THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

A Flexible Domain-Domain Hinge Promotes an Induced-fit Dominant Mechanism for the Loading of Guide-DNA into Argonaute Protein in *Thermus thermophilus*

Lizhe Zhu,[†,‡] Hanlun Jiang,[‡,§] Fu Kit Sheong,[†] Xuefeng Cui,[∥] Xin Gao,[∥] Yanli Wang,[⊥] and Xuhui Huang*[,†,‡,§]

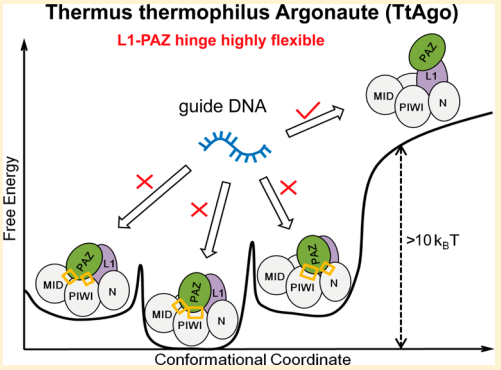
[†]Department of Chemistry, [‡]Center of Systems Biology and Human Health, School of Science and Institute for Advance Study, [§]Bioengineering Graduate Program, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

[∥]Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[⊥]Laboratory of Non-Coding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Supporting Information

ABSTRACT: Argonaute proteins (Ago) are core components of the RNA Induced Silencing Complex (RISC) that load and utilize small guide nucleic acids to silence mRNAs or cleave foreign DNAs. Despite the essential role of Ago in gene regulation and defense against virus, the molecular mechanism of guide-strand loading into Ago remains unclear. We explore such a mechanism in the bacterium *Thermus thermophilus* Ago (TtAgo), via a computational approach combining molecular dynamics, bias-exchange metadynamics, and protein−DNA docking. We show that apo TtAgo adopts multiple closed states that are unable to accommodate guide-DNA. Conformations able to accommodate the guide are beyond the reach of thermal fluctuations from the closed states. These results suggest an induced-fit dominant mechanism for guide-strand loading in TtAgo, drastically different from the two-step mechanism for human Ago 2 (hAgo2) identified in our previous study. Such a difference between TtAgo and hAgo2 is found to mainly originate from the distinct rigidity of their L1-PAZ hinge. Further comparison among known Ago structures from various species indicates that the L1-PAZ hinge may be flexible in general for prokaryotic Ago's but rigid for eukaryotic Ago's.



I. INTRODUCTION

RNA interference (RNAi) is a key cellular mechanism involved in the regulation of gene expression and the defense against viruses throughout all kingdoms of life.[1,2] As the central component of RNAi, the RNA Induced Silencing Complex (RISC) mediates the sequence-specific recognition and inhibition of target mRNA,[3] and foreign nucleic acids.[4,5] As the core component of RISC, Argonaute protein (Ago) loads and utilizes a small guide nucleic acid to selectively capture the target nucleic acid for inhibition or degradation.[3,6−9] Because of its promising pharmaceutical potential in regulating specific gene silencing,[10−12] Ago has received phenomenal attention in the past decades.[6,13]

Despite intensive studies, the mechanism of the very first step of the RISC formation, guide nucleic-acid loading into Ago, remains largely elusive. In vitro[14,15] and, under crystallographic conditions,[9,16−18] spontaneous loading of a single-stranded guide has been observed for various Ago's. In vivo, it is generally accepted that guide-strand loading consists of two-steps: loading of a duplex DNA or RNA from the Dicer protein[19,20] and release of one (passenger) strand, so that only the guide-strand remains in Ago.[21,22] It remains unknown whether apo Ago loads the single-stranded guide in a selective manner.

The guide-strand loading is a molecular recognition process between Ago and the guide nucleic acid. Two extreme mechanistic models are often used in understanding such molecular recognition: conformational selection (CS) and induced fit (IF).[23−26] In CS, the protein is in equilibrium with a number of metastable conformational states and the nuclear acid can readily bind to one of the states.[26,27] In IF, the final bound conformation is inaccessible for the apo protein within the reach of thermal fluctuation, and the nucleic acid has to first bind and form weak interactions with a different conformation and further induces conformational change of the protein toward the final bound form.[25] Both models have found abundant examples in nature. For instance, CS has been suggested in the recognition between lac repressor headpiece and DNA.[28,29] IF seems to play a major role in the recognition between Pancreatic and duodenal homeobox 1 (Pdx1) and

Received: December 18, 2015
Revised: February 23, 2016
Published: February 24, 2016

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

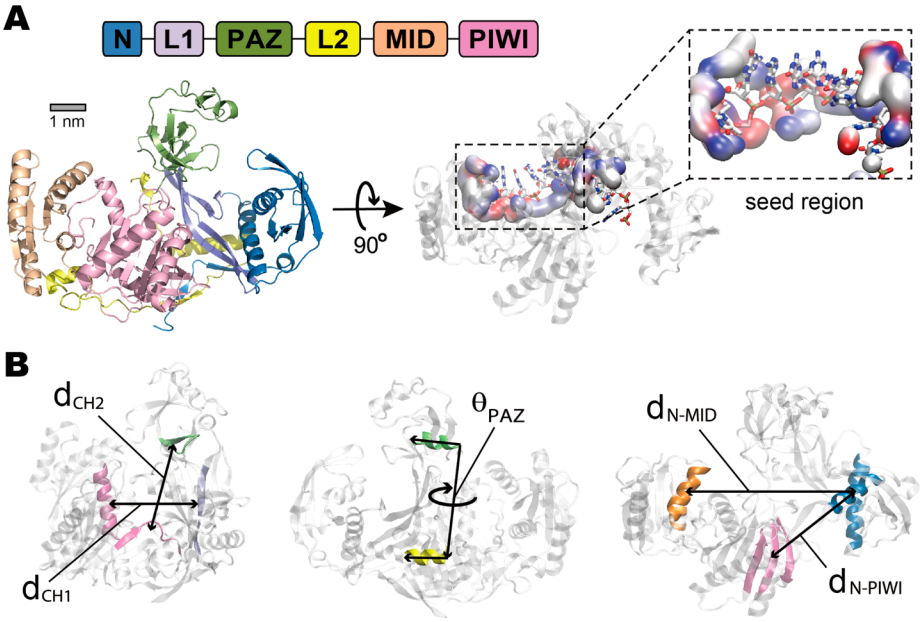


Figure 1. Argonaute protein of the bacterium *Thermus thermophilus* (TtAgo). (A) Domain composition (left) and the nucleic-acid binding channel (right) of TtAgo (structure PDB 3DLH). The seed region (definition in the main text) is enlarged for better visualization in the inset figure. The guide DNA is in licorice representation while TtAgo atoms in contact with the DNA are plotted as surfaces. Atoms with negative and positive charges are in red and blue, respectively. (B) Collective variables (CVs) describing TtAgo conformations: d_{CH1} and d_{CH2} measure the diameter of the nucleic-acid binding channel; θ_{PAZ} describes the torsion angle between PAZ and L2; $d_{N\text{-}MID}$ and $d_{N\text{-}PIWI}$ measure the distance between the N and MID domains and between the N and PIWI domains, respectively. These CVs are used in the enhanced sampling scheme (Methods). Residues defining the CVs are colored as in (A).

DNA,[30] and between p53 and DNA.[31] Due to the necessary conformational change after the initial binding, IF appears to offer higher specificity and fidelity of recognition than CS, especially when similar binding competitors are present.[32] In practice, the recognition mechanism often lies between CS and IF.[33−36]

To understand the molecular mechanism of guide-strand loading, one must consider chemical details and dynamics, using techniques such as molecular simulations,[34] because of the intrinsic complexity of Ago and the guide nucleic acid and the complex interaction between them. Along this direction, we have previously studied miRNA loading into human Ago 2 (hAgo2)[37] via a computational protocol combining all-atom Molecular Dynamics (MD), Markov State Models (MSMs),[38−52] and protein-RNA docking.[53,54] There we proposed a two-step model for guide-strand loading into hAgo2:[37] first, guide strand selectively binds to an open conformation of apo hAgo2 via CS; subsequently, small-scale structural rearrangements of both molecules lead to the final bound complex.

In the present study, we focus on the bacterium *Thermus thermophilus* Ago (TtAgo). TtAgo is among the most extensively investigated Ago proteins.[55−57] In Figure 1A, we display the crystal structure of TtAgo (PDB 3DLH) in its binary form with bound single-stranded DNA (ssDNA). Similar to other Ago proteins, TtAgo consists of 6 functional domains, namely the N (N-terminal) domain, L1 (Linker 1), the PAZ (PIWI-Argonaute-Zwille) domain, Linker 2 (L2), the MID (middle) domain, and the C-terminal PIWI (P element-induced wimpy testis) domain. MID and PAZ bind to the 5′ and 3′ terminus of the guide, respectively, and define the two ends of the nucleic-acid binding channel. Driven mainly by the electrostatic attraction between the positively charged protein residues and the negatively charged backbone atoms of the DNA,[58,59] the 5′ segment of the channel binds nucleotide 2-8 of the guide known as the seed region (right top of Figure 1A).[8] Correct positioning of the seed region in the channel is crucial for subsequent recognition of target nucleic acid in both TtAgo[8] and human Ago.[60,61] The diameter of the nucleic-acid binding channel is regulated by the relative positions of PIWI, L1 and PAZ. Notably, in this binary crystal structure, direct contacts between the PAZ and PIWI domains exist and bury the guide DNA in the binding channel (see SI Figure 1).

Because both TtAgo and the guide DNA are large biopolymers, directly simulating the loading process via unbiased MD simulations is impractical. Therefore, we adopt a stepwise protocol. We first perform All-atom MD simulations and bias-exchange metadynamics (BEMD), an enhanced sampling scheme, to extensively explore the conformational space of apo TtAgo. Large-scale protein−DNA docking simulations are further performed to assess the exposure of the binding channel of guide-DNA in TtAgo conformations. Our results show that apo TtAgo adopts multiple closed conformations that are unable to accommodate a guide DNA, pointing to an IF dominant mechanism for guide-strand loading. Moreover, the rigidity of L1-PAZ hinge is shown to be critical in regulating the exposure of the nucleic-acid binding channel and therefore the guide-strand loading mechanism. The biological meaning of these findings are discussed below.

II. METHODS

1. Molecular dynamics simulation. The initial apo TtAgo conformation was taken from the crystal structures,[7−9] and the missing residues were added using MODELER version 9.11.[62] The apo TtAgo protein was solvated in a dodecahedral box containing 36,349 TIP3P water molecules with 109 Na+ ions and 126 Cl- ions to neutralize the charge. All the MD simulations were performed using the GROMACS 4.5.4

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

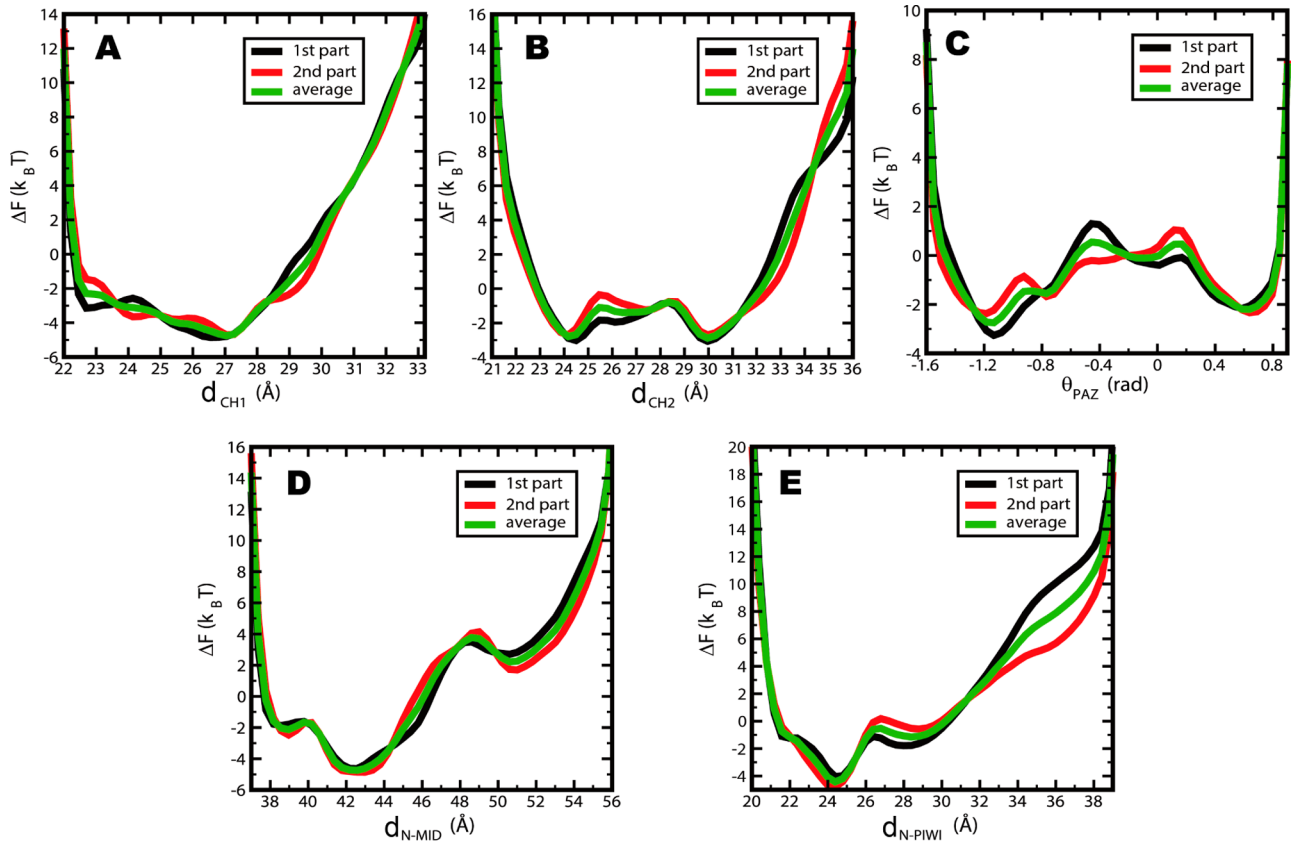


Figure 2. One dimensional free energy profile of 5 collective variables in bias-exchange metadynamics simulations. (A) d_{CH1}; (B) d_{CH2}; (C) θ_{PAZ}; (D) $d_{N\text{-}MID}$; (E) $d_{N\text{-}PIWI}$. The first part of the simulation (black) corresponds to 300−355 ns. The second part of the simulation corresponds to 355−410 ns.

package[63] and the Amber99SB-ILDN force field.[64] Long-range electrostatic interactions were treated with the Particle-Mesh Ewald method.[65,66] Both short-range electrostatic interactions and van der Waals interactions used a cutoff of 10 Å. All bonds were constrained by the LINCS algorithm.[67] Velocity-rescaling thermostat[68] and the Parrinello−Rahman barostat[69] were used for temperature and pressure coupling, respectively. Because *Thermus thermophilus* has its optimal growth rate at 65 °C, all MD simulations were performed at 340 K. The system was first energy minimized with the steepest descent algorithm and equilibrated for 1 ns under NPT ensemble (T = 340 K, P = 1 atm) with the positions of all the heavy atoms restrained. Next, we performed a 5 ns NPT simulation (T = 340 K, P = 1 atm) to further equilibrate the system. Finally, all the production MD simulations were performed under NVT ensemble at 340 K.

2. Enhanced sampling of conformational landscape. BEMD calculations were carried out with PLUMED version 1.3[70] to enhance the sampling of the conformational space of apo TtAgo. In BEMD, several replicas of the same protein system were simulated, each experiencing a different history dependent bias potential that discourages sampling of conformations that has been visited.[71−73] Trial exchanges of the bias potential accumulated in each replica were attempted at a fixed frequency, using a Metropolis Monte Carlo acceptance criterion.[73] Effectively, this procedure flattens the free energy landscape in a set of collective variables (CVs) describing the conformational space of the protein and therefore encourages the protein to explore a large volume of interest in the state space in all replicas. After convergence, the accumulated bias potential were combined and transformed into an unbiased probability distribution via Weighted Histogram Analysis Method (WHAM), resulting in the unbiased free energy landscape in terms of these CVs.[74]

We defined five CVs based on geometrical considerations and physical intuitions to describe TtAgo conformations (Figure 1B). All five CVs were defined using α-Carbon atoms of the corresponding residues. d_{CH1} measures the center-of-mass (c.o.m.) distance between part of L1 (residues 165−174) and one long helix on PIWI (residues 654−668). d_{CH2} is the c.o.m. distance between part of PAZ (residues 192−202) and part of PIWI (residues 572−581). These two distances directly measure the diameter of nucleic-acid binding channel, i.e. the ability of TtAgo to accommodate DNA/RNA. θ_{PAZ} describes the torsion motion between a helix on PAZ (residues 178−185 for TtAgo and 229−236 for hAgo2) and a helix on L2 (residues 283−290 for TtAgo and 371−378 for hAgo2) that has been suggested in previous studies;[56] $d_{N\text{-}MID}$ measures the distance between N (residues 42−53, 105−122) and MID (residues 416−427, 449−456) and encodes the maximum possible length of the nucleic-acid binding channel. Since $d_{N\text{-}MID}$ alone cannot fully describe the position of N, an additional distance $d_{N\text{-}PIWI}$ between the same N residues and part of PIWI (residues 541−574, 618−624) was defined. For direct comparison between TtAgo and hAgo2, $d_{PAZ-PIWI}$ was defined as the c.o.m. distance between PAZ and a helical turn on PIWI (residues 664−667 for TtAgo, 812−815 for hAgo2).

Our BEMD scheme included 8 replicas: 5 replicas with 1-dimensional metadynamics biasing on 5 CVs each, 3 replicas with 2-dimensional metadynamics biasing on 3 pairs of CVs: d_{CH1} and d_{CH2}, d_{CH1} and θ_{PAZ}, d_{CH1} and θ_{PAZ}. These 3 pairs

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

were chosen because d_{CH1}, d_{CH2} and θ_{PAZ} are more directly relevant to exposure of the nucleic-acid binding channel. All replicas initiated from the same relaxed structure of the crystal 3HK2. Boundary quadratic potentials were used to restrict the range of exploration: 22 Å < d_{CH1} < 34 Å, 21 Å < d_{CH2} < 36 Å, 20 Å < $d_{N\text{-}PIWI}$ < 40 Å, 36 Å < $d_{N\text{-}MID}$ < 56 Å, −1.6 rad<θ_{PAZ} < 1.2 rad.

The history-dependent bias potential was progressively developed per replica over the first 300 ns (the filling time) of the BEMD simulation. The height of the bias Gaussians, added every 2 ps, was gradually reduced within 200 ns from 0.04 to 0.0005 kJ•mol^{-1} for 5 replicas with 1-dimensional bias and from 0.4 to 0.05 kJ•mol^{-1} for 3 replicas with 2-dimensional bias. The width of the bias Gaussians was a constant: 0.3 Å for d_{CH1}, 0.4 Å for d_{CH2}, 0.5 Å for $d_{N\text{-}MID}$ and for $d_{N\text{-}PIWI}$ respectively, and 0.05 rad for θ_{PAZ}. Trial exchanges between pairs of replica were attempted every 20 ps. On average, the acceptance ratio of exchanges was ∼30%.

The free energy landscapes of apo TtAgo (in Figure 4−6) were computed from the sampling between 300 and 410 ns per replica via the software METAGUI,[74] following a previously described methodology.[75] The filling time of 300 ns was validated by the reasonable convergence of bias potential in all 5 replicas biasing on 1 CV (Figure 2). WHAM[76,77] was applied to combine the sampling gathered in all replicas and compute the 5-dimensional unbiased free energy landscape. All sampled conformations were divided into microstates on the space of the 5 CVs, with the following bin widths: d_{CH1} 0.3 Å, d_{CH2} 0.4 Å, $d_{N\text{-}PIWI}$ 0.2 Å, $d_{N\text{-}MID}$ 0.5 Å, θ_{PAZ} 0.065 rad. Here the WHAM calculation required, for each replica, the biased probability distribution of all microstates in the 5-dimensional CV space, the time-averaged bias potential accumulated during the simulation (after 300 ns) on each microstate, and the variance of this average. The variance was defined as the mean-squared difference between time averages of two periods, 300−355 ns and 355−410 ns. For a given microstate and replica, the larger the variance is, the less that replica contributes to the unbiased probability of that microstate. We set a cutoff of 2 k_BT in the variance in the calculation of the weights.

For comparing TtAgo and hAgo2, reweighted free energy landscapes (Figure 7,8 and SI Figure 5) were obtained. For TtAgo, the reweighting was based on the microstates in the 5-dimensional CV space and the unbiased probability of all microstates estimated by WHAM (as described above). For hAgo2, the reweighted free energy was computed using the 480 microstates and their equilibrium population in the MSMs.[37]

3. Protein−DNA docking. Protein−DNA docking simulations were performed with HADDOCK 2.1.[53,54] The input TtAgo protein structures were randomly selected from conformations sampled by BEMD with d_{CH1} > 27 Å and d_{CH2} > 25 Å. The input 18-nt DNA structure was modeled using ModeRNA based on DNA fragments in the crystal structure9 (PDB: 3DLH). Specifically, the nucleotides at positions 1-11 and 18-21 in the crystal structure were used to construct the nucleotides 1-11 and 15-18 of our DNA structure.

Using ModeRNA, we modeled DNA structure at position 12-14. Both Ambiguous Interaction Restraints (AIR) and unambiguous distance restraints were defined to drive the docking simulations. For AIR, the "active" protein residues were defined according to the protein−DNA contacts in the crystal structure:[8,9] V167, D168, K191, V193, K213, S222, L223, A228, R251, K252, L277, P282, A414, R418, E428, G429, L430, G453, S576, D609, R611, L638, F647, P653, and K681. All DNA nucleotides were set to be "active". The unambiguous distance restraints were also imposed to guide the docking simulations (see SI Table 1).

The inter- and intramolecular interactions were calculated by CNS 1.3 with PARALLHDG5.4[53,54] and the OPLS-AA force field[78] applied to protein and DNA respectively. For each docking simulation, topologies and coordinates of TtAgo and DNA were generated from the input structures using CNS. The two molecules were initially separated by 25 Å away before the randomization and rigid body energy minimization. Four hundred docking poses were generated after the rigid body docking and ranked by the HADDOCK score[53,54] which is a weighted sum of various energy terms (electrostatic, van der Waals, desolvation, ambiguous interaction restraints and buried surface area). The 200 best-scoring poses were subsequently used for semiflexible simulated annealing and the final solvated refinement.

The quality of the docking models is assessed by a distance metric d_{NT6}: the c.o.m. distance between Nucleotide 6 of the guide DNA and a set of PIWI residues to assess the docking models (Figure 3). A small d_{NT6} (<15 Å) signatures the occupation of the 5′ segment of the binding channel by the seed region of the guide, while a large d_{NT6} (≥15 Å) means otherwise. For reference, d_{NT6} is as small as 6.7 Å in the binary crystal 3DLH (Figure 2).

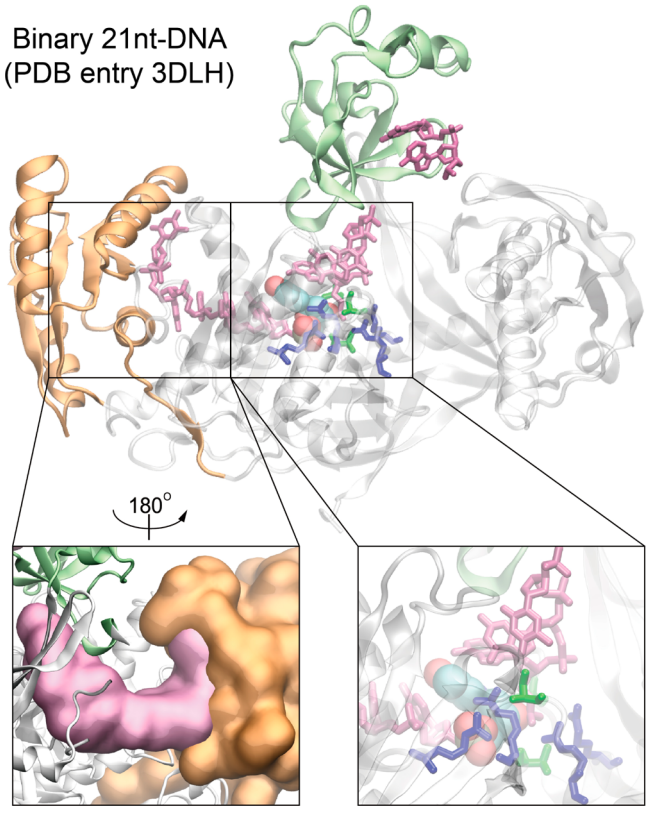


Figure 3. Distance metric d_{NT6} in binary crystal of TtAgo. d_{NT6} is the distance between Nucleotide 6 of the guide DNA and a set of PIWI residues. d_{NT6} is used to assess the quality of docking models in the present work. Color code: green - PAZ, orange - MID, mauve - guide DNA. Nucleotide 6 of the guide DNA is give in the VDW representation. The set of PIWI residues are in blue/green licorice representation.

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

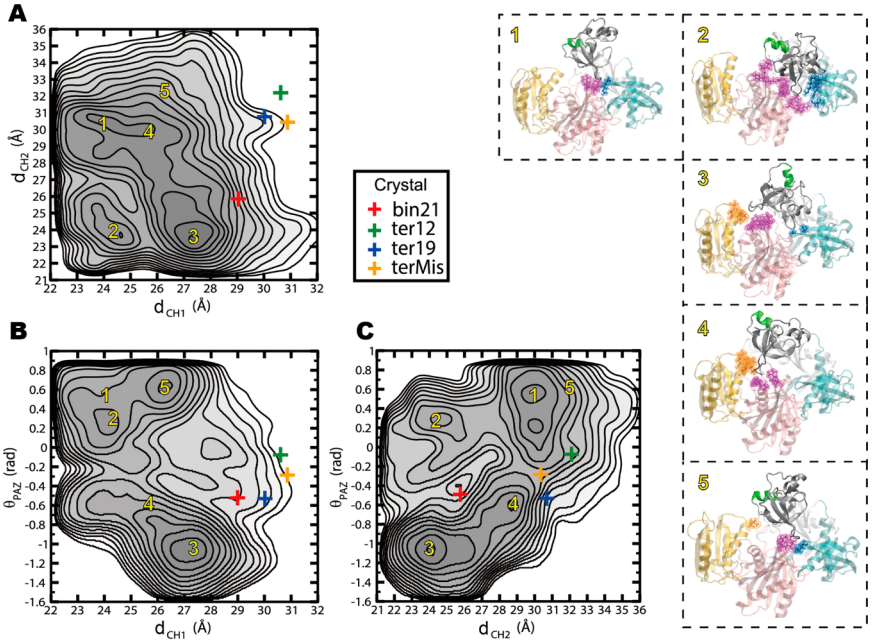


Figure 4. Conformational landscape of apo-TtAgo. (A)−(C) Free energy surfaces computed via bias-exchange metadynamics in collective variable (CV) space: (A) d_{CH1}, d_{CH2}; (B) d_{CH1}, θ_{PAZ}; (C) d_{CH2}, θ_{PAZ} (definition in Figure 1). Darker contours correspond to lower free energy. Contour lines are plotted every 1 kBT (~2.83 kJ/mol at 340 K). Crosses on (A)−(C) correspond to 4 crystal structures bound with DNA: 21nt ssDNA (red, PDB 3DLH), 12nt dsDNA (green, PDB 3HO1), 19nt dsDNA (blue, PDB 3HK2), dsDNA with mismatched base-pairs (orange, PDB 3F73). Representative structures are visualized and marked with yellow numbers 1−5 on (A)−(C). Green helix marks the orientation of PAZ. Residues from N (blue), MID (orange), PIWI (pink) in contact with PAZ (distance <3.5 Å) are highlighted with licorice and cloud-like representation.

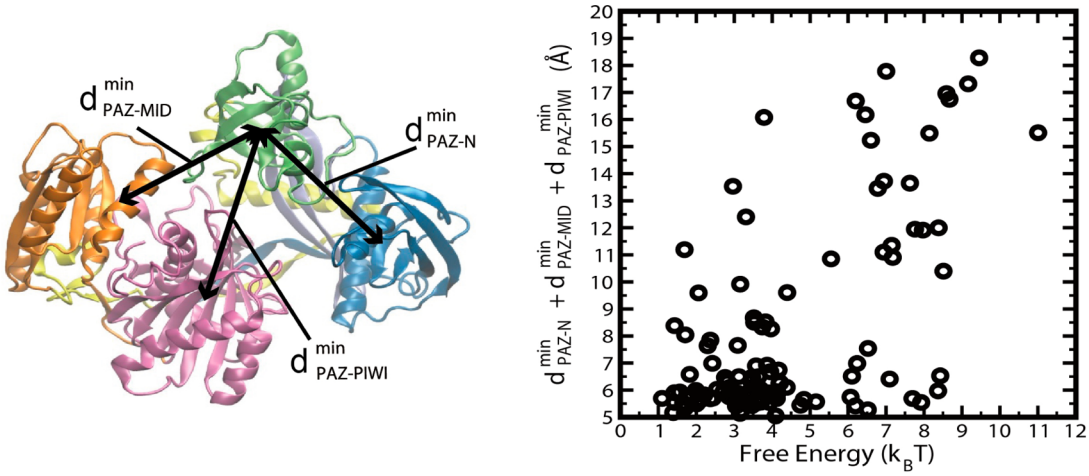


Figure 5. Sum of minimal distances between PAZ and N/MID/PIWI as a function of corresponding free energy. The free energy is estimated by bias-exchange metadynamics.

III. RESULTS

1. TtAgo adopts multiple closed states. We first perform extensive MD simulations to explore the conformational space of apo TtAgo. MD simulations provide dynamical information on atomic level of biomolecules, but are typically limited to those in short time-scale due to finite computing resources. Various recipes exist to tackle this issue including but not limited to metadynamics,[71,72] accelerated MD,[79−81] Temperature or Hamiltonian Replica Exchange MD[82−84] and Markov-State-Models (MSMs).[38−52] In the present work, we choose bias-exchange metadynamics (BEMD),[73,74] a variant of metadynamics that handles large sets of collective variables (CVs), to compute a free energy landscape of apo TtAgo (details in Methods). As shown below, BEMD reveals that it is difficult for apo TtAgo to fully expose its nucleic-acid binding channel (open conformation) via thermal fluctuations. This prohibits observation of sufficient number of transitions via short unbiased MD from the closed conformations to the open ones. Therefore, the MSMs methodology, which uses these transition counts to estimate the thermodynamics and kinetics of a system, is not adopted in this work.

BEMD requires defining a set of CVs that describe the geometrical features of TtAgo. We define 5 CVs for such purpose, shown in Figure 1B. d_{CH1} and d_{CH2} effectively measure the diameter of nucleic-acid binding channel. θ_{PAZ} is the torsion angle between a helix on PAZ and a helical segment on L2 to measure the twisting of PAZ with respect to other domains. $d_{N\text{-}MID}$ and $d_{N\text{-}PIWI}$ measure the relative position of the N

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

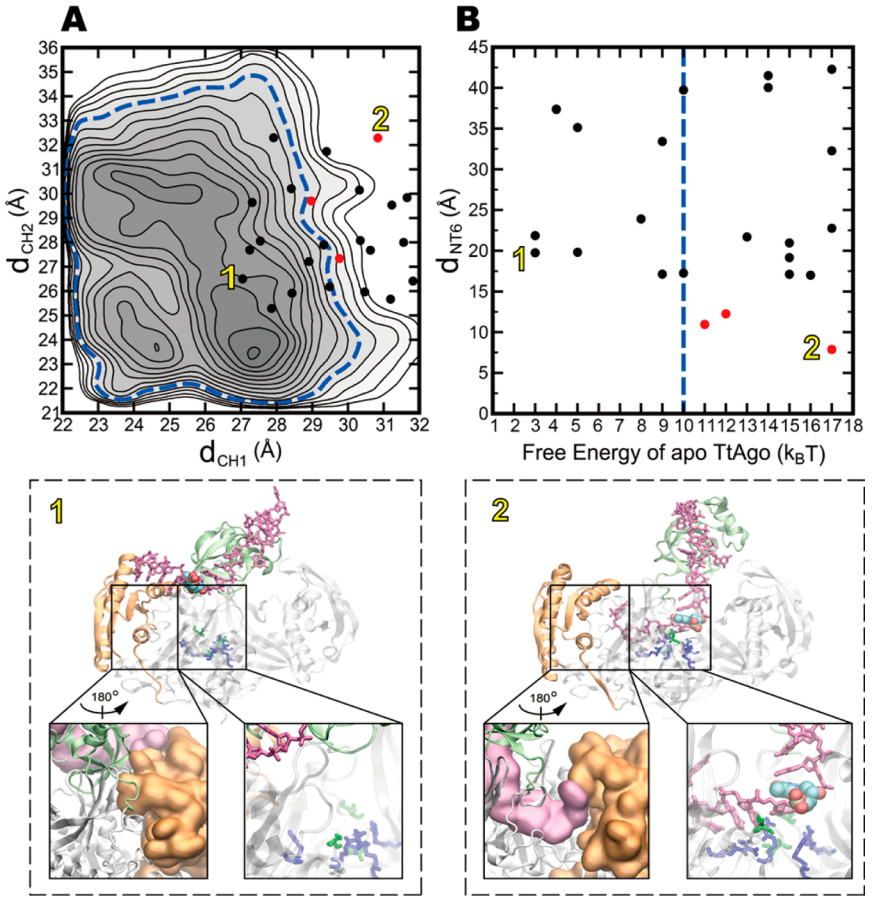


Figure 6. DNA docking simulation on sampled TtAgo conformations. (A) Projection of extracted apo conformations (dots) on the free energy surface of d_{CH1} and d_{CH2} (same as Figure 4A). (B) Best docking models as a function of the free energy of the corresponding apo conformations and the distance d_{NT6} between Nucleotide 6 of the guide DNA and a set of PIWI residues in the docking model. Red and black dots denote successful (d_{NT6} < 15 Å) and unsuccessful ($d_{NT6} \geq$ 15 Å) docking, respectively. Representative structure is visualized for both the unsuccessful (Snapshot 1) and successful (Snapshot 2) docking models. MID, PAZ and the DNA are highlighted in orange, green and mauve, respectively. Nucleotide 6 of the guide DNA and the PIWI residues that define d_{NT6} are plotted in VDW and blue/green licorice representations, respectively. Dashed blue line is the contour corresponding to $10k_BT$ from the most stable state and denotes the border of thermal fluctuation.

domain. θ_{PAZ} is defined because difference in PAZ orientation is detectable among crystal structures[8,9] and large-scale motion of PAZ has been documented.[55−57] $d_{N\text{-}MID}$ and $d_{N\text{-}PIWI}$ are defined because the N domain has been suggested to drive the duplex unwinding process[21] and may also be essential in regulating guide-strand loading.

Our BEMD simulation, consisting of 8 replicas, reaches reasonable convergence after 300 ns. This is evidenced by the similar shape of the 5 one-dimensional bias potential in the low free energy region after this time (Figure 2). Weighted histogram analysis method (WHAM)[76,77] is then used to combine the sampling in all 8 replicas between 300 and 410 ns, resulting in a 5-dimensional free energy landscape of TtAgo. This landscape provides a basis to understand the thermodynamics of TtAgo and its guide-strand loading mechanism.

The free energy landscape computed via BEMD reveals multiple stable states of apo TtAgo but all with closed nucleic-acid binding channel. Figure 4 displays three 2-dimensional projections of such free-energy landscape on d_{CH1}, d_{CH2} and θ_{PAZ}. Clearly, multiple stable states exist in the region with d_{CH1} < 28.7 Å and d_{CH2} < 31.5 Å (see the fifth contour line in Figure 4A). These stable states are featured by direct contacts between PAZ and MID/PIWI/N (Figure 5). The pattern of the contacts of each state depends on the value of θ_{PAZ} (Figure 4B,C). These direct contacts indicate that the nucleic-acid binding channel is closed in all these stable states.

Notably, none of the crystal structures (all with d_{CH1} > 29 Å, crosses in Figure 4) corresponds to stable states of apo TtAgo. This is validated by an additional round of unbiased MD, starting from four different crystal structures. Note that the binary crystal (PDB 3DLH) contains bound ssDNA, while the ternary crystals (PDB 3HO1, 3HK2, 3F73) contain bound duplex DNA. In all these MD trajectories, apo TtAgo relaxes from the crystal structures to regions around the stable states within 50 ns (SI Figure 2). These results suggest that conformations with accessible binding channel to the DNA may only be found with high free energy (see next section).

2. Exposure of binding-channel occurs with small probability. To further search for TtAgo conformations with sufficiently exposed binding channel for the guide DNA, we perform protein−DNA docking simulations using the program HADDOCK[53,54] on a large set of apo TtAgo conformations sampled by BEMD. Docking the complete DNA into TtAgo with exact positioning as in the binary crystal is challenging, not only because of the existence of direct contacts between PAZ and PIWI in the binary crystal structure (see SI Figure 1) and most TtAgo conformations sampled by BEMD (see Figure 4), but also because docking simulations do not consider full

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

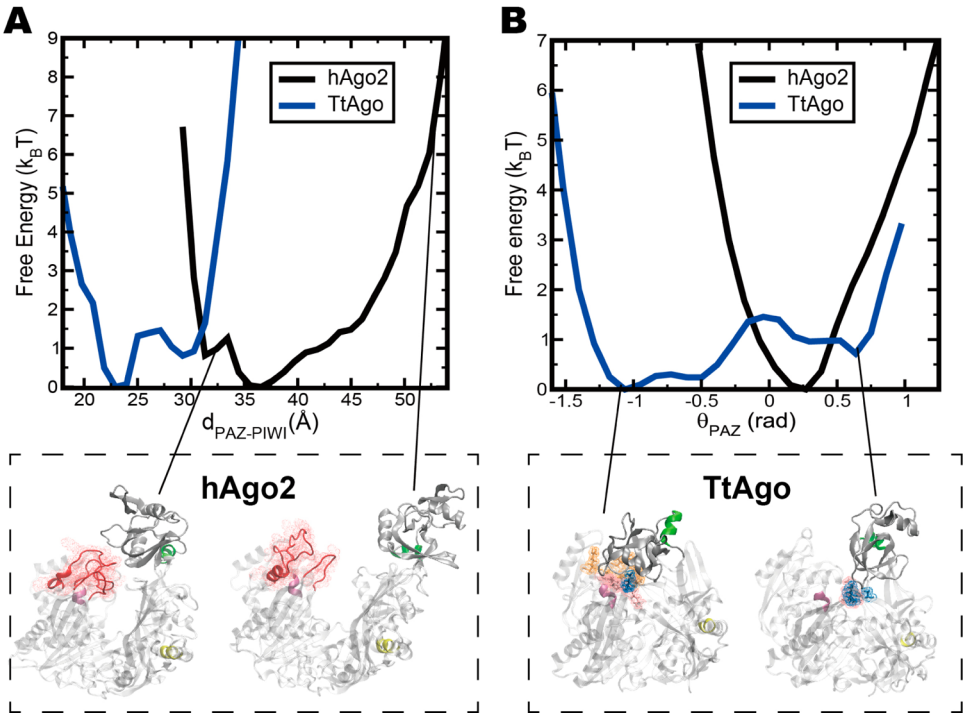


Figure 7. Comparison of PAZ motions between human Argonaute 2 (hAgo2) and TtAgo. One dimensional free energy of (A) $d_{PAZ-PIWI}$ and (B) θ_{PAZ}, obtained from equilibrium population of Markov-State-Models for hAgo2 and reweighted from BEMD for TtAgo. $d_{PAZ-PIWI}$ (details in SI Figure 3) is the center-of-mass distance between PAZ (gray) and a segment of PIWI (magenta). The closed and open conformations of hAgo2 identified in our previous work[37] are visualized. Representative conformations of metastable states of apo TtAgo are visualized. Helices on PAZ (green) and L2 (yellow) defining θ_{PAZ} highlight the relative orientation of PAZ. For hAgo2, two loops not present in TtAgo are in red. For TtAgo, residues from N (blue), MID (orange), and PIWI (pink) in contact with PAZ (distance <3.5 Å) are highlighted.

conformational flexibility of both large biopolymers. However, docking simulations suffice for measuring the exposure of the 5′ segment of the binding channel (corresponding to the crucial seed region of the guide) and for providing additional structural insights than the two CVs d_{CH1} and d_{CH2}.

We first randomly extracted from BEMD trajectories apo conformations with d_{CH1} > 27 Å and d_{CH2} > 25 Å to seed docking simulations (Figure 6A). Conformations with shorter d_{CH1} or d_{CH2} are not extracted since their binding channels are already too narrow to accommodate any guide DNA. Given the importance of correct positioning of the seed region in the 5′ segment of the binding channel,[8,60,61] we further define a distance d_{NT6} to assess the docking models (Figure 3, details in methods). Docking is considered successful if d_{NT6} < 15 Å (the 5′ segment of the binding channel is occupied by the seed region of the guide).

Although the extracted apo TtAgo conformations cover a large free energy range (3−17 k_BT in Figure 6B), only three successful docking models were obtained (red dots in Figure 6). For low free energy conformations (high probabilities to occur), the binding channel is deeply buried, prohibiting the binding of the seed region of the DNA (e.g., Snapshot 1 in Figure 6). For high free energy conformations (small probabilities to occur), seed region binding becomes possible (e.g., Snapshot 2 in Figure 6), but only in conformations with free energy >10 k_BT.

These docking results suggest that, exposure of the nucleic-acid binding channel can only be found with free energy larger than 10k_BT, beyond the reach of thermal motions for apo TtAgo. We therefore conclude that guide-strand loading into TtAgo follows an IF dominant mechanism because large-scale conformational change of TtAgo after the initial encounter with the guide strand is inevitable.

3. Apo TtAgo has a drastically different free energy landscape from human hAgo2. The present results of TtAgo are in clear contrast with our previous results on human Ago 2 (hAgo2).[37] Apo hAgo2 can interconvert among a number of states on microsecond time scale, including an open state readily accessible to the guide miRNA (miRNA conformation similar to the final bound form). It was suggested that once such miRNA encounters the open state Ago, both molecules exhibit small-scale local conformational change leading to the final binary complex. In other words, guide-strand loading into hAgo2 may initiate via CS.[37]

The difference between TtAgo and hAgo2 mainly lies in the mode of motions of their PAZ domain that can be reflected by the distance from PAZ to other domains and the twisting flexibility of PAZ. Figure 7 illustrates this via projection of the free energy landscape of both Ago proteins on two separate order parameters: $d_{PAZ-PIWI}$ and θ_{PAZ}. The former order parameter measures the c.o.m. distance between PAZ and the top of a helix on PIWI (details in Methods and SI Figure 3).

For TtAgo, $d_{PAZ-PIWI}$ rests in small values (18−34 Å, see Figure 7A), consistent with the direct contacts between PAZ and N/MID/PIWI reported in previous sections. For hAgo2, however, $d_{PAZ-PIWI}$ sits between much larger values (30−53 Å), meaning that PAZ is unable to approach its PIWI tip (and N/MID). This is mainly due to the existence of two long loops (residues 601−609 and 818−839) in its PIWI domain that are absent in TtAgo (SI Figure 4). Accordingly, the binding channel of hAgo2 can only be exposed from top between PAZ and these PIWI loops when $d_{PAZ-PIWI}$ > 50 Å. In other words,

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

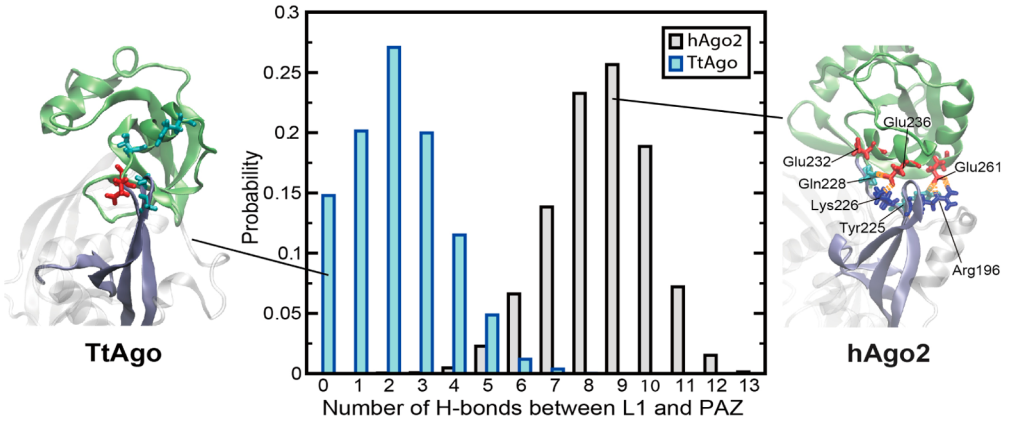


Figure 8. Rigidity of the L1-PAZ hinge for human Argonaute 2 (hAgo2) and TtAgo. Probability of the number of hydrogen bonds between PAZ and L1, obtained from the equilibrium population of Markov-State-Models for hAgo2 and reweighted from BEMD for TtAgo. Representative structures of the L1-PAZ hinge are visualized with PAZ (green), L1 (ice-blue), hydrogen bonds between them (orange), and relevant residues: polar (cyan), positively charged (blue), negatively charged (red).

for hAgo2, the bottleneck for channel exposure is the proximity between PAZ and these PIWI loops. In contrast, for TtAgo without the loops, channel exposure is directly determined by the distance from its PAZ to the nonloop regions of PIWI/N/MID.

More notably, θ_{PAZ} for TtAgo spans a wide range (−1.6−1.0 rad), significantly wider than hAgo2 (−0.6−1.2 rad, see Figure 7B). This indicates that PAZ of TtAgo is considerably more flexible than that of hAgo2 in terms of twisting motions.

4. Rigidity of L1-PAZ hinge controls guide-strand loading. The clear contrast in the twisting motions of PAZ between the two Ago's may originate from different rigidity of the hinge region between their L1 and PAZ. Figure 8 presents, for both Ago's, the probability distribution of hydrogen bonds (H-bonds) between L1 and PAZ, and representative snapshots of the L1-PAZ hinge. As expected, the number of H-bonds between L1 and PAZ of hAgo2 (peak at 9) is considerably larger than that of TtAgo (peak at 2). Moreover, hAgo2 contains two sets of polar contacts on the L1-PAZ interface that do not exist in TtAgo. Because of these two sets of polar contacts (Arg196, Glu261 and Tyr225, Gln228, Glu232, Glu236), PAZ in hAgo2 is largely fixed on the tip of L1 and tends to move concertedly with L1 rather than as independently as in TtAgo.

The difference in rigidity of the L1-PAZ hinge explains why the binding channel of hAgo2 is more exposed to the guide strand than that of TtAgo. For hAgo2, due to its rigid L1-PAZ hinge, exposure of the binding channel is mainly regulated by the motion of L1 and the structure of the two PIWI loops[37] (see a cartoon model in Figure 9). As long as the PIWI loops do not form contacts with PAZ, the binding channel is accessible to the guide. In TtAgo, however, the L1-PAZ hinge is so flexible that, regardless of the motion of L1, PAZ can easily fall from the tip of L1. Given the abundance of charged residues on the surface of Ago proteins, once fallen toward its PIWI, MID and N counterparts, PAZ of TtAgo forms direct contacts with them including a number of salt-bridges (SI Figure 5), tightly blocking the entrance for the guide toward the binding channel.

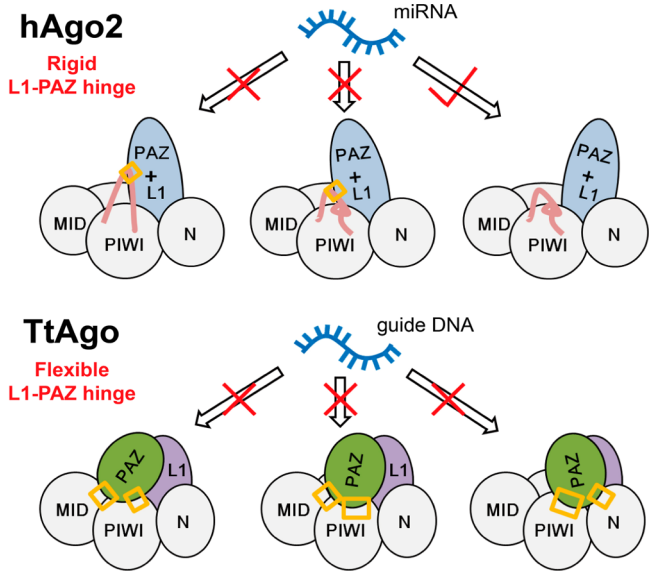


Figure 9. Cartoon models of guide-strand loading for hAgo2 and TtAgo. For hAgo2 whose L1-PAZ hinge is relatively rigid, PAZ moves concertedly with L1 (both colored light blue) and rarely forms direct contacts with N or MID. The exposure of the nucleic-binding-channel is regulated by the long PIWI loops (pink) that can form direct contacts (orange squares) with PAZ. For TtAgo, superflexibility of its L1-PAZ hinge enables large-scale motions of PAZ (green) with respect to L1 (purple). As a result, PAZ is able to form direct contacts with other domains, closing the nucleic-acid binding channel.

IV. DISCUSSION

In the present study, we propose an IF dominant guide-strand loading mechanism for TtAgo based on the free energy landscape of apo TtAgo. Furthermore, through comparison of the conformational landscape of TtAgo and previously studied hAgo2,[37] we identify an essential physical basis underlying the guide-strand loading for Ago in general—the rigidity of the L1-PAZ hinge. It is interesting to note that this difference in the rigidity of the hinge region also exists in various other Ago's from different organisms. For example, Yeast Ago from *Kluyveromyces polysporus* (KpAgo, PDB 4F1N) and human Ago1 (hAgo1, PDB 4KXT) have a similar L1-PAZ interface as hAgo2: two sets of polar contacts can restrain the PAZ motion with respect to L1 from two directions (see right top of Figure 10). In contrast, similar to TtAgo, the Ago's of two other thermophilic bacteria, *Asquifex aeolicus* (AaAgo, PDB 1YVU) and *Pyrococcus furisus* (PfAgo, 1U04) do not possess such polar

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

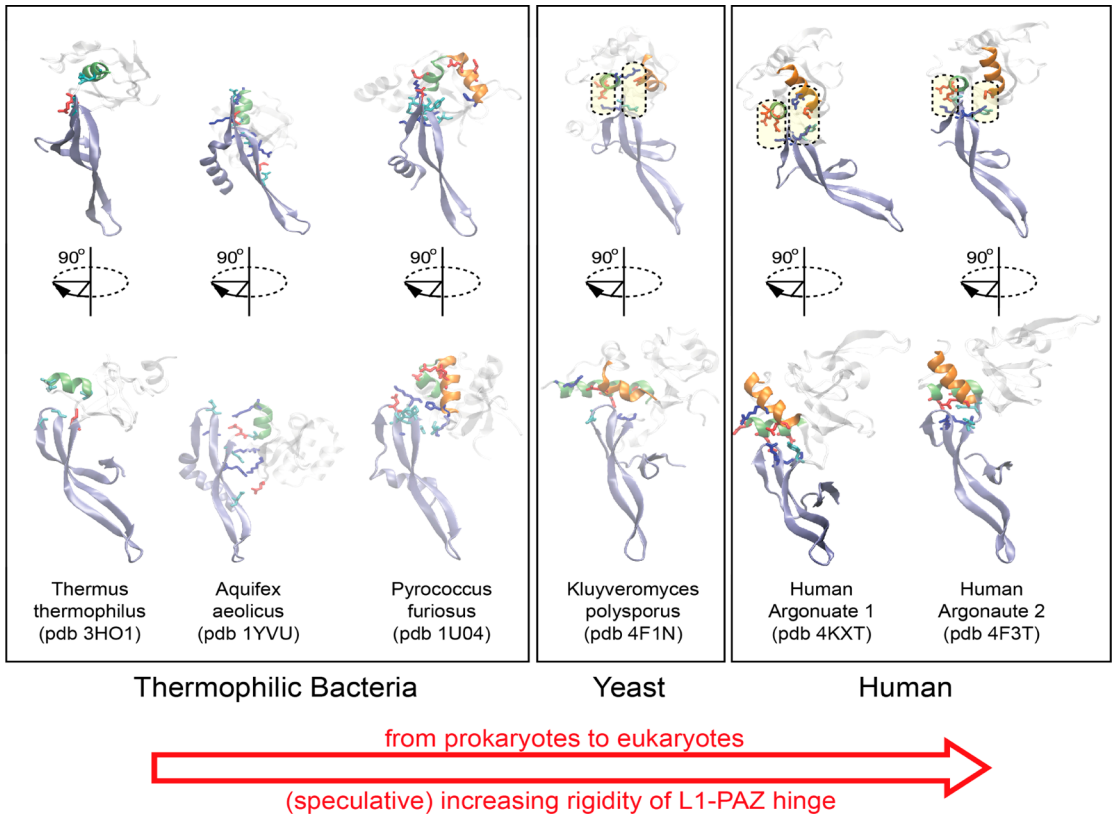


Figure 10. Comparison of L1-PAZ hinge among known crystal Ago structures from 5 species. In hAgo2 (upper right), two helices of PAZ (green and orange) can form two networks of polar contacts with L1 (ice-blue). Residues involved in the contacts are colored. Positively charged (blue), negatively charged (red), polar (cyan). These contacts restrain PAZ to L1 from both directions (yellow bars with dashed outline). The crystal structure of yeast KpAgo shares a similar setup. Unlike eukaryotic Ago, Ago of thermophilic bacteria have no obvious potency for polar contacts between L1 and PAZ. PfAgo PAZ contains two helices but is unlikely to establish polar contacts with L1 because of the position of the polar and charged residues.

contacts at the L1-PAZ hinge. These observations may suggest a distinction of the L1-PAZ hinge rigidity and therefore different guide-strand loading mechanisms between thermophilic bacterial Ago's and eukaryotic Ago's. Living at high temperature, thermophilic bacteria might need a flexible L1-PAZ hinge to make the guide-strand binding channel of its Ago less exposed.

From a physical point of view, a flexible L1-PAZ hinge in TtAgo means the motions of its L1 and PAZ are barely coupled. In contrast, hAgo2 with a rigid L1-PAZ hinge shall witness more correlated motions between its L1 and PAZ. Although such correlation in hAgo2 decays in the long time limit according to the recently discovered aging laws for domain-domain motions of proteins,[85,86] the decaying correlated motions of hAgo2 could mean a more continuous spectrum of tendency (and possibly tunability) for hAgo2 to expose its nucleic-acid binding channel. One speculates that such a continuous spectrum might be particularly necessary for eukaryotes but not required for thermophilic bacteria.

Despite its role in guide-loading, the rigidity of the L1-PAZ hinge may be relevant in other processes of Ago. For example, as suggested in a previous work,[56] a flexible L1-PAZ hinge may enable TtAgo to translate the mismatch or bulge in the target-guide duplex into the large-scale motion of PAZ and therefore play roles in the target recognition. To our knowledge, the existing discussion on the evolution of Ago has largely focused on the nucleic-acid binding pockets in MID, PAZ, and the active site on PIWI and thus the ability of Ago to cleave the target strand.[13] Our finding on the L1-PAZ hinge rigidity may provide another dimension for the classification of Ago proteins and the understanding of their evolution.

The IF dominant mechanism we propose for the guide-strand loading in TtAgo may also be adopted by other bacterial Ago's. One possible explanation for this is that IF could lead to higher specificity and fidelity in a recognition process.[32] We speculate that bacterial Ago's might be required to load their DNAs with higher specificity and fidelity than eukaryotic Ago's and prefer an IF dominant guide-strand loading. In eukaryotes, the exportation of endogenous RNAs out of the nuclear pore is regulated and could induce sufficient specificity in the RNAs before reaching Ago. Therefore, in the cytoplasm, eukaryotic Ago's do not have to load guide-strand with high fidelity. In bacterial cells, because of the absence of the nucleus and therefore the possible regulation during nuclear transport, it could be necessary for Ago's to load their own DNAs with significant fidelity.

The present work has not explicitly considered the possible conformational changes of the guide-strand DNA due to the difficulty in finding optimal sets of CVs *a. priori* for BEMD to describe the complex loading process of DNA into TtAgo. However, our results rule out a pure CS mechanism as the initial loading-step and reveal the remarkable complexity of the overall loading process. The sampled TtAgo conformations in our study could seed a further flexible docking procedure or bidirectional pulling simulations[87] to search for the encounter complex of TtAgo and guide DNA that may be validated by

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

fluorescence resonance energy transfer studies (FRET)[22] or nuclear magnetic resonance (NMR) experiments.[88] The exact pathways of the guide-loading process may further be explored using algorithms such as the string method.[89−91]

V. CONCLUSIONS

Using a computational approach combining molecular dynamics, enhanced sampling, and protein−DNA docking, we explored the guide-strand loading mechanism in bacterium *Thermus thermophilus* Ago (TtAgo). We found that apo TtAgo adopts multiple closed states unable to accommodate guide-DNA and that conformations able to do so are beyond the reach of thermal fluctuations from the closed states. These results suggest an induced-fit dominant mechanism for guide-strand loading in TtAgo, drastically different from the two-step mechanism for human Ago 2 (hAgo2) identified in our previous study. Such a difference between TtAgo and hAgo2 was found to mainly originate from the distinct rigidity of their L1-PAZ hinge. Further comparison of known Ago structures from various species indicated that the L1-PAZ hinge might be flexible in general for prokaryotic Ago's but rigid for eukaryotic Ago's.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b12426.

Direct contacts between the PAZ and PIWI domains in the binary crystal of TtAgo; Unbiased Molecular Dynamics simulation of apo TtAgo started from four crystal structures; Definition of $d_{PAZ-PIWI}$, the distance between the PAZ and PIWI tips; Comparison of the PIWI domain between hAgo2 and TtAgo; Free energy surface of $d_{PAZ-PIWI}$ and number of salt-bridges between PAZ and other domains for TtAgo and hAgo2; and Unambiguous distance restraints for HADDOCK simulations. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: xuhuihuang@ust.hk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

X.H. acknowledges National Basic Research Program of China (973 program 2013CB834703), National Science Foundation of China (No. 21273188), and Hong Kong Research Grants Council (609813, 16304215, 16302214, and AoE/M 09/12). This research made use of the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology. X.C. and X.G. were supported by grants from King Abdullah University of Science & Technology.

■ REFERENCES

(1) Pratt, A. J.; MacRae, I. J. The RNA-Induced Silencing Complex: A Versatile Gene-Silencing Machine. *J. Biol. Chem.* **2009**, *284*, 17897−17901.

(2) Obbard, D. J.; Gordon, K. H. J.; Buck, A. H.; Jiggins, F. M. The Evolution of RNAi as a Defence against Viruses and Transposable Elements. *Philos. Trans. R. Soc., B* **2009**, *364*, 99−115.

(3) Kawamata, T.; Tomari, Y. Making RISC. *Trends Biochem. Sci.* **2010**, *35*, 368−376.

(4) Hur, J. K.; Olovnikov, I.; Aravin, A. A. Prokaryotic Argonautes Defend Genomes against Invasive DNA. *Trends Biochem. Sci.* **2014**, *39*, 257−259.

(5) Swarts, D. C.; Jore, M. M.; Westra, E. R.; Zhu, Y.; Janssen, J. H.; Snijders, A. P.; Wang, Y.; Patel, D. J.; Berenguer, J.; Brouns, S. J. J.; van der Oost, J. DNA-Guided DNA Interference by a Prokaryotic Argonaute. *Nature* **2014**, *507*, 258−261.

(6) Sasaki, H. M.; Tomari, Y. The True Core of RNA Silencing Revealed. *Nat. Struct. Mol. Biol.* **2012**, *19*, 657−660.

(7) Wang, Y.; Juranek, S.; Li, H.; Sheng, G.; Wardle, G. S.; Tuschl, T.; Patel, D. J. Nucleation, Propagation and Cleavage of Target RNAs in Ago Silencing Complexes. *Nature* **2009**, *461*, 754−761.

(8) Wang, Y.; Juranek, S.; Li, H.; Sheng, G.; Tuschl, T.; Patel, D. J. Structure of an Argonaute Silencing Complex with a Seed-Containing Guide DNA and Target RNA Duplex. *Nature* **2008**, *456*, 921−926.

(9) Wang, Y.; Sheng, G.; Juranek, S.; Tuschl, T.; Patel, D. J. Structure of the Guide-Strand-Containing Argonaute Silencing Complex. *Nature* **2008**, *456*, 209−213.

(10) Esquela-Kerscher, A.; Trang, P.; Wiggins, J. F.; Patrawala, L.; Cheng, A.; Ford, L.; Weidhaas, J. B.; Brown, D.; Bader, A. G.; Slack, F. J. The Let-7 MicroRNA Reduces Tumor Growth in Mouse Models of Lung Cancer. *Cell Cycle* **2008**, *7*, 759−764.

(11) Trang, P.; Medina, P. P.; Wiggins, J. F.; Ruffino, L.; Kelnar, K.; Omotola, M.; Homer, R.; Brown, D.; Bader, A. G.; Weidhaas, J. B.; Slack, F. J. Regression of Murine Lung Tumors by the Let-7 MicroRNA. *Oncogene* **2010**, *29*, 1580−1587.

(12) Broderick, J. A.; Zamore, P. D. MicroRNA Therapeutics. *Gene Ther.* **2011**, *18*, 1104−1110.

(13) Swarts, D. C.; Makarova, K.; Wang, Y.; Nakanishi, K.; Ketting, R. F.; Koonin, E. V.; Patel, D. J.; van der Oost, J. The Evolutionary Journey of Argonaute Proteins. *Nat. Struct. Mol. Biol.* **2014**, *21*, 743−753.

(14) Riley, K. J.; Yario, T. A.; Steitz, J. A. Association of Argonaute Proteins and MicroRNAs Can Occur after Cell Lysis. *RNA* **2012**, *18*, 1581−1585.

(15) Deerberg, A.; Willkomm, S.; Restle, T. Minimal Mechanistic Model of siRNA-Dependent Target RNA Slicing by Recombinant Human Argonaute 2 Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17850−17855.

(16) Nakanishi, K.; Weinberg, D. E.; Bartel, D. P.; Patel, D. J. Structure of Yeast Argonaute with Guide RNA. *Nature* **2012**, *486*, 368−374.

(17) Elkayam, E.; Kuhn, C.-D.; Tocilj, A.; Haase, A. D.; Greene, E. M.; Hannon, G. J.; Joshua-Tor, L. The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell* **2012**, *150*, 100−110.

(18) Schirle, N. T.; MacRae, I. J. The Crystal Structure of Human Argonaute2. *Science* **2012**, *336*, 1037−1040.

(19) Wilson, R. C.; Doudna, J. A. Molecular Mechanisms of RNA Interference. *Annu. Rev. Biophys.* **2013**, *42*, 217−239.

(20) Doyle, M.; Jaskiewicz, L.; Filipowicz, W. *Eukaryotic RNases and Their Partners in RNA Degradation and Biogenesis, Part B*; Elsevier: 2012; Vol. 32, pp 1−35.

(21) Kwak, P. B.; Tomari, Y. The N Domain of Argonaute Drives Duplex Unwinding During RISC Assembly. *Nat. Struct. Mol. Biol.* **2012**, *19*, 145−151.

(22) Jung, S.-R.; Kim, E.; Hwang, W.; Shin, S.; Song, J.-J.; Hohng, S. Dynamic Anchoring of the 3′-End of the Guide Strand Controls the Target Dissociation of Argonaute-Guide Complex. *J. Am. Chem. Soc.* **2013**, *135*, 16865−16871.

(23) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35*, 539−546.

(24) Hammes, G. G.; Chang, Y. C.; Oas, T. G. Conformational Selection or Induced Fit: A Flux Description of Reaction Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13737−13741.

(25) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98−104.

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

(26) Ma, B. Y.; Kumar, S.; Tsai, C. J.; Nussinov, R. Folding Funnels and Binding Mechanisms. *Protein Eng., Des. Sel.* **1999**, *12*, 713−720.
(27) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598−1603.
(28) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *305*, 386−389.
(29) Kalodimos, C. G.; Boelens, R.; Kaptein, R. Toward an Integrated Model of Protein−DNA Recognition as Inferred from NMR Studies on the Lac Repressor System. *Chem. Rev.* **2004**, *104*, 3567−3586.
(30) Longo, A.; Guanga, G. P.; Rose, R. B. Structural Basis for Induced Fit Mechanisms in DNA Recognition by the Pdx1 Homeodomain. *Biochemistry* **2007**, *46*, 2948−2957.
(31) Petty, T. J.; Emamzadah, S.; Costantino, L.; Petkova, I.; Stavridi, E. S.; Saven, J. G.; Vauthey, E.; Halazonetis, T. D. An Induced Fit Mechanism Regulates p53 DNA Binding Kinetics to Confer Sequence Specificity. *EMBO J.* **2011**, *30*, 2167−2176.
(32) Savir, Y.; Tlusty, T.; Scalas, E. Conformational Proofreading: The Impact of Conformational Changes on the Specificity of Molecular Recognition. *PLoS One* **2007**, *2*, e468.
(33) Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic Dynamics of Enzymes in the Unbound State and Relation to Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633−640.
(34) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−796.
(35) Zhou, H. X. From Induced Fit to Conformational Selection: A Continuum of Binding Mechanism Controlled by the Timescale of Conformational Transitions. *Biophys. J.* **2010**, *98*, L15−17.
(36) Greives, N.; Zhou, H. X. Both Protein Dynamics and Ligand Concentration Can Shift the Binding Mechanism between Conformational Selection and Induced Fit. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10197−10202.
(37) Jiang, H.; Sheong, F. K.; Zhu, L.; Gao, X.; Bernauer, J.; Huang, X. Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding Followed by Structural Re-Arrangement. *PLoS Comput. Biol.* **2015**, *11*, e1004404.
(38) Silva, D. A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. H. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the Lao Protein. *PLoS Comput. Biol.* **2011**, *7*, e1002054.
(39) Gu, S.; Silva, D. A.; Meng, L. M.; Yue, A.; Huang, X. H. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003767.
(40) Silva, D. A.; Weiss, D. R.; Avila, F. P.; Da, L. T.; Levitt, M.; Wang, D.; Huang, X. H. Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7665−7670.
(41) Da, L. T.; Wang, D.; Huang, X. Dynamics of Pyrophosphate Ion Release and Its Coupled Trigger Loop Motion from Closed to Open State in RNA Polymerase II. *J. Am. Chem. Soc.* **2012**, *134*, 2399−2406.
(42) Huang, X.; Yao, Y.; Bowman, G. R.; Sun, J.; Guibas, L. J.; Carlsson, G.; Pande, V. S. Constructing Multi-Resolution Markov State Models (MSMs) to Elucidate RNA Hairpin Folding Mechanisms. *Pac. Symp. Biocomput.* **2009**, 228−239.
(43) Bowman, G. R.; Huang, X.; Pande, V. S. Using Generalized Ensemble Simulations and Markov State Models to Identify Conformational States. *Methods* **2009**, *49*, 197−201.
(44) Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. Rapid Equilibrium Sampling Initiated from Nonequilibrium Data. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19765−19769.
(45) Gfeller, D.; De Los Rios, P.; Caflisch, A.; Rao, F. Complex Network Analysis of Free-Energy Landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1817−1822.
(46) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.
(47) Buchete, N. V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.
(48) Pan, A. C.; Roux, B. Building Markov State Models Along Pathways to Determine Free Energies and Rates of Transitions. *J. Chem. Phys.* **2008**, *129*, 064107.
(49) Bowman, G. R.; Voelz, V. A.; Pande, V. S. Taming the Complexity of Protein Folding. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4−11.
(50) Chodera, J. D.; Noe, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144.
(51) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular Simulation of Ab Initio Protein Folding for a Millisecond Folder Ntl9(1−39). *J. Am. Chem. Soc.* **2010**, *132*, 1526−1528.
(52) Razavi, A. M.; Wuest, W. M.; Voelz, V. A. Computational Screening and Selection of Cyclic Peptide Hairpin Mimetics by Molecular Simulation and Kinetic Network Models. *J. Chem. Inf. Model.* **2014**, *54*, 1425−1432.
(53) de Vries, S. J.; van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. HADDOCK Versus HADDOCK: New Features and Performance of HADDOCK2.0 on the Capri Targets. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 726−733.
(54) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.
(55) Xia, Z.; Clark, P.; Huynh, T.; Loher, P.; Zhao, Y.; Chen, H.-W.; Ren, P.; Rigoutsos, I.; Zhou, R. Molecular Dynamics Simulations of Ago Silencing Complexes Reveal a Large Repertoire of Admissible 'Seed-Less' Targets. *Sci. Rep.* **2012**, *2*, 569.
(56) Xia, Z.; Huynh, T.; Ren, P.; Zhou, R. Large Domain Motions in Ago Protein Controlled by the Guide DNA-Strand Seed Region Determine the Ago-DNA-mRNA Complex Recognition Process. *PLoS One* **2013**, *8*, e54620.
(57) Wang, Y.; Li, Y.; Ma, Z.; Yang, W.; Ai, C. Mechanism of MicroRNA-Target Interaction: Molecular Dynamics Simulations and Thermodynamics Analysis. *PLoS Comput. Biol.* **2010**, *6*, e1000866.
(58) Beshnova, D. A.; Cherstvy, A. G.; Vainshtein, Y.; Teif, V. B. Regulation of the Nucleosome Repeat Length in Vivo by the DNA Sequence, Protein Concentrations and Long-Range Interactions. *PLoS Comput. Biol.* **2014**, *10*, e1003698.
(59) Cherstvy, A. G.; Teif, V. B. Electrostatic Effect of H1-Histone Protein Binding on Nucleosome Repeat Length. *Phys. Biol.* **2014**, *11*, 044001.
(60) Chandradoss, S. D.; Schirle, N. T.; Szczepaniak, M.; MacRae, I. J.; Joo, C. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **2015**, *162*, 96−107.
(61) Salomon, W. E.; Jolly, S. M.; Moore, M. J.; Zamore, P. D.; Serebrov, V. Single-Molecule Imaging Reveals That Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* **2015**, *162*, 84−95.
(62) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2006**, *Chapter 5*, Unit 5.6, p5.6.110.1002/0471250953.bi0506s15
(63) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(64) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.
(65) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(66) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720

(67) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(68) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(69) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(70) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. Plumed: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(71) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(72) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(73) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(74) Biarnés, X.; Pietrucci, F.; Marinelli, F.; Laio, A. METAGUI. A VMD Interface for Analyzing Metadynamics and Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2012**, *183*, 203−211.
(75) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5*, e1000452.
(76) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(77) Bartels, C. Analyzing Biased Monte Carlo and Molecular Dynamics Simulations. *Chem. Phys. Lett.* **2000**, *331*, 446−454.
(78) Jorgensen, W. L.; Tiradorives, J. The OPLS Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(79) Wang, Y.; Markwick, P. R. L.; de Oliveira, C. A. F.; McCammon, J. A. Enhanced Lipid Diffusion and Mixing in Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2011**, *7*, 3199−3207.
(80) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(81) Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A. Implementation of Accelerated Molecular Dynamics in NAMD. *Comput. Sci. Discovery* **2011**, *4*, 015002.
(82) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(83) Gao, Y. Q. An Integrate-over-Temperature Approach for Enhanced Sampling. *J. Chem. Phys.* **2008**, *128*, 064105.
(84) Galvelis, R.; Sugita, Y. Replica State Exchange Metadynamics for Improving the Convergence of Free Energy Estimates. *J. Comput. Chem.* **2015**, *36*, 1446−1455.
(85) Hu, X.; Hong, L.; Dean Smith, M.; Neusius, T.; Cheng, X.; Smith, J. C. The Dynamics of Single Protein Molecules Is Non-Equilibrium and Self-Similar over Thirteen Decades in Time. *Nat. Phys.* **2015**, *12*, 171−174.
(86) Metzler, R. Protein Physics: Forever Ageing. *Nat. Phys.* **2015**, *12*, 113−114.
(87) Do, T. N.; Carloni, P.; Varani, G.; Bussi, G. RNA/Peptide Binding Driven by Electrostatics—Insight from Bidirectional Pulling Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1720−1730.
(88) Zhu, G. *NMR of Proteins and Small Biomolecules*; Springer: Berlin Heidelberg, 2011.
(89) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
(90) Maragliano, L.; Vanden-Eijnden, E. On-the-Fly String Method for Minimum Free Energy Paths Calculation. *Chem. Phys. Lett.* **2007**, *446*, 182−190.
(91) Pan, A. C.; Sezer, D.; Roux, B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432−3440.

DOI: 10.1021/acs.jpcb.5b12426
J. Phys. Chem. B 2016, 120, 2709−2720