, P. O'Conner, D. W. Noble, and P. H. Von Hippel, *Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: Measurement of DNA-bound Escherichia coli lac repressor in vivo,* Proceedings of the National Academy of Sciences **74**, 4228 (1977).

[24] T. Hu and B. I. Shklovskii, *How a protein searches for its specific site on DNA: The role of intersegment transfer,* Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **76**, 1 (2007).

[25] G.-w. Li, O. G. Berg, and J. Elf, *Effects of macromolecular crowding and DNA looping on gene regulation kinetics,* Nature Physics **5**, 294 (2009).

[26] M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. L. Wuite, and R. Metzler, *Facilitated diffusion with DNA coiling,* Proceedings of the National Academy of Sciences **106**, 8204 (2009).

[27] M. Sheinman and Y. Kafri, *The effects of intersegmental transfers on target location by proteins,* Physical Biology **016003** (2009), 10.1088/1478-3975/6/1/016003.

[28] M. Klein, S. Chandradoss, M. Depken, and C. Joo, *Why Argonaute is needed to make microRNA target search fast and reliable,* Seminars in Cell and Developmental Biology **65**, 20 (2017).

[29] B. van den Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler, and G. J. L. Wuite, *How DNA coiling enhances target localization by proteins,* Proceedings of the National Academy of Sciences **105**, 15738 (2008).

[30] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, *Ionic strength-dependent persistence lengths of single-stranded RNA and DNA,* Proceedings of the National Academy of Sciences **109**, 799 (2012).

[31] T. A. L. I. Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama, *Growth Phase-Dependent Variation in Protein Composition of the,* Society **181**, 6361 (1999).

[32] K. Brogaard, L. Xi, J. P. Wang, and J. Widom, *A map of nucleosome positions in yeast at base-pair resolution,* Nature **486**, 496 (2012).

[33] V. Globyte, S. H. Kim, and C. Joo, *Single-Molecule View of Small RNA – Guided Target Search and Recognition,* Annual Review of Biophysics , 1 (2018).

[34] H. Wang, M. La Russa, and L. S. Qi, *CRISPR/Cas9 in Genome Editing and Beyond,* Annual Review of Biochemistry **85**, 227 (2016).

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

[36] V. Globyte, S. H. Lee, T. Bae, J. Kim, and C. Joo, *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion,* The EMBO Journal **38**, e99466 (2019).

[37] Y. Jeon, Y. H. Choi, Y. Jang, J. Yu, J. Goo, G. Lee, Y. K. Jeong, S. H. Lee, I. S. Kim, J. S. Kim, C. Jeong, S. Lee, and S. Bae, *Direct observation of DNA target searching and cleavage by CRISPR-Cas12a,* Nature Communications **9** (2018), 10.1038/s41467-018-05245-x, arXiv:15334406 .

[38] A. A. Shvets and A. B. Kolomeisky, *Mechanism of Genome Interrogation: How CRISPR RNA-Guided Cas9 Proteins Locate Specific Targets on DNA,* Biophysical Journal **113**, 1416 (2017).

[39] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, and J. A. Doudna, *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9,* Nature **507**, 62 (2014), arXiv:NIHMS150003 .

[40] J. W. Hegge, D. C. Swarts, S. D. Chandradoss, T. J. Cui, J. Kneppers, M. Jinek, C. Joo, and J. van der Oost, *DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute,* Nucleic Acids Research **47**, 5809 (2019).

[41] T. J. Cui, M. Klein, J. W. Hegge, S. D. Chandradoss, J. van der Oost, M. Depken, and C. Joo, *Argonaute bypasses cellular obstacles without hindrance during target search,* bioRxiv (2019).

[42] M. Bauer and R. Metzler, *Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states,* Biophysical Journal **102**, 2321 (2012).

[43] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, *How a*

protein searches for its site on DNA : the mechanism of facilitated diffusion, Journal of Physics A: Mathematical and Theoretical **42**, 1 (2009).

[44] C. Joo and T. Ha, *Single-molecule FRET with total internal reflection microscopy,* Cold Spring Harbor Protocols **7**, 1223 (2012).

[45] Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, and A. Yildiz, *A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9,* Science Advances **3**, 1 (2017).

[46] S. H. Sternberg, B. Lafrance, M. Kaplan, and J. A. Doudna, *Conformational control of DNA target cleavage by CRISPR-Cas9,* Nature **527**, 110 (2015).

[47] M. Yang, S. Peng, R. Sun, J. Lin, N. Wang, and C. Chen, *The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET,* Cell Reports **22**, 372 (2018).

[48] M. D. Biggin, *Animal Transcription Networks as Highly Connected, Quantitative Continua,* Developmental Cell **21**, 611 (2011).

[49] A. Mahmutovic, O. G. Berg, and J. Elf, *What matters for lac repressor search in vivo - Sliding, hopping, intersegment transfer, crowding on DNA or recognition?* Nucleic Acids Research **43**, 3454 (2015).

[50] S. D. Chandradoss, A. C. Haagsma, Y. K. Lee, J.-H. Hwang, J.-M. Nam, and C. Joo, *Surface Passivation for Single-molecule Protein Studies,* Journal of Visualized Experiments , 1 (2014).

[51] H. Pan, Y. Xia, M. Qin, Y. Cao, and W. Wang, *A simple procedure to improve the surface passivation for single molecule fluorescence studies,* Physical Biology **12** (2015), 10.1088/1478-3975/12/4/045006

6

HIGH-SPEED SUPER-RESOLUTION IMAGING USING PROTEIN-ASSISTED DNA-PAINT

Super-resolution imaging allows for visualization of cellular structures on a nanoscale level. DNA-PAINT (DNA Point Accumulation In Nanoscale Topology) is a super-resolution method that depends on the binding and unbinding of DNA imager strands. The current DNA-PAINT technique suffers from slow acquisition due to the low binding rate of the imager strands. Here we report on a method where imager strands are loaded into a protein, Argonaute (Ago), that allows for faster binding. Ago pre-orders the DNA imager strand into a helical conformation, allowing for 10 times faster target binding. Using a 2D DNA origami structure, we demonstrate that Ago-assisted DNA-PAINT (Ago-PAINT) can speed up the current DNA-PAINT technique by an order of magnitude while maintaining the high spatial resolution. We envision this tool to be useful not only for super-resolution imaging, but also for other techniques that rely on nucleic-acid interactions.

6.1. INTRODUCTION

Single-molecule localization microscopy techniques allow researchers to image cellular structures that are not visible through diffraction-limited microscopy methods. Most single-molecule localization techniques rely on the stochastic blinking of fluorescent signal, by using photoswitchable fluorophores as in photoactivated-localization microscopy (PALM) [1] and (direct) stochastic optical reconstruction microscopy ((d)STORM) [2]. An alternative approach to achieve stochastic blinking is through fluorescent probes that transiently bind their target, as in point accumulation in nanoscale topography (PAINT) [3–5]. In DNA-PAINT, a fluorophore is attached to a short DNA oligonucleotide (or imager strand) that specifically binds to a complementary target DNA sequence (or docking strand) [6]. The stochastic blinking of signals is achieved through binding and unbinding of the incoming imager strands to the docking strands and is imaged using total internal reflection fluorescence (TIRF). By changing the length and sequence of an imager strand, one can tune the on- and off-rates of the imager and adjust the specificity. This allows for high multiplexing capabilities since the number of probes is only limited by the number of orthogonal DNA sequences. Furthermore, compared to conventional super-resolution techniques, DNA-PAINT comes with the advantage that imager strands are continuously replenished from the solution and thus photobleaching is circumvented during the imaging process.

A critical limitation of DNA-PAINT, however, is the low binding rate of DNA, which is typically in the order of 10^6 M^{-1} s^{-1}. Given this binding rate, obtaining images with high spatial resolution (5 nm) usually takes several hours [7–9]. Shorter acquisition times can be achieved by increasing concentration of the imager strand. However, single-molecule binding events become unresolvable from the background of unbound imager strands, even when using TIRF. To reduce this acquisition time, DNA-PAINT has recently been combined with single-molecule Förster Resonance Energy Transfer (smFRET)[10, 11]. This however comes at a cost of reduced spatial resolution due to reduced energy transfer efficiency and limited brightness due to limited choice of dyes [12]. Here we describe an alternative approach, in which protein-assisted delivery of imager strands is demonstrated to speed up the acquisition time 10-fold and only requires a single fluorescent channel.

Argonaute proteins (Agos) are a class of enzymes that utilize a DNA or RNA guide to find a complementary target, either to inactivate or to cleave it. In eukaryotes, an RNA guide directs Ago to complementary RNA targets for post-transcriptional regulation [13]. Ago proteins initially bind their target through base pairing with the seed segment of the guide (nucleotides 2-7 for human Ago)[14–16]. Crystal structures have revealed that Ago pre-orders this seed segment into a helical conformation, allowing for the formation of a double helix between guide and target, and hence effectively pre-paying the entropic cost of target binding [17, 18]. This results in binding rates that are near-diffusion limited ($\sim 10^7$ M^{-1} s^{-1}) [19–22]. In prokaryotes, there is a broad diversity of Agos with respect to the identity of their guide (RNA/DNA) and their target (RNA/DNA)[23, 24]. Some well-characterized prokaryotic Ago nucleases (TtAgo, CbAgo) use DNA guides to target single-stranded (ss)DNA [25, 26].

Here we describe a new DNA-PAINT method based on protein-assisted delivery of DNA imager strands, which allows for faster acquisition of super-resolved

nanostructures. In this Ago-PAINT method, we use a wildtype Ago protein from the bacterium *Clostridium butyricum* (CbAgo) to speed up the kinetic binding of DNA imager strands. CbAgo reshapes the binding landscape of the imager strand, resulting in a 10-fold higher binding rate compared to conventional DNA-PAINT. In addition, we show that one can implement Ago-PAINT with minimal imager strand complexity whilst retaining the programmability and specificity of DNA-PAINT, due to the favourable targeting feature of CbAgo [25, 27]. We determine the spatial resolution of Ago-PAINT through the use of 2D DNA origami structures and show that Ago-PAINT generates super-resolution images of diffraction limited structures at least 10-fold faster than conventional DNA-PAINT.

6.2. RESULTS AND DISCUSSION

For high-quality super-resolution images, a PAINT-based method requires more than five transient binding events per localization spot [9], each with a dwell time of at least several hundreds of milliseconds [7–9]. A typical 8-nt DNA-PAINT imager strand exhibits an on-rate of ~ 10^6 M^{-1} s^{-1} and a dwell time (= 1/off-rate) of ~1 s [8]. DNA-PAINT experiments use an imager strand concentration between 1-10 nM. This range is chosen to be high enough to obtain a sufficient number of binding events during the acquisition time, but not too high to avoid cross-talk localization between structures [9]. We determined the on- and off-rates of Ago-PAINT imager strands and compared these to the on- and off-rate of conventional DNA-PAINT with the same imager strands using a smFRET assay (Figure 6.1). Acceptor (Cy5)-labelled ssDNA targets were immobilized through biotin-streptavidin conjugation on a PEGylated quartz slide. Next, either donor (Cy3)-labelled 8-nt DNA-PAINT imager strands or Ago-PAINT imager strands (CbAgo loaded with a Cy3-labelled guide) were injected, and their interactions with the immobilized target strand were probed using TIRF microscopy (Figure 6.1A). The assay was designed to give a high-FRET signal upon specific binding of either DNA imager strand or Ago-guide complex to the complementary target (Figure 6.1B and C). The time between introduction of the imager strands and the first binding event is the arrival time (which is the inverse of the on-rate, kon). The duration of the FRET binding events is the dwell time (Figure 6.1B). For a comparison between Ago-PAINT and DNA-PAINT, we designed an 8-nt DNA-PAINT imager strand (Figure 6.1C) and found that under our experimental conditions the average dwell time of this imager strand is 1.1 ± 0.2 s (Figure 6.1D). Next, we sought to find an Ago-PAINT guide with a similar dwell time. The first nucleotide of an Ago guide is embedded within the protein structure (Supplementary Figure 1A) [17, 18]. Therefore, we determined the dwell time of Ago-guide complexes with different numbers (N) of base pairing with the target starting from the second nucleotide onwards (Supplementary Figure 1B). A guide with N=5 (nt 2-6) base-pairing to the target exhibited a comparable dwell time of 1.2 ± 0.2 s (Figure 6.1E. We observed that for Ago-PAINT the apparent binding rate is influenced by the number of base pairs that are formed between the guide and its target. For N=5 or larger, the on-rate reaches a saturated value (k_{on} = 0.6-1.0 · 10^8 M^{-1} s^{-1}) (Supplementary Figure 1C). Those values are 10 times higher than the typical on-rates for an 8-nt DNA-PAINT imager strand, 8.7 $\pm$ 0.8 ·10^6 M^{-1} s^{-1} (Supplementary Figure 1F).

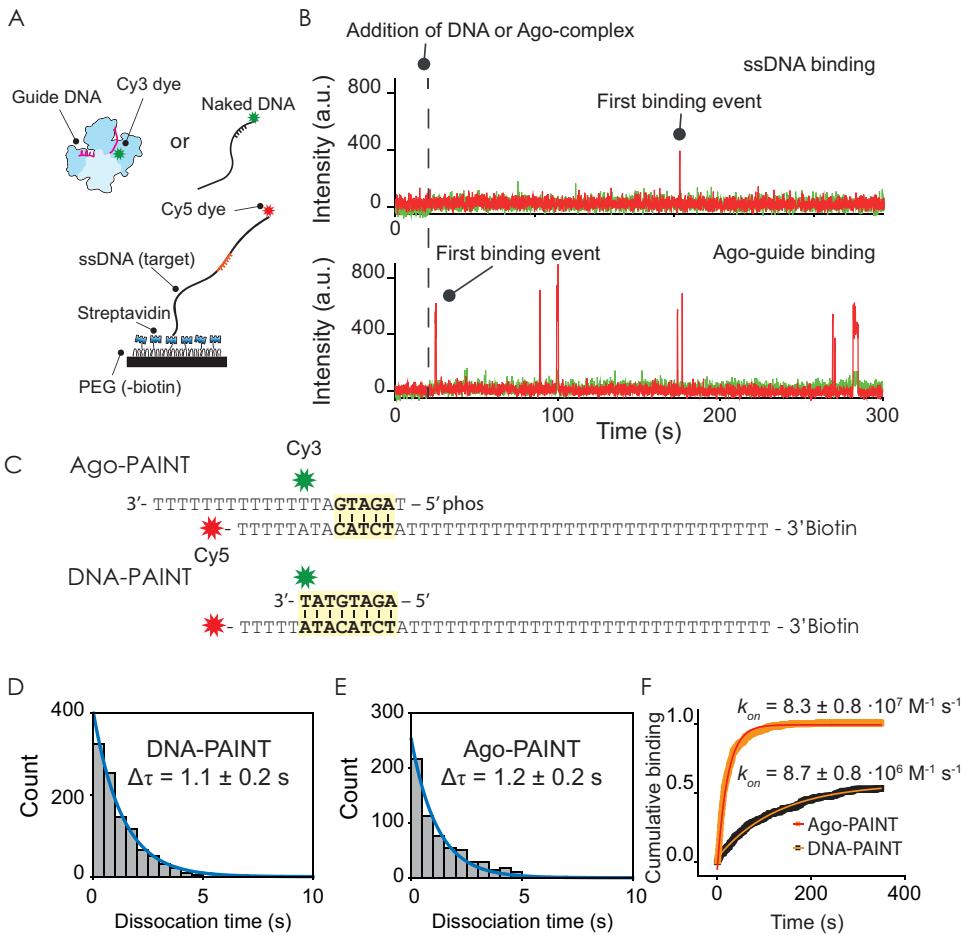To demonstrate the use of Ago-PAINT for super-resolution imaging, we designed a

Figure 6.1: Single-molecule FRET assay to quantify binding kinetics Ago vs DNA-PAINT. (A) A schematic of the single-molecule FRET assay with the target strand immobilized on a PEGylated surface through biotin-streptavidin conjugation. The green and red stars indicate the Cy3 and Cy5 dye respectively. Binding of Ago-guide complex or ssDNA probe to the ssDNA target results in high FRET signal. (B) Representative traces of ssDNA binding (top) and Ago-complex binding (bottom). The dashed line indicates the timepoint at which Ago-guide or DNA is introduced inside the microfluidic chamber. (C) A schematic of the sequences used for Ago-PAINT and DNA-PAINT. Upon binding, both constructs will give rise to a high FRET signal. (D) Dwell-time histogram ($\Delta\tau$) of ssDNA (sequence shown in Figure 6.1C). Maximum likelihood estimation (MLE) gives 1.1 ± 0.2 s as the parameter for a single-exponential distribution (blue line). Number of data-points: 1029. (E) Dwell-time histogram ($\Delta\tau$) of Ago (sequence shown in Figure 6.1A). MLE fitting gives 1.2 ± 0.2 s as the parameter for a single-exponential distribution (blue line). Number of data-points: 696. (F) Cumulative binding event plots of DNA-PAINT (Black) and Ago-PAINT (Orange) vs time. A single-exponential fit is used for DNA-PAINT (red line) and Ago-PAINT (orange line). Errors in (D), (E) and (F) are determined by taking the 95% confidence interval of 10^5 bootstraps.

rectangular 2-dimensional DNA origami structure of 76 nm x 80 nm (Figure 6.2 and Supplementary Figure 2). The DNA origami structure has four docking sites that are
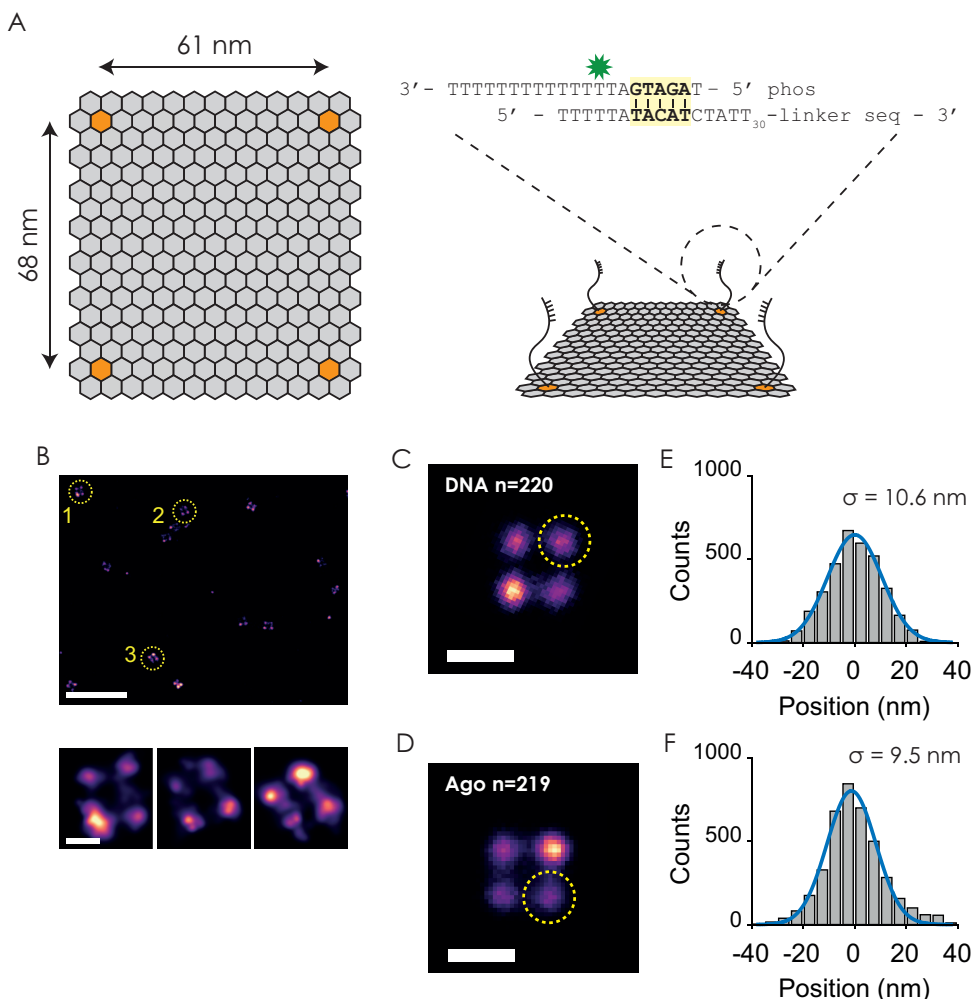
A


```
3'- TTTTTTTTTTTTTTAGTAGAT – 5' phos
        5' - TTTTTATACATCTATT₃₀-linker seq - 3'
```


Figure 6.2: Ago-PAINT enables the same localization precision as conventional DNA-PAINT (A) Left: A schematic design of the 2D-DNA origami structure. The orange honeycombs indicate the approximate locations of binding sites. Right: 3D representation of the imaging scheme with the used docking strand sequence. The green star indicates the position of the Cy3 dye labelled on the backbone of an amino-modified thymine. (B) A representative super-resolution image showcasing binding sites of the 2D-DNA origami structures using Ago-PAINT. Bottom: Super-resolution reconstruction of the four-corner origami structures of the top panel. (C) A summed image of 220 origami structures visualized through the use of DNA-PAINT. (D) A summed image of 219 origami structures made through the use of Ago-PAINT (E) Fitting of a cross-sectional intensity histogram from the yellow encircled area in Figure 6.2(C) to a Gaussian (blue line) shows that a localisation precision of 10.6 nm can be achieved, similar to Ago-PAINT. (F) Fitting of a cross-sectional intensity histogram from the yellow encircled area in Figure 6.2(D) to a Gaussian (blue line) shows that a localisation precision of 9.5 nm is possible under these conditions. Scale bars in (B) indicate 500 nm (top) and 50 nm (bottom three). Scale bars in (C) and (E) indicate 100 nm.

spaced 61 nm x 68 nm apart (Figure 6.2A). To achieve optimal Ago binding to the DNA origami docking strands, we introduced a polyT linker (T_{30}) between the target sequence of Ago and the DNA origami structure (Figure 6.2A, right panel). Next, we sought to compare the localization precision of Ago-PAINT and DNA-PAINT. We tested our Ago-PAINT approach by injecting guide-loaded Ago into our flow cell in which DNA origami structure were immobilized. A super-resolution image could be reconstructed from the Ago-PAINT data which revealed four detectable spots on the origami structures as expected from our assay design (Figure 6.2B). We determined the localization precision by selecting 220 origami structures for DNA-PAINT and 219 structures for Ago-PAINT and created a sum image using the Picasso analysis software [9] (Figure 6.2C and 6.2D). The localization precision was determined by plotting the cross-sectional histogram of one of the four binding sites of the summed DNA origami structure. For DNA-PAINT this resulted in a localization precision of 10.6 nm (Figure 6.2E) and for Ago-PAINT we found a localization precision of 9.5 nm (Figure 6.2F). The histogram demonstrates that Ago-PAINT delivers the same quality of localization precision when compared to the DNA-PAINT approach. Analysis of the data based on nearest neighbour analysis [28] reconfirms this finding since a localization precision of 9.7 nm was found for both Ago-PAINT and DNA-PAINT (Supplementary Figure 3). Furthermore, we determined the possibility to use different linker lengths for Ago-PAINT imaging. With this in mind, we designed DNA origami structures with longer linkers (50 thymines or 100 thymine nucleotides) and found that this did not affect the localization precision of Ago-PAINT (Supplementary Figure 4 and 5), showing that Ago-PAINT is compatible with various linker lengths ($\geq T_{30}$).

Finally, we compared the speed of super-resolution imaging through Ago-PAINT with the conventional DNA-PAINT approach using the 2D DNA origami structures as a testing platform. We evaluated the quality of a super-resolution image after each time-point for both Ago-PAINT and DNA-PAINT (Figure 6.3A). The overall resolution of a single-molecule localization microscopy (SMLM) image is dependent on the number of localizations (binding events) per docking strand ($\sigma_{SMLM} \propto \frac{\sigma_{PSF}}{\sqrt{N}}$) . Therefore, to quantify the speed of imaging we plot the standard error of the localization precision as a function of frame number (Figure 6.3B) where we took the sigma values from Figure 6.2E and 6.2F as the localization precision. We observed that the standard error of the localization precision for Ago-PAINT is smaller than that of DNA-PAINT at each time point, indicating that super-resolved images of identical resolution will be obtained 10x faster through Ago-PAINT compared to DNA-PAINT. This result is further supported by the intensity vs time traces, which shows that our Ago-PAINT method results in more binding events compared to DNA-PAINT approach, under similar conditions with DNA concentrations of 1 nM (Figure 6.3C-E). The on-rates for both Ago-PAINT (k_{on} = 4.4 ± 0.1 $\cdot 10^7$ M^{-1} s^{-1}) and DNA-PAINT (k_{on} = 6.6 ± 0.1 $\cdot 10^6$ M^{-1} s^{-1}) on our DNA-origami structure (Figure 6.3E) are similar to the on-rates that we found in our single-molecule experiments (Figure 6.1F).
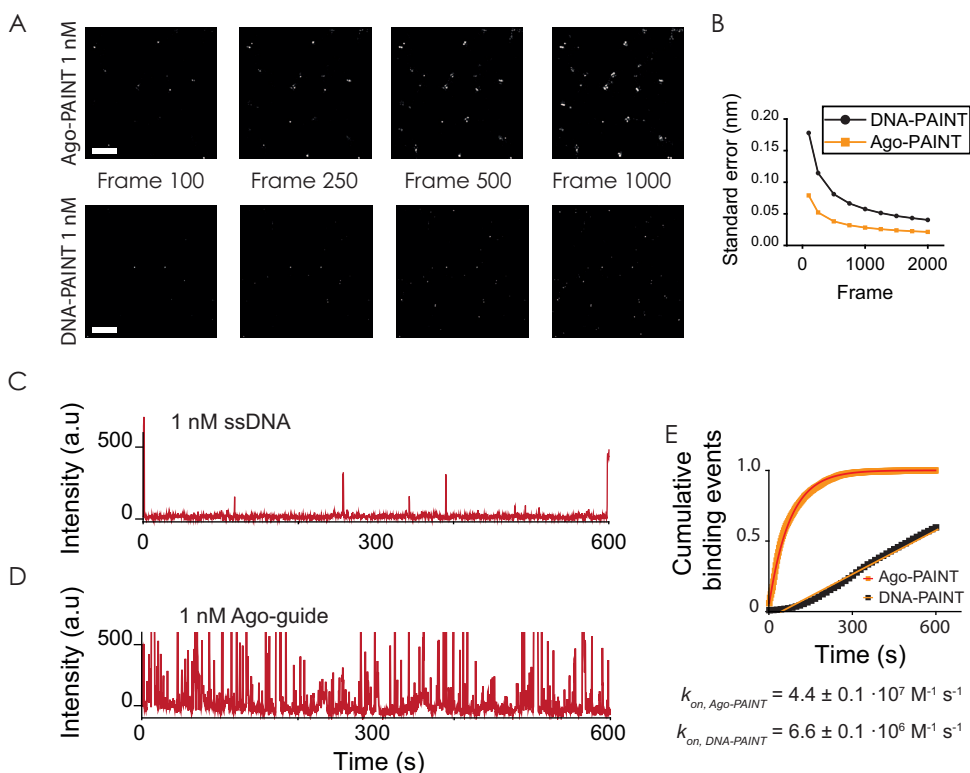
Figure 6.3: Ago-PAINT enables fast imaging of super-resolved structures. (A) Snapshots in time for Ago-PAINT (top) and DNA-PAINT (bottom) showing super-resolution images being formed over time. Exposure time: 0.3 s. The same color scale is used for the intensity in all images. (B) Standard error of Ago-PAINT vs DNA-PAINT plotted versus frame number. (C) Representative intensity vs time data trace of DNA-PAINT at 1 nM DNA concentration shows few binding events occurring within 600 s. The raw data trace is taken from a single origami plate. (D) Representative intensity vs time data trace of 1 nM Ago-guide complex shows binding events occurring frequently within 600 s. The raw data trace is taken from a single origami plate. (E) Normalized cumulative distribution of dark times (the time between binding events) for DNA-PAINT (black, n = 4870) and Ago-PAINT (orange, n = 5793). A single-exponential growth curve (red for DNA-PAINT, orange for Ago-PAINT) is used to estimate the binding rate. Scale bars in (A) indicate 500 nm.

6.3. CONCLUSION

Here we presented a proof-of-concept of Ago-PAINT that allows for rapid super-resolution imaging. We demonstrated that fast Ago-PAINT recording can be used to acquire super-resolution images of nanostructures while retaining the programmability and predictability of DNA-PAINT. In order to fully visualize real-time interactions between multibody cellular components, one would ideally want to look at multiple components at the same time. Effort has been put into temporal [29] or spectral [30] multiplexing of DNA-PAINT technology. In our previous work, we showed that different guide sequences resulted in distinctly different binding kinetics [25]. This kinetic fingerprints will allow for additional freedom when designing Ago-PAINT

[29, 31].

In this study, Ago-PAINT experiments are performed with the wild-type CbAgo protein which substantially increases the probe size compared to conventional DNA-PAINT. However, successful applications of Argonaute proteins for in vivo gene silencing [32, 33] hint that our Ago-PAINT approach could be used in cellular super-resolution imaging. While targeting complex cellular structures in cells could be an issue with a full size CbAgo, it is possible to use truncated versions of Ago. Some truncated versions of approximately half the size (short Agos) exist in nature [24]. We speculate that it will be possible to truncate them further as Ago-PAINT only relies on the property of pre-forming the helix structure of the imager strand and a variant from Kluyveromyces polysporus that contains only the C-lobe was reported to retain almost all the binding properties of the untruncated version [34] . Furthermore, as the imager strand is loaded and protected inside the protein, degradation of the imager strand is less likely to occur over time, unlike oligos that are rapidly digested [35].

In this paper we demonstrated the use of CbAgo for super-resolution microscopy. While this CbAgo targets ssDNA, Agos from other species can target RNA [24]. For example, the Ago from Marinitoga piezophila (MpAgo) [36, 37] targets RNA and one could harness the property of a high association rate for other single molecule imaging applications such as RNA sensing. Recently, dTtAgo has been combined with FISH [38] to allow for labelling of genomic loci in fixed cells. We anticipate the use of RNA guided Agos for a significant speed-up in similar applications for RNA FISH. Lastly, complementary approaches such as qPAINT [39] or crosslinking on single-molecule target using Action-PAINT [40] could be combined with our Ago-PAINT approach. We envision the use of Ago-PAINT as a general toolkit to speed up many current existing applications that rely on base-pairing interactions.

6.4. MATERIALS & METHODS

6.4.1. SINGLE-MOLECULE SETUP

All experiments were performed on a custom-built microscope setup. An inverted microscope (IX73, Olympus) with prism-based total internal reflection is used. In combination with a 532 nm diode laser (Compass 215M/50mW, Coherent). A 60x water immersion objective (UPLSAPO60XW, Olympus) was used for the collection of photons from the Cy3 and Cy5 dyes on the surface, after which a 532 nm long pass filter (LDP01-532RU-25, Semrock) blocks the excitation light. A dichroic mirror (635 dcxr, Chroma) separates the fluorescence signal which is then projected onto an EM-CCD camera (iXon Ultra, DU-897U-CS0-# BV, Andor Technology). A series of EM-CDD images was recorded using custom-made program in Visual C++ (Microsoft). Time traces were extracted from the EM-CDD images using IDL (ITT Visual Information Solution) and further analyzed with Matlab (Mathworks) and Origin (Origin Lab).

6.4.2. SINGLE-MOLECULE DATA ACQUISITION

To avoid non-specific binding of Ago protein to the surface, quartz slides were PEGylated as previously described [41]. Briefly, acidic piranha etched quartz slides were passivated twice with polyethylene glycol (PEG). The first round PEGylation was performed with

mPEG-SVA (Laysan) and PEG-biotin (Laysan), followed by a second round of PEGylation using MS(PEG)4 (ThermoFisher). After assembly of a microfluidic chamber, the slides were incubated with 1 % Tween-20 for 15 minutes. Excess Tween-20 was washed with 100 μL T50 (50 mM Tris-HCl, pH 8.0, 50 mM NaCl) followed by a 2 min incubation of 20 μL streptavidin (0.1 mg/mL, ThermoFisher). Excess streptavidin was removed with 100 μL T50. Next, for single-molecule experiments we immobilized 50 μL of 100 pM Cy5 labelled target DNA for 2 minutes, unbound DNA was washed with 100 μL T50, followed by 100 μL of origami-buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM MnCl$_2$, 5 mM MgCl2). The Ago-guide complex was formed by incubating 10 nM CbAgo with 1 nM of Cy3 labelled DNA guide for 20 minutes at 37 °C in the Ago-buffer. For single-molecule experiments, we injected 50 uL of 1 nM Ago-guide complex or 50 μL of 1 nM DNA-PAINT imager strand in imaging buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM MnCl$_2$, 5 mM MgCl$_2$, 0.8 % glucose, 0.5 mg/mL glucose oxidase (Sigma), 85 μg/mL catalase (Merck) and 1 mM Trolox (Sigma)). The single-molecule FRET experiments for Figure 6.1 were performed at room temperature (23 ± 2 °C). For super-resolution DNA origami experiments, we flushed 50 μL of 200 pM DNA origami structures in a streptavidin coated channel and incubated for 3 minutes to allow for specific immobilization. Unbound DNA-origami was washed with Ago-buffer. Next, 50 μL of 100 pM of Ago-guide complex or 1 nM DNA-PAINT imager strand was injected with imaging buffer.

6.4.3. ASSEMBLY OF DNA ORIGAMI PLATE

The 2D rectangular DNA origami structure was designed by using CaDNAno software based on a square lattice. (SM Douglas, AH Marblestone, S Teerapittayanon, A Vazquez, GM Church, WM Shih. Rapid prototyping of 3D DNA origami shapes with caDNAno. Nucleic Acids Research, 37: 5001-5006 (2009). The 2D rectangular DNA origami was twist corrected and structural behaviour of the origami plate was checked by coarse-grained simulations in CanDo (CE Castro, F Kilchherr, DN Kim, EL Shiao, T Wauer, P Wortmann, M Bathe, H Dietz. A primer to scaffolded DNA origami. Nature Methods, 8: 221-229 (2011) DN Kim, F Kilchherr, H Dietz, M Bathe. Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures. Nucleic Acids Research, 40(7):2862-2868 (2012). The parameters used for simulations are axial rise per base-pair = 0.34 nm, helix diameter =2.25 nm, crossover spacing = 10.5 bp, axial stiffness = 1100 pN, bending stiffness = 230 pN nm^2, torsional stiffness = 460 pN nm^2, nick stiffness factor = 0.01. The 2D rectangular origami structure self-assembled in a total reaction volume of 100 μL containing 10 nM of p8064 scaffold strand (Tilibit), 100 nM core staples (Tilibit), 100 nM Ago-PAINT handles and 100 nM biotin handles in 1x TE folding buffer supplemented with 11 mM MgCl2. The origami structures were annealed using a PCR machine (Bio-rad). First the reaction mixture was heated for 10 minutes at 65°C, then a temperature gradient was applied from 60°C to 40°C with a rate of 1°C/hour. After self-assembly, the origami structures were purified using Amicon spin filter (100K MWCO) and stored in T50 buffer containing 11 mM MgCl$_2$. The purified origami structures were analysed on a 2% agarose gel (Tris-borate-EDTA, 11 mM MgCl$_2$). The gel was run at 90V for 2 hours in ice. After staining the gel with ethidium bromide, the samples were imaged to verify the quality of the folding procedure. Next, the purified

origami sample were checked for rectangular structure by atomic force microscopy (AFM) on mica surface according to AFM imaging procedures. Briefly, 0.01% (w/v) polylysine was incubated 1 min on a freshly cleaved 3 mm (1/8 inch) diameter mica disk. The mica surface was gently washed with mQ water and blow dried with N2. 5 μl of 500 pM DNA origami samples (500 pM) was incubated onto disk for 5 minutes. Washed gently with 1 ml (3x) of folding buffer with 11 $MgCl_2$ to remove any unbound DNA origami structures. Quickly rinsed with mQ water and blow dried with N_2. Dry AFM images were acquired in Bruker Multimode 8 AFM. Sharp AFM tips were used for these measurements (Bruker PeakForce HIRS-F-B) with 0.12 N/m nominal spring constant. AFM images were acquired in tapping mode. Example images of AFM images can be found in Supplementary Fig. 2.

6

6.5. SUPPLEMENTARY INFORMATION
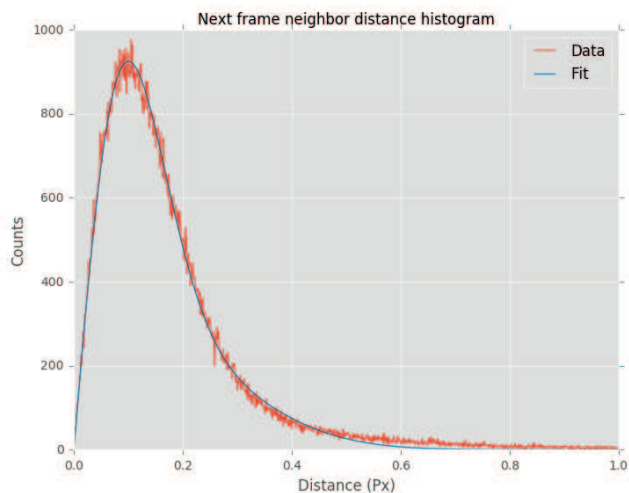### 6.5.1. SUPPLEMENTARY FIGURES
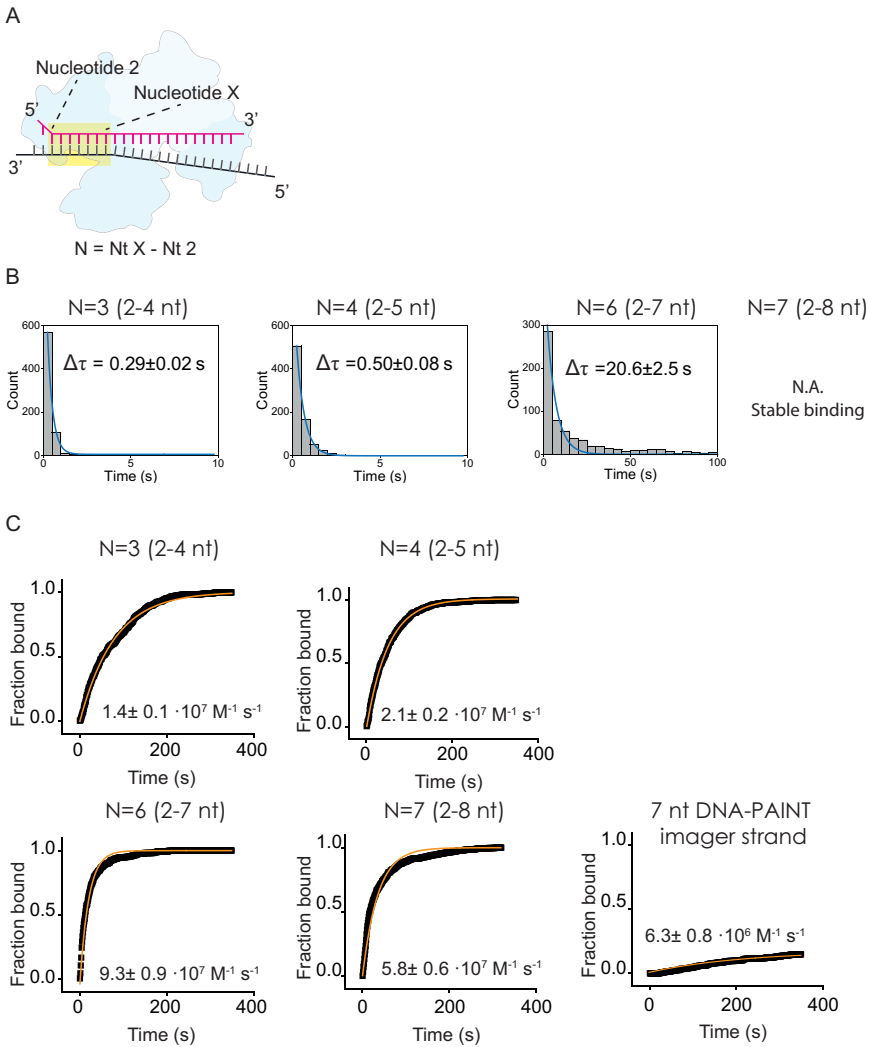
Supplementary Figure 1: Single-molecule binding and unbinding kinetics. (A) Top left: A cartoon figure indicating the base indexes and the definition of N (the number of base pairs). Since the first nucleotide of the imager strand is embedded in Ago, base pairing starts from the second nucleotide. (B) Dwell time histograms for DNA-PAINT for 2-4, 2-5, 2-7 and 2-8 nt base pairing. For 2-8 nt base-pairing, accurate measurements of dwell times were limited by photobleaching.(C) Fractional binding curve for CbAgo-siDNA for 2-4, 2-5, 2-7 and 2-8 nucleotide base pairing with the target sequence. Additionally, a fractional binding curve is shown for 7 nt base pairing with DNA-PAINT. Data was taken on two different days. A single-exponential fit was performed on the data (orange line). Error bars are given by the 95% confidence interval acquired from 10^5 bootstraps.
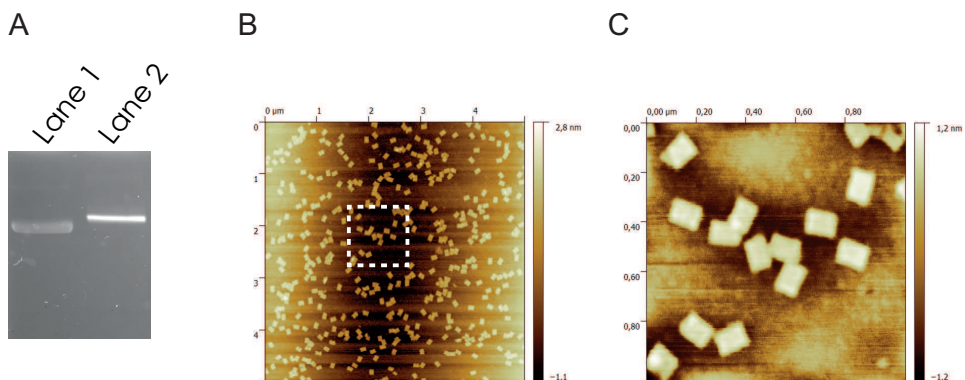
Supplementary Figure 2: Quality control of origami plate assembly. (A) Agarose gel image showing formation of the DNA origami structure. Lane 1: single-stranded M13mp18 p8064 scaffold. Lane 2: annealed origami mixture in 1x TE folding buffer. DNA origami and DNA scaffold were run in a 0.5x TBE + 11 mM MgCl2 buffered 2% agarose gel. (B) An AFM image of 500 pM DNA origami plates deposited on a polylysine treated mica disc. (C) A zoom-in from the white striped square region from (A) shows the rectangular form of the respective plates.
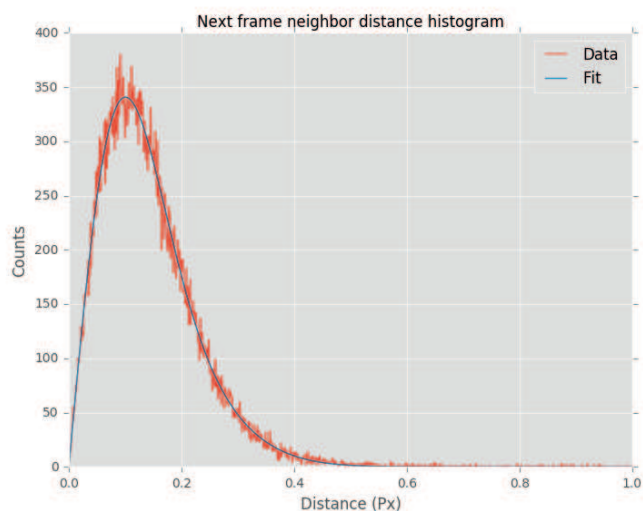
6

A

Ago-PAINT

B

DNA-PAINT

Supplementary Figure 3: Comparison of localization precision between Ago-PAINT and DNA-PAINT through nearest neighbour analysis for a 30x thymine linker docking strand (A) A nearest neighbour in adjacent frame histogram of super-resolution binding sites made through Ago-PAINT. The pairwise displacement fit is given by the blue curve with a NeNA precision 0.09 pixel = 9.7 nm. (B) A nearest neighbour in adjacent frame histogram of super-resolution binding sites made through DNA-PAINT. The pairwise displacement fit is given by the blue curve with a NeNA precision 0.09 pixel = 9.7 nm.
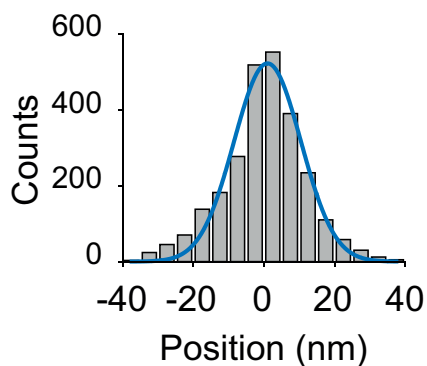
A

B

C
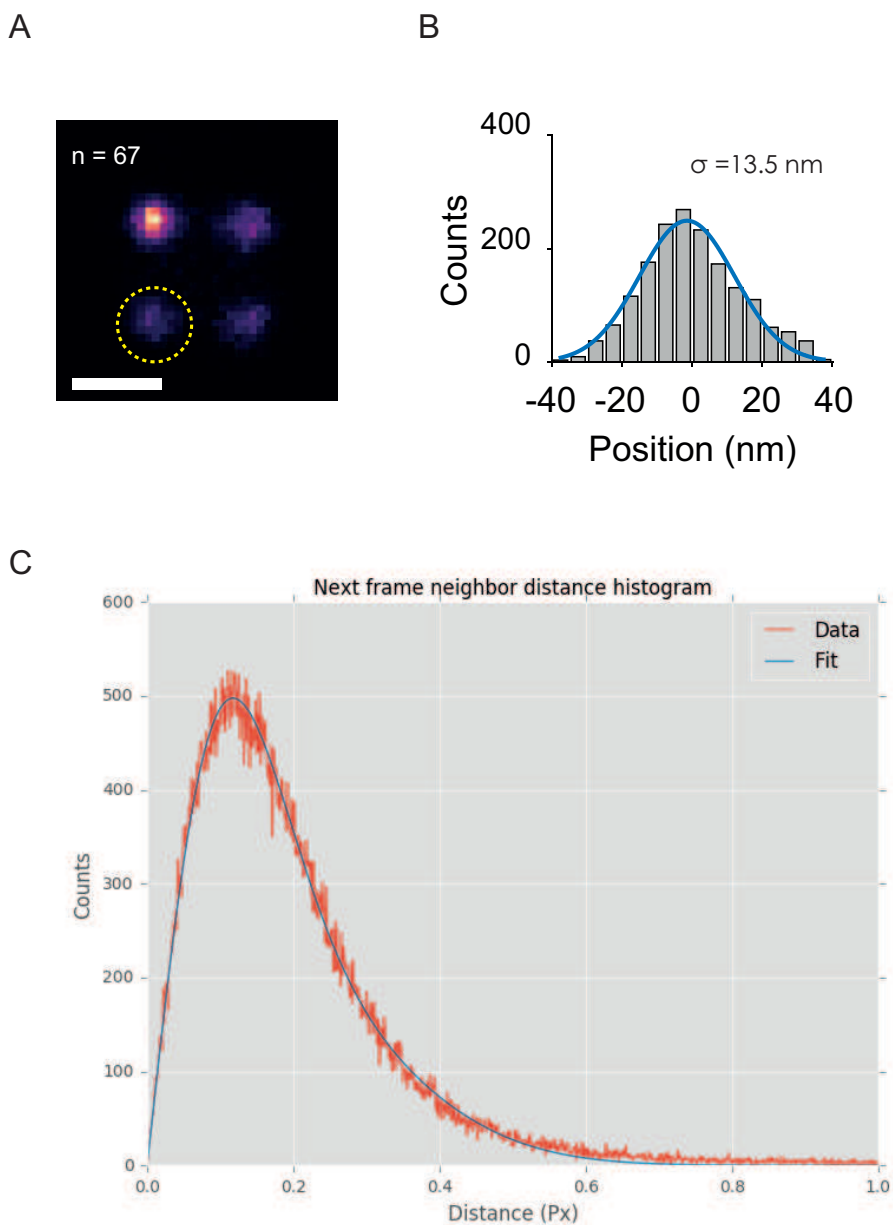
Supplementary Figure 4: Localization precision of 50x thymine linker docking strand (A) A summed image of 67 origami structures made through the use of AGO-PAINT. Scale bar indicates 50 nm. (B) A cross-sectional histogram taken from the yellow encircled area in panel (A). The standard deviation or localization uncertainty is given by σ = 13.5 nm. (C) A nearest neighbour in adjacent frame histogram. The pairwise displacement fit is given by the blue curve with a NeNA precision of 0.1 pixel = 11.3 nm.

A

B

C
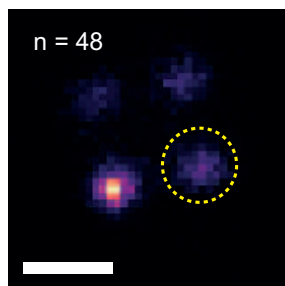
Supplementary Figure 5: Localization precision of 100x thymine linker docking strand (A) A summed image of 48 origami structures made through the use of AGO-PAINT. Scale bar indicates 50 nm. (B) A cross-sectional histogram taken from the yellow encircled area in panel (A). The standard deviation or localization uncertainty is given by σ = 9.8 nm. (C) A nearest neighbour in adjacent frame histogram. The pairwise displacement fit is given by the blue curve with a NeNA precision of 0.1 pixel = 10.8 nm.
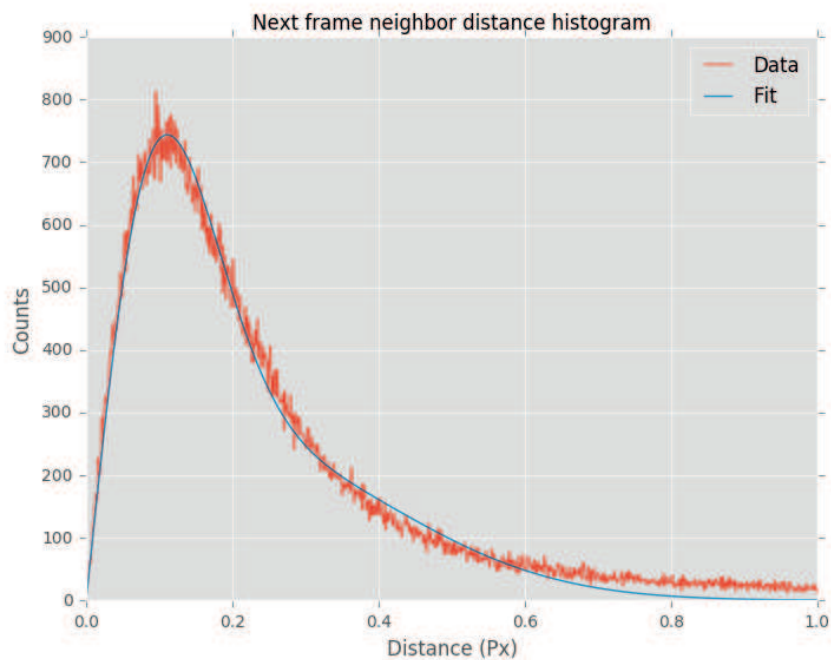
REFERENCES

[1] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, *Imaging Intracellular Fluorescent Proteins at Nanometer Resolution,* Science **313**, 1642 LP (2006).

[2] M. J. Rust, M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),* Nature Methods **3**, 793 (2006).

[3] G. Giannone, E. Hosy, F. Levet, A. Constals, K. Schulze, A. I. Sobolevsky, M. P. Rosconi, E. Gouaux, R. Tampé, D. Choquet, and L. Cognet, *Dynamic Superresolution Imaging of Endogenous Proteins on Living Cells at Ultra-High Density,* Biophysical Journal **99**, 1303 (2010).

[4] I. Schoen, J. Ries, E. Klotzsch, H. Ewers, and V. Vogel, *Binding-activated localization microscopy of DNA l,* Nano Letters **11**, 4008 (2011).

[5] A. Sharonov and R. M. Hochstrasser, *Wide-field subdiffraction imaging by accumulated binding of diffusing probes,* Proceedings of the National Academy of Sciences **103**, 18911 (2006).

[6] R. Jungmann, M. S. Avendaño, J. B. Woehrstein, M. Dai, W. M. Shih, and P. Yin, *Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT,* Nature Methods **11**, 313 (2014).

[7] M. Dai, R. Jungmann, and P. Yin, *Optical imaging of individual biomolecules in densely packed clusters,* Nature Nanotechnology **11**, 798 (2016).

[8] R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, and F. C. Simmel, *Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami,* Nano Letters **10**, 4756 (2010).

[9] J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, and R. Jungmann, *Super-resolution microscopy with DNA-PAINT,* Nature Protocols **12**, 1198 (2017).

[10] A. Auer, M. T. Strauss, T. Schlichthaerle, and R. Jungmann, *Fast, Background-Free DNA-PAINT Imaging Using FRET-Based Probes,* Nano Letters **17**, 6428 (2017).

[11] J. Lee, S. Park, W. Kang, and S. Hohng, *Accelerated super-resolution imaging with FRET-PAINT,* Molecular Brain **10**, 63 (2017).

[12] F. Schueder, J. Stein, F. Stehr, A. Auer, B. Sperl, T. Maximilian, P. Schwille, and R. Jungmann, *An order of magnitude faster DNA-PAINT imaging by improved sequence design and optimized buffer conditions,* Nature Methods , 1.

[13] D. P. Bartel, *MicroRNAs: Target Recognition and Regulatory Functions,* Cell **136**, 215 (2009), arXiv:0208024 [gr-qc] .

[14] D. Baek, J. Villén, C. Shin, F. D. Camargo, S. P. Gygi, and D. P. Bartel, *The impact of microRNAs on protein output,* Nature **455**, 64 (2008).

[15] H. Guo, N. T. Ingolia, J. S. Weissman, and D. P. Bartel, *Mammalian microRNAs predominantly act to decrease target mRNA levels,* Nature **466**, 835 (2010).

[16] M. Selbach, B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, *Widespread changes in protein synthesis induced by microRNAs,* Nature **455**, 58 (2008).

[17] N. T. Schirle and I. J. MacRae, *The Crystal Structure of Human Argonaute2,* Science **336**, 1037 (2012).

[18] Y. Wang, S. Juranek, H. Li, G. Sheng, T. Tuschl, and D. J. Patel, *Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex.* Nature **456**, 921 (2008), arXiv:NIHMS150003 .

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

[20] M. H. Jo, S. Shin, S. R. Jung, E. Kim, J. J. Song, and S. Hohng, *Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs,* Molecular Cell **59**, 117 (2015).

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

[22] C. Yao, H. M. Sasaki, T. Ueda, Y. Tomari, and H. Tadakuma, *Single-Molecule Analysis of the Target Cleavage Reaction by the Drosophila RNAi Enzyme Complex,* Molecular Cell **59**, 125 (2015).

[23] J. W. Hegge, D. C. Swarts, and J. Van Der Oost, *Prokaryotic argonaute proteins: Novel genome-editing tools?* Nature Reviews Microbiology **16**, 5 (2018).

[24] D. C. Swarts, K. Makarova, Y. Wang, K. Nakanishi, R. F. Ketting, E. V. Koonin, D. J. Patel, and J. van der Oost, *The evolutionary journey of Argonaute proteins.* Nature structural & molecular biology **21**, 743 (2014).

[25] J. W. Hegge, D. C. Swarts, S. D. Chandradoss, T. J. Cui, J. Kneppers, M. Jinek, C. Joo, and J. van der Oost, *DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute,* Nucleic Acids Research **47**, 5809 (2019), arXiv:/doi.org/10.1101/534206 [https:] .

[26] D. C. Swarts, M. M. Jore, E. R. Westra, Y. Zhu, J. H. Janssen, A. P. Snijders, Y. Wang, D. J. Patel, J. Berenguer, S. J. J. Brouns, and J. van der Oost, *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature **507**, 258 (2014), arXiv:15334406 .

[27] T. J. Cui, M. Klein, J. W. Hegge, S. D. Chandradoss, J. van der Oost, M. Depken, and C. Joo, *Argonaute bypasses cellular obstacles without hindrance during target search,* Nature Communications **10**, 4390 (2019).

[28] U. Endesfelder, S. Malkusch, F. Fricke, and M. Heilemann, *A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment,* Histochemistry and Cell Biology **141**, 629 (2014).

[29] O. K. Wade, J. B. Woehrstein, P. C. Nickels, S. Strauss, F. Stehr, J. Stein, F. Schueder, M. T. Strauss, M. Ganji, J. Schnitzbauer, H. Grabmayr, P. Yin, P. Schwille, and R. Jungmann, *124-Color Super-resolution Imaging by Engineering DNA-PAINT*

6

Blinking Kinetics, Nano Letters **19**, 2641 (2019).

[30] N. S. Deußner-Helfmann, A. Auer, M. T. Strauss, S. Malkusch, M. S. Dietz, H.-d. D. Barth, R. Jungmann, M. Heilemann, S. Dietz, H.-d. D. Barth, R. Jungmann, and M. Heilemann, *Correlative Single-Molecule FRET and DNA-PAINT Imaging,* Nano Letters **18**, 4626 (2018).

[31] S. Shah, A. K. Dubey, and J. Reif, *Programming Temporal DNA Barcodes for Single-Molecule Fingerprinting,* Nano Letters **19**, 2668 (2019).

[32] R. L. Setten, J. J. Rossi, and S.-p. Han, *The current state and future directions of RNAi-based therapeutics,* Nature Reviews Drug Discovery **18**, 421 (2019).

[33] A. Wittrup and J. Lieberman, *Knocking down disease: a progress report on siRNA therapeutics,* Nature Reviews Genetics **16**, 543 (2015).

[34] D. M. Dayeh, W. A. Cantara, J. P. Kitzrow, K. Musier-Forsyth, and K. Nakanishi, *Argonaute-based programmable RNase as a tool for cleavage of highly-structured RNA,* Nucleic Acids Research **46**, 1 (2018).

[35] S. Tyagi, *Imaging intracellular RNA distribution and dynamics in living cells,* Nature Methods **6**, 331 (2009).

[36] E. Kaya, K. W. Doxzen, K. R. Knoll, R. C. Wilson, S. C. Strutt, P. J. Kranzusch, and J. A. Doudna, *A bacterial Argonaute with noncanonical guide RNA specificity,* Proceedings of the National Academy of Sciences **113**, 4057 (2016), arXiv:arXiv:1408.1149 .

[37] A. Lapinaite, J. A. Doudna, and J. H. D. Cate, *Programmable RNA recognition using a CRISPR-associated Argonaute,* Proceedings of the National Academy of Sciences **115**, 3368 (2018).

[38] L. Chang, G. Sheng, Y. Zhang, S. Shao, Y. Wang, and Y. Sun, *AgoFISH: cost-effective in situ labelling of genomic loci based on DNA-guided dTtAgo protein,* Nanoscale Horizons **4**, 918 (2019).

[39] R. Jungmann, M. S. Avendaño, M. Dai, J. B. Woehrstein, S. S. Agasti, Z. Feiger, A. Rodal, and P. Yin, *Quantitative super-resolution imaging with qPAINT,* Nature Methods **13**, 439 (2016).

[40] N. Liu, M. Dai, S. K. Saka, and P. Yin, *Super-resolution labelling with Action-PAINT,* Nature Chemistry 10.1038/s41557-019-0325-7.

[41] S. D. Chandradoss, A. C. Haagsma, Y. K. Lee, J.-H. Hwang, J.-M. Nam, and C. Joo, *Surface Passivation for Single-molecule Protein Studies,* Journal of Visualized Experiments , 1 (2014)

SUMMARY

In this thesis we used single-molecule FRET to investigate the kinetic properties of a protein called Argonaute. While traditionally one uses bulk methods to investigate the molecular properties of proteins, bulk methods do not confer information that is transient, since that is inherently averaged out. Single-molecule methods, as their name imply, allow one to observe interactions between individual molecules and their substrates. This allows one to see fast kinetics which may otherwise be missed. Furthermore, while the bulk methods give one only the average kinetics, single molecule methods also give the distribution of probabilities to access certain bound or conformational states, which in turn can give the observer information of the nature of the stochastic process. We rely in this thesis on single-molecule FRET, an abbreviation for Förster Resonance Energy Transfer: It's a process where energy is transferred through dipole-dipole interaction from a donor fluorophore to an acceptor fluorophore. The distance between fluorophores determines the efficiency of transfer on a length scale of ~ 10 nm. Since many biological processes take place on this length scale, this technique is exquisitely suitable to study biological processes real-time on the smallest scale, whether it is conformational changes, protein-protein interactions, or in this case, protein target search studies.

The first two chapters are essentially introductions to the material discussed in this thesis, where chapter 1 is mostly a general introduction on life and our interest in Argonaute and chapter 2 focusses more on the target search aspect. Chapter 3 will be focussing on the biological role of a prokaryotic Argonaute (pAgo) whereas chapter 4 and 5 will focus on the target search aspect of the same protein. Lastly, chapter 6 will deal with something different altogether: the use of Ago in to speed up existing methods in super-resolution microscopy.

Chapter 1 provides a general introduction on the biological system of interest. In eukaryotes, the Argonaute is an essential component in a regulatory pathway called RNA interference (RNAi). Using a 20-22 nt long RNA strand as a guide, it finds its target through complementary base pairing. There are several pathways available since RNAi is a broad term for regulation by different small RNAs. In the microRNA (miRNA) pathway RNAs are transcribed inside the nucleus, resulting in primary transcripts with a hairpin called pri-miRNA. After being processed by Drosha, the precursor miRNA (pre-miRNA) is transported outside the nucleus where Dicer cleaves off the hairpin and trims the dsRNA to the right size. The dsRNA is then loaded inside the Argonaute protein. One of the two strands (passenger strand) is ejected here whereas the other one is used as a guide to find a complementary sequence. This complex is called an RNA induced silencing complex (RISC). Finally, gene silencing occurs through RISC binding to the target mRNA. This is followed by recruiting additional factors that are responsible for translational repression and destabilization of the polyA tail. Small interfering RNA (siRNA) are similar to miRNA. Long dsRNAs, which may be exogenous in origin, are

processed by Dicer into 20-24 bp long constructs after which the short dsRNAs are loaded into Ago2. A difference between siRNA and miRNA is that Ago2 can use siRNA to cleave the RNA targets whereas miRNA tolerates bulges and mismatches since it relies on binding to target RNA and recruitment of other factors that induce silencing.

In recent years, people have also turned their focus towards Argonautes from prokaryotes. Contrary to their eukaryotic counterpart, these pAgos were found to target ssDNA in most of the cases, not ssRNA, suggesting that transcriptional gene-regulation is not one of the functions. Host defense has been suggested as a function, but the picture is far from complete and there is surely a host of unknown mechanisms that need be investigated before we can say with certainty what these systems do. In chapter 3 we will take a look at one of these pAgos and discover its properties.

Many cellular processes rely on proteins to bind to their respective RNA/DNA targets in a timely manner. Ranging from gene regulation, DNA repair or host defence, in all these aspects a fast search is paramount to the survival and of the cell. Three-dimensional (3D) collisions are sufficient most of the time for timely recognition, but a few decades ago it was discovered that the binding rate of some proteins vastly exceeded the limit set by 3D diffusion. To account for this paradox, people have come up with the idea of facilitated diffusion. With this mechanism proteins actually not only diffuse freely in solution but may also bind to DNA substrate and translocate laterally along it. In effect, this lateral diffusion reduces the dimensionality of the problem from 3D to 1D. However, this 1D search comes at a cost of becoming redundant very quickly as well, so a balance should exist between the two modes of translocation. Theoretical papers have shown that a combination of 1D and 3D is the best for quickly locating a target. In chapter 2 we provide a brief summary on the current mechanisms that proteins have been found to be using for locating their target sites. The two common mechanisms of lateral search are sliding and hopping, where sliding is characterized by a tight interaction: the protein in continuously in contact with the substrate. Hopping is actually an ill-defined term where the protein undergoes micro-dissociations repeatedly along the DNA strand. Then there is intersegmental jumps and intersegmental transfer, where the protein relies on the proximity of other strands to hop or transfer to that strand. Furthermore, we take a look at the speed-stability paradox and describe how it applies to our Argonaute search system.

In chapter 3 we characterize an Argonaute protein from the mesophile *Clostridium butyricum* (CbAgo). It has been shown that this protein is catalytically active and that it is able to cleave ssDNA at moderate temperatures (37 °C). This is unlike the previously characterized Ago from *Thermus thermophilus* which at operates at high temperatures (≥ 65 °C). Furthermore, nucleotide identity preference has been shown for the first 4 nucleotides of the seed region of the guide. Lastly, it has been shown that using a pair of CbAgo-siDNA complexes, it's in principle possible to target double stranded DNA. This shows that the use of pAgos may potentially open up new gene-editing applications.

In chapter 4 we focus on the target search aspect. Here, we harness the remarkable biophysical properties of CbAgo with ssDNA and use single molecule FRET to disentangle the different target search mechanisms that it uses. We find that CbAgo is able to use a sliding-like mechanism on short (<20 nt) length scales. Unlike what has been previously assumed, this sliding actually consists of weak interactions between

protein-complex and DNA strand, which allows it to bypass junctions in ssDNA or ssRNA. We see that for larger distances between target sites (greater than 30 nt), the time to reach the other side does not increase as expected for lateral diffusion: It actually deviates from lateral diffusion theory and shows a much slower than expected increase in shuttling time as the target distances increase up to 120 nt. We show through dsDNA block assays that this is most likely caused by intersegmental jumps. Lastly we show that the complex uses this to bypass proteins and to cover large distances efficiently.

The findings in chapter 4 have significance for understanding how optimized this system is in finding its target. At the same time, it may serve as a model-system for nucleic acid guided proteins. In literature, facilitated diffusion mechanisms are often described as optimal if equal time is spent sliding along the DNA as is spent on 3D diffusion. In chapter 5 we develop a target search model based on our experimental observations where Argonaute undergoes a new lateral diffusion movement called skip-n-slide. As its name implies, the movements contain not only sliding motions, but these motions are always followed by a hopping/skipping like motion towards another segment close by. We compare these theoretical results with the experimental findings of CbAgo and hAGO2. From theory and simulations we can calculate back the occurrence distribution between slides and hops based on the experimental data. Furthermore we arrive at a new theoretical optimum for search time based on this skip-n-slide mechanism besides the traditional sliding-mode. This new mode of lateral diffusion may open up a new area of target search for nucleic acid guided proteins.

Lastly, in chapter 6 we turn to the application side of Ago. Argonaute pre-arranges the seed of the guide so that the bases undergo helical stacking. This in turn allows the seed of the guide to rapidly bind to the target, at near-diffusion limited speed, with a binding rate ($\sim 10^7$ M^{-1} s^{-1}). This is typically an order of magnitude higher than naked DNA or RNA strands only. Here we utilize this fast binding property of Ago as a tool to speed up existing DNA-PAINT-based techniques. As a proof-of-concept experiment we assembled a DNA origami plate with four binding sites that are otherwise not resolvable through conventional microscopy methods. We compare the localisation precision of summed origami structures of DNA-PAINT with Ago-PAINT, our method of choice, and show that there is indeed no difference between the two. Following this, we compared our approach with DNA-PAINT in terms of image formation. We notice that on average Ago-PAINT exhibited a 10x faster binding rate, which allows one to probe nanostructures on a much faster timescale. Furthermore, in cases where a high density of binding sites becomes limiting for probe concentration, one can make use of a lower concentration of probe resulting in a lower signal-to-noise ratio.

The results of this thesis hold significance for the RNAi field as well as for the biophysics field since for a long time not much was known about the search process of Argonaute. How is it able to quickly find the correct sequence in a vast collection of off-targets? Techniques such as DNA curtains may be good in visualizing target search processes on a larger length scale across thousands of basepairs, but the transient movements that occur on tens/few nucleotide length scale are not easily uncovered from these diffraction limited methods. Seemingly three-dimensional diffusion driven collisions may actually contain many more complicated movements hidden within the short length scales. Through single-molecule FRET, we discovered a new mode

of interaction for Argonautes, an important class of proteins for (post)transcriptional regulation. This loose interaction with the substrate allows Ago to bypass secondary structures with ease while staying associated. Furthermore, increased physical understanding of the process shows us that there is more than meets the eye. Through experimental assays and physical modelling we uncovered a new mechanism of lateral diffusion, a combination of skips and slides. This finding may result in a different optimal distribution of mechanisms compared to what was known before. The assay developed here may also function as a way to test target search methods of nucleic acid guided proteins, in order to see to which extent the search behaviours of different proteins are guided by the same principles. It will also be of interest to see if and how the search mechanism provides a speed-up in target search time in actual experiments.

6

SAMENVATTING

In deze thesis gebruiken wij enkele-molecuul FRET om de kinetische eigenschappen van een Argonaute eiwit te onderzoeken. Hoewel men traditioneel gezien bulkmethodes gebruikt om de moleculaire eigenschappen van eiwitten te onderzoeken, geven bulk methodes niet de kortstondige informatie door, aangezien deze er inherent uit wordt gemiddeld. Enkele-molecuul methodes, zoals hun naam impliceert, zorgen ervoor dat men in staat is om interacties te bestuderen tussen individuele moleculen en hun substraten. Dit zorgt ervoor dat men snelle kinetiek kan observeren, welk anderzijds niet gezien kan worden. Naast het feit dat bulk methodes slechts gemiddelde kinetische waarden geven, is het met behulp van enkel-molecuul methoden ook mogelijk de kansverdelingen te extraheren voor het eiwit om in bepaalde gebonden- of conformatietoestanden gevonden te worden. Dit kan op zijn beurt de waarnemer weer informatie geven met wat voor soort stochastisch proces we te maken hebben. We gebruiken in deze thesis enkel-molecuul FRET, een afkorting voor Förster Resonance Energy Transfer: Het is een proces waar energie wordt overgedragen van een donor fluorofoor naar een acceptor fluorofoor via een dipool-dipool interactie. De afstand tussen fluoroforen bepaalt de efficientie van energie-overdracht op een lengteschaal van ~10 nm. Aangezien veel biologische processen plaatsvinden op deze lengteschaal, is deze techniek bijzonder geschikt om biologische processen real-time te bestuderen op de kleinste schaal, of het nu conformatieveranderingen zijn, eiwit-eiwit interacties of zoals in dit geval, eiwit-zoekprocessen.

De eerste twee hoofdstukken zijn in essentie introducties voor de materie die in deze thesis wordt behandeld, waarin in hoofdstuk 1 hoofdzakelijk een algemene introductie wordt gegeven op de basis van het leven en onze interesse in Argonaute en in hoofdstuk 2 wordt de focus nader gelegd op het doelwit-zoekproces door eiwitten. Hoofdstuk 3 zal zich focussen op de biologische rol van een prokaryotische Argonaute (pAgo), terwijl in hoofdstuk 4 en 5 de focus gelegd wordt op het zoekproces van hetzelfde eiwit. Tenslotte zal in hoofdstuk 6 iets compleet anders worden behandeld: het gebruik van Ago om bestaande super-resolutie methodes te versnellen.

Hoofdstuk 1 geeft een algemene introductie met betrekking tot het bestudeerde biologische systeem. In eukaryoten is de Argonaute een essentieel component in een regulerend proces dat RNA interferentie (RNAi) heet. Met behulp van een 20-22 nt lang RNA als gids, vind het eiwit zijn target door het vormen van complementaire baseparen. Er zijn verschillende processen beschikbaar aangezien RNAi een brede term is voor post-transcriptionele regulatie door verschillende kleine RNAs. In het microRNA (miRNA) proces vindt transcriptie plaats binnen de nucleus, wat resulteert in primaire RNA transcripties met haarspeldvormige lussen, dit wordt ook wel pri-miRNA genoemd. Nadat het verwerkt wordt door het eiwit Drosha, wordt het pre-miRNA vervoerd naar buiten de nucleus waar Dicer de haarspeld eraf knipt en het dubbelstrengs RNA naar de juiste lengte trimt. Het dsRNA wordt dan vervolgens geladen in het Argonaute eiwit.

Een van de twee strengen (de passenger) wordt afgestoten waarna de andere streng wordt gebruikt als gids om een complementaire sequentie te vinden. Dit complex wordt een RNA induced silencing complex (RISC) genoemd. Uiteindelijk vindt gene-silencing plaats door het binden van RISC aan het doelwit-mRNA. Vervolgens worden aanvullende factoren gerekruteerd; deze onderdrukken translatie en destabiliseren de poly(A)-staart. Small interfering RNA (siRNA) lijken op miRNA. Lang dubbelstrengs RNA, wat mogelijk van buiten de cel kan komen, wordt door Dicer verwerkt tot 20-24 basepaar lange constructen welke geladen worden in Ago2. Een verschil tussen siRNA en miRNA is dat Ago2 siRNA kan gebruiken om RNA doelwitten doormidden te knippen terwijl miRNA uitstulpingen en mismatches kan tolereren, aangezien deze hier afhankelijk is van stabiele binding aan het doel-RNA om andere factoren te rekruteren die silencing induceren.

In recente jaren hebben mensen hun focus ook verplaatst naar Argonauten van prokaryoten. In tegenstelling tot hun eukaryotische variant, waren deze pAgos vooral enkelstrengs DNA aan het targeten in de meeste gevallen, niet enkelstrengs RNA, wat suggereert dat transcriptionele genregulatie niet een van de functies is van dit eiwit. Er is geopperd dat de functie hier host-verdediging is, maar het plaatje is verre van compleet en er is zeker een hoeveelheid onbekende mechanismen die eerst begrepen moeten worden voordat wij met zekerheid kunnen zeggen wat deze systemen doen. In hoofdstuk 3 kijken we naar een van deze pAgos en ontdekken zijn eigenschappen.

Veel cellulaire processen zijn afhankelijk van het tijdig binden van eiwitten aan hun RNA of DNA doelwitten. Of het nu genregulatie, DNA reparatie of verdediging van het gastheerorganisme is, in al deze aspecten is het van opperste belang om snel het doelwit te lokaliseren voor het voortbestaan van de cel. Drie-dimensionale (3D) botsingen zijn meestal voldoende om doelwitten tijdig te herkennen, echter zijn enkele decennia geleden eiwitten gevonden die veel sneller bonden dan de harde limiet die 3D diffusie opstelt. Om deze paradox op te lossen hebben sommigen een theorie ontwikkeld van gefaciliteerde diffusie, waar eiwitten niet alleen vrij in oplossing diffunderen maar ook kunnen binden aan DNA substraat en erlangs kunnen transloceren. Feitelijk wordt het aantal dimensies van het probleem gereduceerd van drie naar een dimensie. Deze 1D zoekmodus heeft wel als minpunt dat het heel snel overbodig wordt, er moet dus een balans zijn tussen de twee manieren van translocatie. Theoretische werken hebben laten zien dat een combinatie van 1D en 3D het beste is om een doelwit snel te vinden. In hoofdstuk 2 geven wij een korte samenvatting van de huidige ontdekte mechansimen die eiwitten gebruiken om hun doelwit te vinden. De twee algemene mechanismen van het zoeken langs het DNA zijn het glijden en het hoppen. Bij het glijden komt er een sterke interactie voor: het eiwit is continu in contact met het substraat. Hoppen is eigenlijk een slecht-gedefinieerde term waar het eiwit herhaaldelijk micro-dissociaties ondergaat langs het DNA-streng. Dan zijn er ook intersegmentele sprongen en intersegmental transfer, waar het eiwit afhankelijk is van de nabijheid van andere DNA segmenten om ernaartoe te hoppen of te gaan. Daarnaast kijken wij ook naar de snelheid-stabiliteitsparadox en hoe dit wordt toegepast op ons Argonaute zoeksysteem.

In hoofdstuk 3 karakteriseren wij een Argonaute eiwit van het mesofiel *Clostridium butyricum* (CbAgo). Er is laten zien dat dit eiwit in staat is om enkelstrengs

DNA bij gematigde temperaturen (37 °C) kan knippen, wat verschilt van de eerder gekarakteriseerde Ago van *Thermus thermophilus* welke werkt op hoge temperaturen (≥ 65 °C). Daarnaast is laten zien dat het eiwit een voorkeur heeft voor bepaalde nucleotides voor de eerste vier posities van het "seed"gebied van de guide DNA. Deze studie laat zien dat bepaalde pAgos wellicht gebruikt kunnen worden voor potentiele toepassingen van genbewerking.

In hoofdstuk 4 focussen wij op het zoekproces van Ago. Hier maken wij gebruik van de opmerkelijke biofysische eigenschappen van CbAgo met ssDNA en gebruiken wij enkel-molecuul FRET om de verschillende zoekmechanismen van elkaar te ontkoppelen. We komen erachter dat CbAgo een soort glij-mechanisme gebruikt op korte lengte schalen (< 20 nt). In tegenstelling tot wat eerder is aangenomen, bestaat dit glijden uit zwakke interacties tussen het eiwit-complex en het DNA, wat ervoor zorgt dat het langs knooppunten kan in enkelstrengs DNA of enkelstrengs RNA. We zien dat voor grotere afstanden tussen de doelwitten (> 30 nt) de tijd die het eiwit erover doet om de andere kant te bereiken, niet zodanig toeneemt zoals men zou verwachten voor laterale diffusie. Het wijkt juist af van laterale diffusie theorie en laat juist een veel langzamere toename zien van shuttling time wanneer de afstand tussen doelwitten toeneemt tot 120 nt. We laten zien door middel van dsDNA blokkade assays dat dit zeer waarschijlijk veroorzaakt wordt door intersegmentele sprongen. Tenslotte laten we zien dat dit complex dit gebruikt om langs eiwitten te komen en om lange afstanden efficiënt te overbruggen.

De bevindingen in hoofdstuk 4 zijn belangrijk voor ons begrip van hoe geoptimaliseerd dit systeem is om z'n doelwit te vinden. Tegelijkertijd kan het wellicht dienen als een model-systeem voor nucleïnezuur-geleide eiwitten. In de literatuur worden gefaciliteerde diffusie mechanismen vaak beschreven als optimaal wanneer evenveel tijd wordt gespendeerd aan glijden langs het DNA alswel aan 3D diffusie. In hoofdstuk 5 ontwikkelen we een zoekproces model gebaseerd op onze experimentele observaties, waar Argonaute een nieuwe lateraal diffusieve beweging maakt genaamd een skip-n-slide. Zoals de naam impliceert, bevatten deze bewegingen niet alleen glij-bewegingen, maar ook zogenaamde skips naar andere segmenten die dichtbij zijn. We vergelijken deze theoretische resultaten met de experimentele bevindingen van CbAgo en hAgo2. Van theorie en simulaties kunnen we berekenen wat de verdeling is van glij-bewegingen en hops, welke gebaseerd zijn op experimentele data. Daarnaast ontdekken we een nieuwe theoretische optimum voor de zoektijd gebaseerd op dit skip-n-slide mechanisme naast de bestaande glij-modus. Deze nieuwe modus van laterale diffusie kan een nieuw veld zoekprocessen openen voor nucleinezuur-geleide eiwitten.

Tenslotte kijken wij in hoofdstuk 6 naar de toepassingskant van Ago. Argonaute structureert van tevoren de seed van de guide zodat de bases als een spiraal op elkaar stapelen. Dit zorgt er op zijn beurt voor dat de seed van de guide heel snel kan binden aan het doelwit, met bijna diffusie gelimiteerde snelheden met een bindingssnelheid van (~ 10^7 M^{-1} s^{-1}). Dit is typisch een ordergrootte sneller dan met slechts naakte DNA of RNA strengen. Hier gebruiken wij deze snelle bindingseigenschap van Ago als gereedschap om bestaande DNA-PAINT gebaseerde technieken te versnellen. Als een proof-of-concept experiment hebben wij een DNA origami plaat geassembleerd

met vier bindingsplekken die anderzijds niet te onderscheiden zijn met behulp van conventionele microscopie technieken. We vergelijken de localisatie precisie van gesommeerde origami structuren van DNA-PAINT met Ago-PAINT, onze gekozen methode, en laten zien dat er inderdaad geen verschil is tussen de twee. We vergelijken vervolgens onze aanpak met DNA-PAINT in termen van beeldvorming. We merken dat gemiddeld gezien Ago-PAINT een 10 maal zo snel bindt, wat het mogelijk maakt om nanostructuren op een veel snellere tijdschaal te onderzoeken. Daarnaast kan men lagere concentraties van fluorescente probes gebruiken wanneer de dichtheid van bindingsplekken te hoog is, wat resulteert in een lagere signaal-ruisverhouding.

De resultaten in deze thesis zijn significant voor het RNAi onderzoeksgebied als wel als voor de biofysiche onderzoeksgebieden aangezien voor lange tijd niet veel bekend was over het zoekproces van Argonaute. Hoe kan het snel de correcte sequentie vinden in een enorme verzameling van niet-doelwit sequenties? Technieken zoals DNA-gordijnen kunnen toepasselijk zijn om zoekprocessen op een grotere lengte schalen over duizenden baseparen, maar de kortstondige bewegingen die plaatsvinden op enkele/tientallen nucleotiden lengteschalen zijn niet makkelijk bloot te leggen met deze diffractie-gelimiteerde methodes. Schijnbaar drie-dimensionaal gedreven diffusieve botsingen kunnen eigenlijk veel meer complexere bewegingen verborgen hebben binnen deze korte lengteschalen. Met behulp van enkel-molecuul FRET hebben wij een nieuwe manier van interactie ontdekt voor Argonaute, een belangrijk eiwit voor (post)transcriptionele regulatie. De zwakke interactie met het substraat zorgt ervoor dat Ago gemakkelijk langs secundaire structuren kan terwijl het gebonden blijft aan het substraat. Beter fysisch begrip laat zien dat er meer is dan op het eerste gezicht. Door middel van experimentele assays and fysisch modelleren hebben wij een nieuw mechanisme ontdekt van laterale diffusie, een combinatie skips en glijden. Deze bevinding kan resulteren in een andere optimale verdeling van mechanismen vregelijken met wat eerder bekend was. De assay die hier ontwikkeld is kan ook gebruikt worden in het algemeen om zoekprocessen van nucleinezuur-geleide eiwitten te testen, om te zien in welke mate de zoek strategiëen door dezelfde principes bepaald worden. Het zal ook interessant zijn om te kijken óf en zo ja, in welke mate deze zoekmechanismen zorgen voor een kortere zoektijd in experimenten.

CURRICULUM VITÆ

Tao Ju Cui

04-11-1990   Born in Leiden, the Netherlands.

EDUCATION

2002–2009   Secondary Education
Fioretti College, Lisse, the Netherlands

2009–2012   B.Sc. in & Applied Physics
Delft University of Technology, the Netherlands

2012–2014   M.Sc. in Applied Physics
Delft University of Technology, the Netherlands

2015–2019   PhD. in Biophysics
Delft University of Technology, the Netherlands
Thesis: A Random Walk Towards the Golden Fleece:
Single-Molecule Investigations of Argonaute
Target Search
Promotor: Dr. C. Joo
Co-Promotor: Dr. M. Depken

LIST OF PUBLICATIONS

10. **Cui, T.J.**[†], Filius, M.[†], Ananth, A.N., Doctor, M.W.W., Joo, C., (2019) *Fast and accurate FRET-Based imaging through protein-assisted probes, (In review)*

9. Klein, M., **Cui, T.J.**, Joo, C., Depken, M., (2019) *Optimal target search using frequent skip-n-slides, (In preparation)*

8. Ruijtenberg, S.[†], Sonneveld, S.[†], **Cui, T. J.**, Logister, I., de Steenwinkel, D., Xiao Y., MacRae, I. J., Joo, C., Tanenbaum, M. E., (2019), *mRNA structural dynamics and heterogeneity shapes Argonaute-target interactions, (Submitted)*

7. Gilboa, B., Jing, B., **Cui, T.J.**, Sow, M., Plochowietz, A., Mazumder, A., and Kapanidis, A. N., (2019) *Confinement-Free Wide-Field Ratiometric Tracking of Single Fluorescent Molecules.*, Biophys. J. *(In press)*

6. **Cui, T.J.**, Klein, M., Hegge, J.W., Chandradoss, S.D., van der Oost, J., Depken, M., and Joo, C. (2019). *Argonaute bypasses cellular obstacles without hindrance during target search.* Nat. Commun. **10**, 4390.

5. **Cui, T.J.**, and Joo, C. (2019). *Facilitated diffusion of Argonaute-mediated target search.* RNA Biol. **16**, 1093–1107.

4. Hegge, J.W.[†], Swarts, D.C.[†], Chandradoss, S.D., **Cui, T.J.**, Kneppers, J., Jinek, M., Joo, C., and van der Oost, J. (2019). DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute. Nucleic Acids Res. **47**, 5809–5821.

3. Kim, S.H., Ahn, T., **Cui, T.J.**, Chauhan, S., Sung, J., Joo, C., and Kim, D. (2017). RecA filament maintains structural integrity using ATP-driven internal dynamics. Sci. Adv. **3**.

2. Dulin, D., Yu, Z., **Cui, T.J.**, Berghuis, B.A., Depken, M., and Dekker, N.H. (2016). *Real-time observation of replicative helicase assembly onto single-stranded DNA.*, bioRχiv.

1. Dulin, D., **Cui, T.J.**, Cnossen, J., Docter, M.W.W., Lipfert, J., and Dekker, N.H. (2015). *High Spatiotemporal-Resolution Magnetic Tweezers: Calibration and Applications for DNA Dynamics.* Biophys. J. **109**, 2113–2125.

[†] denotes equal contribution