Sampling the Structure of the Noncanonical Lin-4:Lin-14 microRNA:mRNA Complex by Molecular Dynamics Simulations

Chandramouli Balasubramanian,[†,‡] **Rajendra Prasad Ojha,**[§] **Souvik Maiti,***[†] **and Alessandro Desideri***[,‡]

Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India; Department of Biology, University of Rome "Tor Vergata" Via della Ricerca Scientifica 1, 00133, Rome, Italy; and Biophysics Unit, Department of Physics, DDU Gorakhpur University, Gorakhpur 273 009, India

siRNA and microRNA (miRNA) are two classes of noncoding RNAs that carry out post-transcriptional gene regulation by interacting with the target mRNA. The structural features of siRNA/miRNA−mRNA complex play a crucial role in gene silencing. Here we have used computer modeling and simulation approach to (i) elucidate the possible structures of the partially complementary lin-4:lin-14 miRNA-mRNA complex and (ii) compare the structural features with the fully complementary lin-4:lin-14 siRNA-mRNA complex. The fully complementary siRNA complex maintains a canonical helical shape while the presence of mismatch base pairs and bulge induce a kinked structure in the miRNA complex. Docking of the simulated duplex models on to the argonaute protein gives a direct indication that in the siRNA−mRNA structure, the scissile phosphate of the mRNA strand is more accessible to the catalytic site as compared to the miRNA−mRNA structures, providing an explanation for the less prevalence of the cleavage in the later case.

Introduction

Small RNA-mediated gene regulatory pathways, broadly referred to as RNA interference (RNAi) or RNA silencing, have gained a great deal of attention in the recent past as they have a crucial role in several cellular processes in a multitude of organisms.[1−3] The three basic steps of RNAi include (i) the production of a small RNA duplex (∼20−30 nucleotides) from long dsRNA precursors by a RNaseIII enzyme Dicer;[4−7] (ii) incorporation of one strand of the small RNA duplexes into a protein complex known as the RISC (RNA induced silencing complex); and (iii) preferential cleavage of a target mRNA by an endonuclease called argonaute in the RISC assembly directed by the loaded RNA strand (usually referred to as the guide strand) that acts as a specificity determinant.[8,9] The two primary classes of small RNAs that are involved in RNAi are the short interfering RNA (siRNA) and microRNA (miRNA). Although both types of molecules can be functionally equivalent, they are distinguished by their modes of biogenesis.[10−13]

The key issue in the post-transcriptional gene silencing by miRNA/siRNA is the specificity of their interaction with the target mRNA that can lead to different downstream consequences. siRNA is typically known to guide the RISC to perfectly complementary mRNA targets, which are then cleaved by the argonaute protein.[14−16] This process is defined as "slicer" activity and the mRNA is precisely cleaved at the phosphodiester linkage between the bases that are paired to the 10 and 11 bases from the 5′ end of the siRNA.[17−22] As gene regulators, miRNAs

work by two different modes and the degree of miRNA−mRNA complementarity has been deemed as a determining factor of the regulatory mechanism. In plants, miRNAs like siRNAs, form base pairs with perfect or nearly perfect complementarity that leads to the argonaute-catalyzed cleavage of the mRNA strand, but in animals, miRNAs bind with partial complementarity producing mismatches and bulges that do not permit the cleavage by argonaute, promoting the repression of mRNA translation.[23−25] There are numerous studies on the functional characterization of the miRNAs, but limited information is presently available on the structural features of the miRNA−mRNA duplexes. To our knowledge, only the NMR structure of the let-7:lin-41 (miRNA−mRNA) mimic has been reported so far,[26] likely because the partial complementarity gives a high conformational flexibility that can hurdle the experimental determination of their structures.

In the present work, we have used molecular modeling and a simulation approach to investigate the structural features of the well studied and experimentally validated lin-4:lin-14 pair from *C.elegans*. The lin-4 miRNA has been the first member of the miRNAs family that has been shown to regulate the timing of gene expression during *C.elegans* development.[27] The lin-4 miRNA interacts with the 3′-UTR of lin-14 mRNA transcript that contains seven binding sites with the well-conserved 9nt core element CUCAGGGAA, complementary to the 5′ end bases of the lin-4 miRNA. The lin-4 miRNA acts as a repressor, negatively regulating the lin-14 mRNA, by repressing the translation of the lin-14 transcript during the end of the first larval stage L1.[28,29] We constructed a three-dimensional model for the lin-4:lin-14 duplex as depicted schematically in Figure 1, following the coupling described in the Tarbase database, which is a repository of experimentally validated miRNA−mRNA pairs.[30] The lin-4:lin-14 pair has a characteristic internal loop with G:A and C:A mismatch base pairs. Mismatches can form alternative patterns of hydrogen bonds

* To whom correspondence should be addressed. Tel: +39-06-72594376; Fax: +39-06-2022798; E-mail: desideri@uniroma2.it (A.D.). Tel: +91-11-2766-6156; Fax: +91-11-2766-7471; E-mail: souvik@igib.res.in (S.M.).
† Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR.
‡ Department of Biology, University of Rome "Tor Vergata" Via della Ricerca Scientifica.
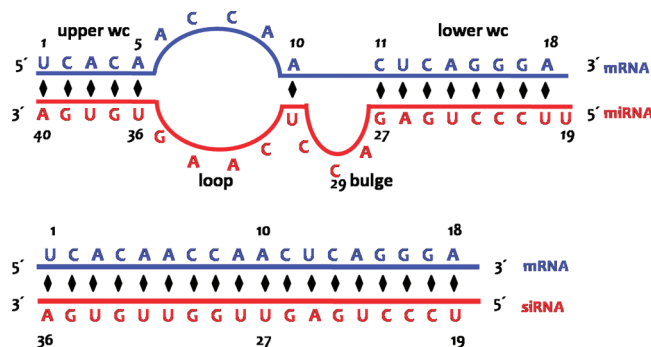§ Biophysics Unit, Department of Physics, DDU Gorakhpur University.

Figure 1. Schematic picture of the partially complementary miRNA:mRNA (lin4:lin14) duplex (top) and of the complete complementary siRNA:mRNA duplex (bottom). The mRNA strand is shown in blue, while the miRNA/siRNA strand is shown in red. The diamond symbol indicates the Watson−Crick base pairs.

that have been taken in to consideration by carrying out simulations on starting miRNA models (referred to as A−D in the text) differing in their hydrogen bond patterns. An additional simulation has been done on the fully complementary lin-4:lin-14 duplex (referred to as siRNA duplex), that has been constructed by considering the mRNA sequence as a template strand (Figure 1).

The simulation results show that the presence of mismatches and bulges greatly perturbs the global geometry of the miRNA−mRNA duplex structures. All of the simulated miRNA duplexes take up a kinked global structure. In contrast, the complementary siRNA model remains in a canonical helical geometry, due to the stability of the W−C base pairing. The local distortion induced by the presence of mismatched base pairs in the miRNA duplexes affects the groove dimensions and the stacking of bases. The docked binary complex obtained by docking the simulated RNA duplex and the argonaute protein indicates that the presence of non-canonical secondary structure elements alters the accessibility of the scissile phosphate of the mRNA strand from the catalytic site of the protein, providing an explanation for the absence of cleavage.

Materials and Methods

Duplex Modeling and Simulation Setup. The starting structure of the miRNA duplex was modeled based on the schematic binding picture of lin-4:lin-14 (miRNA−mRNA) complex, shown in Figure 1 (top panel) as adapted from Tarbase database.[30] Initial modeling was performed using the Insight II

modeling package (Accelrys Inc., San Diego, CA). The internal loop in the miRNA complex (Figure 1), has a G:A and several C:A mismatch base pairs, that can form alternative hydrogen bonds in reality.[31] The possible hydrogen bonding schemes for the G:A and C:A base pairs with at least two hydrogen bonds, is shown in Figure 2. For G:A mismatch, there are four such schemes that include (i) GO6-AN6, GN1-AN7 (Figure 2A); (ii) GO6-AN6, GN1-AN1 (Figure 2B); (iii) GN3-AN6, GN2-AN7 (Figure 2C); and (iv) GN3-AN6, GN2-AN1 (Figure 2D). The C:A mismatch has two such schemes (i) CN3-AN6, CN4-AN7 (Figure 2E); and (ii) CN3-AN6, CN4-AN1 (Figure 2F).

All of the possible hydrogen bonding combinations of G:A and C:A mismatch base pairs and the presence of three C:A mismatch pairs demand the generation of 32 different models. In order to reduce the computational cost, we decided to concentrate our attention on four models, with the three C:A base pairs constrained to have the identical CN3-AN6 hydrogen bond, while the G:A base pair is left to have all the possible hydrogen bonding patterns, giving rise to four starting models (referred as A, B, C, and D in the text), with model A representing the hydrogen bonding scheme of G:A base pair defined in Figure 2A, model B depicting the pattern in Figure 2B, models C and D, depicting the schemes in Figure 2C and 2D, respectively. We have then carried out four independent simulations over the different models. A fifth simulation has also been carried out on the fully complementary siRNA duplex (Figure 1, bottom panel), modeled using nucgen program in Amber9.[32]

Simulation Protocol. All simulations were performed using sander module in Amber9 with ff99 force field.[33] The RNA duplexes have been neutralized with Na^+ ions. The neutralized systems were immersed in TIP3P[34] tetrahedron water box with a dimension extending up to 8 Å from the solute. The periodic boundary condition has been applied using the particle mesh Ewald (PME) method to treat long-range electrostatics.[35] Hydrogen bonds have been constrained using SHAKE.[36] A time step of 2 fs and a direct-space nonbonded cutoff of 10 Å were used. All simulations were performed at a constant temperature of 300 K. The Langevin coupling with a collision frequency of 1.0 has been used for temperature regulation.[37] The Berendsen's barostat was used to maintain a constant pressure (1 atm) with a relaxation time of 2 ps.

The equilibration procedure involves (1) three rounds of minimizations (1500 iteration each) and dynamics (30 ps each) of solvent and counterions with the solute constrained with a force constant of 500, 100, 300, 50, 100, 25 kcal/(mol Å²); (2)
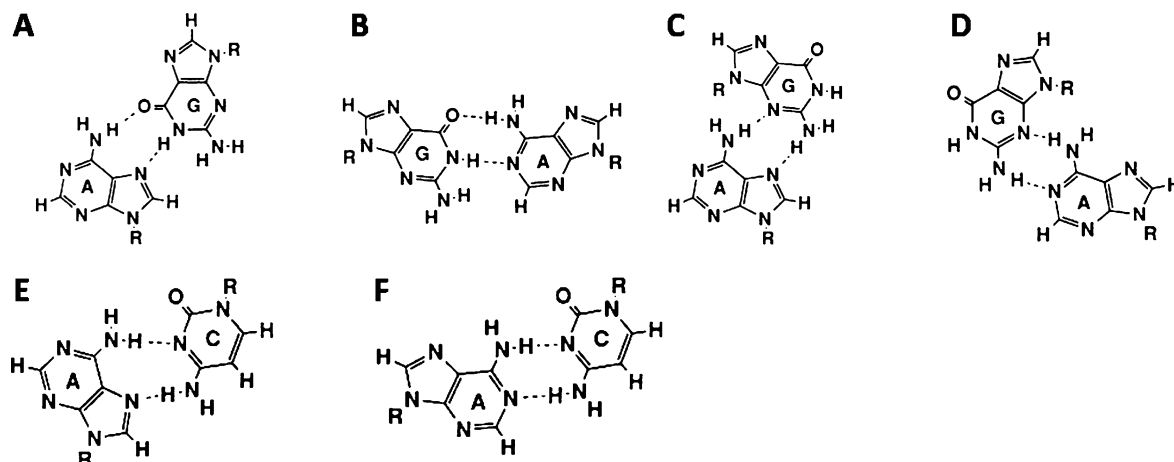

Figure 2. Possible hydrogen bonding schemes of G:A (A−D) and C:A (E,F) mismatch base pairs with at least two hydrogen bonds.

Sampling the Structure of Lin-4:Lin-14 miRNA:mRNA

J. Phys. Chem. B, Vol. 114, No. 49, 2010 **16445**

two rounds of 3000 steps of minimization of whole system where the solute restraint has been kept as 15 and 5 kcal/(mol Å^2); and (3) an unrestrained minimization of the whole system. Finally, the system was heated to 300 K at constant volume and has been equilibrated for 2 ns at a constant pressure of 1 atm and a temperature of 300 K, restraining the distances and angles on the G35:A6 and C:A base pairs. After equilibration, the restraints have been removed and the snapshots in the trajectory have been collected at an interval of 2 ps.

Analysis of Trajectories. All of the analysis on the RNA trajectories has been carried out on the last 7 ns of the trajectory. The root-mean square deviation (rmsd), two-dimensional root-mean square deviation (2drms), hydrogen bonds analysis, and interstem angle calculations were done using the Ptraj program of Amber9.[32] The hydrogen bond calculation has been done by considering a donor−acceptor distance lower than 3.5 Å. The interstem angle was calculated by defining three centers of mass that includes the nucleotide segments (i) 3−4/37−38 in the upper WC region, (ii) 10,31 at the center, and (iii) 16−17/21−22 in the lower WC region, respectively. Clustering analysis was carried out with the K-clust program of MMTSB toolset.[38] A clustering radius of 3 Å was chosen after testing of different radii (from 1 Å to 4 Å). Smaller radii resulted in a number of less populated clusters and larger radii produced single cluster. The snapshot having the least rmsd from the cluster centroid were chosen as the representative configuration of the cluster. The solvent accessibility calculation was done with the Naccess program.[39] The structural parameters were calculated with a curves program.[40] The figures were generated with UCSF chimera software[41] and plots were generated using matplotlib library.[42]

RNA−Argonaute Modeling. The crystal structure of full length *T. thermophilus* argonaute protein (pdb id: 3f73) crystallized with a 21 base pair DNA−RNA duplex has been used to generate the binary complex.[43] We chose the representative configuration of the siRNA and all of the four miRNA duplexes belonging to the most populated cluster (cluster gray in all cases) to generate the binary complex with the argonaute protein. Since automatic docking of such a huge RNA duplex is not computationally feasible, we utilized the DNA−RNA duplex present in the crystal structure as a template to fit the representative conformers from the simulations within the argonaute protein. Before docking, the RNA duplexes have been truncated after the first 12 base pairs. The binary models have been subjected to energy optimization and dynamics in explicit water and counterion. The simulation protocol is explained in the Supporting Information.

Results

RNA Duplex Simulations: rmsd and Cluster Analysis. The stability of the simulations has been monitored by calculating the root-mean square deviation (rmsd) of the conformations evolved during the dynamics from both the equilibrated and the average structures. The heavy atom rmsd as a function of simulation time is reported in Supporting Information Figure S1. For siRNA duplex (Figure S1,S), the rmsd from both the equilibrated (black) and average structure (gray) is well stabilized, as expected due to the presence of fully complementary W−C base pairs, that help to maintain a stable structure during the dynamics. For the miRNA duplexes (Figure S1, A−D), the rmsd from the equilibrated structure (black line) is larger when compared to that from the average structure (gray line). This result indicates that the evolved conformations are reasonably stable around the average structure. To guarantee
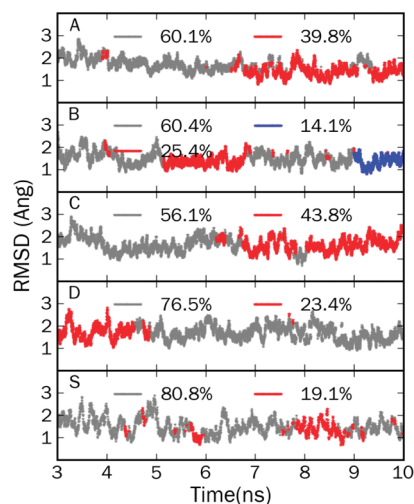

Figure 3. Evolution of conformational clusters for the miRNA (A−D) and siRNA (S) duplexes. The legend values in the plots represent percentage occurrence of each cluster. Panels A−D represent the clustering of miRNA models A, B, C, and D, respectively. Panel S represents the clustering of siRNA duplex.

the investigation over a stabilized part of the trajectory, all of the analysis has been carried out omitting the first three nanoseconds, i.e., over the last 7 ns.

To understand the structural dynamics of the RNA duplexes and to pick up the representative configurations that can best explain the structural characteristics, the conformations in the trajectories were clustered using a rmsd threshold of 3 Å (refer to the Methods section). The evolution of the conformational clusters of the RNA duplexes as a function of simulation time is depicted in different colors in Figure 3, with the percentage of occurrence of each cluster. For the siRNA duplex, two clusters are observed (gray and red in Figure 3, panel S) with the gray cluster present for more than 80%. For all of the miRNA duplexes except model B, two major clusters are observed with the most representative ones (gray cluster) occurring in the range of 56−76% of the total simulation time. In the case of model B, three clusters are observed with the most representative one (gray cluster) present for 60%, the other two being characterized by small percentages (Figure 3, panel B). The evolution of the clusters in general follows a similar trend as that of the rmsd and hence to elucidate structural changes during the dynamics, we chose representative configurations for the miRNA and siRNA duplexes from each cluster. These representatives showed in Figure 4 permits us to outline some qualitative structural features of the duplexes. The fully complementary siRNA duplexes maintain a canonical A-type helical structure (Figure 4, model S), while the miRNA duplexes (Figure 4, models A, B, C, and D) are characterized by a kinked structure, with the upper and lower stems, orienting in different directions with respect to the central part. The upper stem has lost the characteristic helical shape in all miRNA duplexes, while the lower stem retains the helical shape.

Structural Comparison of miRNA Duplexes. To examine the global similarity of the representative configurations obtained from the simulation of different miRNA models, we calculated the pairwise rmsd between the conformers. The 2drms matrix is depicted in Figure 5 as a hinton diagram. The rmsd values <2 Å are shown as red square, while the rmsd >2 Å are shown as black squares. It is clear that no representative conformer from one simulation is close to a representative from another simulation by less than 2 Å. Also, except for the case of the miRNA model A, the representatives from the same simulations
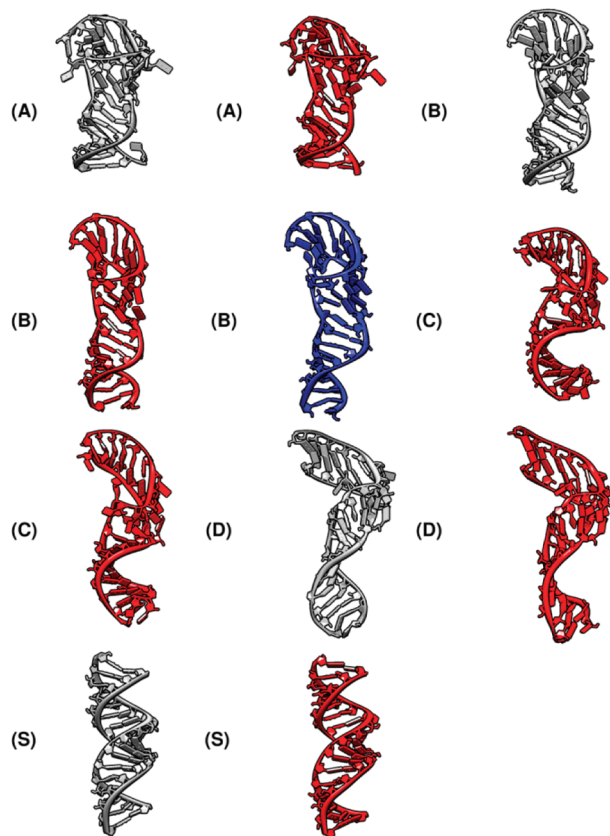
Figure 4. Representative cluster configurations of the miRNA (A−D) and siRNA (S) duplexes. The color of the conformers represents the color of cluster as in Figure 3. The model name before each structure is indicated for clarity.
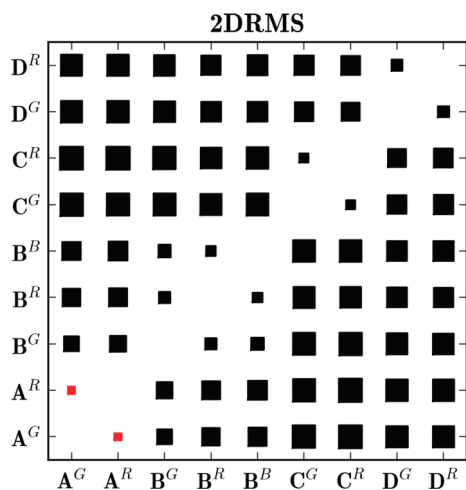

Figure 5. Hinton diagram depicting the pairwise rmsd matrix for the cluster representatives of the miRNA duplex models A−D. Representative configurations having rmsd values <2 Å are shown in red while the rmsd values >2 Å are shown in black. The size of each square in the diagram is proportional to the magnitude of the rmsd. The superscripts above each miRNA model names A to D indicate the cluster color. G-gray, R-red, and B-blue.

have a structural deviation of more than 2 Å. To further characterize the general structural features and base stacking mode, we have calculated the interstand phosphate distance (Figure S2, top panel of the Supporting Information (SI)); base pair (Figure S2, middle and bottom panel of the SI) and base step (Figure S3 of the SI) structural parameters for the representatives of miRNA and siRNA models. Compared to the

TABLE 1: Inter-Stem Angle (Degrees) of the Representative Conformers of the miRNA Models

cluster color	model A	model B	model C	model D
gray	120.3	128.6	143.6	125.5
red	115.8	145.3	149.5	131.9
blue		139.7		

siRNA representatives, the miRNA configurations have smaller interphosphate distances (Figure S2, top SI), the difference being more significant in the upper WC and internal loop regions. The values of base pair and base step parameters for the miRNA models (shown in Figures S2 and S3 of the SI) are different from that of siRNA models and within the four miRNA models, the main difference are localized mainly in the internal loop and the base pairs flanking the loop. In order to have a rough estimate of the difference in the orientation of the stem regions with respect to the central (A10:U31) base pair, the angle between the center of mass of the central base pair and of the upper and lower WC stem regions has been calculated. The plot of the angle as a function of simulation time, reported in Figure S4 of the SI, shows that the value is always less than 180°, indicating that the stem regions strongly deviate from the linear axis. This is confirmed by the average value calculated for the last 7 ns of the trajectories (Figure S4 of the SI) and the interstem angle value calculated for each representative conformer (Table 1).

Solvent Exposure of the Bases. To understand the solvent exposure of the bases, we have calculated the average solvent accessible surface area (sasa) that is correlated to the stacking of bases in the structure. The average sasa values for miRNA duplexes are represented in Figure 6 as a function of base number. For comparison, the sasa value of bases in the siRNA duplex is shown in SI Figure S5. In the case of the siRNA duplex, all of the bases have a uniform solvent exposure (Figure S5 of the SI) of ~120 Å^2 except the bases at the terminal ends, as expected due to the canonical helical shape of the duplex with well-defined grooves. In the case of miRNA models, in addition to the bases at the terminal ends, also some bases forming the bulge and internal loop have higher solvent
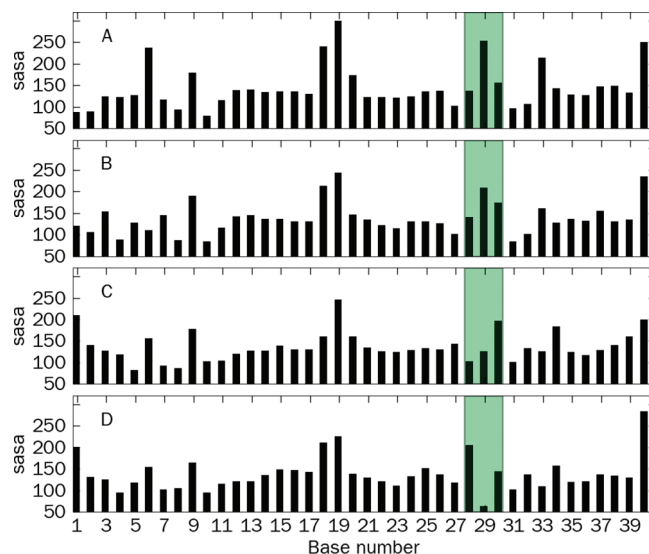

Figure 6. Average solvent accessible surface area (Å^2) as a function of base number calculated for the miRNA models A−D are represented in plots A, B, C, and D, respectively. The green shaded portion indicates the bulge region.
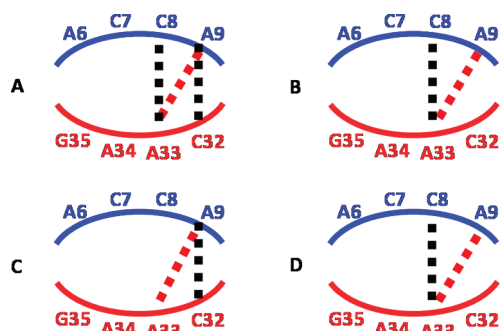
Sampling the Structure of Lin-4:Lin-14 miRNA:mRNA

J. Phys. Chem. B, Vol. 114, No. 49, 2010 **16447**

Figure 7. Schematic picture of the triplet base pairing within the internal loop of the miRNA duplexes. The black dotted line represents the hydrogen bonding of the C:A base pair and the maroon dotted line represents the A9:A33 hydrogen bond interaction. The bases of the mRNA and miRNA strands are shown in blue and red colors, respectively. Letters A−D denote the miRNA models A, B, C, and D.

TABLE 2: Distance (Å) of the Scissile Phosphate from Catalytic ASP Residues[a]

model	ASP478-P		ASP546-P		ASP660-P	
	Min	max	min	max	min	max
siRNA	6.1	6.8	4.9	5.8	5.4	6.1
A	8.5	10.0	7.6	9.4	5.5	6.7
B	6.3	7.2	5.5	6.3	5.1	5.8
C	7.8	9.1	5.3	7.0	8.3	9.6
D	7.1	8.6	5.4	6.8	7.6	9.5

[a] Row 1 corresponds to the siRNA−argonaute complex, Rows 2−5 corresponds to miRNA (A, B, C, D)-argonaute complexes. The minimum and maximum distances were calculated for the trajectory omitting the thermalization phase.

accessibility values. This includes A6, C29 in model A; A9, C29 in model B; A9, C30, A34 in model C; and A28, A34 in model D (Figure 6). The level of solvent exposure of these bases can be visualized in the representative conformers of the miRNA models (SI Figures S6−S9), where these bases protrude from the grooves, suggesting that they can play a role in the recognition of molecular partners.[44]

Hydrogen Bond Analysis. To examine the stability of the base pairs, we have calculated the percentage of occupancy of the hydrogen bonds between the pairing bases in each cluster. For the siRNA duplex, the hydrogen bonds between all of the base pairs are highly stable, their persistence ranging from 97 to 100% of the simulation time (data not shown). For the miRNA duplexes, the percentage of the hydrogen bond occupancy is reported in SI Tables ST1−ST3, for the upper WC, lower WC, and internal loop regions, respectively. These regions can be identified in the schematic diagram of the miRNA duplex depicted in Figure 1 (top panel). In the upper stem, all of the W−C base pairs are stable in models B and D, while end fraying is observed for the terminal base pairs in model A and C (Table ST1 of the SI). Base A40 of the miRNA strand is found to be splayed out of the structure in both cases. The central W−C base pair (A10:U31) is highly stable in models A and D as compared to models B and C. In the latter cases, base A28 in the bulge region of the miRNA strand is involved in hydrogen bonding interaction with A10 of mRNA strand. In the lower stem, the WC base pairs are reasonably stable in all miRNA models except the terminal pair in model C (Table ST2 of the SI).

In the loop region, among the four kinds of starting G:A base pairs (refer to Figure 2A−D) between the bases A6 and G35, the one in model A is completely lost during the dynamics (Table ST3 of the SI). As a result, A6 is splayed out of the structure. Visual inspection of the representative conformers does not show any transformation of the starting hydrogen bonding pattern to other forms. The stability of the C:A base pairs is different among each of the miRNA models, although it follows a similar trend in all clusters evolved from the same trajectory (Table ST3 of the SI). Despite these differences, a common feature is found in the internal loop of all of the miRNA models; the occurrence of a triplet hydrogen bond interaction between bases A9 and A33 and one of the C (C8 or C32) base located in the opposite strand (refer to the schematic representation in Figure 7), providing a stabilization to the internal loop. Such a triplet pairing has also been experimentally observed in the internal loop of the let7-lin41 miRNA−mRNA complex investigated by NMR spectroscopy, suggesting that it

represents a common structural feature of the bases present in the internal loop.[26]

RNA−Argonaute Complex. The Argonaute protein is an essential component of the RISC assembly and is involved in the guide−strand (siRNA/miRNA) mediated recognition and cleavage of the target mRNA. The cleavage site is usually located between the 10th and 11th steps of the target mRNA as counted from the 5′ end of the siRNA/miRNA strand.[4,5] The cleavage activity requires a WC base pairing between the guide and target strand in the seed segment (position 2−8 of guide strand). To examine how the canonical helical siRNA and kinked miRNA duplex structures can be accommodated within the argonaute protein, we selected five representative snapshots from the MD trajectories of the five simulated models (four for the miRNA and one for the siRNA duplexes) to be docked against the argonaute protein. In each case, the representative configuration was selected from the most populated cluster. The docked binary complexes were then subjected to an energy optimization and short dynamics (refer to Supporting Information for MD protocol) in explicit water and counterions. The SI Figure S10 depicts the distance of the scissile phosphate from the catalytic ASP residues of the argonaute protein as function of simulation time. The distance profile shows that in the case of miRNA model A (gray line), C (magenta line), and D (blue line), two of the distances between the scissile phosphate and the catalytic residues are significantly (∼2.0 Å) longer than that of the siRNA−argonaute complex, while the third one is only slightly longer than that found in the siRNA−argonaute complex. In miRNA model B (red line), although the distances are relatively close to that in the siRNA model, a longer distance range is also observed, as confirmed by Table 2, where the maximum and minimum distances observed in the five RNA−argonaute binary complex simulations are reported. These results indicate that the canonical A-type structure of the siRNA duplex permits a better accessibility of the mRNA strand for cleavage activity as compared to miRNA duplexes, where the presence of noncanonical secondary structure elements such as bulge and loop induce a kink at/near the catalytic site, thereby altering the accessibility of the phosphate group.

Discussion

RNA interference (RNAi) involves the regulation of gene expression by small RNA molecules (siRNA and miRNA) through their interaction with the target mRNA. These small RNAs, especially the miRNA, are known to interact with the target mRNA in an intricate manner with different degrees of complementarity, resulting in the formation of loops and bulges. Investigation of the structural features of such complex duplexes is instrumental in understanding their biological role in structural context. The prediction of the three-dimensional structure of

RNA is a great challenge of modern biology, since the existing predictors work well for canonical RNA structure, while are less accurate for RNA with noncanonical structural elements.[45,46] In the present work, we have utilized a computational approach to examine the structural characteristics of lin4-lin14 (miRNA−mRNA) and siRNA-lin14 duplexes, carrying out 5×10 ns unrestrained different molecular dynamics simulations. For siRNA, a single model has been built while for miRNA, multiple models have been constructed by considering different hydrogen bond patterns of the mismatch base pair (A6:G35) at the level of the internal loop (Figure 12 and Methods).

The results highlight that the siRNA duplex, due to the presence of a complementary Watson−Crick base pairs (Figure 4) is highly stable and maintains a canonical A-type helical structure throughout the simulation. Irrespective of the starting structure, all the lin4-lin14 miRNA duplexes forms a kinked global structure (Figure 4). The presence of noncanonical secondary structure elements such as internal loop and bulge in the miRNA duplexes makes it more flexible, resulting in alterations in the local structure, groove dimensions, and stacking of bases (Figures 5 and S2 and S3 of the SI). As an outcome, some of the bases in the bulge and loop tend to be exposed to the solvent, likely playing role in the RNA−protein recognition (Figure 6). In the two Watson−Crick stem regions, the hydrogen bonds between base pairs are stable (Tables ST1,2 of the SI), despite the loss of canonical helical structure of the stems. The degree of stability of the G:A and C:A base pairs, present in the internal loop are different among the miRNA duplexes. This leads to a disorder in the stacking of bases suggesting that in the real biological scenario, such loops can sample different conformations (Figures S2 and S3 of the SI). In the internal loop, all of the miRNA models give rise to a three-base interaction, where an A interacts with another A and C bases located on the opposite strand (Figure 7). Such a structural feature has also been experimentally observed by NMR spectroscopy on an miRNA−mRNA duplex having a stem-loop structure,[26] suggesting that it represents a common strategy to confer stabilization to the internal loop.

The docking results evidence the importance of the duplex geometry in influencing the extent of accessibility of the scissile phosphate group to the catalytic residues of the argonaute protein. Experimental studies have shown that the interaction of the lin4 miRNA with lin14 mRNA (model system investigated in our study) blocks the translation of lin14 protein without affecting the mRNA stability,[28,29] suggesting that the base complementarity between the guide miRNA and target mRNA influences the cleavage activity of argonaute. Here we show that the presence of internal loop and bulge regions modulates the structure of the miRNA duplex, altering the local flexibility and the accessibility of the mRNA cleavage site of the protein catalytic site (Figure S10 of the SI, Table 2). The binary protein-RNA models in this work have been generated with preformed RNA duplexes but we expect a similar behavior when the duplex is formed in the presence of the protein. In agreement, in the crystal structure of the *T. thermophilus* Argonaute solved in the presence of 21 nucleotides DNA−RNA hybrid, the presence of mismatch base pairs at the 10th and 11th positions prevents the enzyme mediated cleavage.[43]

Conclusions

In the present computational study, we have used MD simulations on multiple models of partially complementary lin-4:lin-14 (miRNA:mRNA) duplexes to examine their structural features. We also performed an additional simulation on a fully complementary siRNA duplex. The present study has provided information on the role of base complementarity in modulating the global structure of the miRNA duplexes and influencing the cleavage of mRNA cleavage by argonaute protein. The main results are summarized as follows:

(i) The lin-4:lin-14 miRNA duplex takes up an unusual kinked structure in all of the simulated duplexes, since the presence of mismatch base pairs induces changes in the groove dimension and base stacking.

(ii) The local structural alterations result in the protrusion of nonterminal bases that are less exposed in a canonical helical siRNA duplex, suggests they have a role in the recognition of the protein in the RISC assembly.[44]

(iii) Both siRNA and miRNA act via similar mechanism, but their nature of interaction with the target mRNA leads to different modes of gene regulation. Docking of representative RNA structures on the argonaute protein provides explanations for the different cleavage ability on a fully or partially complementary RNA substrate: the cleavage site of the siRNA model is better reached by the catalytic site as compared to the miRNA models (Table 2).

Supporting Information Available: Figure S1: plot of root-mean square deviation (Å) as a function of simulation time. Figure S2: Interstrand phosphate distance and base pair parameters calculated for the representative conformers. Figure S3: Base step parameters as a function of nucleotide number. Figure S4: Plot of interstem angle vs simulation time. Figure S5: Average solvent accessible surface area (Å^2) of the bases in the siRNA model. Figures S6−S9: Solvent exposure of the bases in the representative conformers of miRNA model A, B, C and D, respectively. Tables ST1−ST3: Percentage occupancies of hydrogen bonds between the base pairs in the upper WC stem, lower WC stem and the loop, respectively. Figure S10: Plot of the distance (Å) of scissile phosphate from the catalytic ASP residues as a function of simulation time. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **1998**, *391*, 806–811.

(2) Stefani, G.; Slack, F. J. Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 219–230.

(3) Ding, S. W.; Voinnet, O. Antiviral immunity directed by small RNAs. *Cell* **2007**, *130*, 413–426.

(4) Meister, G.; Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **2004**, *431*, 343–349.

(5) Tomari, Y.; Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **2005**, *19*, 517–529.

(6) Bernstein, E.; Caudy, A. A.; Hammond, S. M.; Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **2001**, *409*, 363–366.

(7) MacRae, I. J.; Doudna, J. A. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr. Opin. Struct. Biol.* **2007**, *17*, 138–145.

(8) Hutvagner, G.; Simard, M. J. Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 22–32.

(9) Höck, J.; Meister, G. The Argonaute protein family. *Genome Biol.* **2008**, *9*, 210.

(10) Carthew, R. W.; Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **2009**, *136*, 642–655.

(11) Chua, J. H.; Armugam, A.; Jeyaseelan, K. MicroRNAs: biogenesis, function, and applications. *Curr. Opin. Mol. Ther.* **2009**, *11*, 189–199.

Sampling the Structure of Lin-4:Lin-14 miRNA:mRNA

J. Phys. Chem. B, Vol. 114, No. 49, 2010 **16449**

(12) Siomi, H.; Siomi, M. C. On the road to reading the RNA-interference code. *Nature* **2009**, *457*, 396–404.

(13) Kim, V. N. Small RNAs: classification, biogenesis, and function. *Mol Cells.* **2005**, *19*, 1–15.

(14) Parker, J. S.; Barford, D. Argonaute: A scaffold for the function of short regulatory RNAs. *Trends Biochem. Sci.* **2006**, *31*, 622–630.

(15) Tolia, N. H.; Joshua-Tor, L. Slicer and the argonautes. *Nat. Chem. Biol.* **2007**, *3*, 36–43.

(16) Liu, J.; Carmell, M. A.; Rivas, F. V.; Marsden, C. G.; Thomson, J. M.; Song, J. J.; Hammond, S. M.; Joshua-Tor, L.; Hannon, G. J. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **2004**, *305*, 1437–1441.

(17) Elbashir, S. M.; Lendeckel, W.; Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **2001**, *15*, 188–200.

(18) Elbashir, S. M.; Martinez, J.; Patkaniowska, A.; Lendeckel, W.; Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **2001**, *20*, 6877–6888.

(19) Song, J. J.; Smith, S. K.; Hannon, G. J.; Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **2004**, *305*, 1434–1437.

(20) Parker, J. S.; Roe, S. M.; Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **2004**, *23*, 4727–4737.

(21) Ma, J. B.; Yuan, Y. R.; Meister, G.; Pei, Y.; Tuschl, T.; Patel, D. J. Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **2005**, *434*, 666–670.

(22) Parker, J. S.; Roe, S. M.; Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **2005**, *434*, 663–666.

(23) Pillai, R. S.; Artus, C. G.; Filipowicz, W. Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA* **2004**, *10*, 1518–1525.

(24) Pillai, R. S.; Bhattacharyya, S. N.; Artus, C. G.; Zoller, T.; Cougot, N.; Basyuk, E.; Bertrand, E.; Filipowicz, W. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **2005**, *309*, 1573–1576.

(25) Gebauer, F.; Hentze, M. W. Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 827–835.

(26) Cevec, M.; Thibaudeau, C.; Plavec, J. Solution structure of a let-7 miRNA:lin-41 mRNA complex from *C. elegans*. *Nucleic Acids Res.* **2008**, *36*, 2330–2337.

(27) Lee, R. C.; Feinbaum, R. L.; Ambros, V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **1993**, *75*, 843–854.

(28) Wightman, B.; Ha, I.; Ruvkun, G. Post-transcriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **1993**, *75*, 855–862.

(29) Olsen, P. H.; Ambros, V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **1999**, *216*, 671–680.

(30) Sethupathy, P.; Corda, B.; Hatzigeorgiou, A. G. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* **2006**, *12*, 192–197.

(31) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(32) Case, D. A.; Darden, T. A.; Cheatham, T. E., III, Simmerling, C. L.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, 2006.

(33) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(35) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald-an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(37) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phy.* **2001**, *114*, 2090–2098.

(38) Feig, M.; Karanicolas, J.; Brooks, C. L. MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph Model.* **2004**, *22*, 377–395.

(39) Hubbard, S. J.; Thornton, J. M. *"NACCESS", Computer Program. Department of BIochemistry and Molecular Biology*; University College: London, 1993.

(40) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929.

(41) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(42) John, D. Hunter, Matplotlib: A 2D Graphics Environment. *Comp. Sci. Eng. Achieve.* **2007**, *9*, 90–95.

(43) Wang, Y.; Juranek, S.; Li, H.; Sheng, G.; Tuschl, T.; Patel, D. J. Structure of an Argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **2008**, *456*, 921–926.

(44) Lingel, A.; Simon, B.; Izaurralde, E.; Sattler, M. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* **2004**, *11*, 576–577.

(45) Das, R.; Karanicolas, J.; Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* **2010**, *4*, 291–294.

(46) Shapiro, B. A.; Yingling, Y. G.; Kasprzak, W.; Bindewald, E. Bridging the gap in RNA structure prediction. *Curr. Opin. Struct. Biol.* **2007**, *17*, 157–165.