

## Critical role of backbone coordination in the mRNA recognition by RNA induced silencing complex

Lizhe Zhu<sup>1,2</sup>, Hanlun Jiang<sup>3,4</sup>, Siqin Cao<sup>2,5</sup>, Ilona Christy Unarta<sup>3,5</sup>, Xin Gao<sup>6</sup> <sup>✉</sup> & Xuhui Huang<sup>2,3,5</sup> <sup>✉</sup>

Despite its functional importance, the molecular mechanism underlying target mRNA recognition by Argonaute (Ago) remains largely elusive. Based on extensive all-atom molecular dynamics simulations, we constructed quasi-Markov State Model (qMSM) to reveal the dynamics during recognition at position 6-7 in the seed region of human Argonaute 2 (hAgo2). Interestingly, we found that the slowest mode of motion therein is not the gRNA-target base-pairing, but the coordination of the target phosphate groups with a set of positively charged residues of hAgo2. Moreover, the ability of Helix-7 to approach the PIWI and MID domains was found to reduce the effective volume accessible to the target mRNA and therefore facilitate both the backbone coordination and base-pair formation. Further mutant simulations revealed that alanine mutation of the D358 residue on Helix-7 enhanced a trap state to slow down the loading of target mRNA. Similar trap state was also observed when wobble pairs were introduced in g6 and g7, indicating the role of Helix-7 in suppressing non-canonical base-pairing. Our study pointed to a general mechanism for mRNA recognition by eukaryotic Agos and demonstrated the promise of qMSM in investigating complex conformational changes of biomolecular systems.

<sup>1</sup>Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, China. <sup>2</sup>Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. <sup>3</sup>Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. <sup>4</sup>Department of Biochemistry, Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. <sup>5</sup>Center of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. <sup>6</sup>Computational Bioscience Research Center, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. ✉email: [xin.gao@kaust.edu.sa](mailto:xin.gao@kaust.edu.sa); [xuhuihuang@ust.hk](mailto:xuhuihuang@ust.hk)

Small non-coding RNAs (miRNA, siRNA) are critical for post-transcriptional regulation of human gene expression<sup>1–4</sup>. These RNAs are loaded into the Argonaute (Ago) protein, forming the RNA-induced silencing complex (RISC) which recognizes and inhibits target messenger RNA (target mRNA) in a highly sequence specific manner<sup>5–12</sup>. RISC regulates over 50% of human genes<sup>13,14</sup> and is involved in numerous normal physiological functions<sup>15,16</sup> and disease progression including cancer<sup>17,18</sup>. Accordingly, elucidating the mechanism of the recognition of target mRNA by RISC is expected to inspire the development of next-generation RNA-based therapeutics for cancer and other human diseases<sup>19–21</sup>.

Recent structural and single-molecule studies have pointed to a step-wise model<sup>22–26</sup> of the recognition between target mRNA and the RISC complex formed by gRNA and the best studied human Argonaute human Ago 2 (hAgo2)<sup>13,27–31</sup>. The recognition initiates with the base-pairing process between the gRNA and target mRNA at position 2–8 (g2–g8) known as the seed region<sup>14,27</sup> (Fig. 1). RISC first conducts a rapid diffusion-controlled search for the target mRNA sites that are complementary to the first half of the seed region (g2–g5). Subsequently, base pairs are formed at the second half of the seed region (g6–g8), accompanied by substantial conformational changes of hAgo2 that fully expose the gRNA g6–g8 for recognition. Though for certain gRNA sequences, base-pairing at g13–g16, known as the 3' supplementary site, is also necessary, the seed base-pairing remains the most decisive step for the successful recognition and the translational repression of the target mRNA.

Existing crystal structures of the human RISC show that gRNA bases at g2–g6 are fully solvent-exposed and directly accessible to the mRNA while g7–g8 are much more buried by the PAZ and L2 domains of hAgo2<sup>28</sup>. In particular, Helix-7 (residues L356–T368 of L2 in hAgo2, conserved in eukaryotic Agos<sup>32</sup>) from L2 has direct contact with the mRNA at g6–g7 and causes a kink in the corresponding region in the gRNA–mRNA duplex<sup>28,33</sup>. This suggests that concerted motions of the PAZ and L2 are required to expose g7–g8 for the mRNA recognition, as validated in part by

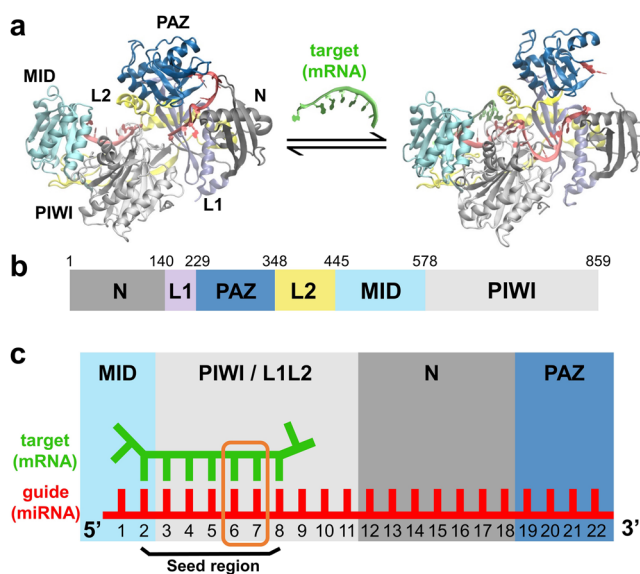
a recent biochemical and crystallographic effort<sup>33</sup>. However, due to the challenges for experimental techniques to resolve dynamics at the atomic level, fine details of the conformational dynamics of the RISC–mRNA complex, particularly hAgo2, during the seed recognition at g6–g7 have not been explicitly elucidated. Accordingly, the key protein residues that modulate the recognition process are also obscure.

Molecular dynamics (MD) simulations offer a valuable tool to investigate the conformational dynamics of large biomolecules at the atomic resolution. Previous MD studies at sub-microsecond timescales have demonstrated the impact of miRNA and double strand RNA on the conformational stability of the Ago complex<sup>34–38</sup>. However, solely using MD to study the target mRNA recognition at g6–g7 faces tremendous challenges due to the gap between the experimental timescale (at millisecond or longer) and that of MD simulations (at microsecond). The Markov state model (MSM) has been a popular framework to bridge this timescale gap<sup>39–52</sup>. In an MSM, we coarse grain both time and the conformational space into the lag-time  $\Delta t$  and a number of metastable states simultaneously, such that fast motions are integrated out. When  $\Delta t$  is longer than the intra-state relaxation time, the model becomes Markovian, i.e. the probability for the system to visit a conformational state at the next time step ( $t + \Delta t$ ) is only determined by its location at the current time step  $t$ . If the model is Markovian, we can model the long timescale dynamics using the first order master equation. In recent years, MSM has been widely applied to study conformational dynamics of molecular recognition<sup>53–59</sup> and aggregation<sup>60,61</sup>, including the mechanism of guide strand loading into hAgo2<sup>62</sup>. Since the lag-time ( $\Delta t$ ) in an MSM must be long enough to allow Markovian interstate transitions, the estimation of transition probabilities in MSMs of the slow RISC–mRNA recognition could still be limited by the upper bound in the affordable length of the MD simulations. To address this issue, we recently developed the quasi-MSM (qMSM) method based on the generalized master equation formalism, which encodes non-Markovian dynamics into memory kernel functions<sup>63</sup>. qMSM has been applied to the study of bacterial RNA-polymerase<sup>64</sup> and provides a promising approach to study the conformational changes during the RISC–mRNA recognition.

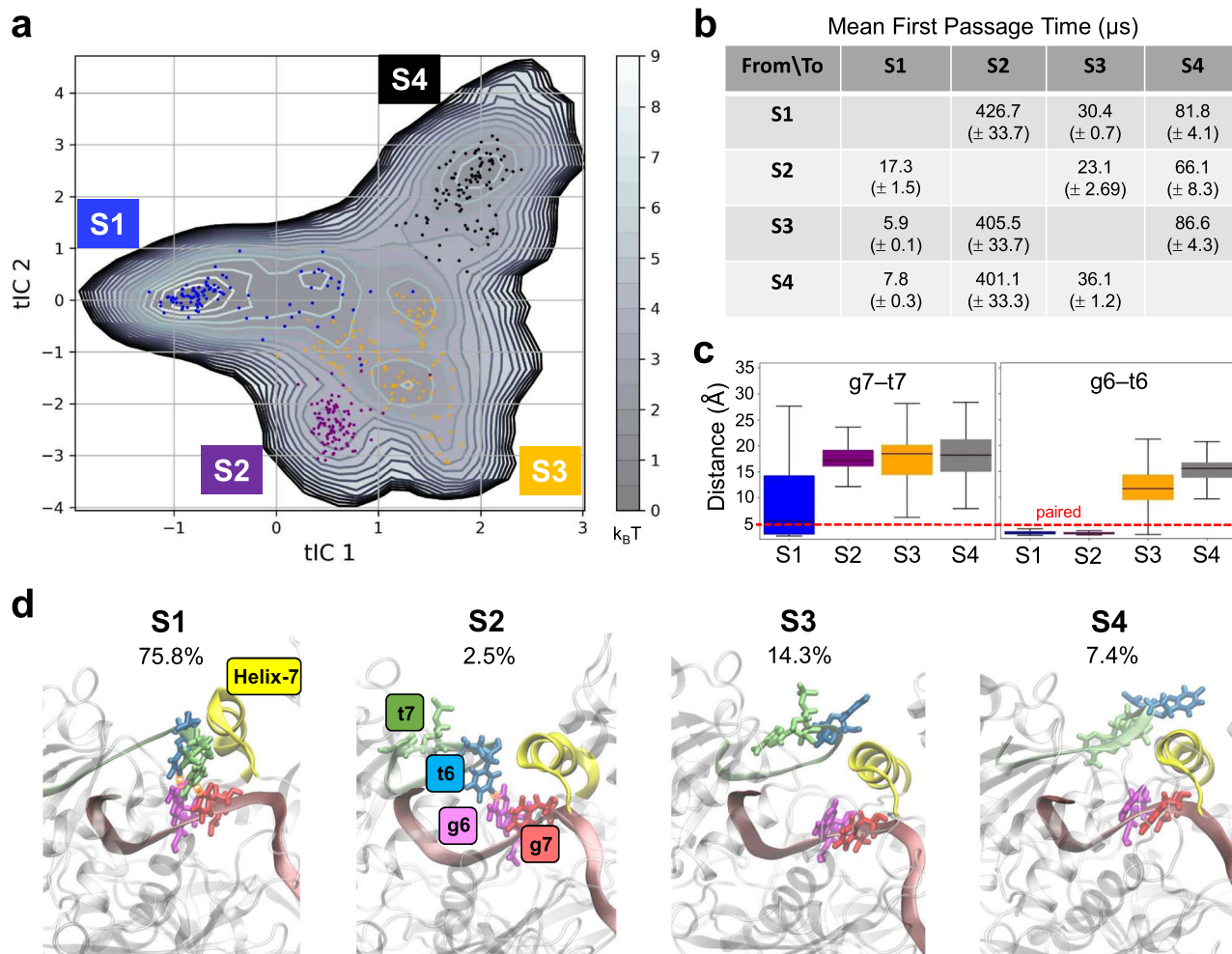
Therefore, we performed MD simulations that amounts to 84 microseconds and constructed a 4-state qMSM based on such a dataset. Our qMSM revealed that the slowest mode of motion of target mRNA recognition is not the gRNA–mRNA base-pairing, but the coordination between the phosphate groups of the mRNA and the positively charged residues of hAgo2 (R554, K550, K525, K355). Moreover, the positioning of Helix-7 is essential in facilitating this recognition: when Helix-7 approaches PIWI and MID, the effective 3D-space that needs to be explored by the target mRNA is reduced, therefore facilitating the backbone coordination and base-pairing. Further mutant simulations via metadynamics revealed that Alanine mutation of the negatively charged D358 on Helix-7 induced a conformational state that traps the system prior to the recognition at g7, which decelerated mRNA loading. A similar trap state was also observed when UG wobble-pairs are introduced at g6–g7. Altogether, our results suggest a backbone-coordination dominant and Helix-7 assisting mechanism for the target recognition by hAgo2, and thus highlight the vital role of protein–RNA interactions in this important biological process.

## Results and discussion

**qMSM reveals four conformational states during the target mRNA recognition.** Our qMSM contains four macrostates S1–S4. In Fig. 2a, we present the free energy landscape as a function of the



**Fig. 1** Function, domain decomposition and structure of the human Argonaute 2 (hAgo2). **a** Crystal structures of the RISC or hAgo2–miRNA complex (left, PDB ID: 4W5N) and RISC–mRNA complex (right, PDB ID: 4W5O). **b** Structural domains of hAgo2. **c** Cartoon illustration of target (mRNA) recognition by guide (miRNA) in hAgo2. The seed region (g2–g8) of guide is the labeled. Recognition at position g6–g7 studied in this work is highlighted in orange box.



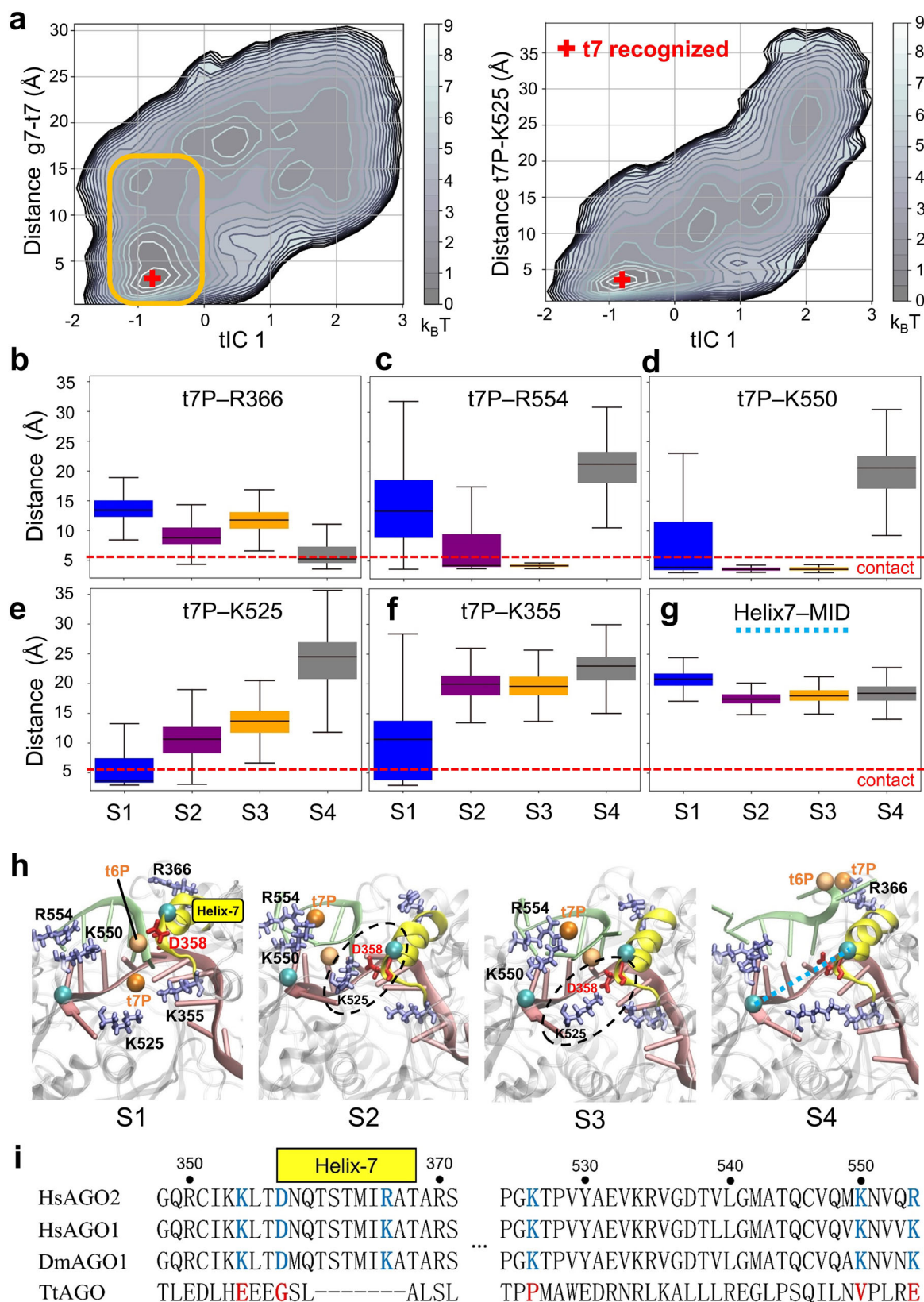
**Fig. 2** Markov State Models of target mRNA recognition at g6–g7. **a** the free energy landscape of the first two tICs. Samples from four macrostates (S1–S4) are labeled by blue, purple, orange and black, respectively. **b** The mean-first-passage-time (MFPTs) among the four macro-states. **c** Boxplot of the nucleobase distances at g6 and g7 for the four macrostates. Statistics were made on all structures in each macrostate. **d** Representative structure and the population of each macrostate. The MFPTs and population of the four macrostates are obtained from our qMSM.

first two time-lagged independent components (tICs). Obtained by the tICA analysis<sup>65,66</sup>, the tICs approximate the slowest modes of motions in our simulation data (see Methods). Macrostate S1, with a population of 75.8%, mainly corresponds to the recognized state where both base-pairs of g6–t6 and g7–t7 are formed (see statistics of the base distances in Fig. 2c) and Helix-7 moves away from MID and PIWI to accommodate the nucleotide t7 (first panel of Fig. 2d). Macrostate S3 and S4 denote two different unrecognized state (both base pairs broken, see Fig. 2c) with a population of 14.3 and 7.4% respectively. In State S3, the target mRNA stays close to MID (third panel in Fig. 2d) and the sidechain of t7 faces the solvent. In S4, the target mRNA locates far away from MID with a more extended conformation pointing to the hAgo2 protein (fourth panel in Fig. 2d). Macrostate S2, populated at 2.5%, corresponds to a metastable state where g6–t6 is formed while g7–t7 is not (see Fig. 2c and second panel in Fig. 2d).

Kinetically, the mean-first-passage-times (MFPTs) for the transitions to the mostly populated state S1 from other states (corresponding to the recognition process) are on timescales at several to tens of microseconds (see the first column of Fig. 2b), with significantly slower reverse transitions (see the first row of Fig. 2b). Meanwhile, the transitions to S2 from other states are all at  $\sim 400 \mu\text{s}$  (see the second column of Fig. 2b), significantly slower than the reverse transition (see the second row of Fig. 2b). This is consistent with the least population of S2 among all states and

indicates that S2, though with t6 recognized, is kinetically less accessible from all other states. By contrast, the unrecognized states S3 and S4 are not only more populated than S2 but also kinetically close to S1 and to themselves. The recognition processes S3-to-S1 and S4-to-S1 only take  $5.9 \mu\text{s}$  and  $7.8 \mu\text{s}$  respectively, 2–3 times faster than the S2-to-S1 transition. Meanwhile, the interconversion between S3 and S4 occur on the same timescales (tens of microseconds) as the transition from S1 to them. These results indicate that the recognition from S3/S4 to S1 takes several microseconds and the t6-recognized and the least populated state S2 is an off-pathway intermediate state.

**Backbone phosphate coordination is rate-limiting for target recognition.** To identify the slowest mode(s) of motions during target mRNA recognition, we examined the details of the RNA–RNA and protein–RNA interactions therein. As the first tIC from the tICA analysis approximates the direction of the slowest motion in our model, we projected the qMSM data on the first tIC (tIC1) and various geometric measurements. As shown in Fig. 3a, no clear correlation can be found between the tIC1 and the nucleobase distance g7–t7, since the distribution of this distance is nearly orthogonal to that of tIC1 (see the orange box in Fig. 3a) at the regions close to the recognized state (tIC1  $\sim -0.8$ ). Instead, the change of the distance between the t7 phosphate



(t7P) and the positively charged residue K525 exhibits notable correlation with tIC1 (right panel of Fig. 3a). These results suggest that it is the coordination between the nucleotide backbone phosphate with the positively charged protein residues that represents the rate-limiting step for the target mRNA recognition, rather than the formation of the base pairs at g6 and g7.

To further verify this observation, we made statistics on all MD conformations within each macrostate over the distances between t7P and all positively charged hAgo2 residues near the RNA loading channel (R366 / R554 / K550 / K525 / K355, Fig. 3b–f, definition in Table S1). Different macrostates have different phosphate-protein interactions. In macrostate S4, t7P is only in

**Fig. 3 Key protein-RNA interactions during target recognition revealed by qMSMs.** **a** Projection of the MSM data on the first tIC and the distances between the base of g7 and t7 (left), and between the phosphate of t7 (t7P) and the protein residue K525 (right). Orange box highlights the orthogonal distribution of the g7-t7 distance with respect to tIC1 near the fully recognized state. **b-g** Boxplot of various geometric features for the macrostates, including the distance between t7P and positively charged residues R366 / R554 / K550 / K525 / K355 and the distance between Helix-7 and MID domain (cyan dashed line in **c**). Atoms used for defining the distances are listed in SI Table S1. Statistics were made on all structures in each macrostate. **h** Representative structures of the four macro-states. The backbone phosphate t7P (orange sphere), the positively charged residues (ice-blue), Helix-7 (yellow), particularly the negatively charged D358 (red) on Helix-7 are highlighted. Cyan spheres in the S4 structure (right) are the C $\alpha$  atoms of N359 and Q527 used to define the Helix-7-MID distance. Black dashed ellipses highlight the contact between D358 and K525 in S2/S3. **i** Sequence alignment for hAgo2, hAgo1, DmAgo1 and TtAgo. K355 / D358 / R366 / K525 / K550 / R554 are conserved in eukaryotic Agos.

contact with R366 (Fig. 3b and fourth panel of Fig. 3h), whereas in S3, t7P is coordinated with R554 and K550 (Fig. 3c, d and third panel of Fig. 3h). In macrostate S2, t7P also forms contact with R554/K550 (second panel of Fig. 3h) but the R554 contact appears weaker than in state S3 (wider distribution in Fig. 3c). In the recognized state S1 (first panel of Fig. 3h), t7P turns to coordinate with K525 (Fig. 3e) and sometimes with K355 (Fig. 3f). Altogether, these results suggest that the transfer of t7P from R554/K550 or R366 to K525 appears a decisive step for target recognition (the S3-to-S1 and S4-to-S1 transitions). In fact, visual inspection of a MD trajectory where near-complete recognition is observed (Movie S1) shows that during recognition, t7P coordinates with the positively charged residues in a sequential manner R554- > R550- > K525- > K355. Projection of this trajectory on the first two tICs can be found in Fig. S1a. Detailed order of events can be found in Table S2 and Fig. S2 (see SI Note 1 for more discussion). Interestingly, all of these identified residues are highly conserved in eukaryotic Agos (see sequence alignment of hAgo2, hAgo1, DmAgo1 and TtAgo in Fig. 3i).

**Helix-7 facilitates target nucleotide positioning before recognition.** As crystallographic studies have documented Helix-7 to pose a steric barrier for target mRNA recognition beyond g5<sup>28,33</sup>, we measured the distance between Helix-7 and the MID domain to investigate the role of Helix-7 in the recognition process (see definition in dashed cyan line in the fourth panel of Fig. 3h and Table S1). As expected, Helix-7 is 21.5 Å away from MID in the recognized state S1, significantly more distant than the 17–18.5 Å in state S2–S4 (Fig. 3g). Interestingly, in states S2 and S3, the negatively charged residue D358 on Helix-7 is even able to form a salt-bridge with K525 to keep Helix-7 proximal to MID/PIWI, partially closing the RNA-loading channel before recognition. As D358 is also highly conserved in eukaryotic Agos<sup>32</sup> (Fig. 3i), we anticipate that the proximity between Helix-7 and MID/PIWI may play a pre-requisite role for the target mRNA recognition.

Given the aforementioned importance of phosphate coordination, we further hypothesized that the proximity of Helix-7 (e.g. driven by the salt-bridge between D358 in Helix-7 and K525 of hAgo2) could reduce the effective conformational space for t6 and t7 to explore and therefore facilitating the t7P coordination with K525 and the base-pairing. To examine this hypothesis, we designed two protein mutants K525A and D358A that disrupt the K525-D358 salt-bridge. In particular, we performed 1  $\mu$ s PCV-MetaD simulation for each mutant system to examine their impact on the target mRNA recognition. We identified one MD trajectory, in which a near-complete loading process is observed (see Fig. S1a and Movie S1) and extracted 36 structures from this MD trajectory to form a reference path. We then defined a PCV on this path and performed 1  $\mu$ s PCV-MetaD biasing on PCV-s and PCV-z for the wild-type (WT) protein and the two mutants (see Methods for details). In all the MetaD, the recognized state was revisited after complete unloading at least twice (Fig. S3), indicating sufficient sampling.

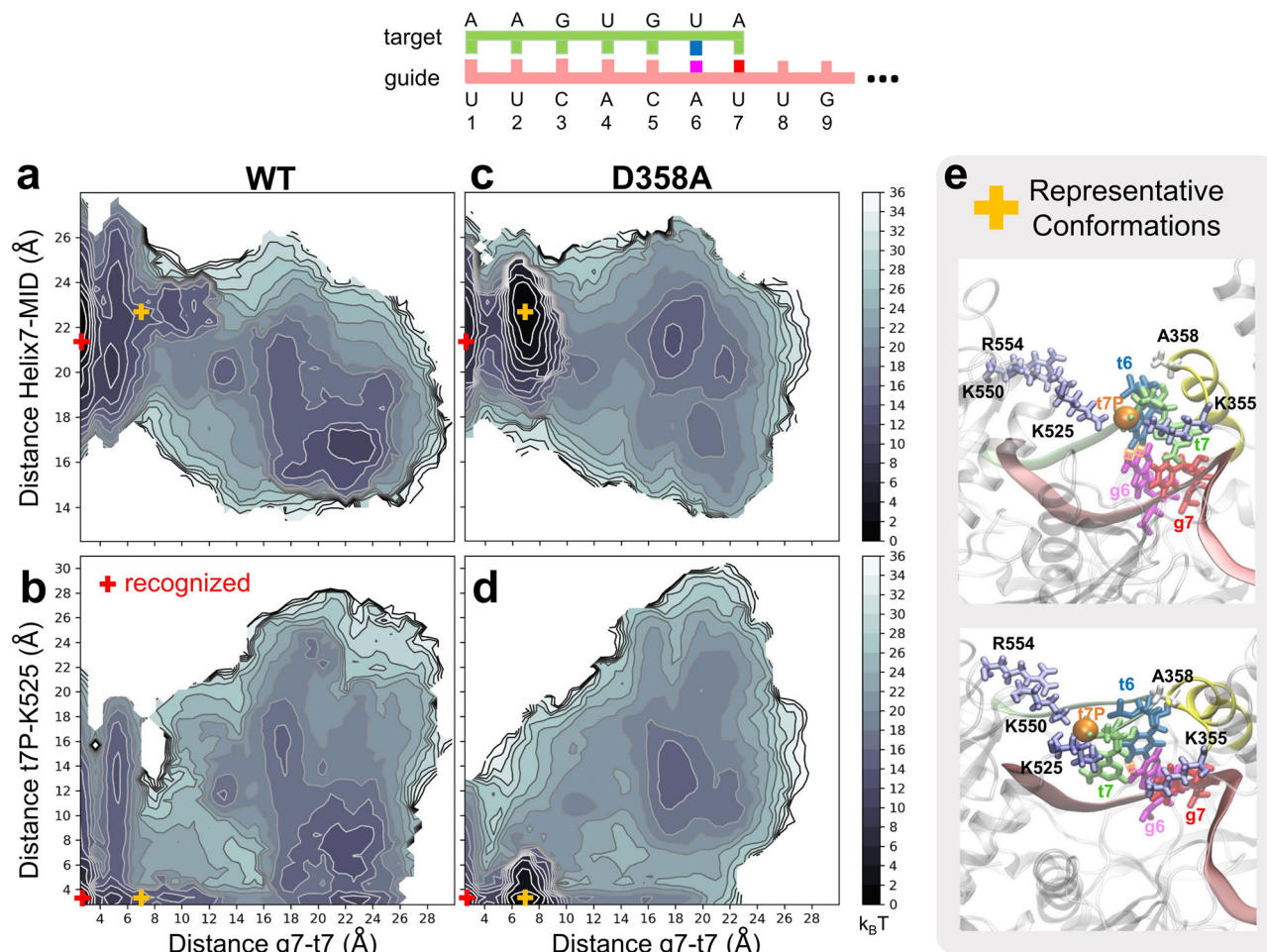
In Fig. 4a–d, we illustrate, for WT and the D358A mutant, the reweighted free energy landscape from the PCV-MetaD simulations as functions of three distances: (i) the Helix7-MID distance (*y*-axis of Fig. 4a, c) measuring the proximity of Helix-7 to MID/PIWI; (ii) the t7P-K525 distance (*y*-axis of Fig. 4b, d) that is statistically correlated to tIC1 (right panel of Fig. 3a); (iii) the g7-t7 base distance (*x*-axis in Fig. 4a–d) measuring the progress of t7-recognition (fully unrecognized >10 Å, in the recognition process 4.5–10 Å, fully recognized 3–4.5 Å). The results of WT PCV-MetaD simulations are qualitatively consistent with our qMSM data (Fig. S4).

For D358A, as expected, Helix-7 is distant (21–22 Å) from MID throughout the recognition process (Fig. 4c). As a result, a new trap metastable state that does not exist for WT emerged in D358A (see the orange cross in Fig. 4a–d and their conformations in Fig. 4e) right before the full recognition (g7-t7 distance at 7 Å) and therefore slowed down the recognition for D358A. We note that this new trap state is structurally distinct from any of the macrostates S1–S4 from qMSM of the WT system. In this trap state, Helix-7 is 23 Å away from MID (Fig. 4c), the t7P-K525 contact is formed (*y*-axis of Fig. 4d), and base-pairing at g6 is formed (Fig. 4e). Similar trap state was also observed for the other mutant of K525A that disrupts the K525-D358 salt-bridge, due to the discouraged proximity of Helix-7 to MID/PIWI (Fig. S5). But the trap state in this mutant is less populated because A525 is no longer able to form stable contact with t7P.

Altogether, the above observations suggest a space-controlling role of Helix-7 in facilitating target mRNA recognition, rather than a previous hypothesis of Helix-7 pre-organizing the guide strand in the seed region<sup>33</sup>. Removal of negative charge on D358 widens the entrance of the RNA-loading channel, creates a trap state where only t6 is recognized, and consequently slows down the overall recognition.

**Helix-7 decelerates wobble pairing at g6-g7.** Why is it necessary for Helix-7 to assist the target mRNA recognition in hAgo2? A previous FRET study has found that the facilitating role of Helix-7 was reduced if wobble pairs are introduced, i.e. the presence of Helix-7 discourages off-target recognition<sup>33</sup>. Therefore, we performed additional PCV-MetaD on WT and the D358A mutant with two wobble pairs at g6/g7 (wb67). We chose UG wobble pairs since it is known that the pairing free energy for a UG pair and a canonical pair are highly similar; the only difference is that a UG pair is ~1 Å longer than a canonical pair<sup>67</sup>, requiring larger space for accommodation. This introduced minimal alteration to the energetics and helped dissect the entropic effect of Helix-7 on the wobble pairs.

As shown in Fig. 5a, b, the results for the WT-wb67 system are considerably different from WT (Fig. 4a, b), but similar to that of D358A (Fig. 4c, d). Not only Helix-7 has to be at least 24 Å away from MID to accommodate two wobble pairs (*y*-axis of Fig. 5a), a trap state (orange cross in Fig. 5a–d, representative conformations in Fig. 5e) similar to that of D358A (Fig. 4e) also emerges, indicating that the recognition of wobble pairs for the WT hAgo2 is indeed slower than the canonical pairs, consistent with previous experimental results<sup>33</sup>. For D358A-wb67 (Fig. 5c, d), this trap state is



**Fig. 4 Metadynamics simulations of the wild type protein mutant D358A.** **a–d** Reweighted free energy landscape of the g7-t7 distance, the distance between MID and Helix-7, and the distance between t7-phosphate and K525 are presented by contour plots. **e** Structures on the right are representative conformations for the D358A mutant extracted from the orange cross on the contour plots. The t7-phosphate (orange sphere), the positively charged residues (ice-blue), Helix-7 (yellow) and the A358 (white) are highlighted.

similarly populated to the D358A system (Fig. 4c, d), despite the lacking of base-pairing at g6 in the wobble pairs (Fig. 5e). These results suggest that Helix-7 can induce an off-pathway trap state when the wobble base pairs are present, and help hAgo2 to prevent the off-target recognition. Our observations provide a reasonable explanation for previous experimental findings<sup>33</sup>.

Argonaute is known to accelerate the formation of duplex between the gRNA and target mRNA through protein-RNA interactions<sup>68</sup>. Here we showed that backbone coordination of its target mRNA with the positively charged residues (K355 / R366 / K525 / K550 / K554) on hAgo2 is rate-limiting for the target mRNA recognition. The conservation of these residues in eukaryotic Argonautes (Fig. 3i) indicates a common recognition mechanism for them. Yet such backbone-coordination alone is not sequence dependent and thus cannot distinguish canonical base-pairs from uncanonical ones. By utilizing Helix-7, particularly the negatively charged D358 (also highly conserved in eukaryotic Argonautes only, see Fig. 3i), eukaryotic Argonautes, unlike their prokaryotic counterparts, developed an additional mechanism for fidelity control, i.e. regulating space at the entrance to the RNA-loading channel for the mRNA such that only the canonical base-pairing but not wobble-pairing is facilitated.

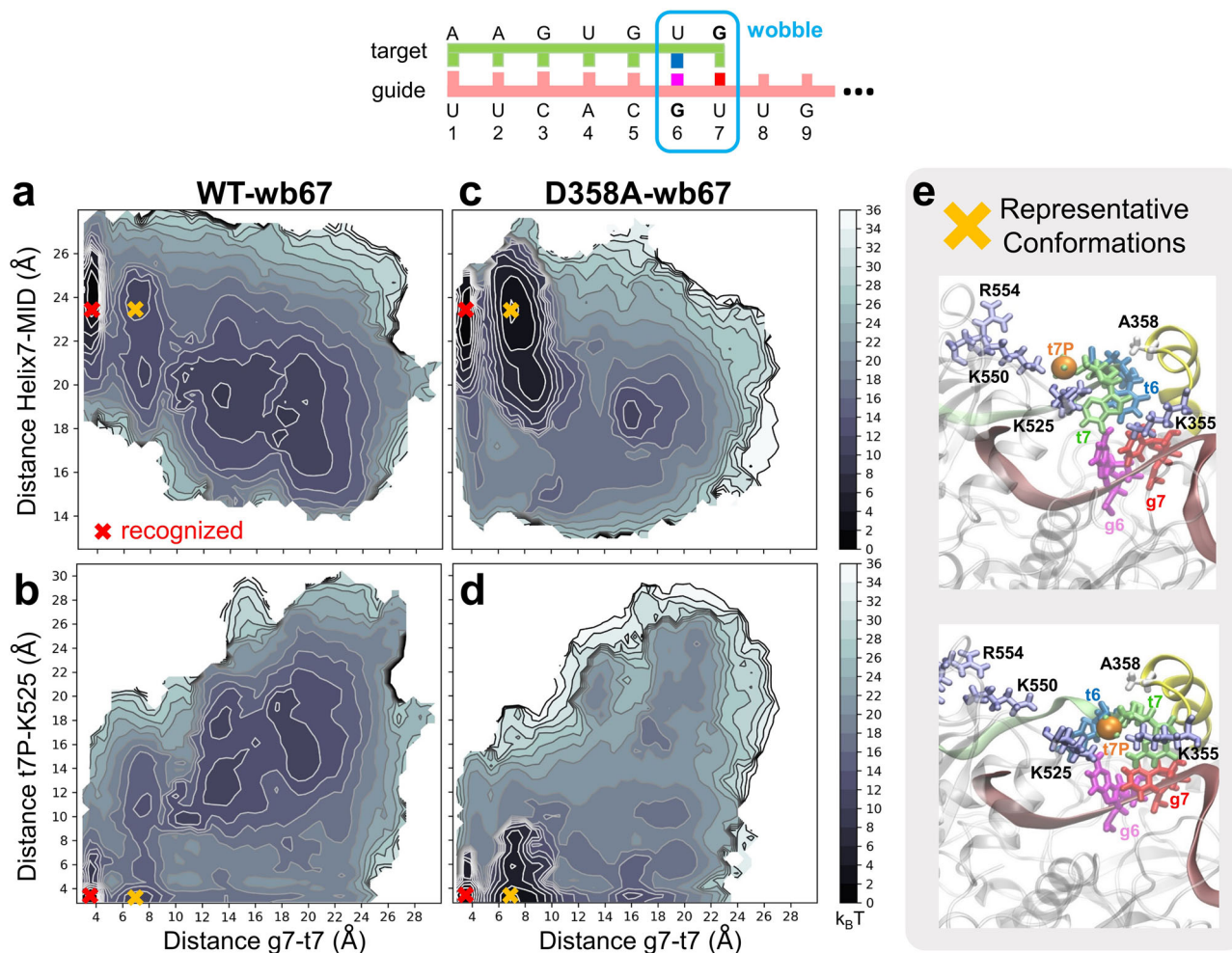
## Conclusion

We have constructed qMSM based on an extensive MD simulation dataset (84  $\mu$ s in total) to elucidate the mechanism of target

mRNA recognition by the hAgo2-gRNA complex at the second half of the seed region (g6-g7). Our qMSM revealed that the slowest mode for target mRNA recognition is not the gRNA-mRNA base-pairing, but the mRNA backbone phosphate coordination with the positively charged residues of hAgo2 (R366, R554, K550, K525, K355). Moreover, the positioning of Helix-7 (conserved in eukaryotic Agos) facilitates the recognition through controlling the effective conformation space to be explored by target mRNA. Alanine mutation of the negatively charged D358 on Helix-7 and UG wobble-pair mutants at g6-g7 both created a trap state prior to recognition at g7 and therefore decelerated the overall recognition. These results clearly illustrated the vital role of protein-RNA interaction in target mRNA recognition by eukaryotic Agos.

## Methods

**Structural modeling and molecular dynamics simulations.** We built the structural models of hAgo2-gRNA-mRNA complex based on a crystal structure (PDB id: 4W5O) containing 9 base pairs (g2-g9)<sup>38</sup>. Homology modeling was performed to fill in missing part of the crystal structure via Modeller v.9.10<sup>69</sup>. We removed the nucleotides from both RNA strands at g8 and g9. This structure was solvated in a dodecahedron box with ~47,000 TIP3P waters, 144 Na<sup>+</sup> and 143 Cl<sup>-</sup> ions, resulting in a total of ~156,000 atoms in the simulation box. The AMBER 99SB-ILDN force-field<sup>70</sup> was used to describe the interactions in the system. Although the  $\chi$ OL3-correction of AMBER 99SB<sup>71–73</sup> could have offered improved RNA backbone dynamics, we found that the  $\chi$ OL3-correction did not alter the transition state and therefore the recognition mechanism revealed in this work (see SI Note 2 and Fig. S6 for details). We used the software GROMACS v5.0.4 for MD



**Fig. 5** Metadynamics simulations of the wild type and mutant D358A protein with two UG wobble pair at g6 and g7. **a–d** Reweighted free energy landscape of the g7-t7 distance, the distance between MID and Helix-7, and the distance between t7-phosphate and K525 are presented by contour plots. **e** Structures on the right are representative conformations for the D358A mutant extracted from the orange cross on the contour plots. The t7-phosphate (orange sphere), the positively charged residues (ice-blue), Helix-7 (yellow) and the A358 (white) are highlighted.

simulations<sup>74</sup>. Energy minimization was performed for 10,000 steps by the steepest descent algorithm and then by the conjugate gradient algorithm. Then a 100 ps NVT simulations using the Berendsen thermostat<sup>75</sup> was performed at 300 K for solvent equilibration, followed by a 1 ns NPT equilibration to 1 atm using the Berendsen barostat<sup>75</sup>. All MD simulations were performed in the NPT ensemble with a time-step of 2 fs, the *v*-rescale thermostat<sup>76,77</sup> and the Parrinello-Rhmann barostat<sup>78</sup>. Long-range electrostatic interactions were treated by the Particle-Mesh Ewald method<sup>79</sup>. The short-range electrostatic and van der Waals interactions both used a cut-off of 10 Å. All bonds were constrained by the LINCS algorithm<sup>80</sup>.

#### Initial sampling of base-pair disruption at g6-g7 via metadynamics simulations.

Directly simulating the target mRNA recognition via conventional MD simulations is extremely challenging because hAgo2, miRNA and mRNA are all large and flexible biomolecules. We have applied metadynamics (MetaD) to obtain sufficient initial sampling<sup>81,82</sup>. Starting from the structural model of RISC-mRNA complex with base pairs formed at g2-g7, we performed MetaD simulations via PLUMED<sup>83</sup>, with bias potentials on the base distances at g6-g7 to sample the disruption and re-formation of the two base-pairs (details in SI Note 3). Four independent MetaD simulations were performed. In all four simulations, multiple rounds of breaking and re-formation of the two base-pairs were observed (Fig. S7), indicating sufficient initial sampling. Note that no breaking of the base-pairs or base-stacking at g2-g5 was observed during the metadynamics or subsequent unbiased simulations (see SI Note 4 and Fig. S8 for details).

**Microstate-MSM construction and validation.** We then performed K-centers clustering on the four MetaD trajectories and obtained 84 clusters. Starting from the geometric center conformations of these 84 clusters, we performed unbiased MD simulations for MSM construction, resulting in 84 trajectories with each in length of 1  $\mu$ s, with a saving interval of 200 ps. Based on this MD dataset, we

followed our previously published protocol<sup>49,84</sup> to construct microstate-MSM to investigate the RISC-mRNA recognition. Time-lagged Independent Component Analysis (tICA)<sup>65,66</sup> was first used for dimensionality reduction. Spectral oASIS<sup>85</sup> was used to reduce the number of input features (Fig. S10). The APLoD algorithm<sup>86</sup> was then used to define microstates in the reduced tIC space. Throughout this procedure, we applied the Generalized Matrix Rayleigh Quotient (GMRQ)<sup>87</sup> score to evaluate the quality of the model (Fig. S9c–e). The final microstate MSM was built on the first 4 tICs with 81 microstates (details in SI Note 5–7), which is validated by the Chapman-Kolmogorov test (Fig. S11b, details in SI Note 7). Construction of the microstate MSM was performed using our in-house python code based on MSMbuilder version 3.8.0<sup>88</sup>. The free energy landscapes for the microstate MSM data were visualized by MSMExplorer<sup>89</sup>.

**Macrostate quasi-MSM construction and validation.** To assist the interpretation of target-recognition mechanisms, we applied our recently developed qMSM<sup>63</sup> approach to construct a model containing 4 macrostates (state S1–S4). To obtain this qMSM, we first performed the kinetic lumping to group 81 microstates into 4 macrostates using the PCCA + algorithm<sup>90,91</sup> implemented in PyEMMA version 2.5.2<sup>92</sup>, because a stable gap is observed between 3rd and 4th slowest implied timescale (details in SI Note 7, 8). Our qMSM applies the generalized master equation formalism to encode the non-Markovian dynamics in time-dependent memory kernels ( $K(\tau)$ ):

$$\dot{T}(t) = T(t)\dot{T}(0) + \int_0^{\min\{t, \tau_K\}} T(t-\tau)K(\tau)d\tau, \quad (1)$$

where  $\tau_K$  corresponds to the memory kernel relaxation time where  $K(t > \tau_K) \approx 0$ , and  $T(t)$  refers to the TPM. We validated We validated our final qMSM via the Chapman-Kolmogorov test and compute MFPTs among the four macrostates via the transition path theory<sup>93,94</sup> (details in SI Note 8, 9).

**Setup of the Path-Collective-Variable metadynamics simulations for mutants.**

To verify the predictions from the qMSM, we designed several protein mutants and a RNA mutant with two UG-wobble-pairs at position 6 and 7. To perform conformational sampling of these mutant systems, we conducted 1  $\mu$ s long well-tempered<sup>82,95</sup> Path-Collective-Variable (PCV) MetaD simulations<sup>96</sup> for each mutant as well as the wild-type. PCV-MetaD is a MetaD simulation biasing on two PCV-s and PCV-z, denoting the progress along and the average distance from a high dimensional reference path respectively, given a pre-defined distance metric. Our reference path was extracted from one of the 84 MD simulation trajectories, in which a near-complete recognition process can be observed (Fig. S1a and Movie S1). The reference path consisted of 36 nodes with a RMSD of  $\sim 1.4$  Å between neighbor nodes (28 nodes for the H7G5 mutant, where Helix-7 was replaced by 5 glycine residues). RMSD is measured by two atom-sets: (i) structural alignment was performed on the C $\alpha$  atoms of the MID and PIWI domain; (ii) RMSD was calculated using the C $\alpha$  of Helix-7, all heavy atoms from target nucleotides t6/t7 and the sidechains of R554, K550, Q548, K525, K355. The widths of Gaussian hills for PCV-s and PCV-z were chosen as 0.5 and 3 Å<sup>2</sup> respectively. These Gaussian hills of height 1.25 kJ/mol were deposited every 200 ps with a bias-factor of 15 at reference temperature of 310 K. To ensure efficient sampling, we imposed a wall potential at PCV-z = 36 Å<sup>2</sup>, below which the majority of MSM samples were encompassed (Fig. S1b). Analysis of the free energy landscapes over the collective variables (CVs, with physical meanings) other than s and z were obtained via a standard reweighting procedure<sup>97</sup>. To ensure the smoothness of the reweighted landscape, values of the CVs were recorded every 10 ps.

**Statistics and reproducibility.** Though MD simulations at constant temperature are stochastic in nature, the statistics are reproducible if sufficient sampling, under the framework of MSM and qMSM as in this manuscript, is achieved. We performed homology modeling via the software MODELLER (<https://salilab.org/modeller/>). MD simulations were performed via GROMACS version 5.0.4 (<http://www.gromacs.org>). The mutant metadynamics simulations were performed via the PLUMED plugin (<https://www.plumed.org>). The MSM and qMSM were built via the MSMbuilder version 3.8.0 (<http://msmbuilder.org>) and PyEMMA version 2.5.2 (<http://emma-project.org/latest/>).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

Data underlying main figures is presented in Supplementary Data 1. The authors declare that all other data supporting the findings of this study are available within the paper and its supplementary information files, or are available from the corresponding author upon reasonable request.

**Code availability**

All in-house code is available from the corresponding authors upon reasonable request.

Received: 17 May 2021; Accepted: 26 October 2021;

Published online: 30 November 2021

**References**

- Pratt, A. J. & MacRae, I. J. The RNA-induced silencing complex: a versatile gene-silencing machine. *J. Biol. Chem.* **284**, 17897–17901 (2009).
- 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* **364**, 99–115 (2009).
- Joshua-Tor, L. & Hannon, G. J. Ancestral roles of small RNAs: an ago-centric perspective. *Cold Spring Harb. Perspect. Biol.* **3**, a003772 (2011).
- Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).
- Li, Z. & Rana, T. M. Molecular mechanisms of RNA-triggered gene silencing machineries. *Acc. Chem. Res.* **45**, 1122–1131 (2012).
- Kawamata, T. & Tomari, Y. Making RISC. *Trends Biochem. Sci.* **35**, 368–376 (2010).
- Sasaki, H. M. & Tomari, Y. The true core of RNA silencing revealed. *Nat. Struct. Mol. Biol.* **19**, 657–660 (2012).
- Wang, Y. et al. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
- Wang, Y. et al. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461**, 754–761 (2009).
- Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008).
- Ipsaro, J. J. & Joshua-Tor, L. From guide to target: molecular insights into eukaryotic RNA-interference machinery. *Nat. Struct. Mol. Biol.* **22**, 20–28 (2015).
- Gorski, S. A., Vogel, J. & Doudna, J. A. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* **18**, 215–228 (2017).
- Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* **14**, 447–459 (2013).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Chivukula, R. R. et al. An essential mesenchymal function for miR-143/145 in intestinal epithelial regeneration. *Cell* **157**, 1104–1116 (2014).
- Wang, X. W., Heegaard, N. H. H. & Orum, H. MicroRNAs in liver disease. *Gastroenterol* **142**, 1431–1443 (2012).
- Esquela-Kerscher, A. et al. The let-7 microRNA reduces tumor growth in mouse models of lung cancer. *Cell Cycle* **7**, 759–764 (2008).
- Trang, P. et al. Regression of murine lung tumors by the let-7 microRNA. *J. Thorac. Oncol.* **29**, 1580–1587 (2009).
- Broderick, J. A. & Zamore, P. D. MicroRNA therapeutics. *Gene Ther.* **18**, 1104–1110 (2011).
- Li, Z. & Rana, T. M. Therapeutic targeting of microRNAs: current status and future challenges. *Nat. Rev. Drug Discov.* **13**, 622–638 (2014).
- Parsel, S. M., Grandis, J. R. & Thomas, S. M. Nucleic acid targeting: towards personalized therapy for head and neck cancer. *Oncogene* **35**, 3217–3226 (2016).
- Chandrasekhar, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A dynamic search process underlies MicroRNA targeting. *Cell* **162**, 96–107 (2015).
- Jo, M. H., Song, J.-J. & Hohng, S. Single-molecule fluorescence measurements reveal the reaction mechanisms of the core-RISC, composed of human Argonaute 2 and a guide RNA. *BMB Rep.* **48**, 643–644 (2015).
- Jo, MyungH. et al. Human Argonaute 2 has diverse reaction pathways on target RNAs. *Mol. Cell* **59**, 117–124 (2015).
- Herzog, V. A. & Ameres, S. L. Approaching the golden fleece a molecule at a time: biophysical insights into argonaute-structured nucleic acid interactions. *Mol. Cell* **59**, 4–7 (2015).
- 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. USA* **110**, 17850–17855 (2013).
- Elkayam, E. et al. The structure of human argonaute-2 in complex with miR-20a. *Cell* **150**, 100–110 (2012).
- Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science* **346**, 608 (2014).
- Schirle, N. T. et al. Structural analysis of human Argonaute-2 bound to a modified siRNA guide. *J. Am. Chem. Soc.* <https://doi.org/10.1021/jacs.6b04454> (2016).
- Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F. R. & MacRae, I. J. Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *EMBO J.* **38**, e101153 (2019).
- Baronti, L. et al. Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. *Nature* **583**, 139–144 (2020).
- Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037–1040 (2012).
- Klum, S. M., Chandrasekhar, S. D., Schirle, N. T., Joo, C. & MacRae, I. J. Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition. *EMBO J.* **37**, 75–88 (2018).
- Willkomm, S. & Restle, T. Conformational dynamics of Ago-mediated silencing processes. *Int. J. Mol. Sci.* **16**, 14769–14785 (2015).
- Wang, Y., Li, Y., Ma, Z., Yang, W. & Ai, C. Mechanism of microRNA-target interaction: molecular dynamics simulations and thermodynamics analysis. *PLoS Comput. Biol.* **6**, e1000866 (2010).
- Xia, Z. et al. Molecular dynamics simulations of Ago silencing complexes reveal a large repertoire of admissible ‘seed-less’ targets. *Sci. Rep.* **2**, 569 (2012).
- 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* **8**, e54620 (2013).
- Zhu, L. et al. A flexible domain-domain hinge promotes an induced-fit dominant mechanism for the loading of guide-DNA into Argonaute protein in *Thermus thermophilus*. *J. Phys. Chem. B* **120**, 2709–2720 (2016).
- Noe, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **18**, 154–162 (2008).
- 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.* **126**, 155101 (2007).
- Buchete, N. V. & Hummer, G. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **112**, 6057–6069 (2008).



42. Pan, A. C. & Roux, B. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129**, 064107 (2008).
43. Prinz, J. H. et al. Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
44. Gfeller, D., De Los Rios, P., Caflisch, A. & Rao, F. Complex network analysis of free-energy landscapes. *Proc. Natl Acad. Sci. USA* **104**, 1817–1822 (2007).
45. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
46. Bowman, G. R., Huang, X. & Pande, V. S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49**, 197–201 (2009).
47. Huang, X. et al. Constructing multi-resolution markov state models (msms) to elucidate RNA hairpin folding mechanisms. *Pac. Symp. Biocomput.* **15**, 228–239 (2010).
48. Huang, X., Bowman, G. R., Bacallado, S. & Pande, V. S. Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl Acad. Sci. USA* **106**, 19765–19769 (2009).
49. Wang, W., Cao, S., Zhu, L. & Huang, X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *WIREs Comput. Mol. Sci.* **8**, e1343 (2017).
50. Noe, F., Schutte, C., Vanden-Eijnden, E., Reich, L. & Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl Acad. Sci. USA* **106**, 19011–19016 (2009).
51. Silva, D. A. et al. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl Acad. Sci. USA* **111**, 7665–7670 (2014).
52. Vanatta, D. K., Shukla, D., Lawrenz, M. & Pande, V. S. A network of molecular switches controls the activation of the two-component response regulator NtrC. *Nat. Commun.* **6**, 7283 (2015).
53. 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.* **134**, 2399–2406 (2012).
54. Chodera, J. D. & Noe, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).
55. Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl Acad. Sci. USA* **108**, 10184 (2011).
56. Silva, D.-A., Bowman, G. R., Sosa-Peinado, A. & Huang, X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput. Biol.* **7**, e1002054 (2011).
57. Plattner, N., Doerr, S., De Fabritiis, G. & Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **9**, 1005–1011 (2017).
58. Zhu, L. et al. Understanding the core of RNA interference: the dynamic aspects of Argonaute-mediated processes. *Prog. Biophys. Mol. Biol.* **128**, 39–46 (2017).
59. Malmstrom, R. D., Kornev, A. P., Taylor, S. S. & Amaro, R. E. Allosteric through the computational microscope: cAMP activation of a canonical signalling domain. *Nat. Commun.* **6**, 7588 (2015).
60. Zhu, L., Sheong, F. K., Zeng, X. & Huang, X. Elucidation of the conformational dynamics of multi-body systems by construction of Markov state models. *Phys. Chem. Chem. Phys.* **18**, 30228–30235 (2016).
61. Zeng, X. et al. Elucidating dominant pathways of the nano-particle self-assembly process. *Phys. Chem. Chem. Phys.* <https://doi.org/10.1039/c6cp01808d> (2016).
62. Jiang, H. et al. 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.* **11**, e1004404 (2015).
63. Cao, S., Montoya-Castillo, A., Wang, W., Markland, T. E. & Huang, X. On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *J. Chem. Phys.* **153**, 014105 (2020).
64. Unarta, I. C. et al. Role of bacterial RNA polymerase gate opening dynamics in DNA loading and antibiotics inhibition elucidated by quasi-Markov State model. *Proc. Natl. Acad. Sci. USA* (in press).
65. Schwantes, C. R. & Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000–2009 (2013).
66. Naritomi, Y. & Fuchigami, S. Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* **139**, 215102 (2013).
67. Vendéix, F. A. P., Munoz, A. M. & Agris, P. F. Free energy calculation of modified base-pair formation in explicit solvent: a predictive model. *RNA* **15**, 2278–2287 (2009).
68. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides. *Cell* **162**, 84–95 (2015).
69. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5.6.1–5.6.37 (2016).
70. Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
71. Banáš, P. et al. Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849 (2010).
72. Zgarbová, M. et al. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
73. Tan, D., Piana, S., Dirks, R. M. & Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl Acad. Sci. USA* **115**, E1346 (2018).
74. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
75. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
76. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
77. Bussi, G., Zykova-Timan, T. & Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **130**, 074101 (2009).
78. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
79. Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N·Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
80. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
81. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12562–12566 (2002).
82. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
83. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
84. Wang, X., Unarta, I. C., Cheung, P. P.-H. & Huang, X. Elucidating molecular mechanisms of functional conformational changes of proteins via Markov state models. *Curr. Opin. Struct. Biol.* **67**, 69–77 (2021).
85. Litzinger, F. et al. Rapid calculation of molecular kinetics using compressed sensing. *J. Chem. Theory Comput.* **14**, 2771–2783 (2018).
86. Liu, S., Zhu, L., Sheong, F. K., Wang, W. & Huang, X. Adaptive Partitioning by Local Density-peaks (ALoD): An efficient density-based clustering algorithm for analyzing molecular dynamics trajectories. *J. Comput. Chem.* **38**, 152–160 (2017).
87. McGibbon, R. T. & Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **142**, 124105 (2015).
88. Beauchamp, K. A. et al. MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* **7**, 3412–3419 (2011).
89. Carlos, X. H., Matthew, P. H., Mohammad, M. S. & Vijay, S. P. MSMEExplor: data visualizations for biomolecular dynamics. *J. Open Source Softw.* **2**, 188 (2017).
90. Deuffhard, P. & Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.* **398**, 161–184 (2005).
91. Rblitz, S. & Weber, M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data Anal. Classification* **7**, 147–179 (2013).
92. Scherer, M. K. et al. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
93. Metzner, P., Schütte, C. & Vanden-Eijnden, E. Transition path theory for Markov jump processes. *Multiscale Model. Simul.* **7**, 1192–1219 (2009).
94. Weinan, E. & Vanden-Eijnden, E. Towards a theory of transition paths. *J. Stat. Phys.* **123**, 503–523 (2006).
95. Dama, J. F., Parrinello, M. & Voth, G. A. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **112**, 240602 (2014).
96. Branduardi, D., Gervasio, F. L. & Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).
97. Bonomi, M., Barducci, A. & Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **30**, 1615–1621 (2009).

## Acknowledgements

We thank Dr. Fu Kit Sheong for fruitful discussions about the Argonaute proteins. This work was supported by the Hong Kong Research Grant Council [16303919, 16307718,

and 16318816] to X.H., Shenzhen Science and Technology Innovation Committee [JCYJ20200109150003938] to L.Z., the National Science Foundation of China General Fund [31971179] to L.Z., research fund from Warshel Institute for Computational Biology and Longgang District Shenzhen to L.Z., and King Abdullah University of Science and Technology (KAUST) under award number FCC/1/1976-04-01, URF/1/4098-01-01, URF/1/4352-01-01, and URF/1/4379-01-01 to X.G.. This research made use of the resources of the Supercomputing Laboratory at KAUST.

### Author contributions

L.Z., H.J., X.G. and X.H. conceived this study. H.J. initiated the study and performed the initial sampling. L.Z. and H.J. performed the MD simulations. L.Z., S.C. and I.C.U. analyzed the data and constructed the microstate MSMs and the qMSM. L.Z. performed the mutant simulations. L.Z., S.C. I.C.U., X.G. and X.H. wrote the manuscript. All authors are involved in discussion and finalization of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02822-7>.

**Correspondence** and requests for materials should be addressed to Xin Gao or Xuhui Huang.

**Peer review information** *Communications Biology* thanks Abhishek Singharoy and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Audrone Lapinaite and George Inglis.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021