Received: 18 January 2022 | Revised: 25 May 2022 | Accepted: 5 June 2022

DOI: 10.1002/pro.4377

FULL-LENGTH PAPER

The dynamics of t1 adenosine binding on human Argonaute 2: Understanding recognition with conformational selection

Silvia Rinaldi[1] | **Giorgio Colombo**[2] | **Antonella Paladino**[3]

[1]CNR- Institute of Chemistry of OrganoMetallic Compounds (ICCOM), Sesto Fiorentino (FI), Italy

[2]Department of Chemistry, University of Pavia, Pavia, Italy

[3]CNR- Institute of Biostructures and Bioimaging, Naples, Italy

Correspondence
Silvia Rinaldi, CNR- Institute of Chemistry of OrganoMetallic Compounds (ICCOM), Via Madonna del Piano 10, Sesto Fiorentino (FI) 50019, Italy.
Email: silvia.rinaldi@iccom.cnr.it

Antonella Paladino, CNR- Institute of Biostructures and Bioimaging, via Pietro Castellino 111, 80131 Naples, Italy.
Email: antonella.paladino@cnr.it

Review Editor: John Kuriyan

Abstract

The control of expression in genetic regulation is a fundamental process for cell life. In RNA-mediated silencing, human Argonaute-2 protein (hAgo2) uses sequence information encoded in small RNAs (guide) to identify complementary sites in messenger RNAs (target) for repression. The specificity of this molecular recognition lies at the basis of the mechanisms that control the expression of thousands of genes, which necessarily requires a fine tuning of complex events. Among these, the binding of the first nucleotide of the target RNA (t1) is emerging as an important modulator of hAgo2-mediated machinery. Using atomistic molecular dynamics-derived analyses, we address the mechanism behind t1-dependent regulation and study the impact of different t1 nucleotides (t1A, t1C, t1G, t1U) on the conformational dynamics of both hAgo2 and guide–target RNAs. Only when an adenine is found at this position, t1 directly interacts with a specific hAgo2 binding pocket, favoring the stabilization of target binding. Our findings show that hAgo2 exploits a dynamic recognition mechanism of the t1-target thanks to a modulation of RNA conformations. Here, t1-adenine is the only nucleobase endowed with a dual binding mode: a T-shape and a co-planar conformation, respectively, orthogonal and parallel to the following base-pairs of guide–target duplex. This triggers a composite set of molecular interactions that stabilizes distinctive conformational ensembles. Our comparative analyses show characteristic traits of local and global dynamic interplay between hAgo2 and the RNA molecules and highlight how t1A binding acts as a molecular switch for target recognition and complex stabilization. Implications for future mechanistic studies are discussed.

KEYWORDS

hAgo2, molecular dynamics, RNA, RNA silencing, t1 nucleotide

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

Human Argonaute-2 proteins (hAgo2) represent the functional core of the RNA-induced silencing complexes (RISCs) that mediate RNA silencing in eukaryotes: Argonaute proteins bind small regulatory RNAs—that is, small interfering RNAs (siRNAs) and microRNAs (miRNAs)—and use the encoded sequence information to recognize and silence their complementary target messenger RNAs (mRNA). This task is achieved either by direct cleavage via the endonucleolytic "slicing" processes or by recruitment of additional silencing factors.[1–5] Structural studies revealed that hAgo2 adopts a two-lobe architecture shaped by four different domains: the (N)-terminal domain, the Piwi–Argonaute-Zwille (PAZ) domain, the middle (MID) domain, and the P-element-induced wimpy testes (PIWI) domain connected by two flexible linkers (L1 and L2). The 3' microRNA-guide binds PAZ domain while its 5′ end anchors the MID and PIWI domains (Figure 1), resulting in a functional complex that engages a dynamic interplay with the complementary mRNA.[6–8] Target recognition occurs primarily through a Watson–Crick (WC) base pairing at position g2-g8 of the regulatory-guide RNA referred as seed sequence (Figure 1). This complementarity is the primary determinant of efficacy and specificity, with over 80% of guide–target interactions occurring through seed pairing.[9,10] However, it has been demonstrated that additional regions outside the seed can enhance or reinforce target-guide interactions[5,11–14]; very extensive pairing to the 3′ region of miRNA guide, for example, can compensate for a wobble or mismatch at the seed position, contributing to target recognition.[15] Besides, RNA-induced conformational rearrangements at different key hotspots of hAgo2 domains have been shown to play roles as molecular switches for target recognition/stabilization: in this framework, the 3' end of the seed (g6–g8) can adopt different conformations, whose stabilization is controlled by hAgo2 helix-7 that pivots to enforce rapid making and breaking of miRNA:target base pairs, destabilizing off-targets interactions. The resulting big picture reveals a fine, dynamic, and multifactorial mechanism of hAgo2 regulation that underpins the precise control of mRNA expression with crucial implications for RNA-interference (RNAi) or gene therapy.

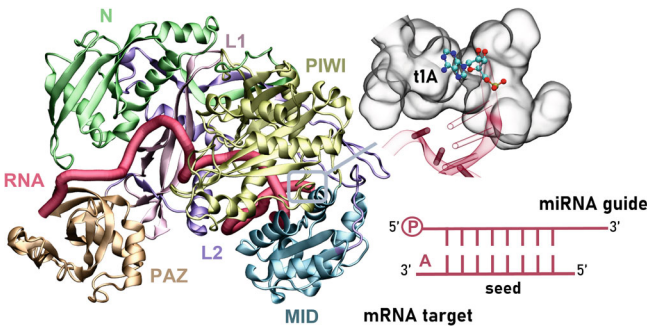

FIGURE 1 Structure of hAgo2–guide–target complex. Focus of the 3' RNA target at the MID-PIWl interface. Schematic representation of guide–target RNA pairing

In this scenario, the role of an adenosine nucleotide in the t1 position of mRNA target (t1A) (Figure 1) raises intriguing mechanistic questions: the nucleotide is conserved among many vertebrate target sites[16] and has been shown to stabilize target binding beyond pairing to the seed region,[17] conferring enhanced repression independently of the identity of miRNA-guide nucleotide-1.[11,18–20] Indeed, t1 adenine is recognized through a mechanism that differs from the WC base pairing but directly interacts with a specific binding pocket within hAgo2 proteins. Using crystallographic investigations, Schirle and colleagues have demonstrated that t1A nucleotides binds to a solvent-accessible pocket at the interface of hAgo2 L2 and MID domains (Figure 1).[21] This interaction anchors and stabilizes seed-matched sites on target RNAs, playing a major role for targeting efficacy and specificity control. In their study, the authors identified t1A N6 amine as the key determinant of t1A specificity, whose methylation in fact hampers target recognition; they argued that t1A role of molecular switch is fulfilled through a specific water-mediated network between the adenosine N6 amine and hAgo2 protein. Despite such an impact on hAgo2 fine regulation and consequently on RNA silencing mechanisms, the contacts between hAgo2 and t1A are mainly non-specific. Moreover, the t1A binding pocket is large enough to accommodate any of the four natural RNA bases.[9,21] These observations, together with the complex dynamics interplay between hAgo2 and its RNA targets, suggest that possible additive recognition/regulation mechanisms, not yet characterized, could contribute to t1A-mediated control of target specificity. An additional important way by which proteins recognize their RNA targets involves a conformational readout of RNA 3D structure,[22–24] in which RNA different conformations affect the recognition process. In SRF2 splicing factor, for example, RNA recognition is achieved thanks to a modulation of RNA conformations: flipping the bases of the two consecutive C or G nucleotides into either *anti* or *syn* conformation allows SRF2 to bind two different RNA sequences equally well without changing the protein conformation.[25] Protein Hfq, a key factor in the RNA-mediated control of gene expression in most known bacteria, has been proposed to utilize dynamic recognition of RNA substrates by the selective stabilization of different conformations of adenosines bound to the same pocket.[26]

Herein, using extensive molecular dynamics-derived analyses, we demonstrate that t1A exploits a composite set of molecular interactions between hAgo2 and its RNA target to populate distinctive conformational ensembles. The t1A-specific conformations may facilitate hAgo2 local-to-global conformational rearrangements that act as molecular switches for target recognition and stabilization on the one hand, while controlling downstream RNAs flexibility on the other hand.

Since target binding affinity is considered a key factor in mRNA repression by miRNAs,[17] the characterization of the possible role of t1A as a dynamic trigger may contribute to shed light on the fine-tuned recognition mechanism behind hAgo2-mediated regulation of gene expression. Our study provides new mechanistic insights on guide–target communication mechanisms that occur during RISC regulation. Molecular recognition at hAgo2 t1-pocket and the onset of interdependent structural adjustments echoing on the biomolecular assembly aids the definition of novel specific structure-guided determinants in RNA-interference that may be exploited in therapeutic gene silencing technology.

2 | RESULTS

2.1 | Models of hAgo2–RNA complexes with different t1 nucleotides

Argonaute proteins exist in a dynamic ensemble of structural states which are controlled by the interactions with several binders at various steps of the RNA silencing pathway.[27,28] Among these, small RNA loading (and only minimally target recognition) is shown to stabilize and rigidify hAgo2 in a relatively open conformation[6,29,30] In this framework, MD simulations can provide an excellent tool to investigate the impact of different nucleotides at t1 (3′) of the target RNA on the structural "adaptation/response" of the protein from a local to global perspective. Building upon the recently reported crystal structures of hAgo2 bound to a guide RNA and short, seed-paired target RNAs with different nucleotides a t1 position,[21] the full-length hAgo2–guide–target complexes were modeled, using the protein bound to t1C target RNA (PDB entry: 4Z4C) as a template. This is indeed the solved structure with the minimal number of structural gaps: consistently, coordinates of the t1C in the t1 binding cavity were used to model the other nucleobases (adenine, uracil, guanine). The t1 binding pocket is large enough to accommodate purines as well as pyrimidines[21]; therefore, it may host and stabilize many different conformations. For example, t1-cytosine stacks over the double-stranded RNA in a co-planar conformation with respect to the seed region and adopts an orthogonal orientation with respect to the conformation of t1-adenosine (PDB entry 4w5o, Figure S1). The t1C co-planar binding pose was thus used to explore different binding modes, establishing the most stable interaction pattern, and providing an unbiased starting point.

In summary, four independent hAgo2 complexes were built, hereafter indicated as ***t1A, t1G, t1U, t1C*** with respect to the t1 nucleotide at the target RNA (Figure 1). In the simulated timescale (500 ns per 3 replicas = 1,5 μs per hAgo2 complex), all systems reached a satisfactory level of stability (root mean square deviations—RMSDs—computed on backbone atoms stabilize at an average value of 3.0 Å).

Microsecond MD simulations of the four systems disclosed a complex variety of structural fluctuations at either the nucleic acid or the protein levels. Specifically, at the binding pocket, we observed an enhanced conformational mobility of the unpaired t1A nucleotide compared both to the smaller pyrimidines and to t1-guanine that are stably packed to the following base (t2) (see Figure S2). Indeed, cluster analysis carried out on the backbone atoms of residues within 4 Å from t1 nucleotide highlights two main structural poses for the t1-adenosine: a T-shape and a co-planar conformation, respectively, orthogonal and parallel to the following base-pairs of guide–target duplex (Figure 2) that co-exist along the simulation time (Figure S3). Interestingly, the T-shape conformation well fits into the omic map (contoured at 1.5 σ) of the crystal structure of the hAgo2–guide–target structure (PDB entry: 4w5o) bearing t1A nucleotide (RMSD = 1.27 Å by aligning 629 Cα atoms) as well as into the electron density map of the t1DAP (2,6-diaminopurine) nucleotide bound to the hAgo2 structure (PDB entry: 4z4f) (RMSD = 1.28 Å on 628 Cα atoms) as shown in Figure S4.

2.2 | t1 nucleotides impact on pocket conformation

The analysis of the interactions established by t1A in the first cluster (T-shape conformation) reveals a close correspondence with the hydrogen bond network described in the crystal structure[21] (Figure 3) confirming the viability of the modeled complex. Herein, t1A engages a direct interaction with Ser-561 side chain via its N6 amine of t1A matching the crystallographic data. Furthermore, an additional network of hydrogen bonds mediated by water molecules favors t1A binding. In particular, main chain carbonyls of Met437, Lys440, and Ile477 are engaged in hydrogen bonds with water molecules that in turn bind the t1 nucleotide.

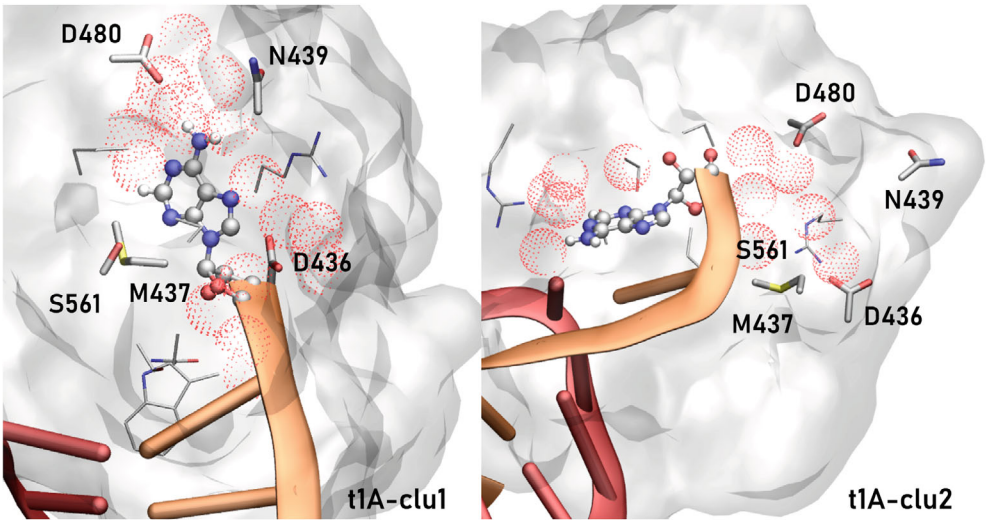

FIGURE 2 Representative structures obtained from MD simulations. The top two most populated clusters show the dual t1A binding mode. Residues within 4 Å from t1 nucleotide are shown in sticks, bold licorice is used for amino acids involved in significative interactions. Dotted red spheres represent waters while adenosine is rendered in ball and sticks. Ghost surface is used for hAgo2 t1-binding pocket. See also Figure S2

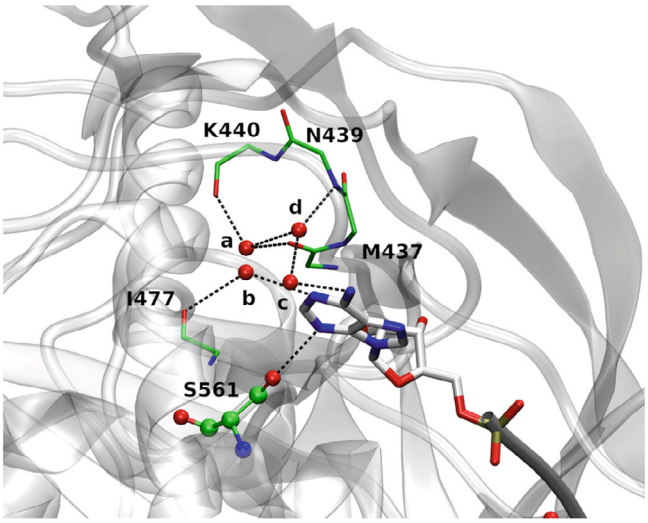

FIGURE 3 Hydrogen bonds network within the t1-binding pocket shown by the first cluster of t1A system. Only backbone atoms of the interacting amino acids are shown and labeled for clarity. Image is given in the same orientation as in the Schirle et al.[21]

Because of the dynamic nature of hydrogen bonding and of the variable association strength, a dynamic description is required to provide insight into the mechanisms that drive the specific recognition of t1A nucleobase. It is worth noting here that the dynamic character of the recognition is also due to the exposed nature of this pocket, which favors exchange of water molecules. Analysis of the hydrogen bonds engaged by t1 nucleotide and detectable at least for 5% of the simulated time pinpoints for t1A a distinctive pattern with respect to the remaining systems (see Table S1). Indeed, t1A stably engages a larger network of hydrogen bonds, in particular with Asp436 and Asp480 via strong O–H ··· O and N–H ··· O interactions, respectively, that anchor the binding pocket to the t1A T-shape configuration. Remarkably, Asp480 engages t1A through N6 amine that has been reported to act as key determinant of t1A specificity.[21] Notably, sequence alignment performed over a non-redundant set of argonaute proteins from any source organism obtained by querying UniRef50[31] confirms a large conservation for the binding pocket region and, specifically, for all those residues (Asp436, Met437, Asp480, Ser561) outlined as important hotspots of t1A stabilization in our analyses (Figure S5).

The superposition of a guanine onto the T-shape conformation adopted by t1A helps shed light on the reasons behind t1A selectivity. The presence of a guanine leads to an increased steric hindrance and electrostatic repulsion mainly caused by the proximity of Asp480 (carboxylic moiety) toward the t1G carbonyl at C6 position (Figure 4). Moreover, the hydrogen bond mediated by the Ser561-OH with adenine N3 atom is disfavored by the steric occupancy of the amine group at C2 in the guanine nucleotide.

Finally, to further compare the hydrogen-bonding landscape of the t1-binding pocket in the four complexes, we studied the conformations adopted by the t1 nucleotide in relation to the position of Ala481, which resides at the back of the t1-binding pocket and acts as sensor/modulator of the hAgo2-directed solvent interactions with the t1 adenosine. Indeed, a mutation to threonine in this position perturbs the water placement leading to a sever impairment of t1A recognition.[21] Figure 5 shows the conformational ensembles sampled by the t1A, t1C, t1G, and t1U complexes along the two significative coordinates (RC) defining t1 and Ala481 conformations: RC1 depicts the distance between the centres of mass (COM) computed on t1-nucleobase and Ala861 sidechain; RC2 represents a combined coordinate that describes t1 conformation with respect to the ds-RNA and is obtained by summing the distance between the COM computed on t1 and t2

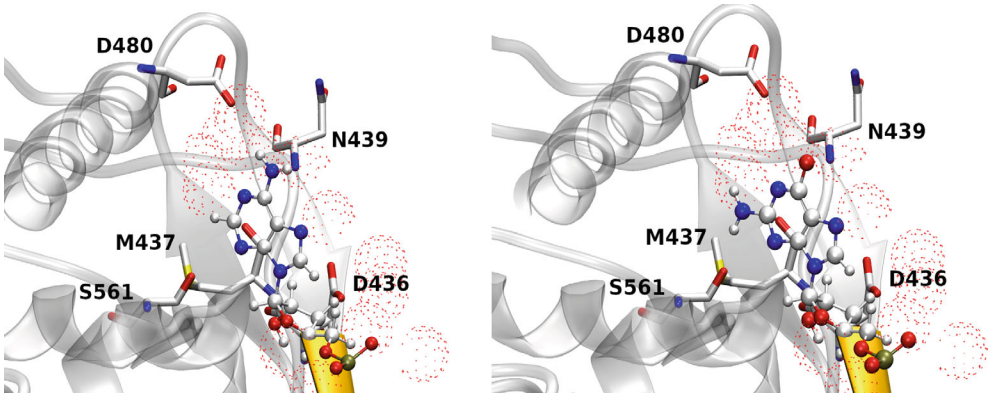

FIGURE 4 Comparison at the t1 nucleotide. Occupancy of guanine (right) at t1-binding pocket obtained by the superposition onto the conformation that t1A assumes in the first cluster (T-shape) (left)

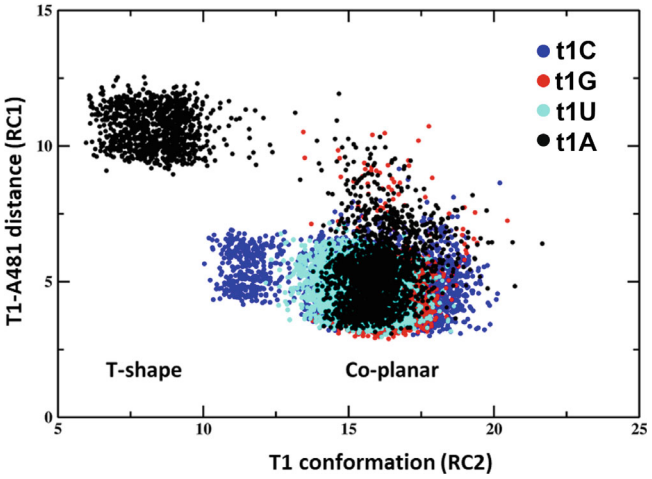

FIGURE 5 Conformational ensemble distribution for t1A, t1C, t1U and t1G systems along two significative coordinates: x-axis represents a linear combination of geometrical parameters that describe t1 versus t2 conformations; y-axis reports the distance between the center of mass of A481 residue and t1A. RC1 describes the distance between t1 and Ala861, while RC2 combines the distance between t1 and t2 with a proxy of t1-t2 stacking (i.e., cosine of inter-base angle). The contribution of the t1-t2 angle in terms of cosine function to the RC2 value, though minimal, is meant to emphasize the modulation of t1 conformers

nucleobases to the cosine of the angle spanned by the z-axis of t1 and t2 nucleobases. The results show that t1A is the only system where a T-shape conformer of the t1 nucleotide is explored, herein Ala481 is close enough to impact and modulate the hydrogen bond network of the t1-binding pocket. The remaining systems experience a different pattern that stabilize co-planar t1 configurations.

2.3 | t1 nucleotides impact on RNA conformation

Next, to unravel possible alterations induced by t1 binding on the whole nucleic molecules, the conformational mobility for the different systems was studied. The root-mean-square deviation on atoms of the guide miRNA shows similar patterns regardless of the nature of t1 nucleotide (see Figure S6). Notable differences, on the contrary, arise from fluctuations on target mRNA. When a purine is bound to the RNA backbone at t1 position, larger RMSD values are populated, and a bimodal distribution of the conformational variations is observed (Figure 6). This behavior is significantly enhanced in the presence of t1A, that allows the nucleic acid to populate conformations that are unaccessible to the other t1 systems and that are linked to the structural drift of t1A from and to the co-planar and T-shape conformations (Figure S3). Among the four systems, while the RMSD values computed on the central target RNA segment (t2–t6) are rather invariant (data not shown), larger dynamics fluctuations of the peripheral nucleotides differentiate the complexes, suggesting a t1-dependent pattern (Figure S7). Interestingly, the increased flexibility of the peripheral positions are shown to act as driving forces that mediate the recognition process.[5,13] To support this model, the root-mean-square fluctuation (RMSF) of each system was calculated. RMSF plots confirmed enhanced fluctuations for the nucleotides of mRNA target in the presence of t1A, especially at t1 position (Figure 6). This fits the degree of fluctuations required for the flexible alignment of regions beyond the seed, that in our systems are accessible only when t1A binds the protein.

Conformational plasticity (deformation) at the single nucleotide level can impact on the global motions of RNA molecules. For instance, it has been demonstrated that RNA bending is mostly due to a collective change in roll and slide along the sequence resulting into a global curvature,[32] where these parameters describe respectively rotational and translational motions of base pairs with respect to the RNA helix. The distribution of slide and roll values averaged upon the seed region base pairs for the four systems is reported in Figure 7. While for t1C, t1U, t1G complexes the analyses return average values compatible with a predominant typical A-conformation

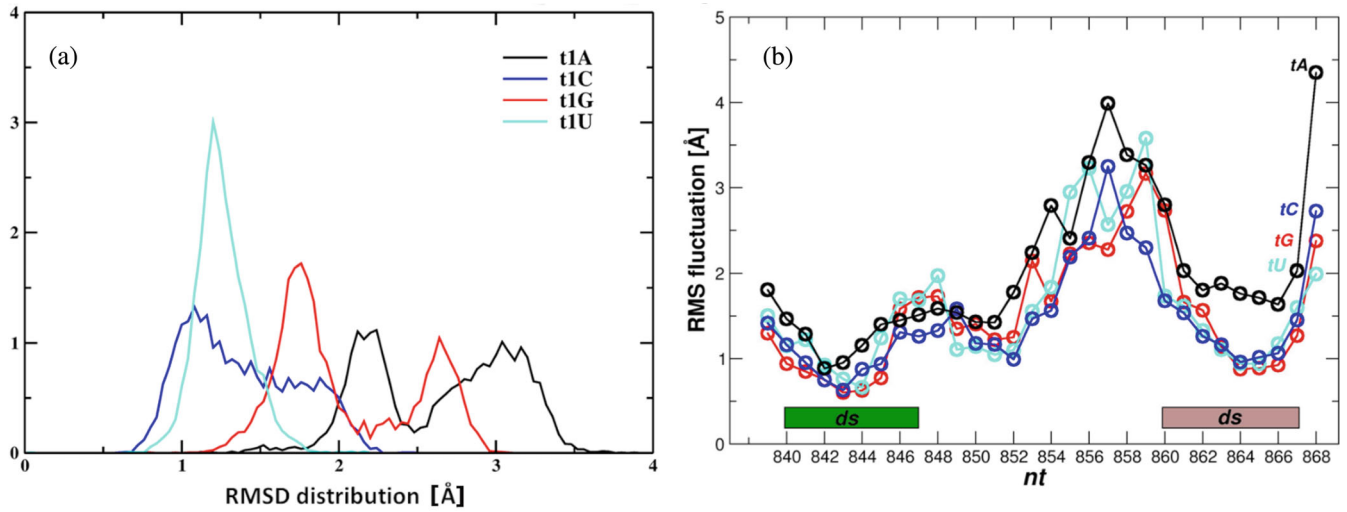

FIGURE 6 Structural flexibility analysis for t1A, t1C, t1G, and t1U systems. Time-averaged (a) RMSD distributions of the full-length target mRNAs and (b) RMSF per nucleotide of the guide–target RNAs. Ds boxes indicate double-stranded RNAs, namely U840::A867; C841::G866; A842::U865; C843::G866; A844::U863; U845::A862; U846::A861; G847::C860

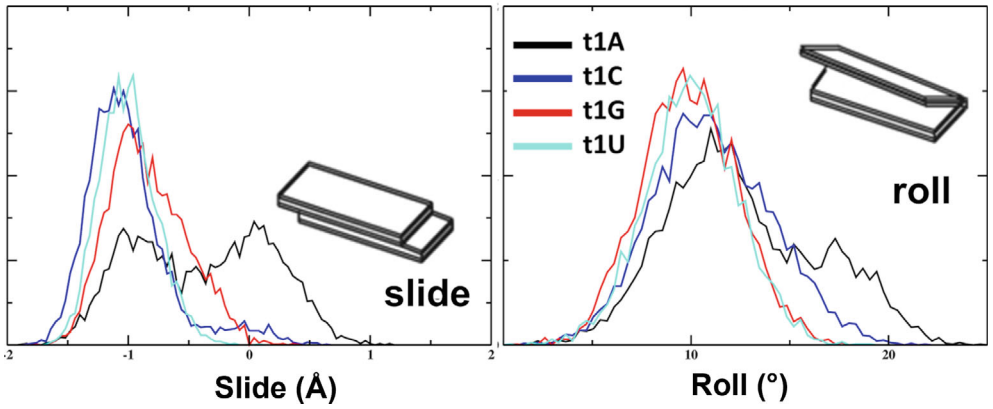

FIGURE 7 Statistical distribution of average slide (Å) and roll (degree) values calculated on double stranded RNAs, namely U840::A867; C841::G866; A842::U865;C843::G866; A844::U863; U845::A862; U846::A861; G847::C860. On the y-axis, values are normalized to 1

(slide = −1.6 Å, roll = +9°[33]), in the presence of t1A a second population is observed characterized by increased values of slide and roll. A more acute base pair inclination allows RNA to adopt a more compact conformation with a narrower major groove (and a shallower minor groove), and consequently reduced helical extension.[34] These slow motions/transitions cannot be exhaustively caught by classic MD techniques and by current force fields.[35] In this framework the application of a comparative strategy is aimed at minimizing eventual caveats. Thus, the analysis of torsional angles at 3′ of the seed reveals characteristic t1A-dependent angles distribution (see Figure S8). In particular, delta angle values were used to discriminate between the C3' and C2' endo puckers, distinctive, respectively, of RNA and DNA conformations. Delta angles analysis was used due to its ability to grasp significant differences, even small ones, that might be missed in a global analysis such as the pucker in line with other similar studies.[36] This analysis demonstrated that a bimodal dihedral distribution is uniquely observed when an adenosine is present at t1.

2.4 | t1 nucleotides impact on protein conformation

The dynamic evolution of the local perturbations of hAgo2 triggered by the different nucleotides at t1 position shows a characteristic modulation by t1A binding. To clarify if a local-to-global conformational effect is detectable, we next analyzed the dynamic pattern of hAgo2 domains. To gain insight into the main protein motions explored in the different cases, we performed a principal component analysis on hAgo2 protein along a unique meta-trajectory, obtained by the concatenation of all the single MD runs from the four systems (see Methods and Figure S9). A common essential space allows to collectively describe and compare (and possibly differentiate)

the specific behavior of each system with respect to the principal motions shared by the four setups, as reported for similar cases.[37,38] In line with previous results, this analysis reveals that, when t1 pocket is occupied by an adenine, hAgo2 explores peculiar conformational ensembles that are not visited by the remaining complexes. The concerted motions associated with the largest fluctuations define a remarkable rearrangement of PAZ-MID domains (Figure S9). Specifically, in Figure 8, the distributions of the distance between the center of mass and the angle spanned by PAZ and MID are plotted. This characterization demonstrates that the presence of the t1A favors the opening of the two domains, stabilizing a conformation where their reciprocal tilt is reduced. This in turn shifts PAZ toward the N-domain and impacts on the N-PAZ channel where the guide is anchored to the 3′.

To dissect protein regions whose internal dynamics could be altered in the presence of t1A as well as pinpoint distinctive elements, the study of long-range communication networks was computed by means of distance fluctuations (DF) between residue pairs: the results of the analysis can be represented as matrices.[39,40] Any two residues are defined to be quasi-rigidly coordinated if their DF is low (dark color in Figure 9). Generally, when a purine is bound to the RNA skeleton at t1 position, the coupling of L2 and PAZ dynamics is increased. This pattern is remarkably enhanced in the presence of t1A nucleotide (white square in Figure 9). This is particularly interesting for the key role played by α7-helix (Figure 8). This helix belongs to L2 domain which is adjacent to PAZ and represents a crucial hotspot for target recognition that propagates the conformational changes from the duplex to the protein domains.[9,41] α7-helix forms with PAZ a single dynamics body that moves in concert with respect to the rest of the protein and behaves as a molecular wedge interacting with the minor

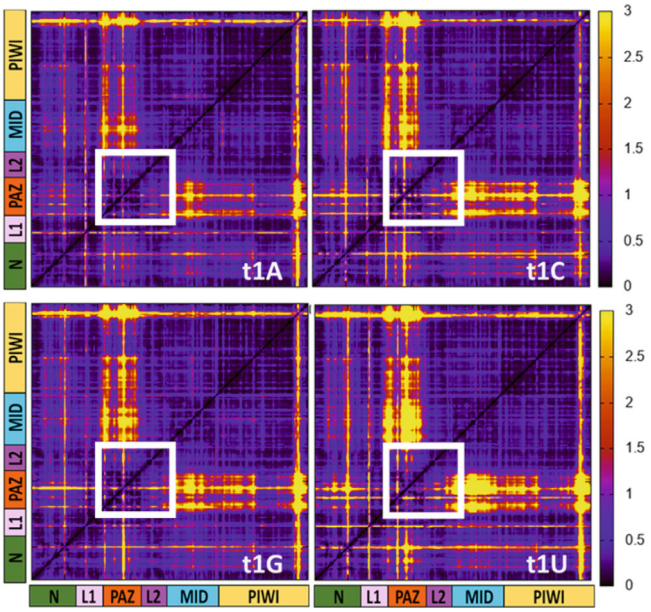

FIGURE 9 DF matrices of t1A, t1C, t1G, t1U complexes. The white square highlights PAZ-L2 domains. The yellow and dark blue areas in the matrices are associated with flexible and rigid regions, respectively. The color bar on the right reports the intensity (Å^2) of the fluctuations

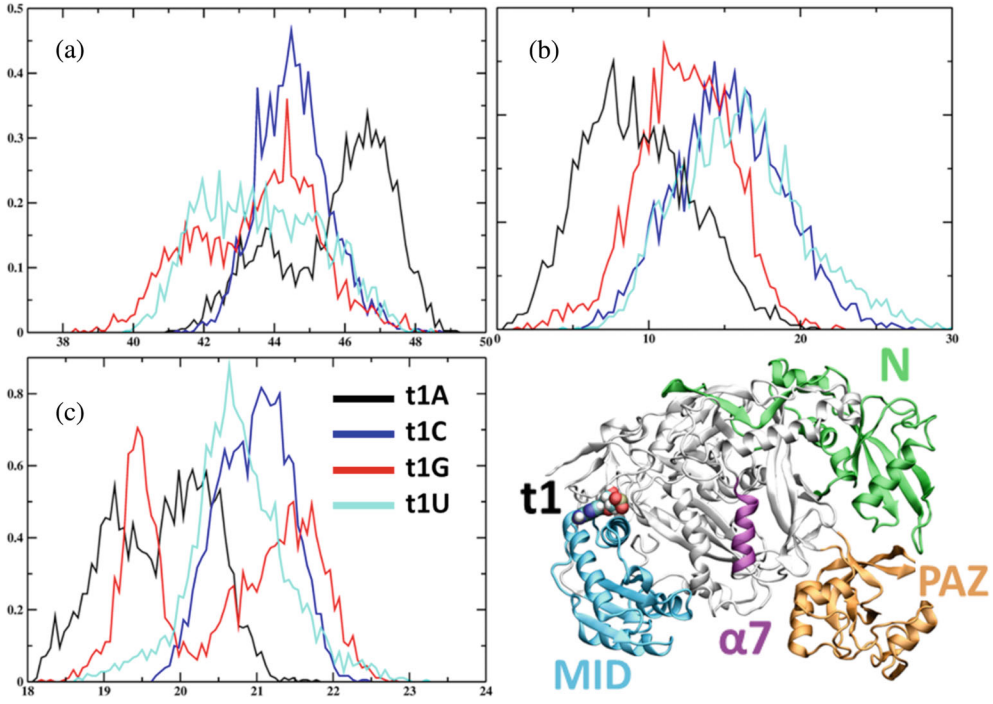

FIGURE 8 Statistical distribution of geometrical parameters along hAgo2 simulation time. (a). Distances (Å) between the center of mass (COM) calculated on MID and PAZ domain. (b). Angle (degree) between axes spanning MID and PAZ domains. (c). Distances (Å) between the centers of mass calculated on N and PAZ domain. On the y-axis, values are normalized to 1

groove formed by the target and guide double strand RNA by numerous interactions with the seed region.

3 | DISCUSSION

hAgo2 specifically recognizes adenosine nucleotides by a solvent exposed pocket defined by the MID and L2 domains with almost three-fold higher affinity than equivalent targets with U, G, or C t1 nucleotides.[9] The mechanistic reason behind this nucleotide-dependent affinity has been ascribed to a distinctive hydrogen bond network that is stabilized in the presence of an adenine.[21] Here, we used MD simulations to characterize the impact of different nucleotides on the conformational dynamics of both the protein and the nucleic acid ligand. Starting from the characterization of local perturbations tuned by the binding of different nucleotides at t1 position, we aimed to shed light on the differentiating details on internal dynamics that can ultimately be linked to the perturbation of specific functional states and to the triggering of large-scale changes. In this framework, we showed that t1A system adopts a distinctive dynamics pattern. Our findings showed that t1A-hAgo2 explores a wider conformational variability of its binding pocket. As a consequence, t1A is the only nucleobase endowed with a dual binding mode characterized by a dynamic interplay between a T-shape and a co-planar conformation, respectively, orthogonal and parallel to the guide–target duplex. Interestingly, the T-shape conformation is able to fit the omics map of the crystallized protein in the presence of t1A nucleotide as well as of the t1DAP (2,6-diaminopurine) nucleotide. In the latter, the diamino substitutions at N6 and N2 determined a twofold increase of the binding affinity shown by t1A and forced the nucleobase in the same T-shape conformation as t1A.[21] The analysis of the interactions defined by the T-shape conformation with its binding pocket contributed to shed light on its binding specificity. In fact, t1A T-shape is the only conformation that forms a stable network of hydrogen bonds involving Asp437 and Asp480. In particular, the N6 amine in Asp480 has been reported to play a role in t1A specificity as its methylation impairs target recognition.[21] Interestingly, these hotspots are largely conserved among Argonaute proteins from different organisms. Despite the similar molecular structure, the adoption of the T-shape conformation by t1 guanine is hampered by an increased steric hindrance and electrostatic repulsion, helping to explain the selective binding of the two purines. The specificity of the t1A-binding interactions reverberates on the dynamic modulation of the RNA molecules. The presence of t1A favors larger dynamic fluctuations of nucleotides of mRNA-target, in particular on peripheral positions (t1, t7-t9), that have shown to act as hotspots that mediate the recognition process.[4,10] On one hand, as already discussed, t1 binding affinity affects target specificity. On the other hand, t9 (that in our structures, as it is in the starting PDB, is the last modeled nucleotide of the target sequence, see Figure S7) belongs to a paired region proven to be bulged by clashes within hAgo2 cleft. The conformational flexibility of the aforementioned region is an additional mechanism hAgo2 exploits to tune RISC activity.[14,42] Indeed, full complementarity to t9 and t10 can actively destabilize the RISC pairing, likely by the selective stabilization of unfavorable conformations. However, the nature of the guide–target interaction remains to be clarified.[5,13] Our data suggest that t1A binding may favor an increased flexibility of the central region, that in turn assists the seed association. We showed that this pattern results in t1A-distinct rearrangements of RNA target as demonstrated by the distributions of both slide and roll parameters and of torsional angles of 3' end nucleotides. Only when an adenine is a t1 position, the double strand can adopt a wider range of conformations, spanning from a regular A-like form to a more bent helix, where the interactions with the protein might be maximized thanks to a shallower and more accessible minor groove, that is, the major mode of recognition exploited by hAgo2.[9] Interestingly, this is consistent with recent reports, where the target-guide pairing at the seed region is divided into two functional ends: a highly complementary 5′ end (g2-g5) that assumes a rigid A-form to promote the WC pairing, and a more dynamic and flexible 3′ of seed (g6-g8) that moves between kinked and A-form conformations,[41] stabilized by specific interaction of the surrounding protein with the minor groove of the double strand. Finally, a local-to-global conformational tuning modulated by the nature of t1 nucleotide is detectable. When t1 pocket is occupied by an adenine, hAgo2 adopts peculiar conformations that stabilize the opening of PAZ and MID domains which reverberates on the N-PAZ channel where the guide is anchored. Moreover, DF analysis shows that t1A enhances the dynamics coupling of L2 and PAZ domains, impacting on the motion α7, a central helical segment of L2 that is reported to accelerate target binding.[41] α7 motion is, indeed, necessary to avoid steric clashes that prevent guide–target pairing in the seed downstream region, thus its removal reduces both target binding and release rates. Experimental data on hAgo2 mutants have shown that an increased mobility of α7 likely uncouples its motions from the PAZ domain and L2 stalk and this ultimately leads to a structural disorganization of the 3′ half of the unpaired seed.[41] We demonstrated that the presence of t1A enhances the internal coordination among these two protein regions

and corresponds to an improved mechanical rigidification of distal domains. Here, the t1A-induced effect could be explained by either the stabilization of t1 local shell of coordination, by the structural modulation of the downstream ds-RNA, or a combination of the two.

4 | CONCLUSIONS

MicroRNAs regulate gene expression by guiding the Argonaute containing RNA-induced silencing complex (RISC) to specific target mRNA molecules. The stable association of RISC to the correct target mRNA, that is, the target binding affinity, is the key factor that drives hAgo2 high silencing efficiency and on-target specificity.[10] Besides Watson–Crick base-pairing at the 5′ miRNA seed region, hAgo2 exploits a combination of mechanisms to fine tune miRNA–target interactions. This can enhance target-guide affinity/specificity. Recently, it has been reported that an ubiquitous adenosine nucleotide in the t1 position of mRNA target (t1A) plays a crucial role in this extra-seed regulation of the target binding, regardless to the identity of miRNA-guide nucleotide-1. Indeed, t1 adenine is able to directly interact with a specific binding pocket within hAgo2 proteins. Herein, using extensive molecular mechanics-derived investigations, we depicted the atomistic and dynamics mechanism behind t1A-dependent regulation of target binding affinity. By using a bottom-up approach, we were able to associate small local perturbations derived by the substitution of a single chemical moiety (i.e., replacing the nature of the t1 nucleotide) to global motions and structural rearrangements of the surrounding macromolecules. The resulting outcomes indicated that t1A binding to hAgo2 pocket acts as a molecular trigger that activates a complex network of both structural and dynamics interactions that appear tightly interdependent. Consistent with X-ray findings, we observed that the presence of t1A in t1-binding pocket uniquely activates a specific H-bonds network, previously reported to be the determinant of t1A specificity with respect to the other nucleotides. In addition, our data suggest a possible additive mechanism of regulation that also exploits differential nucleotides conformations to modulate target recognition, in line with other examples of RNA-proteins interplay reported in literature.[24–26] We demonstrated that the t1A-specific conformations are used by hAgo2 for a dynamic recognition of RNA substrates, that ultimately tune the target affinity. Indeed, although t1 pocket is able to stably bind all the four nucleotides, adenosine is the only nucleobase endowed with a dual binding mode. To our opinion, this conformational enrichment carries a supplementary level of selection among different dynamical ensembles that may explain the increased dwell time associated to t1A nucleotides on target sites experimentally observed.[10] In this framework, the selective stabilization of the T-shape conformation of t1A promotes local-to-global dynamical rearrangements on both the full-length guide–target RNA and the hAgo2 protein. Accordingly, our findings show that the t1A system adopts a unique functional mechanism, characterized by patterns of interactions and modulations that ultimately determine the onset of the sequential conformational changes required by the fully active hAgo2 to trigger efficient RNA silencing. We believe that our study offers an additional point of view upon guide–target communication during RISC regulation and its relevance for gene silencing processes.

5 | METHODS

5.1 | hAgo2–guide–target building

The starting X-ray structure of human Argonaute-2 protein bound to the guide–target duplex (838 residues) was retrieved from the Protein Data Bank. Specifically, coordinates from the most complete structure (PDB entry 4Z4C) of this data collection was selected to build all simulated complexes. In the present structure, hAgo2-guide complex is bound to target RNA, bearing t1C. Protein refinements were carried out using Maestro suite.[43]

In the seed region, nucleotides 840–847 of the guide RNA pair to 867–860 of the target RNA.

Guide RNA: 5′ P-UUCACAUUGCCCAAGUCUCUU 3′;

Target RNAs: t1C → 5′ CAAUGUGAC 3′; t1U → 5′ CAAUGUGAU 3′; t1A → 5′ CAAUGUGAA 3′; t1G → 5′ CAAUGUGAG 3′.

5.2 | MD settings

The MD simulation package Amber[44] v12 was used to perform computer simulations by applying the Amber-ff99SB force field.[45] The systems were solvated, in a simulation box of explicit water molecules (TIP3P model),[46] counterions were added to neutralize the system, and periodic boundary conditions were imposed in the three dimensions. After minimizations, systems were subjected to an equilibration phase where water molecules and protein heavy atoms were position restrained, and then, unrestrained systems were simulated in a NPT ensemble; a Berendsen thermostats was used to keep constant temperature (300 K) and pressure (1 atm). Electrostatic energies were evaluated by the particle mesh Ewald method[47] and Lennard-Jones forces by a cut-off of 8 Å. All bonds involving hydrogen atoms were constrained

using the SHAKE algorithm.[48] An ionic strength of 0.150 M NaCl was reproduced based on experimental settings.

5.3 | Analyses

Principal component analysis (PCA) has been applied to MD simulations in order to extract functionally relevant movements.[49] The largest collective fluctuations that account for the largest conformational variation were recovered by the principal eigenvectors (essential modes) of the covariance matrix of the given dynamic ensemble. PCA analysis was carried out on Cα atoms of hAgo2 protein along a meta-trajectory, resulting from the concatenation of the t1A, t1C, t1G, t1U simulations.

Cluster analysis was carried out using the *g_cluster* module of Gromacs[50] to evaluate the structural effects induced by the binding of the different nucleotides on protein-RNAs complexes. Clustering of the trajectories was obtained with the *gromos* method fitting backbone atoms of the residues shaping the 4 Å t1A binding pocket (RMSD cutoff of 0.15 nm). On the same cavity, hydrogen bonds analysis was performed by means of the *ptraj* module of AMBER12. Each residue of t1A-binding cavity that established a hydrogen bond for at least 5% of the whole simulation time was reported.

To study the reverberation of the t1-dependent local perturbations on protein global motions and structural rearrangements, the evolution of the distance between each hAgo2 domains centroids was analyzed by the VMD tools package in the t1A, t1C, t1U and t1G systems.[51] Centroids were defined as the center of mass of the different hAgo2 domains, namely N-domain (aa. 36–166), L1 (aa. 176–226), PAZ (aa. 231–365), L2 (aa. 374–420), MID (aa. 429–511), and PIWI (aa. 496–797). Similarly, the angles distribution between the axes of inertia spanning each domain along the simulated systems was analyzed by the VMD tools package.

To correlate conformational rearrangements with the internal dynamics of each protein system, distance fluctuations analysis was used to describe the dynamic coordination between any two residues.

Distance fluctuation DF_{ij} is defined as the time-average mean square fluctuation of the distance r_{ij} between Cα atoms of residues i and j:

$$DF_{ij} = \left\langle \left(\left(r_{ij} - \langle r_{ij} \rangle \right)^2 \right) \right\rangle$$

where brackets indicate the time-average over the trajectory. Low DF values indicate highly coordinated residues.[52]

Finally, RNA structural analyses and parameters were obtained by using the *ptraj* module of AMBER12.

Figures were analyzed and created using VMD, Pymol, and Xmgrace.[51,53,54]

AUTHOR CONTRIBUTIONS

Silvia Rinaldi: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); project administration (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Giorgio Colombo:** Conceptualization (supporting); funding acquisition (lead); resources (lead); software (lead); writing – original draft (supporting). **Antonella Paladino:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); project administration (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENT

Open Access Funding provided by Consiglio Nazionale delle Ricerche within the CRUI-CARE Agreement.

ORCID

Silvia Rinaldi https://orcid.org/0000-0002-1088-7253
Antonella Paladino https://orcid.org/0000-0002-9397-1572

REFERENCES

1. Nakanishi K. Anatomy of RISC: How do small RNAs and chaperones activate Argonaute proteins? Wiley Interdisciplinary Reviews: RNA, 2016;7(5):637–660.
2. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell. 2004;15(2):185–197.
3. Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature. 2008;456:921–926.
4. Wang Y, Juranek S, Li H, et al. Nucleation, propagation and cleavage of target RNAs in ago silencing complexes. Nature. 2009;461(7265):754–761.
5. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell. 2012;151(5):1055–1067.
6. Elkayam E, Kuhn CD, Tocilj A, et al. The structure of human argonaute-2 in complex with miR-20a. Cell. 2012;150(1):100–110.
7. Schirle NT, MacRae IJ. The crystal structure of human argonaute2. Science (80-). 2012;336(6084):1037–1040.
8. Chandradoss SD, Schirle NT, Szczepaniak M, Macrae IJ, Joo C. A dynamic search process underlies MicroRNA targeting. Cell. 2015;162:96–107.

9. Schirle NT, Sheu-Gruttadauria J, MacRae IJ. Structural basis for microRNA targeting. Sci. 2014;346(6209):608–613.
10. Sheu-Gruttadauria J, Xiao Y, Gebert LF, MacRae IJ. Beyond the seed: Structural basis for supplementary micro RNA targeting by human Argonaute2. EMBO J. 2019;38(13):e101153.
11. Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. Mol Cell. 2007;27(1):91–105.
12. Moore MJ, Scheel TKH, Luna JM, et al. MiRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. Nat Commun. 2015;6(May):1–17.
13. Salomon WE, Jolly SM, Moore MJ, Zamore PD, Serebrov V. Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. Cell. 2015;162(1):84–95.
14. Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. Pairing beyond the seed supports MicroRNA targeting specificity. Mol Cell. 2016;64(2):320–333.
15. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4:1–38.
16. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15–20.
17. Becker WR, Ober-Reynolds B, Jouravleva K, Jolly SM, Zamore PD, Greenleaf WJ. High-throughput analysis reveals rules for target RNA binding and cleavage by AGO2. Mol Cell. 2019;75(4):741–755.e11.
18. Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. RNA. 2007;13(11):1894–1910.
19. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 2008;455(7209):64–71.
20. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 2008;455(7209):58–63.
21. Schirle NT, Sheu-Gruttadauria J, Chandradoss SD, Joo C, MacRae IJ. Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. Elife. 2015;4:e07646.
22. Julien KR, Sumita M, Chen PH, Laird-Offringa IA, Hoogstraten CG. Conformationally restricted nucleotides as a probe of structure function relationships in RNA. RNA. 2008;14(8):1632–1643.
23. Wilson R, Doudna JA. Molecular mechanisms of RNA interference a BIOLOGICAL VIEW OF RNA INTERFERENCE • small regulatory RNAs in cellular function and dysfunction HHS public access. Annu Rev Biophys. 2013;42:217–239.
24. Kligun E, Mandel-Gutfreund Y. The role of RNA conformation in RNA-protein recognition. RNA Biol. 2015;12(7):720–727.
25. Daubner GM, Cléry A, Jayne S, Stevenin J, Allain FH-T. A *syn - anti* conformational difference allows SRSF2 to recognize guanines and cytosines equally well. EMBO J. 2012;31(1):162–174.
26. Krepl M, Dendooven T, Luisi BF, Sponer J. MD simulations reveal the basis for dynamic assembly of Hfq-RNA complexes. J Biol Chem. 2021;296:100656.
27. Kawamata T, Tomari Y. Making RISC. Trends Biochem Sci. 2010;35:368–376.
28. Meister G, Landthaler M, Peters L, et al. Identification of novel argonaute-associated proteins. Curr Biol. 2005;15(23):2149–2155.
29. Tsuboyama K, Tadakuma H, Tomari Y. Conformational activation of Argonaute by distinct yet coordinated actions of the Hsp70 and Hsp90 chaperone systems. Mol Cell. 2018;70:722–729.e4.
30. Rinaldi S, Colombo G, Paladino A. Mechanistic model for the Hsp90-driven opening of human Argonaute. J Chem Inf Model. 2020;60(3):1469–1480.
31. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31(6):926–932.
32. Zacharias M, Sklenar H. Conformational deformability of RNA: a harmonic mode analysis. Biophys J. 2000;78(5):2528–2542.
33. Liu JH, Xi K, Zhang X, Bao L, Zhang X, Tan ZJ. Structural flexibility of DNA-RNA hybrid duplex: stretching and twist-stretch coupling. Biophys J. 2019;117(1):74–86.
34. Lavery R, Sklenar H. Defining the structure of irregular nucleic acids: conventions and principles. J Biomol Struct Dyn. 1989 Feb [cited 2021 Dec 14;6(4):655–667.
35. Sponer J, Bussi G, Krepl M, et al. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. Chem Rev. 2018;118(8):4177–4338.
36. Paladino A, Zangi R. Propensities for loop structures of RNA & DNA backbones. Biophys Chem. 2013;180–181:110–118.
37. Rinaldi S, Assimon VA, Young ZT, et al. A local allosteric network in heat shock protein 70 (Hsp70) links inhibitor binding to enzyme activity and distal protein-protein interactions. ACS Chem Biol. 2018;13(11):3142–3152.
38. D'Abrosca G, Paladino A, Baglivo I, et al. Structural insight of the full-length Ros protein: a prototype of the prokaryotic zinc-finger family. Sci Rep. 2020;10(1):1–10.
39. Paladino A, Civera M, Belvisi L, Colombo G. High affinity vs. native fibronectin in the modulation of αvβ3 integrin conformational dynamics: insights from computational analyses and implications for molecular design. PLoS Comput Biol. 2017;13(1):e1005334.
40. Rinaldi S, Gori A, Annovazzi C, Ferrandi EE, Monti D, Colombo G. Unraveling energy and dynamics determinants to interpret protein functional plasticity: the Limonene-1,-2-epoxide-hydrolase case study. J Chem Inf Model. 2017;57(4):717–725.
41. Klum SM, Chandradoss SD, Schirle NT, Joo C, MacRae IJ. Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition. EMBO J. 2018;37(1):75–88.
42. Baronti L, Guzzetti I, Ebrahimi P, et al. Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. Nature. 2020;583(7814):139–144.
43. Schrödinger Release 2016–4. New York, NY: Schrödinger, LLC, 2016.
44. Case DA, Darden TA, Cheatham TE III, et al. Amber 12. University of California, San Francisco, 2012.
45. Lindorff-Larsen K, Piana S, Palmo K, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins Struct Funct Bioinforma. 2010;78(8):1950–1958.
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926–935.

47. Darden T, York D, Pedersen L. Particle mesh Ewald: An W log(N) method for Ewald sums in large systems. J Chem Phys. 1993;98(12):10089–10092.
48. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23(3):327–341.
49. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins Struct Funct Genet. 1993;17(4):412–425.
50. 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(3):435–447.
51. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33–38.
52. Morra G, Neves MAC, Plescia CJ, et al. Dynamics-based discovery of allosteric inhibitors: selection of new ligands for the C-terminal domain of Hsp90. J Chem Theory Comput. 2010;6(9):2978–2989.
53. L DeLano W. The PyMOL molecular graphics system (2002). Palo Alto, CA: DeLano Scientific, 2002. http://www.pymol.org.
54. Turner: XMGRACE, Version 5.1. 19.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rinaldi S, Colombo G, Paladino A. The dynamics of t1 adenosine binding on human Argonaute 2: Understanding recognition with conformational selection. Protein Science. 2022;31(8):e4377. https://doi.org/10.1002/pro.4377