UC Irvine
UC Irvine Previously Published Works

Title

Kink turn sRNA folding upon L7Ae binding using molecular dynamics simulations

Permalink

Journal

ISSN

Authors

Ye, Wei
Yang, Jingxu
Yu, Qingfen
et al.

Publication Date

DOI

Copyright Information

Peer reviewed

PCCP

PAPER

Kink turn sRNA folding upon L7Ae binding using molecular dynamics simulations†

Wei Ye,[a] Jingxu Yang,[a] Qingfen Yu,[a] Wei Wang,[a] Joseph Hancy,[b] Ray Luo*[c] and Hai-Feng Chen*[ad]

The kink-turn sRNA motif in archaea, whose combination with protein L7Ae initializes the assembly of small ribonucleoprotein particles (sRNPs), plays a key role in ribosome maturation and the translation process. Although many studies have been reported on this motif, the mechanism of sRNA folding coupled with protein binding is still poorly understood. Here, room and high temperature molecular dynamics (MD) simulations were performed on the complex of 25-nt kink-turn sRNA and L7Ae. The average RMSD values between the bound and corresponding apo structures and Kolmogorov–Smirnov P test analysis indicate that sRNA may follow an induced fit mechanism upon binding with L7Ae, both locally and globally. These conclusions are further supported by high-temperature unfolding kinetic analysis. Principal component analysis (PCA) found both closing and opening motions of the kink-turn sRNA. This might play a key role in the sRNP assembly and methylation catalysis. These combined computational methods can be used to study the specific recognition of other sRNAs and proteins.

Introduction

Small nucleolar ribonucleoprotein particles (snoRNPs) in eukaryotes and small ribonucleoprotein particles (sRNPs) in archaea guide posttranscriptional modification of precursor ribosomal RNAs (rRNAs) to maturation, a key activity in the ribosome synthesis and translation process.[1–5] Two types of modifications of nucleotides in precursor rRNAs have been discovered:[6,7] (1) the conversion of specific uridines to pseudouridines (Ψ); and (2) the addition of methyl groups to specific nucleotides, mostly to O2′ of ribose (Nm). 2′-O-methylation is catalyzed by a conserved class of snoRNPs which contains a box C/D RNA and several protein components.[8–11] In archaea, box C/D RNP consists of a piece of sRNA with a conserved motif called the kink-turn (K-turn motif).[11–14] L7Ae protein binds directly to box C/D elements,[8,14] Nop5p binds to L7Ae, and fibrillarin serves as the core catalytic component.[15–17] This conserved structure is homologous to the box C/D snoRNP in eukaryotes,[16] which contains a K-turn snoRNA and four protein components.[18,19] Both sRNA and protein components are required in the methylation process.[20,21] K-turn's combination with L7Ae initializes the organization of other subunits into sRNPs.[11,22,23] Previous experiments also have revealed that several similar K-turn RNAs can specifically bind with 15.5 kDa homologous proteins.[11,14] Mutations on the K-turn motif, especially on the linking bulge, could disrupt the binding.[11,14,23–25] These observations indicate that the folding of K-turn sRNA is a highly dynamic process, and the binding between the K-turn RNA and the protein is a highly specific and selective procedure.

A canonical K-turn structure contains two bent stems linked by a 3-nt bulge with an extruded nucleotide.[26] It has been observed that the K-turn sRNA would have two distinct tertiary structures, *e.g.* it has plasticity to some extent.[27,28] Recently, the applications of fluorescence resonance energy transfer (FRET)[29,30] have shed light on the conformation of K-turn RNAs. Many observations show that the unbound K-turn RNA exists in an extended conformation, and the metal ions could induce it to fold with its chaperon protein (L7Ae homolog proteins).[31,32] Recently, this tightly kinked geometry was shown to be very stable under long time observation.[33,34] It was also reported that even in the absence of metal ions, the sRNA could be induced to fold into tightly kinked conformation upon the binding of L7Ae and other related proteins.[35] On the other hand, crystallography is another powerful tool for researching the binding details of sRNP, which could help us to understand the relationship between the structure of sRNP and its biological functions.

a State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai, 200240, China

b Shanghai New Oriental School, 108 Guoquan Road, Shanghai, 200433, China

c Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, USA

d Shanghai Center for Bioinformation Technology, 1278 Keyuan Road, Shanghai, 200235, China. E-mail: haifengchen@sjtu.edu.cn, rluo@uci.edu; Fax: +86-21-34204348; Tel: +86-21-34204348

Crystal structure of eukaryotic U4 snRNA bound with 15.5 kDa protein[36] shows similar conformation with the archaeal box C/D sRNA bound with L7Ae protein,[19,37] which confirmed the tight kinking of sRNAs in the bound state.

However, the mechanism of sRNA folding coupled with binding, which is hard to gain insight from crystal structures, is still poorly understood. Fortunately, molecular dynamics (MD) simulation is a powerful tool for analyzing the structural and dynamic features of biomacromolecules. Several MD simulations of apo-sRNA or their complex have been made to reveal the recognition mechanism and the conformational change,[38–40] which also indicated that the flexibility of the V-shaped K-turn sRNA and the tertiary interactions between the sRNA and its chaperon protein play critical roles in the binding procedure.[38,39,41–43] Among these previous observations, most attention was focused on the interactions between sRNA and its chaperon, while the internal folding kinetics of sRNA upon L7Ae-binding were poorly understood. In the current study, we compared the hinge motion and folding kinetics of L7Ae-bound and apo sRNA to further discover the mechanism of recognition between sRNA and L7Ae using MD simulations. In addition, a putative mechanism of the hinge motion of sRNA in the complex was proposed to infer its biological functions.

Several crystal structures of bound box C/D s(no)RNP have been released.[19,36,37] Among these structures, the archaeal box C/D sRNP bound with archaeal L7Ae[19] (PDB code: 1RLG) was selected as a starting structure in our simulations (Fig. 1). In this structure, a synthetic 25-mer K-turn sRNA forms a common K-turn motif containing a canonical bent stem (C-stem) and a noncanonical stem (NC-stem) with a GUG bulge (G17-U18-G19) linked. Because of the steric hindrance upon U18, the orientation of the base of U18 is flipped, so that U18 protrudes its base into the cavity of L7Ae protein (Fig. 1). The interactions are responsible for the strong binding stability between the sRNA and L7Ae.

Recently, Xue *et al.* reported a larger and more complete crystal structure for archaeal sRNP[37] (PDB code: 3NMU). In the

crystallized asymmetrically assembled structure, K-turn sRNA is bound with L7Ae and Nop56/58. Fibrillarin, acting as the catalytic core which transfers the methyl group from SAM to substrate rRNA, binds with Nop56/58 with the opposite orientation. Based on the electron microscopy results,[44] a diRNP model[37] was proposed and shown in Fig. S1A (ESI†) with manual docking and energy minimization. In this model, fibrillarin has two positions: (i) the "*cis*" position far away from the methylation site of the rRNA of the same sRNP and (ii) the "*trans*" position closer with rRNA of the other sRNP. However, in this model, fibrillarin could still hardly come in contact with the substrate rRNA. Consequently, a cross-RNP model (Fig. S1B and C, ESI†) was proposed in the assembly and catalysis process.[37] Fibrillarin needs to reposition from the "*trans*" to "*Af*" position and take the place of L7Ae. Based on previous experiments, tight folding of sRNA is induced upon L7Ae's binding. Then an interesting question if the flexibility contributes to the release of L7Ae when making room for fibrillarin's reposition would be naturally asked. In this study, we also tried to answer this question to get insight into the relationship between the sRNA's conformational change and the catalytic cascade.

Materials and methods

Molecular dynamics simulations

The atomic coordinates of bound box C/D sRNA were obtained from the Protein Data Bank, with accessory code 1RLG. The structure was then split into sRNA and L7Ae, as starting structures of apo-sRNA and apo-protein, respectively. Mutagenesis on U18 to G18 and subsequent minimization were conducted with SYBYL. All simulations and most analysis procedures were conducted using the AMBER11 software package.[45] Hydrogen atoms were added using the LEaP module of AMBER11. Counter-ions were used to maintain system neutrality. All systems were solvated in a truncated octahedron box of TIP3P water with a buffer of 10 Å. Particle Mesh Ewald (PME)[46] was employed to treat long-range electrostatic interactions with the default setting in AMBER11. The improved parm99SBildn force field[47] was used for the intramolecular interactions. The SHAKE algorithm[48] was used to constrain bonds involving hydrogen atoms. 2000-step steepest descent minimization was performed to relieve any structural clash in the solvated systems. This was followed by heating up and brief equilibration for 20 ps in the NVT ensemble at 298 K with PMEMD of AMBER11. Langevin dynamics with a time step of 2 fs were used in the heating and equilibration runs with a friction constant of 1 ps^{-1}.

To study the folded state of each solvated system, 10 independent trajectories in the NPT ensemble at 298 K were simulated for 10.0 ns. High temperature simulations were presented for the apo-sRNA, apo-L7Ae, and sRNA–L7Ae complex in the NVT ensemble at 498 K to investigate the unfolding kinetics of each solvated system, in all 10 independent trajectories.
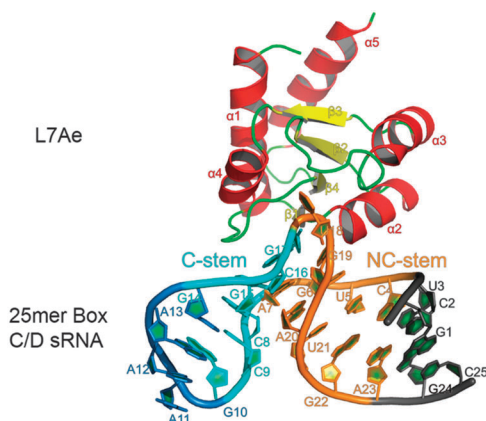

Fig. 1 A cartoon representation of the 25-mer box C/D sRNA and L7Ae crystal structure (pdb code: 1RLG). The sRNA links to L7Ae with U18 protruding into the cavity of L7Ae. The C-stem (orange), NC-stem (cyan) and the GAAA tetramer (blue) are shown in colors, with the two termini in black.

Table 1 Simulation condition for all the simulation systems

System	Counter ions	Waters	Temperature (K)	Trajectories	Simulation time (ns)
Bound 25-mer box C/D sRNA with L7Ae	27Na$^+$	8978	298	10	10
			498	10	30
Apo-sRNA	2Na$^+$	4223	298	10	10
			498	10	20
Apo-L7Ae	25Na$^+$	4747	298	10	10
			498	10	10
Mutant complex with U18G	26Na$^+$	10 802	298	1	10
			498	1	15

As mentioned before, one of our missions is to show the folding process. To date, however, the time scales of current MD simulations are restricted to microsecond magnitude at room temperature, which is significantly shorter than the folding half times of most biomolecules at room temperature (at least 1 ms).[49] Fortunately, the rate of unfolding can be accelerated approximately by six orders of magnitude at high temperature.[50] Most biomacromolecules unfold in the nanosecond timescale at this temperature. Furthermore, experiments have confirmed the transition state for folding and unfolding is expected to be the same from the principle of microscopic reversibility.[51] Based on the previous work, unfolding simulations at high temperature were used in this study.

Duration time of high temperature simulations for the four systems was determined according to the conformational stability reflected with RMSD and Q_f (the fraction of tertiary native contacts) during the simulations. A cumulation of 925 ns of trajectories were collected for three systems at both 298 K and 498 K, taking about 80 000 CPU hours on the Xeon (3.0 GHz) cluster. Detailed simulation conditions are listed in Table 1.

Data analysis

Tertiary contact assignment was handled using in-house software.[52–61] Residues and nucleotides are in hydrophobic contact when mass centers of their side chains are closer than 6.5 Å for the complex. A previous study has shown that charge-to-charge interactions of up to 11 Å were found to contribute to protein–RNA binding free energies. Thus, electrostatic (*i.e.* charge–charge) interactions are assigned when the distance between the mass center of the positively charge residue and the dsRNA phosphate backbone is less than 11 Å. A hydrogen bond is defined such that the distance between two polar heavy atoms either of which has a hydrogen atom is less than 3.5 Å. The non-adjacent nucleotides and residues are in contact (termed "native contact") when their side chains are closer than 7.5 Å. All the 3D molecular representations were visualized and rendered using PyMOL 1.5.[62]

The energy landscape was mapped by calculating normalized probability from a histogram analysis, and plotted using Origin 8.5, which was also employed to plot all the diagrams in this paper. For each simulation, sampling was conducted every 10 ps (10 000 snapshots for 10 × 10 ns simulations). Radius of gyration (R_g) and root mean standard deviation (RMSD) were

both separated into 8 bins. The energy landscape was plotted among these 64 (8 × 8) bins. Average structures were extracted from the structure ensembles of lowest energy. Landscapes were also used on distance variations to detect the structural adjustments between apo and bound structures. The distance between every pair of C5′ (sRNA) or Cα (protein) was calculated in both apo and bound structures, respectively. Then, the landscape of base-to-base (or residue-to-residue) distances in the apo-structure minus the corresponding distances in bound structures was plotted. A positive region indicates a structural contraction, while a negative region means a structural extension. In order to exclude the effect of thermal fluctuations, we averaged all the distances in the whole 10 trajectories of 10 ns simulations instead of using an averaged structure.

Principal component analyses

The two V-shaped stems of the K-turn sRNA constitute a highly flexible motif which has intimate contacts with L7Ae and plays key roles in stabilizing the complex. Previous MD simulations have confirmed this flexibility.[43] In our study, principal component analysis (PCA)[63–66] was employed to discover the principal movements of these two stems. The ptraj module of AMBER11 was used to solve eigenvalues and corresponding eigenvectors from a covariance matrix and to calculate the contribution of each principal component. Three-dimensional structural snapshots along each principal component were projected with VMD[67] and its plugin NMWiz.[68] The final results of PCA were visualized using PyMOL with 25 snapshots along each principal component.

Binding mechanism evaluation

The mechanism of specific binding is one of the most attractive issues in our study. There are two mainstream hypotheses to explain ligand binding: (1) the "Induced fit" model,[69–71] emphasizing the conformation of the receptors is induced to change upon the ligand binding (binding first); and (2) the "Conformational selection" model,[70–78] where the ligand will directly select the most optimal conformation of receptor among several different conformations. If the binding obeys a mechanism of induced fit, a significant conformational change in the receptor must occur upon the ligand binding, especially near the binding site. Otherwise, if the conformational change is not so significant, the binding tends to obey a conformational selection mechanism. Following this concept, induced fit could be calculated from the RMSD between a bound structure and its most similar apo-structure (which has the lowest RMSD from the bound structure). In the induced-fit model, relative magnitude of conformational selection could be characterized from the average RMSD between this most similar apo-structure and the other apo-structures.[79] RMSD for induced fit of 10 trajectories was given as a function of distance from the binding site. Because the RMSD values are not normally distributed, which is not suitable for a t test, a standard two-sample Kolmogorov–Smirnov (KS) test was used to evaluate the statistical significance of RMSD variations. Recent studies have demonstrated that the binding process may employ a mixed mechanism, selecting an optimal conformation globally and inducing the local structure of the ligand to be adjusted.[80–82]

The equation below was used to quantitatively describe relative differences between conformational selection and induced fit.

$$\Delta = \frac{\sum_{X_i, f_i \in \text{CS}} X_i f_i}{N_{\text{CS}}} - \frac{\sum_{X_j, f_j \in \text{IF}} X_j f_j}{N_{\text{IF}}}$$

X is the RMSD value and f is the frequency for a particular RMSD value region in a distribution of conformational selection (CS) or induced fit (IF), and N_{CS} and N_{IF} are the numbers of data points for the distribution of CS or IF, respectively.

As previously mentioned, U18 protrudes its base into a cavity on L7Ae. The base of this nucleotide was defined as the binding site in the binding process. Atomic RMSDs of all the bases between aligned bound and apo-structures within every distance range from the binding site were first calculated. A KS P test was employed to indicate the significance of the difference between RMSDs of atoms included within a certain distance from the binding site and the atomic RMSDs of the whole structure. Thus, the high significance shows that, upon binding, the structural change in this distance is different from the structural change of the whole structure. A more detailed description on the identification of the binding mechanism can be found in ref. 79.

Transition state identification

The transition state reflects a representative snapshot at the free energy maxima of the unfolding and folding pathways. The structures at the free energy maxima constitute the transition state ensemble (TSE). TSE structures can either fold or unfold, and the transition probability (P) will be 50%. To search for the transition state for each simulation system, snapshots along high temperature simulations in every trajectory were clustered into groups[83] through global multidimensional scaling (MDS) analysis[79] on atomic RMSD values. Values between any two snapshots were supposed to be the distances between any two points in a three-dimensional space. MDS was used to minimize point-to-point distances through a steepest descent procedure. In order to reduce dimensionality so that all the snapshots could be converted into 3D space and clustered, the Sammon stress

function[84] was used to optimize the mapping among all the N snapshots, by minimizing the error E as defined as:

$$E = \frac{1}{\sum_{i=1}^{N} \sum_{i<j}^{N} D_{ij}} \sum_{i=1}^{N} \sum_{i<j}^{N} \frac{(D_{ij} - d_{ij})^2}{D_{ij}}$$

where i and j are two snapshots, D_{ij} is the RMSD between these two snapshots, d_{ij} is the distance between the corresponding two points in the 3D space after a minimizing iteration step. When E tends to be stable, minimization will stop.

Results

To collect enough snapshots for statistically meaningful structural analysis, up to ten trajectories of 10.0 ns were collected for apo and bound sRNA to analyse their structural properties. Average Cα RMSD, distance landscapes, KS test, and PCA were used to study the conformational changes, folding kinetics and the hinge motion of the sRNA. Average Cα root mean square deviations (RMSDs) with respect to simulation time (shown in Fig. S2, ESI†) indicate that 10 ns' simulation time at 298 K was sufficient for the three systems to equilibrium in solvated environment. Bound K-turn sRNA shows a lower RMSD than the other two structures, which reflects a stabilizing role of the binding between the two components. This stabilization could also be observed in RMS fluctuations (RMSF) of bases and residues. Fig. 2 shows RMSF for all the C5′ and Cα atoms in three systems, as well as secondary structures labelled in the same colors as Fig. 1, which indicate lower fluctuations in bound complex than in apo-structures, especially at the binding interface (U18 between the NC-stem and C-stem, loops between β1 and α2 and between α4 and β4). RMSF differences (bound minus apo, histograms in Fig. 2) would clearly show these changes upon binding. Additionally, distinct fluctuations were detected at two termini of the C-stem and NC-stem of sRNA.

Mechanisms of specific binding

The mechanism of specific binding is one of the most attractive issues in our study. Between the two mainstream hypotheses
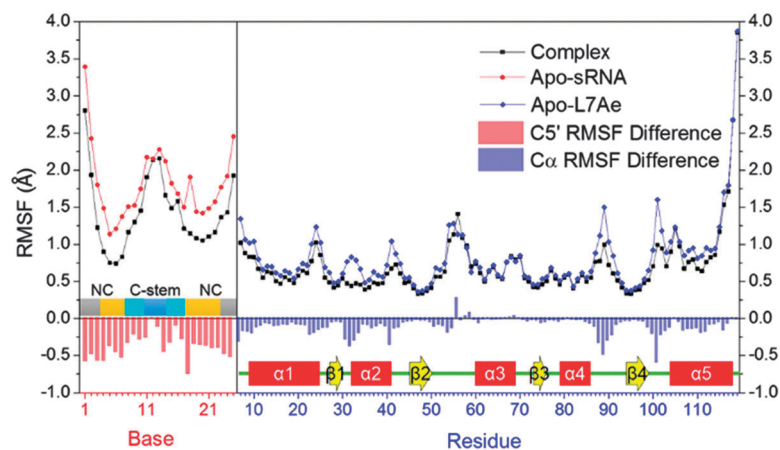

Fig. 2 The RMS fluctuations for C5′ and Cα atoms for bound and apo-structures. Cα atoms in L7Ae were numbered as the original number in pdb structure 1RLG. RMSF differences (bound minus apo) are shown with lower histograms. Also, secondary structures were labelled in according to Fig. 1.
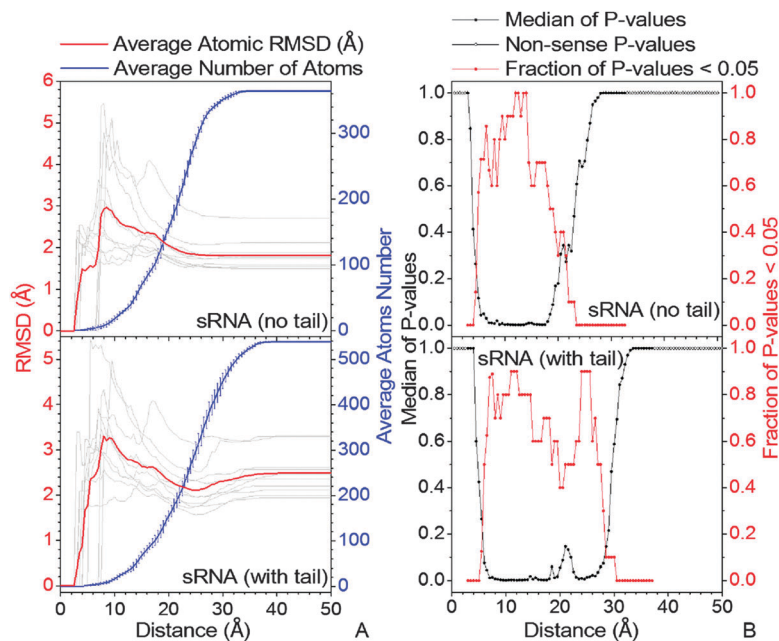
Fig. 3 Atomic RMSD (red lines for average and grey ones for each trajectory) and number of included atoms (blue lines) *vs.* distance from binding site, between each bound structure and its most similar apo-structure, aiming for induced fit calculation. Significance of local structural changes was elaborated by the Komogorov–Smirnov test, shown with median of *P* values (solid black circles with non-sense *P* values labelled in empty black circles) and fraction of *P* < 0.05 (solid red circles). (A) Atoms number and RMSD for bound and apo-sRNA; (B) KS test for bound and apo-sRNA.

explaining ligand binding, "induced fit"[69–71] and "conformational selection",[70–78] recent studies have demonstrated that the binding process may employ both mechanisms: binding site with the former and the whole molecule with the latter.[80–82] Structures of the last snapshots from 10 trajectories at room-temperature of the three systems were extracted for recognition mechanism evaluation. Average RMSD between bound K-turn sRNA and its most similar apo-structure (detailed information listed in Table S1, ESI†) among 10 trajectories *vs.* distance from binding site (base of U18) is shown in Fig. 3A. As stated in RMSF analyses, termini of the C-stem and NC-stem have high fluctuations, which might overestimate the magnitude of conformational selection. In order to observe the influence of these two highly flexible termini, average RMSDs for sRNA with and without these termini were both calculated.

The RMSD gradually decreases for sRNA including or excluding two termini (tails), then keeps equilibration. This suggests that the area of conformational change for sRNA is focused on the binding site of L7Ae. In order to investigate the statistical significance for the local conformational deviations, the two-sample Kolmogorov–Smirnov (KS) *P* value test was analyzed. Note that the KS test, as a nonparametric test, is a good choice for this study because the distributions of magnitudes of atomic deviations do not fit well to any distribution used for parametric tests. As shown in Fig. 3B, the conformational differences are statistically significant up to 20 Å away from the L7Ae binding site, with the median *P* values typically less than 0.05 and the fraction of typical *P* values greater than 0.5. In summary, the specific recognition between sRNA and L7Ae is

more likely to obey an induced fit mechanism based on the MD simulation and the subsequent analysis.

Relative magnitude for induced fit and conformational selection

The average RMSD and KS *P* test analysis indicate that sRNA may follow an induced fit at the binding site of L7Ae. The next natural question to ask is if the conformational selection mechanism exists and the relative magnitudes of the both mechanisms in the sRNA–L7Ae recognition. To pursue this question, conformational changes were quantitatively compared through histograms of RMSD counts for both mechanisms.[79] Relative magnitudes were calculated and shown in Fig. 4. Δ_1 represents the relative difference between conformational selection and global induced fit in whole structure and Δ_2 represents the difference from conformational selection and local induced fit (binding site). In the case excluding the tails, Δ_2 equals −1.02, confirming sRNA undergoes a local induced fit mechanism upon binding to L7Ae, which is consistent with RMSD and KS *P* tests; in addition, Δ_1 of −0.30 also shows a global induced fit mechanism for the whole structure. Interestingly, in the case of the tails included, the conformational selection appears to be somewhat more dominant than induced fit on average. This is consistent with the previous work.[79]

Same analyses were also conducted on protein L7Ae (listed and shown in Table S1 and Fig. S3, ESI†). This indicates that L7Ae obeys a mechanism of conformational selection upon sRNA binding.
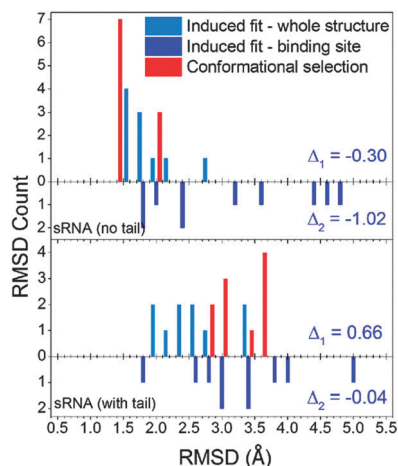
Fig. 4 Comparison between conformational selection and induced fit, represented in histograms that show counts of quantified RMSDs between: bound structures and most similar apo-structures in whole molecule (global induced fit, blue bars); bound structure and most similar apo-structure near binding site (local induced fit, light blue bars); and the most similar apo-structure and other apo-structures (conformational selection, red bars). Δ_1 is calculated from $CS - IF_{global}$, while Δ_2 means $CS - IF_{binding\,site}$. "No tail" and "with tail" mean the calculations for sRNA excluding and including the tails of two stems with high fluctuations, which would overestimate conformational selection.

Distance difference landscapes

In order to evaluate the conformational adjustment of sRNA, the distance difference landscapes between apo and bound states were used to visualize this point. Fig. 5A illustrates the
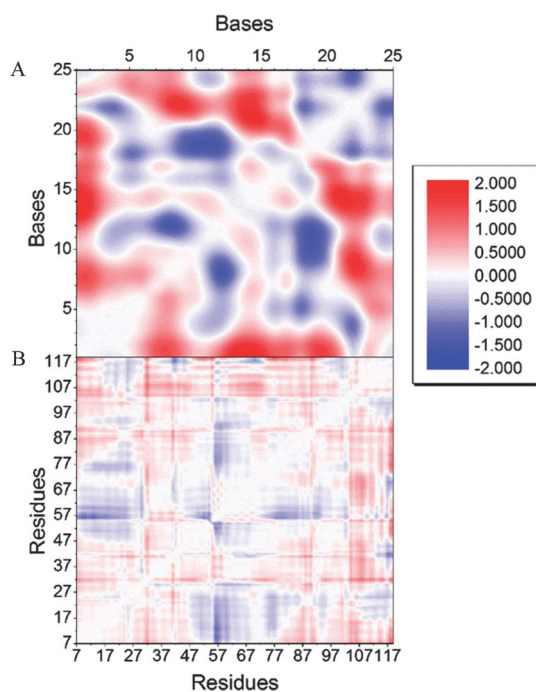

Fig. 5 Average distances difference landscape for sRNA and L7Ae. Red regions mean that the average distance between related atoms pair (C5′–C5′/Cα–Cα) in the apo-structure is larger than in bound structure (compaction upon binding); correspondingly, blue regions mean extension upon binding.

conformational adjustment of sRNA upon binding of L7Ae. Regions in blue represent negative differences which indicate that relative C5′ atoms separated farther upon binding (extension); while regions in red indicate the relative atoms are induced to be closer (compaction). Except the flexible termini, blue regions show that distances between C8 and A12, and between [C9–G10–A11–A12] and [U18–G19] are increased. These observations indicate that the C-stem tends to uncoil and become extended. Red regions reveal that distances between [U21–G22] and [C8–C9], and between [U21–G22] and [G14–G15] are reduced, indicating that C-stem and NC-stem got closer upon binding to L7Ae. Consistent observations could also be found in the NC stem motion with PCA.

The landscapes of distance difference for L7Ae are shown in Fig. 5B. The conformational changes are less significant than those of sRNA. Extensions are outnumbered by compactions and mainly distributed between residue no. $\sim$60 and (70–90) where is extruded by the protruded U18 of sRNA. Distances between residues (30–40) and most other residues are reduced upon binding, probably because this region is extruded by the NC-stem of sRNA.

Driving forces of conformational adjustment

In order to reveal the driving forces for these conformational adjustments, the interactions between sRNA and L7Ae are shown in Fig. 6 and listed in Table S2 (ESI†). 25 hydrogen bonds were found with population higher than 40%, mainly spreading in the two regions: (i) the protruded U18 and (ii) U17 and U19. The opposite orientations of these two hydrogen bonding regions provide the main driving force for the local induced fit. Electrostatic interactions are long-range and non-specific interactions between phosphates and positively-charged amino acids, mainly around C4-A7 with K29, K30, K37 and R41. Another notable electrostatic interaction was found between K79 and the phosphate group of G17 and U18, which may be important in the recognition of the two components. A stable hydrophobic core was also found between C16 and G17 and three hydrophobic residues (I88, V90 and P91), which may enhances the binding affinity between sRNA and L7Ae. In summary, these stable interactions play key roles in conformational adjustments of sRNA.

Closing–opening motion on two stems of sRNA

Besides local conformational adjustments, a closing motion of two stems was also observed, which is in agreement with previous experiments and MD simulations.[32,35,40] To quantitatively identify this motion, principal component analysis (PCA) was carried out on the bound sRNA separately for each trajectory. In general, the three most principal components (named PC1, PC2, and PC3) represent over 80% of the overall fluctuations for ten trajectories. And the V-shaped stems contribute the majority of all the fluctuations. To clearly display the motion of the sRNA, structural projections along the first principal component ($\sim$55%) of trajectory 3 were shown in Fig. 7A, fluctuating, from blue to red, which shows C-stem fluctuates vertically to the binding interface, and NC-stem rotates in the plane parallel to the binding interface. Average
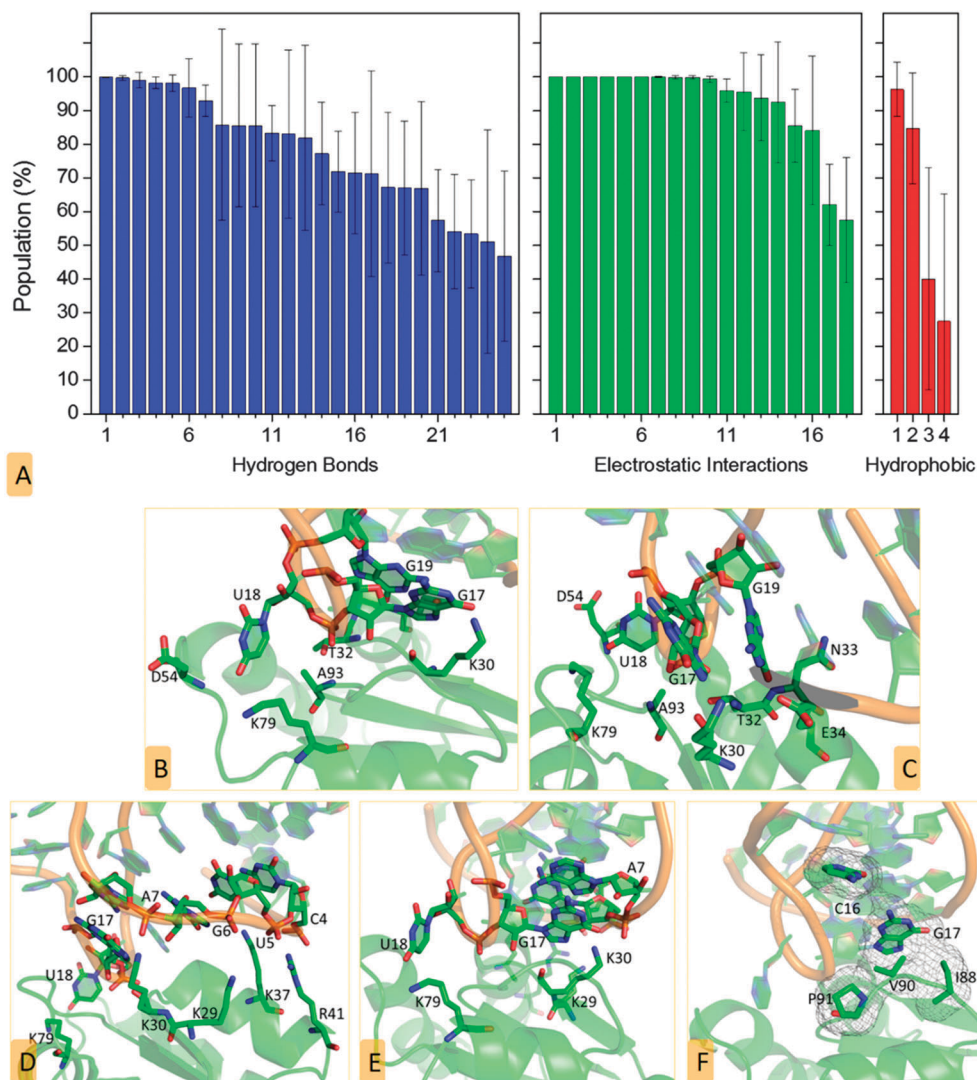
Fig. 6 Interactions between K-turn sRNA and L7Ae. (A) Histogram of interaction in the bound structure. Blue for hydrogen bonds, green for electrostatic interactions, and red for hydrophobic interactions; (B and C) detailed views for hydrogen bonds; (D and E) detailed views of electrostatic interactions; and (F) hydrophobic core within G17 and I88-V90-P91.

time of this motion among all the 10 trajectories was 7.37 ± 1.36 ns. Surprisingly, the motions of first PCs for 3 trajectories out of 10 were reversed with trajectory 3, and shown in Fig. 7B (taking trajectory 1 for instance). This motion has never been observed in bound K-turn sRNA before. Therefore, the principal motion for K-turn sRNA might include both V-shaped stem closing and the stem opening.

For comparison, PCA was also carried out on apo-sRNA. No significant conformational changes were observed, partly because the kink-turn opened at an early phase in room-temperature simulations. This could be indicated in time evolutions of average kink angle between the C-stem and NC-stem, defined as ∠[A11, A13] [G17, U18, G19, G6, A7] [G1, C25], as shown in Fig. 7C for bound and apo-sRNA, respectively. The angle of bound sRNA had a propensity of decrease, *i.e.* bound sRNA tended to be folded more tightly, which is in agreement with the distance difference landscape analysis. However, the same

angle of apo-sRNA increased during MD simulation. This angle in trajectories 1 and 3 was also plotted in black and blue lines, which could also indicate the movement of the two stems.

Unfolding kinetics of bound and apo states

In order to investigate the sRNA's folding mechanism upon L7Ae binding, high temperature MD simulations at 498 K were performed. The fraction of native tertiary contacts (Q_f) and the fraction of native binding contacts (Q_b) were applied to reveal unfolding and unbinding kinetics. Time evolutions of R_g, Q_f of bound sRNA, Q_f of bound L7Ae, and Q_b of the sRNA–L7Ae complex are shown in Fig. 8, which suggests that the tertiary unfolding and unbinding could be fitted well by a single exponential function (red lines), indicating first order kinetics in the NVT ensembles at high temperature (498 K). The fitted kinetics data are listed in Table 2. The kinetics analysis shows that the tertiary unfolding half-time was 3.605 ± 0.025 ns for
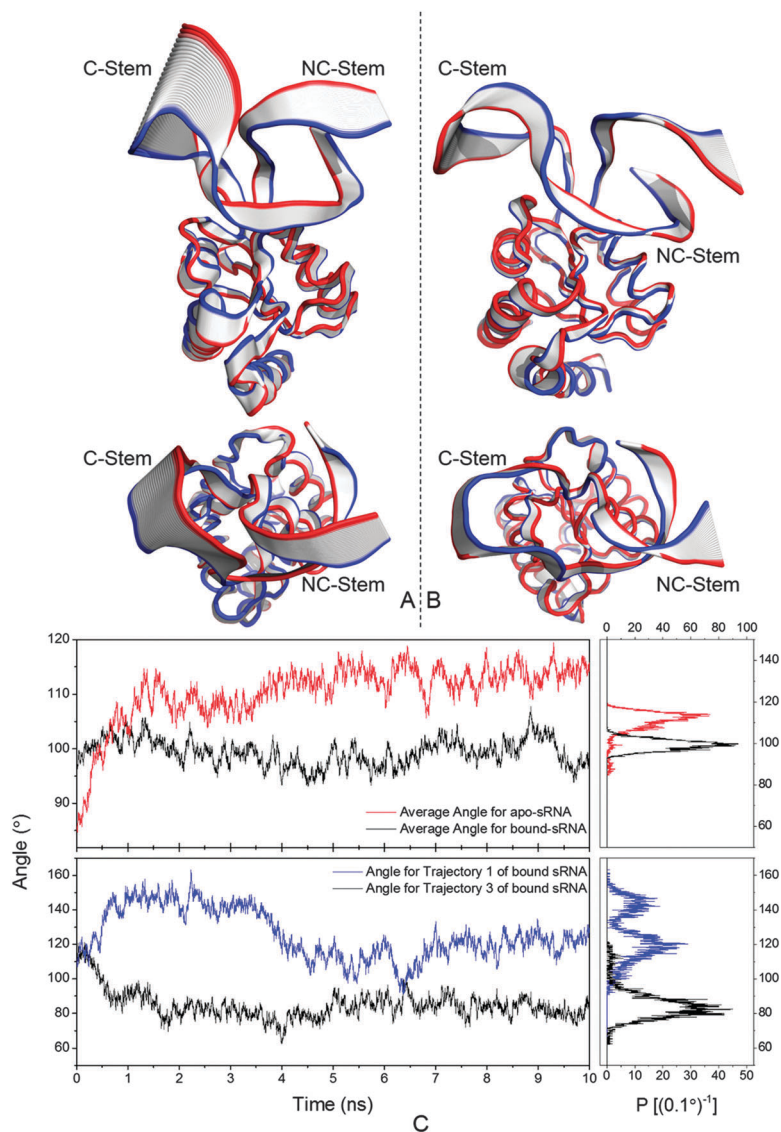
Fig. 7 Closing–opening motion from PCA analysis. (A and B) Front and top sights of the movements in the first principal component, A for trajectory 3, and B for trajectory 1. Structures fluctuate from blue to red, representing the reverse orientations of the movements in the 10 ns' simulation, for these two trajectories, respectively. (C) Angles between the 2 stems vs. time. The average angle separately for bound and apo-sRNA and this angle for bound trajectory 1 and 3 (blue and black, representing PCA results) are both shown. Corresponding normalized probability distributions are also shown.

bound sRNA, and 6.294 ± 0.09 ns for bound L7Ae. The unbinding process was rather slow, with a half-time of 20.932 ± 0.423 ns. Kinetic parameters indicate that the tertiary unfolding of sRNA and L7Ae was much faster than the unbinding between sRNA and L7Ae, respectively. Tertiary unfolding for apo-sRNA and apo-L7Ae was also faster than the bound states (1.609 ± 0.006 ns and 1.699 ± 0.027 ns, respectively). This suggests that the tertiary unfolding of bound sRNA and L7Ae is significantly postponed upon the formation of complex.

Transition state

The unfolding kinetics for all three systems suggest that the complex structure unfolds via a two-state process. Thus, a transition state that corresponds to the free energy maximum along each of their reaction coordinates exists in the unfolding simulations.

Experiments have supported that the transition state moves closer to the native state with kinetic and thermodynamic distortion under the high-temperature simulation conditions.[85–87] Fig. 9A and B show the conversion from RMSD to clusters[83] for a representative trajectory of bound K-turn sRNA. For this 30 ns trajectory, 1000 snapshots (30 ps per snapshot) were extracted and analyzed. There are three RMSD plateaus along the simulation. The first one consisted of original rapid structural deviations, appearing at an early period and lasting for a long time (1–20 ns, but not for all trajectories or other types of biomacromolecules). After a sharp RMSD increase (from 4 Å to 14 Å), the second plateau emerged, representing a structure distorting from the native state. The last plateau region was mainly due to the structure entering into a fully unfolded state.
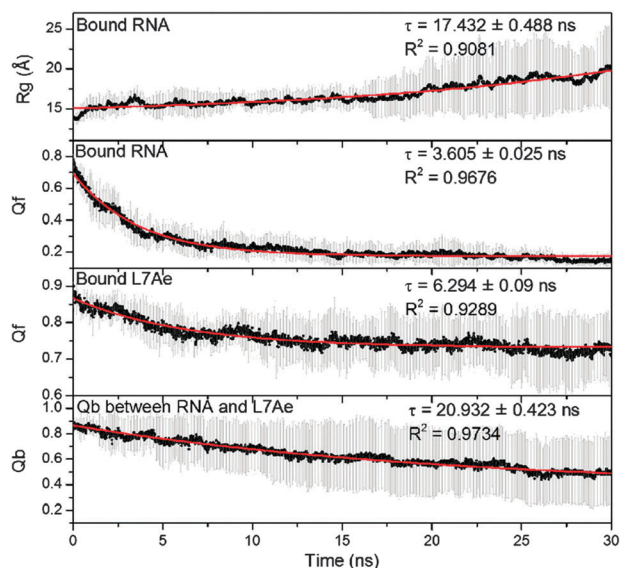
This journal is © the Owner Societies 2013

Phys. Chem. Chem. Phys., 2013, **15**, 18510–18522 | **18517**

Fig. 8 Unfolding kinetics of time evolutions on R_g, Q_f, Q_f and Q_b of bound sRNA, bound sRNA, bound L7Ae and complex, respectively. Fitting for every index in the first order exponential function is represented by the red line. Unfolding half times and correlation coefficients are also listed.

Using MDS analysis[79,84] on snapshot-to-snapshot RMSD (a 1000 × 1000 distance matrix generated), 1000 snapshots were gathered into 3 clusters in 3D space and shown in Fig. 9B. The structures around the red point (which has just left cluster 1) were chosen as transition state ensemble (TSE) for these trajectories.

Discussion

Comparison with experiments and previous MD simulations

The structural analysis has shown that hydrogen bonds are important interactions and focused on U18 and G19 of sRNA in the crystal structure.[19] Three stable hydrogen bonds were found in our room temperature simulation for U18/K79, G19/N33, and G19/N34. This result is in good agreement with the structure analysis that U18 and G19 form important interactions with L7Ae.[19] Furthermore, for the hydrogen bond between U18 and D54, there are two trajectories with strong hydrogen bonds with population higher than 90%, while 6 trajectories are rather weak (0%). This indicates D54 must have more than one low-energy conformation which could not be observed through experimentation.

Two *trans* sugar-Hoogsteen G:A pairs which have notable hydrogen bonds were determined to be very important to the kinking of sRNA.[88] In current study, the high stability among these bases was also observed to form a hydrogen bonding network. In room-temperature MD simulations, all these four H bonds have populations over 95%. At high temperature, this hydrogen bonding network disrupted along the unfolding process of the K-turn sRNA.

Another important motif which stabilizes the sRNA is the GAAA (G10-A13) tetramer. In this study, we used ff99SBildn force field to simulate bound and apo-sRNA. For bound sRNA, 1 disruption and 2 vertical oriented trajectories out of 10 were observed for this motif. For apo-sRNA, the vertical orientation was found in 4 trajectories. This suggests that bound sRNA has a higher stability than the apo-structure. Spackova et al.[43] discussed the choice of force fields in MD simulations. It could not be properly described by ff99SB/ff99bsc0 either with the A12 flipped to the vertical orientation or with these 4 nucleotides totally disrupted.

The unfolding kinetics of bound and apo-sRNA were distinctly different (Q_f half-time of ∼3.6 ns and ∼1.6 ns), suggesting K-turn conformation was stabilized upon L7Ae binding. In our study, K-turn sRNA might obey the induced fit mechanism both locally and globally, calculated with the variations of RMSD vs. distance. Globally induced conformational change could also be illustrated through PCA, which could only be found in bound sRNA. These results are in agreement with the previous reports[32,35,40] that K-turn RNAs fold tightly upon protein binding and might undergo a protein-induced fit mechanism through both FRET experiments and MD simulations. As shown in Fig. 7C, the average angle between C-stem and NC-stem (defined as previously mentioned) in apo-sRNA is ∼15° larger than that in bound sRNA. In trajectory 3, which has the most significant principal movement, this angle difference is up to ∼35° (data not shown). This is consistent with the previous reports[40,43,89] that the larger and more flexible angle in apo-sRNA confirmed bound sRNA with the stabilization upon L7Ae binding.

Biological function of sRNA motion in specific 2′-O-methylation

Gagnon et al.[90] also confirmed the induced fitting mechanism of K-turn sRNA upon binding with L7Ae and Nop56/58. The assembly and sRNA's structure change are temperature-dependent, indicating the key role of the sRNA's flexibility in sRNP assembly. Furthermore, mutant experiments indicate the impact of sRNA's conformation during the sRNP assembly pathway. Crystal structure suggests that

Table 2 Unfolding kinetics of bound sRNA, apo-sRNA, and apo-L7Ae

Property		τ (ns)	A	B	R^2
Bound sRNA	Q_f	3.605 ± 0.0254	0.516 ± 0.0021	0.177 ± 0.0005	0.968
	R_g	17.432 ± 0.4875	1.034 ± 0.0596	14.022 ± 0.0758	0.908
Bound L7Ae	Q_f	6.294 ± 0.0900	0.136 ± 0.0007	0.731 ± 0.0006	0.929
Complex	Q_b	20.932 ± 0.4230	0.494 ± 0.0055	0.372 ± 0.0058	0.973
Apo-sRNA	Q_f	1.609 ± 0.0060	0.755 ± 0.0017	0.054 ± 0.0003	0.988
Apo-L7Ae	Q_f	1.699 ± 0.0272	0.067 ± 0.0005	0.800 ± 0.0002	0.743
Mutant complex	Q_b	1.496 ± 0.0485	0.378 ± 0.0075	0.683 ± 0.0015	0.575

Q_f and Q_b are fitted by $y = Ae^{(-t/\tau)} + B$. R_g is fitted by $y = Ae^{t/\tau} + B$.

18518 | *Phys. Chem. Chem. Phys.*, 2013, **15**, 18510–18522

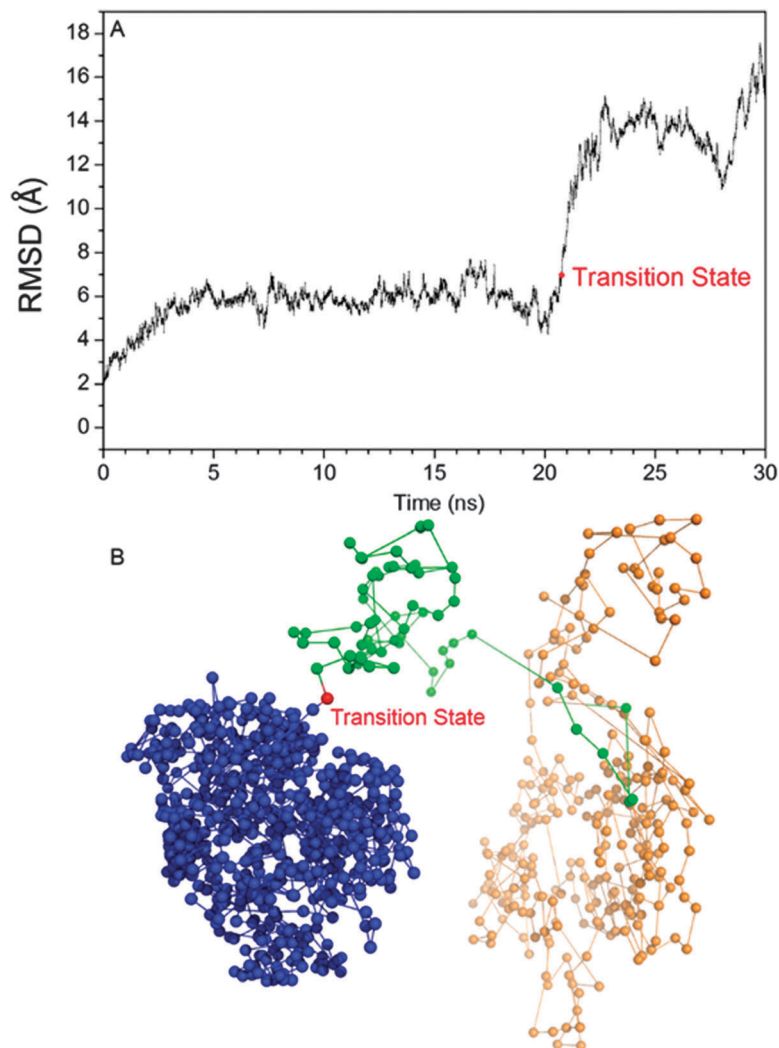This journal is © the Owner Societies 2013

Fig. 9 Transition state search for a representative trajectory. RMSDs in high temperature simulation *vs.* time are shown in (A). Results of multidimensional scaling (MDS) analyses are shown in (B). Blue points for the first RMSD plateau, red point for TS ensembles, green points for the second RMSD plateau, and orange ones for the third plateau.

the GAEK motif in Nop56/58's α9A closely interacts with the last 2 nucleotides of sRNA[37] (PDB code: 3NMU, orange in Fig. S3, ESI†), which contains a 34-mer model K-turn sRNA, an L7Ae, a fibrillarin, a Nop56/58 protein, and a 2′-O-methylation substrate rRNA fragment (12-mer). The deletion of these 2 nucleotides might induce the disorder of the GAEK motif. Consequently, because of the key role of the GAEK motif in the placement of the substrate rRNA, only 3 nt of the rRNA were observed in the crystal structure and determined to be oriented vertically to it in the full complex (PDB code: 3NVK). Additionally, the structure of a complex lacking the substrate rRNA was also identified (PDB code: 3NVI). Three structures were aligned based on Cα atoms with the starting structure used in this study (Fig. S3, ESI†). Four structures have motion propensities similar to those shown in our principal component analysis. The structure with the deletion of the last 2 nucleotides (cyan) is in an environment most similar to our MD system (no interactions with Nop56/58), and this complex has the most significant trend of

fluctuation in the alignment of the structures. The motion detected in the principal component analysis might be important for the two chaperons. As the two stems fluctuate, the interactions between sRNA and L7Ae may change. To investigate the relationship between the two stems' closing and the specific recognition, the distances between atoms that form hydrogen bonding interactions were analyzed. As shown in Fig. S4 (ESI†), hydrogen bond (O6 of G19 and N of N33, in red) and hydrogen bond (N7 of G19 and N of N33, black) were rather weak for trajectory 3. The threshold of hydrogen bond (3.5 Å) is labeled with a blue line. As the two stems closing, $HB_{G19O6\text{-}N33N}$ is stable, while $HB_{G19N7\text{-}N33N}$ changes from ~ 5 Å to 3.5 Å. This confirms the interactions with the dynamics adjustment. The stability of the specifically bound complex might be optimized during this change. As mentioned above, a diRNP model[37] was proposed and demonstrated in Fig. S1A (ESI†). The release of L7Ae off its original site would facilitate the binding of rRNA and the methylation. Summarizing from the models above, the

large motion of sRNA may have key biological functions. Binding between sRNA and L7Ae initiates the assembly of sRNP.[11,22,23] As previously mentioned, the binding may be optimized with the closing of sRNA. While, in the catalytic process, sRNA's flexibility and the V-shaped stems' opening motion, which was also observed in PCA, enable the transition of fibrillarin and its separation from L7Ae so that the methylation could proceed (shown in Fig. S1B and C, ESI†).

Unfolding and folding pathways

According to the unfolding kinetics and the transition state analysis, the unfolding/unbinding pathways for sRNA and L7Ae were constructed and are shown in Fig. 10. At the tertiary unfolding half time of bound sRNA, there were 7 out of 24 (29.2% in folded state) native contacts within sRNA, 153 out of 164 (93.3% folded state) native contacts within L7Ae, and 15 out of 16 (93.8% folded state) native binding contacts between sRNA and L7Ae. The average structure of the transition state is more native-like than the sRNA Q_f half time structure, with the same number of interface native contacts. At the tertiary unfolding half time of bound L7Ae, there were 4, 145, and 14 native contacts, respectively within sRNA, L7Ae, and between the two components. At the unbinding half time, native contacts within sRNA and between the binding interfaces had almost been lost. At the unfolded state, there were still 138 native contacts within L7Ae, one tertiary contact between G17 and G19 within sRNA, and two native contacts between sRNA and L7Ae remaining. Interestingly, two native contacts of G17/K30 and G17/V90 always exist during the unfolding of bound sRNA. Furthermore, they are also included in the TSE of bound sRNA. These base–residue pairs might be a nucleus and play a key role in the folding of bound sRNA.

U18G mutagenesis analyses

Previous mutagenesis experiments show that U18G mutation can disrupt the binding between the K-turn sRNA and protein L7Ae.[11,14,23–25] A 10 ns MD simulation was performed on U18G

Fig. 10 Unfolding pathway for bound sRNA and L7Ae. (A) Folded state; (B) transition state; (C) tertiary unfolding of sRNA; (D) tertiary unfolding of L7Ae; (E) tertiary unbinding of bound sRNA and L7Ae; (F) unfolded state. Spheres represent bases/residues in native contacts. Stable electrostatic interaction between G17 and K30 and stable hydrophobic interaction between G17 and V90 are labelled in green spheres. Other native contacts are in pink.

mutant complex. The loop between β2 and α3 was pushed away by the steric effect of G18 (shown in Fig. S6, ESI†). Interactions between U18G mutant sRNA and L7Ae were also calculated. Comparison with WT sRNA, a strong hydrogen bond between O4 of U18 (O6 of G18 in mutant sRNA) and Asp54 disappeared for U18G complex. Furthermore, two electrostatic interactions of U18–K79 and U18–K30, as well as a hydrophobic interaction between C16 and Pro91, were disrupted. These interactions might play key roles in the stability of WT complex.

The unbinding kinetics of the U18G mutant complex was also analyzed. The unbinding kinetics of the mutant complex is shown in Fig. S5 (ESI†). The unbinding half-time of the U18G mutant complex was 1.496 ± 0.049 ns, which was much faster than that of the WT complex. This suggests that this transversion mutation from U to G might decrease the stability of the complex. Average structure of transition state (TS) ensembles is shown and aligned with the WT complex in Fig. S7 (ESI†). It is less native-like than the WT complex.

Further discussion on the specific binding mechanism

As discussed on interactions between K-turn sRNA and L7Ae and unfolding pathways, a putative mechanism for specific recognition and binding between the two components was deduced. Firstly, since an electrostatic interaction is stable in the high-temperature simulations, this non-specific interaction might be the driving force for the attraction between sRNA and L7Ae. As the two chaperones get closer to each other, further binding builds a stable hydrophobic core, which in turn stabilizes the complex. Meanwhile, global conformation would be induced fit the binding interface (global induced fit). Finally, locally induced fitting near the binding site occurs under hydrogen bonding interactions, enhancing the binding stability. Additionally, sRNA would fluctuate between closing and opening to have a hinge-like motion. The fluctuation optimizes the interactions and keeps the complex stable. Furthermore, this hinge motion might play key roles in the following repositioning of fibrillarin and L7Ae for methylation catalytic procedure.

Conclusion

In the current study, multiple-trajectory MD simulations showed a tightly kinking archaeal box C/D sRNA in complex with protein L7Ae at 298 K and the unfolding process of it at 498 K. Comparison between bound and apo-structures confirmed induced fitting might contribute a lot in specific binding of the two chaperones. Hydrogen bonds were considered as the main driven force for the local induced fit. Principal component analysis not only shows the global conformational changes of sRNA, but also an opening and closing motion of bound sRNA, which could be thought to be critical in the previously proposed diRNP model in the sRNP assembly and the methylation process. Through high-temperature unfolding pathways, the binding procedure could be summarized as (1) tertiary binding, (2) protein tertiary folding and (3) RNA tertiary folding. A putative mechanism that could explain the selection specificity is from non-specific attraction (electrostatic interaction) to specific interaction
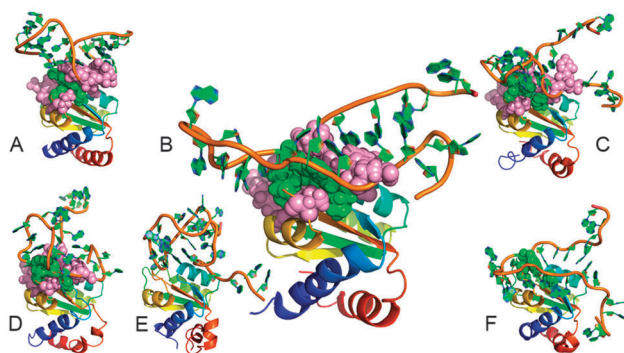
(hydrogen bonds), with an intermediate complex stabilized by a hydrophobic core.

Acknowledgements

References

1 E. S. Maxwell and M. J. Fournier, *Annu. Rev. Biochem.*, 1995, **64**, 897–934.

2 J. P. Bachellerie and J. Cavaille, *Trends Biochem. Sci.*, 1997, **22**, 257–261.

3 J. P. Bachellerie, J. Cavaille and A. Huttenhofer, *Biochimie*, 2002, **84**, 775–790.

4 P. P. Dennis and A. Omer, *Curr. Opin. Microbiol.*, 2005, **8**, 685–694.

5 S. L. Reichow, T. Hamma, A. R. Ferre-D'Amare and G. Varani, *Nucleic Acids Res.*, 2007, **35**, 1452–1464.

6 W. A. Decatur and M. J. Fournier, *Trends Biochem. Sci.*, 2002, **27**, 344–351.

7 W. A. Decatur and M. J. Fournier, *J. Biol. Chem.*, 2003, **278**, 695–698.

8 Z. Kiss-Laszlo, Y. Henry, J. P. Bachellerie, M. Caizergues-Ferrer and T. Kiss, *Cell*, 1996, **85**, 1077–1088.

9 Z. Kiss-Laszlo, Y. Henry and T. Kiss, *EMBO J.*, 1998, **17**, 797–807.

10 D. A. Samarsky and M. J. Fournier, *Mol. Cell. Biol.*, 1998, **18**, 3431–3444.

11 N. J. Watkins, A. Dickmanns and R. Luhrmann, *Mol. Cell. Biol.*, 2002, **22**, 8342–8352.

12 N. J. Watkins, V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, M. Wilm, M. Rosbash, C. Branlant and R. Luhrmann, *Cell*, 2000, **103**, 457–466.

13 D. J. Klein, T. M. Schmeing, P. B. Moore and T. A. Steitz, *EMBO J.*, 2001, **20**, 4214–4221.

14 J. F. Kuhn, E. J. Tran and E. S. Maxwell, *Nucleic Acids Res.*, 2002, **30**, 931–941.

15 A. Niewmierzycka and S. Clarke, *J. Biol. Chem.*, 1999, **274**, 814–824.

16 A. D. Omer, T. M. Lowe, A. G. Russell, H. Ebhardt, S. R. Eddy and P. P. Dennis, *Science*, 2000, **288**, 517–522.

17 A. D. Omer, S. Ziesche, H. Ebhardt and P. P. Dennis, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5289–5294.

18 L. B. Weinstein and J. A. Steitz, *Curr. Opin. Cell Biol.*, 1999, **11**, 378–384.

19 T. Moore, Y. Zhang, M. O. Fenley and H. Li, *Structure*, 2004, **12**, 807–818.

20 D. L. Lafontaine and D. Tollervey, *Mol. Cell. Biol.*, 2000, **20**, 2650–2659.

21 A. Fatica and D. Tollervey, *Nat. Struct. Biol.*, 2003, **10**, 237–239.

22 N. M. Cahill, K. Friend, W. Speckmann, Z. H. Li, R. M. Terns, M. P. Terns and J. A. Steitz, *EMBO J.*, 2002, **21**, 3816–3828.

23 R. Rashid, M. Aittaleb, Q. Chen, K. Spiegel, B. Demeler and H. Li, *J. Mol. Biol.*, 2003, **333**, 295–306.

24 L. B. Szewczak, S. J. DeGregorio, S. A. Strobel and J. A. Steitz, *Chem. Biol.*, 2002, **9**, 1095–1107.

25 E. J. Tran, X. Zhang and E. S. Maxwell, *EMBO J.*, 2003, **22**, 3930–3940.

26 K. Ye, R. Jia, J. Lin, M. Ju, J. Peng, A. Xu and L. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13808–13813.

27 A. H. Antonioli, J. C. Cochrane, S. V. Lipchock and S. A. Strobel, *RNA*, 2010, **16**, 762–768.

28 S. A. Strobel, P. L. Adams, M. R. Stahley and J. Wang, *RNA*, 2004, **10**, 1852–1854.

29 R. M. Clegg, *Curr. Opin. Biotechnol.*, 1995, **6**, 103–110.

30 T. Ha, *Methods*, 2001, **25**, 78–86.

31 T. A. Goody, S. E. Melcher, D. G. Norman and D. M. Lilley, *RNA*, 2004, **10**, 254–264.

32 A. K. Wozniak, S. Nottrott, E. Kuhn-Holsken, G. F. Schroder, H. Grubmuller, R. Luhrmann, C. A. Seidel and F. Oesterhelt, *RNA*, 2005, **11**, 1545–1554.

33 K. T. Schroeder, P. Daldrop and D. M. Lilley, *Structure*, 2011, **19**, 1233–1240.

34 J. Wang, T. Fessl, K. T. Schroeder, J. Ouellet, Y. Liu, A. D. Freeman and D. M. Lilley, *Biophys. J.*, 2012, **103**, 2541–2548.

35 B. Turner, S. E. Melcher, T. J. Wilson, D. G. Norman and D. M. Lilley, *RNA*, 2005, **11**, 1192–1200.

36 I. Vidovic, S. Nottrott, K. Hartmuth, R. Luhrmann and R. Ficner, *Mol. Cell*, 2000, **6**, 1331–1342.

37 S. Xue, R. Wang, F. Yang, R. M. Terns, M. P. Terns, X. Zhang, E. S. Maxwell and H. Li, *Mol. Cell*, 2010, **39**, 939–949.

38 F. Razga, J. Koca, A. Mokdad and J. Sponer, *Nucleic Acids Res.*, 2007, **35**, 4007–4017.

39 J. Curuksu, J. Sponer and M. Zacharias, *Biophys. J.*, 2009, **97**, 2004–2013.

40 V. Cojocaru, S. Nottrott, R. Klement and T. M. Jovin, *RNA*, 2005, **11**, 197–209.

41 F. Razga, J. Koca, J. Sponer and N. B. Leontis, *Biophys. J.*, 2005, **88**, 3466–3485.

42 F. Razga, M. Zacharias, K. Reblova, J. Koca and J. Sponer, *Structure*, 2006, **14**, 825–835.

43 N. Spackova, K. Reblova and J. Sponer, *J. Phys. Chem. B*, 2010, **114**, 10581–10593.

44 F. Bleichert, K. T. Gagnon, B. A. Brown, 2nd, E. S. Maxwell, A. E. Leschziner, V. M. Unger and S. J. Baserga, *Science*, 2009, **325**, 1384–1387.

45 D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvááry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 11*, University of California, San Francisco, 2010.

46 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

47 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

48 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

49 D. Baker, *Nat. Struct. Biol.*, 1998, **5**, 1021–1024.

50 J. E. Shea and C. L. Brooks, 3rd, *Annu. Rev. Phys. Chem.*, 2001, **52**, 499–535.

51 A. R. Fersht and V. Daggett, *Cell*, 2002, **108**, 573–582.

52 H. F. Chen and R. Luo, *J. Am. Chem. Soc.*, 2007, **129**, 2930–2937.

53 H. F. Chen, *J. Chem. Theory Comput.*, 2008, **4**, 1360–1368.

54 H. F. Chen, *PLoS One*, 2009, **4**, e6516.

55 F. Qin, Y. Chen, Y. X. Li and H. F. Chen, *J. Chem. Phys.*, 2009, **131**, 115103.

56 Y. Chen, Y. J. He, M. Wu, G. Yan, Y. Li, J. Zhang and H. F. Chen, *Biopolymers*, 2010, **93**, 578–586.

57 F. Qin, Y. Chen, M. Wu, Y. Li, J. Zhang and H. F. Chen, *RNA*, 2010, **16**, 1053–1061.

58 F. Qin, Y. Jiang, Y. Chen, M. Wu, G. Yan, W. Ye, Y. Li, J. Zhang and H. F. Chen, *Phys. Chem. Chem. Phys.*, 2011, **13**, 1407–1412.

59 F. Qin, W. Ye, Y. Chen, X. Chen, Y. Li, J. Zhang and H. F. Chen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 538–545.

60 W. Ye, Y. Chen, W. Wang, Q. Yu, Y. Li, J. Zhang and H. F. Chen, *PLoS One*, 2012, **7**, e36382.

61 W. Ye, F. Qin, J. Zhang, R. Luo and H. F. Chen, *PLoS One*, 2012, **7**, e43788.

62 The PyMOL Molecular Graphics System, Schrödinger, LLC, 2010.

63 A. Amadei, A. B. Linssen and H. J. Berendsen, *Proteins*, 1993, **17**, 412–425.

64 A. Kitao and N. Go, *Curr. Opin. Struct. Biol.*, 1999, **9**, 164–169.

65 H. J. Berendsen and S. Hayward, *Curr. Opin. Struct. Biol.*, 2000, **10**, 165–169.

66 R. Baron and N. A. Vellore, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 12509–12514.

67 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

68 A. Bakan, L. M. Meireles and I. Bahar, *Bioinformatics*, 2011, **27**, 1575–1577.

69 D. E. Koshland, *Proc. Natl. Acad. Sci. U. S. A.*, 1958, **44**, 98–104.

70 D. D. Boehr and P. E. Wright, *Science*, 2008, **320**, 1429–1430.

71 P. Csermely, R. Palotai and R. Nussinov, *Trends Biochem. Sci.*, 2010, **35**, 539–546.

72 S. Kumar, B. Ma, C. J. Tsai, N. Sinha and R. Nussinov, *Protein Sci.*, 2000, **9**, 10–19.

73 B. Ma, S. Kumar, C. J. Tsai and R. Nussinov, *Protein Eng.*, 1999, **12**, 713–720.

74 B. Ma, M. Shatsky, H. J. Wolfson and R. Nussinov, *Protein Sci.*, 2002, **11**, 184–197.

75 C. J. Tsai, S. Kumar, B. Ma and R. Nussinov, *Protein Sci.*, 1999, **8**, 1181–1190.

76 C. J. Tsai, B. Ma and R. Nussinov, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9970–9972.

77 C. J. Tsai, B. Ma, Y. Y. Sham, S. Kumar and R. Nussinov, *Proteins*, 2001, **44**, 418–427.

78 T. R. Weikl and C. von Deuster, *Proteins*, 2009, **75**, 104–110.

79 T. Wlodarski and B. Zagrovic, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 19346–19351.

80 D. Bucher, B. J. Grant and J. A. McCammon, *Biochemistry*, 2011, **50**, 10530–10539.

81 N. J. Anthis, M. Doucleff and G. M. Clore, *J. Am. Chem. Soc.*, 2011, **133**, 18966–18974.

82 D. A. Silva, G. R. Bowman, A. Sosa-Peinado and X. Huang, *PLoS Comput. Biol.*, 2011, **7**, e1002054.

83 M. Levitt, *J. Mol. Biol.*, 1983, **168**, 621–657.

84 J. Sammon Jr, *IEEE Trans. Comput.*, 1969, **C–18**, 401–409.

85 A. Matouschek, J. T. Kellis, Jr., L. Serrano and A. R. Fersht, *Nature*, 1989, **340**, 122–126.

86 A. Matouschek and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 7814–7818.

87 D. E. Otzen, L. S. Itzhaki, N. F. elMasry, S. E. Jackson and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10422–10425.

88 B. Turner and D. M. Lilley, *J. Mol. Biol.*, 2008, **381**, 431–442.

89 V. Cojocaru, R. Klement and T. M. Jovin, *Nucleic Acids Res.*, 2005, **33**, 3435–3446.

90 K. T. Gagnon, X. Zhang, P. F. Agris and E. S. Maxwell, *J. Mol. Biol.*, 2006, **362**, 1025–1042.