



# NIH Public Access

## Author Manuscript

*J Phys Chem B.* Author manuscript; available in PMC 2013 June 14.

Published in final edited form as:  
*J Phys Chem B.* 2012 June 14; 116(23): 6819–6831. doi:10.1021/jp212614b.

## Assembly of the Five-Way Junction in the Ribosomal Small Subunit Using Hybrid MD-Gō Simulations

Ke Chen<sup>†</sup>, John Eargle<sup>‡</sup>, Jonathan Lai<sup>‡</sup>, Hajin Kim<sup>¶</sup>, Sanjaya Abeysirigunawardena<sup>§</sup>, Megan Mayerle<sup>||</sup>, Sarah Woodson<sup>§</sup>, Taekjip Ha<sup>†,‡,¶</sup>, and Zaida Luthey-Schulter<sup>\*,†,‡,¶</sup>

<sup>†</sup>Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL

<sup>‡</sup>Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL

<sup>¶</sup>Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL

<sup>§</sup>Department of Biophysics, Johns Hopkins University, Baltimore, MD

<sup>||</sup>Program in Cell, Molecular, Developmental Biology and Biophysics, Department of Biology, Johns Hopkins University, Baltimore, MD

### Abstract

The assembly of the bacterial ribosomal small subunit (SSU) begins with the folding of the five-way junction upon interaction with the primary binding protein S4. This complex contains the largest contiguous molecular signature, which is a conserved feature in all bacterial 16S rRNAs. In a previous study, we used all-atom molecular dynamics simulations to demonstrate that the coevolving signature in the N-terminus of S4 is intrinsically disordered and capable of accelerating the binding process through a fly-casting mechanism. In this paper, we capture the simultaneous folding and binding of the five-way junction and r-protein S4, by use of a structure-based G<sub>0</sub>-potential implemented within the framework of the all-atom molecular dynamics CHARMM force field. Comparisons between the all-atom MD simulations and FRET experiments identify multiple metastable conformations of the naked five-way junction without the presence of S4. Using the hybrid MD-G<sub>0</sub> simulations, different folding pathways are observed for the refolding of the five-way junction upon partial binding of S4. Our simulations illustrate the complex nature of RNA folding in the presence of a protein binding partner and provide insight into the role of population shift and the induced fit mechanisms in the protein:RNA folding and binding process.

### Keywords

SHAPE; single-molecule FRET; molecular dynamics simulations; transferable force field; structure-based G<sub>0</sub>-potential; ribosomal assembly

---

\*To whom correspondence should be addressed.

### Supporting Information

Supporting information includes: further details about the hybrid MD-G<sub>0</sub> simulations (Figure S1); a summary of all simulations done in this paper (Table S1); example time traces from the FRET experiments showing the heterogeneity of the RNA conformations (Figure S2); and additional figures (Figure S3 – S5) validating the RNA conformational clustering in the MD and hybrid MD-G<sub>0</sub> simulations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Introduction

The assembly of the ribosome from over 50 individual components is a highly coordinated process that is crucial for cell growth. The study of ribosomal assembly first drew attention 40 years ago, when Nomura *et.al* summarized the hierarchy of the binding of ribosomal proteins (r-proteins) to the ribosomal RNA (rRNA) using reconstitution experiments. The results showed that the assembly of the *E. coli* ribosomal small subunit (SSU) was dependent on the presence of S4 and five other primary binding proteins, which are required to be in place before the remaining r-proteins can be incorporated.<sup>1</sup> Two of the primary binding proteins, S4 and S7, were later identified as the only assembly initiator proteins based on their noncooperative binding during the onset of the assembly process.<sup>2</sup> More recently, pulse-chase and quantitative mass spectrometry measurements of the binding kinetics for all r-proteins showed that the 5'-domain binding proteins, especially S4, bind more quickly than proteins in the central or 3' domain, indicating a 5' to 3' directionality in the assembly process.<sup>3-5</sup>

Progress in biophysical approaches allows the study of ribosomal assembly to include both protein binding and rRNA folding in the absence or presence of r-proteins. Förster resonance energy transfer (FRET) has been used to probe large conformational changes of the three-way junction in the central domain of 16S rRNA that binds the primary binding protein S15.<sup>6-8</sup> It was shown that the RNA molecule went through a broad distribution of conformations at intermediate Mg<sup>2+</sup> concentrations, however, the binding of S15 induced the three-way junction into its native folded state. Chemical footprinting of RNA further pushes the description of RNA-folding to nucleotide resolution. Particularly, time-resolved hydroxyl radical footprinting of the 5'-domain of the *E. coli* 16S rRNA identified two folding stages in the folding of this domain.<sup>9,10</sup> The early stage proceeds through multiple parallel pathways while the later is cooperative and induced by protein binding. More recently, the use of selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) confirmed the role of the primary binding protein S4 in inducing and rearranging the five-way junction conformation during folding by measuring local RNA structure flexibility at single nucleotide resolution.<sup>11</sup>

Computational methods, in principle, may yield a more comprehensive view of the assembly process such that both protein and RNA behaviors can be taken into account simultaneously at a high structural and energetic resolution. Methods and force fields to simulate peptide and protein folding have made considerable advances since the pioneering studies of Scheraga and coworkers,<sup>12-14</sup> which were among the first to capture structural transitions. Although the availability of ribosome crystal structures has allowed all-atom molecular dynamics (MD) simulations to investigate functions of the intact ribosomes, computational studies of ribosomal assembly have been mostly based on reduced representations due to the number of particles as well as the time-scale of the assembly process. Studies of ribosome stability, including coarse-grained Monte Carlo simulation,<sup>15</sup> implicit solvent model,<sup>16</sup> and self-consistent pair contact probability approximation,<sup>17</sup> provide insights into the correlation between order of protein binding with electrostatic energy or flexibility of the rRNA binding site.

In this work, we used a structure-based hybrid MD-G model to estimate the assembly pathway of a rRNA:protein complex which occurs on a millisecond experimental timescale. G models were developed in the context of protein folding according to the energy landscape theory.<sup>18</sup> The theory states that evolution has shaped the folding landscape of proteins in the shape of a rugged funnel. Frustration and barriers to folding come from competing interactions in the energy function which one can minimize by including stabilizing forces from native contacts and thereby reducing the probability of being trapped

in non-native structures. Early studies with simple G<sub>0</sub> models showed that the topology of the native structure of a protein determines to a large extent its folding mechanism<sup>19–22</sup> as well as its binding to other proteins.<sup>23–25</sup> Simple G<sub>0</sub> models for RNA have also been able to differentiate the folding mechanisms of relatively small RNA pseudoknots.<sup>26</sup>

In order to incorporate sequence-specific energetic contributions of the side-chains and base pairing, all-atom G<sub>0</sub> models have been developed for protein folding<sup>27–29</sup> and transitions in nucleic acids. Specifically with the RNA models, one has been able to study transitions in riboswitches<sup>30</sup> and accommodation of tRNA in the ribosome.<sup>31</sup>

Using an all-atom structure-based G<sub>0</sub> model, we study the simultaneous folding and binding of the initiator r-protein S4 and its minimal binding site on the five-way junction in 16S rRNA.<sup>32</sup> This system is essential to the assembly process not only because the five-way junction is the part of the 5'-domain that folds first during ribosomal SSU assembly, but also due to its critical role in the ribosomal evolution. Helix 16 (h16) of the five-way junction is the largest bacterial structure signature (idiosyncrasies in the rRNA that are characteristic of the individual domain of life<sup>33,34</sup>) in the ribosomal SSU, and it is in contact with a bacterial protein signature in the S4 N-terminus (S4N).<sup>35</sup> The stability of h16 and intrinsic disorder of S4N revealed by MD simulations suggested that the interactions between this pair of coevolving signatures were key to the rapid recognition between the protein and RNA.<sup>36</sup> Here, we showed results from new MD simulations of the five-way junction, and identify multiple stable states of the naked RNA highlighting the rugged nature of its folding landscape. Starting from selected stable states of the five-way junction, binding of S4 was initiated from a large distance away using the structure-based G<sub>0</sub> model. To our knowledge, this is the first example of simultaneous folding and binding of RNA and protein molecules together using all-atom structure-based G<sub>0</sub> model. With statistics obtained from hundreds of replicate simulations, we were able to reconstruct the binding pathway of S4 to the five-way junction. In addition, it also provides molecular details on how the intrinsically disordered S4N accelerates the binding process.

## Methods

### RNA preparation

In the fluorescence resonance energy transfer (FRET) experiments, rRNA containing the *E. coli* 16S rRNA (nts 21–46, 395–556) was used to construct the minimal five-way junction according to Bellur et al.,<sup>32</sup> with the sequence CUCAA capping the gap between nucleotide G46 and C395. Two extensions were added to target h16 and h3 for fluorescent labeling. First, a 26-nucleotide insertion (5'-CGCGUCGCCAGACCAGACGCCUCCGCG-3') was placed between C419 and G424, which extends h16 by 4 GC pairs for higher stability. Then a 15-nucleotide RNA primer labeled with the Cy3 dye (5'-GAGCGUCUGGUCCUGG-Cy3-3') (Integrated DNA Technologies, Inc) was hybridized to the h16 insertion from the 8th to the 22nd position. The 3' end of h3 (C556) was extended by 38 nucleotides (5'-AGGACGACACACUUUUGGACAGGGACACACAGGACACAGG-3') with a 37-nucleotide complementary DNA strand labeled with Cy5 and biotin (5'-biotin-CCTGTGTCCTGTCAAAGTGTGTCGTCC-Cy5-3') (Integrated DNA Technologies, Inc).

Molecular models for the simulations were created using crystal structure of the five-way junction and S4 from the *E. coli* ribosomal SSU (PDB code: 2I2P). The five-way junction was modeled according to the minimal S4 binding system suggested by Bellur et al.<sup>32</sup> A five-membered loop (CUCAA) was added to cap the end of h4 producing a single, continuous RNA molecule, and h17 which normally extends away from the S4 binding site was truncated and capped with a seven-membered loop (UUUUGC). These two additional

loops were taken from the *E. coli* SSU (PDB ID 2I2P, residues 618–622) and the anticodon loop of tRNA<sup>Gln</sup> (PDB ID 2RD2, residue 932–938) with two mutations (C934U and A937C), respectively.

Results from the selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) experiment performed on the five-way junction RNA<sup>11</sup> were used to compare with simulation data in the paper. The 3' end of the native five-way junction was extended with the sequence 5'-aaagcaatatagactcgggatgaggaCATCATGCCCTTACGA-3' for primer extension in the SHAPE experiment. The 3' extension is shown not affect RNA folding or S4 binding. Secondary structures of the RNA molecule used in SHAPE and FRET experiments, as well as in simulations are depicted in Figure 1.

### Fluorescence resonance energy transfer (FRET) measurement

The RNA was immobilized on the surface with the biotin on the h3 primer, using the same method as previously reported in Hohng et al.<sup>37</sup> Single molecule FRET traces were measured using a TIR-based fluorescence microscope and analyzed using conventional methods.<sup>37</sup> The imaging solution contained 80 mM K-HEPES at pH 7.6, 300 mM KCl, the desired amount of MgCl<sub>2</sub>, saturated Trolox of 2 mM, 0.6% glucose, 0.1 mg/ml glucose oxidase (Sigma), and 0.02 mg/ml catalase (Sigma). For the data analysis, each FRET histogram was built by accumulating 100 ms data bins of 300–500 molecules.

### Molecular dynamics simulations

The five-way junction and the r-protein S4 were neutralized with Na<sup>+</sup> and/or Mg<sup>2+</sup> placed by ionize,<sup>38</sup> and hydrogens were added to the system by psfgen in VMD.<sup>39</sup> Water molecules were placed to complete the primary solvation shells for all Mg<sup>2+</sup> atoms. 400 mM NaCl was added to the final solution to match the corresponding experimental conditions. Solvent was prepared in two phases: first Solvate<sup>40</sup> was run to place the first two solvation layers and then the solvate plugin to VMD was used to complete the water box.

MD simulations were run using NAMD2<sup>41</sup> with the CHARMM27 force field.<sup>42</sup> Stepwise minimization and equilibration were performed according to the protein:RNA simulation protocol established in Eargle et al.<sup>43</sup> Production runs used periodic boundary conditions and the NPT ensemble with pressure set to 1 atmosphere and temperature set to 298 K. Multiple time-stepping was used to calculate bonded interactions at 1 fs, van der Waals (vdW) interactions every 2 fs, and electrostatic forces every 4 fs. Particle mesh ewald summation was used to evaluate electrostatic interactions, and the vdW force calculations used a cutoff of 12 Å and a switching distance of 10 Å.

A total of 400 nanoseconds of MD simulations on the naked five-way junction are presented in this paper, including 100 ns of the unfolding simulations without Mg<sup>2+</sup> ion; 100 ns of equilibrium simulation from the crystal structure with Mg<sup>2+</sup> ions; 100 ns refolding simulations with Mg<sup>2+</sup> from each of the two conformations selected from the unfolding trajectory (see Results and Discussion). In all three simulations with Mg<sup>2+</sup>, 25 Mg<sup>2+</sup> ions were placed according to the electrostatics of the five-way junction starting conformation, among which only 4 were placed in direct contact with the RNA molecule. R-protein S4 was not included in these MD simulations. In addition, to compare nucleotide flexibility in simulation with the SHAPE reactivity, one 100 ns MD simulation of the five-way junction bound with S4 was performed.

### The structure-based Gō model

In our heavy-atom hybrid MD-G<sub>0</sub> models, all bonded terms (including bonds, angles, dihedrals and impropers) are described by the transferrable CHARMM27 force field<sup>42</sup> with

additional improper angle potentials to maintain chirality (see Supporting Material). All non-bonded potentials (vdW and electrostatics) were replaced by the knowledge based G - potential. The G -potential is formulated with respect to a reference structure, which can either be the equilibrated native structure for the folding of RNA/protein or a target conformation that is important for the function of the biomolecules. For atom pairs closer than 4 Å within a molecular chain or pairs closer than 4.4 Å between the RNA and protein chains in the reference structure, the pair is defined as a native contact, and is subject to a Lennard-Jones-style potential:

$$V_{native}(\sigma_{ij}^{native}, r_{ij}) = 4\epsilon^{native} \left[ \left( \frac{\sigma_{ij}^{native}}{r_{ij}} \right)^a - \left( \frac{\sigma_{ij}^{native}}{r_{ij}} \right)^b \right] \quad (1)$$

where  $\sigma_{ij}^{native}$  is determined by the native pairwise distance,  $r_{ij}^{native}$ , between atoms  $i$  and  $j$  in the reference structure:

$$\sigma_{ij}^{native} = \left( \frac{b}{a} \right)^{\frac{1}{b-a}} r_{ij}^{native} \quad (2)$$

such that the potential reaches the minimum at  $\sigma_{ij}^{native} \cdot r_{ij}$  is the instantaneous distance between atoms  $i$  and  $j$  in the simulation,  $\epsilon^{native}$  specifies the depth of the potential, and the exponents –  $a$ ,  $b$  – determine the shape of the potential. All non-native atom pairs with a distance longer than 4 Å in the reference structure experience a repulsive potential of the form:

$$V_{nonnative}(r_{ij}) = \epsilon^{nonnative} \left( \frac{\sigma^{nonnative}}{r_{ij}} \right)^c \quad (3)$$

where the non-native  $\epsilon^{nonnative}$  scales the strength of the potential,  $\sigma^{nonnative}$  determines where the potential equals  $\epsilon^{nonnative}$ , and the exponent,  $c$ , determines the long range decay.

The above described hybrid MD-G model is implemented within the newest version of NAMD2, and our implementation allows for the study of complex systems in which molecules with different properties are present at the same time. It is designed so that each molecule may have its own parameters set independently of the others. In the particular system that we are studying in this paper, we set  $a = 12$  and  $b = c = 6$  for the folding of both protein and nucleic acid. The native interaction strength  $\epsilon^{native}$  is set to 0.1 kcal·mol<sup>-1</sup> and 0.23 kcal·mol<sup>-1</sup> for RNA and protein, respectively, to ensure that S4 and the five-way junction have similar transition temperatures. All contacts within protein and RNA, irrespective of being involved in sidechain-sidechain, sidechain-backbone, backbone-backbone, sugar-sugar, sugar-base, or base-base interactions, are treated equally. To facilitate binding initiated from large separations, protein-nucleic interaction potential is set with exponents  $a = 8$  and  $b = 4$  and  $\epsilon^{native} = 0.15$  kcal·mol<sup>-1</sup>. All non-native parameters are set to  $\epsilon^{nonnative} = 0.01$  kcal·mol<sup>-1</sup>, and  $\sigma^{nonnative} = 2.5$  Å in this study.

### Hybrid MD-Gō folding simulations

The hybrid MD-G simulations were prepared without water molecules or ions, and only the heavy atoms of the five-way junction and S4 protein were used. Two representative states of the partially unfolded five-way junction (discussed in Results and Discussion) were chosen from the all-atom MD simulations as the starting conformations for the simultaneous folding and binding simulations. Given the high degree of secondary structure stability of S4 especially in its globular C-terminal domain,<sup>44,45</sup> we chose a slightly unfolded structure with

an overall RMSD of ~5.2 Å in which the disordered S4 N-terminus has a RMSD of 7.2 Å and the globular S4 C-terminal domain has a RMSD of 4.5 Å. The unfolded S4 and five-way junction were placed ~38 Å (center-of-mass distance) away from each other, such that the closest contacts between them are just under the non-bonded interaction cutoff (12 Å). A 2 fs time step was used for all hybrid MD-G simulations. Each simulation was run for 2,250,000 steps, such that if the five-way junction and S4 were not bound within this time, they were likely to diffuse away from each other. 100 replicates starting from each chosen five-way junction conformation were performed with parameters introduced in the previous section, and statistics were generated. Supporting Material contains a table summarizing all the simulations presented in this paper.

### Analysis of the simulated trajectories

**The angle between RNA helices** is calculated as the angle between the principal axes of inertia extending through each helix, and the resulting angle value is translated into the range [0°, 180°]. With this definition, we measured angles between every pair of helices in the five-way junction such that a ten-member angle vector can be obtained for every frame to depict the five-way junction conformation captured during simulation. Pairwise distances between frames were then calculated based on these angle vectors to get a distance matrix. This matrix was processed using affinity propagation<sup>46</sup> for clustering of the five-way junction conformations. The same matrix was entered into the multi-dimensional scaling method<sup>47</sup> to reduce data dimensionality and to visualize all frames from the trajectories as coordinates in a two-dimensional space.

**Three domains in the r-protein S4** are defined in order to characterize the binding behavior in detail. The S4 N-terminal domain (S4N), which is unresolved in solution NMR experiments<sup>48</sup> and fully structured<sup>49</sup> but still highly fluxional<sup>36</sup> upon interaction with the five-way junction in the crystal structure, is taken from residue 1 to 40. The C-terminal domain of S4 (S4C) is split into two sub-domains: the C1-domain (residue 41 to 96 and 185 to 205) interacts with the lower junction between h18 and h4, while the C2-domain (residue 97 to 184) has contacts with the upper junction between h16 and h17. We focus our analysis on the S4N and S4 C2-domain since these two make the most interactions with the five-way junction. The five-way junction binding sites for these three domains are defined to include nucleotides that are within 5 Å of each domain, as well as any nucleotides base-pairing to them. According to this definition, the number of nucleotides involved in the binding sites are 45, 20 and 28 for the S4N, S4 C1- and S4 C2-domain, respectively.

**Number of native interface contacts** is defined using a cutoff distance of 4.4 Å. This results in 660 native interface contacts between S4 and the rRNA five-way junction, among which 320 are contacts between the S4N and the five-way junction. The C1- and C2-domains of S4 have 98 and 242 native contacts with the five-way junction, respectively. During simulation, if the distance between the two atoms forming a native contact gets within 1.2 times the distance in the reference structure, the corresponding native contact is said to be formed. **Q** is defined as the fraction of all native contacts that is formed in one particular frame.

**Radius of gyration ( $R_G$ )** is a measure of how unfolded (or more precisely, extended) the molecule is. Since it is independent of any pre-alignment of the molecule to remove the translational and rotational degree of freedom, it can be useful to describe the conformational state of the molecule in addition to the conventional RMSD value. Furthermore, in the case of h16 and h18 of the five-way junction,  $R_G$  gives a more sensitive description of the degree of extension of the structure, while it is positively correlated with the angle between the two stems so that the conformational assignment remains consistent.

## Results

### Dynamics of the coevolving signatures of the five-way junction and S4

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) measures local RNA structure flexibility at single nucleotide resolution, and has been shown to successfully predict RNA secondary structure as well as some tertiary interactions.<sup>50–53</sup> Although dynamics information is embedded in the SHAPE chemistry, SHAPE reactivities have only been compared to structural characteristics of static crystal structures or theoretical models of the corresponding RNA molecules. Here, we report all-atom molecular dynamics simulations with explicit solvent and hybrid MD-G<sub>0</sub>G<sub>1</sub> simulations of the five-way junction with S4 bound that directly visualize these motions. As shown in Figure 2, the dynamical flexibility of each nucleotide in the five-way junction extracted from simulations on the *E. coli* system compared very well with the SHAPE reactivity obtained for five-way junction bound to *B. stearothermophilus* S4.<sup>11</sup> The slightly higher nucleotide flexibility in simulations is most likely due to the lower binding affinity of the *E. coli* S4 compared to that of the *B. stearothermophilus* S4 observed in earlier experiments.<sup>54</sup>

The flexibility of each nucleotide was measured as the root mean square deviation (RMSD) of heavy atoms from their average positions during simulation. Two alignment schemes were employed to remove the translational and rotational degrees of freedom. Each frame in the trajectory was either aligned back to the whole five-way junction to account for tertiary interaction change, or aligned to each individual helix to account for secondary structure fluctuations. The overall Pearson correlation coefficient between SHAPE reactivity and nucleotide fluctuations are 0.505 and 0.468 for tertiary and secondary structure alignment in all-atom MD simulation; and 0.618 and 0.604 in the hybrid MD-G<sub>0</sub>G<sub>1</sub> simulations respectively. The slightly higher correlation achieved when aligning the whole RNA molecule supports previous evidence that the SHAPE chemistry measures nucleotide stability not only constrained by base-pair interactions but also by long-range tertiary contacts.<sup>55</sup>

Among the five helices of the five-way junction, h17 shows the largest deviation between simulated nucleotide flexibility and the SHAPE reactivity. However, this discrepancy is expected due to the large structural difference arising from the truncation of h17 in the simulations (see Methods) as compared to the native h17 used in the SHAPE experiments. Differences in the signal intensities for the 5' and 3' end of h3 can be attributed to the “end effects” that strongly affect RNA folding, as well as the use of an extended 3' sequence in SHAPE studies. Despite these difficulties, we still observe the same trend in signal fluctuation (relative intensity among residues) between calculated flexibility and SHAPE reactivity.

Two major reasons have been proposed to account for the high experimental SHAPE reactivity for rigid nucleotides in the structure. First, since the nucleophilicity of the ribose 2' hydroxyl group is sensitive to the electronic influences from nearby chemical groups, a larger than average distance between 2'-OH and the 5' phosphate oxygen atoms (O1P or O2P) of the next nucleotide, or the purine N3 atom or pyrimidine O2 atom may increase the SHAPE reactivity. Second, reactivity might be enhanced by the C2'-endo pucker conformation of the sugar ring instead of the C3'-endo commonly seen in the A-form RNA. These suggested mechanisms were examined and confirmed in our simulations. For example, high reactivities of G517, A520, and G530 can be explained by relatively large average distances (> 4.2 Å) between 2'-OH and the N3 atoms on their bases; and that of A520 and U552 should be a result of large average distances between 2'-OH and the phosphate oxygens. On the other hand, the low reactivity of U421 is likely due to close adjacency between 2'-OH and phosphate oxygens when its flexible base flipped out into the

solvent. One single nucleotide, A533, is seen to adopt a C2'-endo pucker conformation in the crystal structure and through the hybrid MD-G<sub>0</sub> simulations. Note that none of these proposed mechanisms is sufficient to account for all the difference between experimentally measured SHAPE reactivity and computationally calculated flexibility. While the distance argument is supplementary to nucleotide flexibility, sugar pucker conformation has only minimal contribution to the SHAPE reactivity. It is pointed out by Gherghe *et al.* that the C2'-endo pucker conformation is useful in determining the kinetic behavior of the slow folding nucleotides.<sup>56</sup>

In general, nucleotide flexibility averaged over multiple hybrid MD-G<sub>0</sub> simulations (Figure 2 bottom) gives a better prediction of the SHAPE reactivity compared to that from a single MD simulation (Figure 2 top). This is expected, since SHAPE chemistry measures averaged nucleotide flexibility in an ensemble of RNA conformations. If one particular conformation dominates the RNA sample, as in the case of five-way junction bound to r-protein S4, high correlation between measured reactivity and calculated nucleotide fluctuation is expected, and discrepancies may be explained by the dynamics of the structure. On the other hand, if multiple conformations exist in the sample, especially when the relative population of each conformation is hard to estimate, poor correlations will probably result.

Also seen in Figure 2, the dynamics of each nucleotide in MD and hybrid MD-G<sub>0</sub> simulations share the same characteristics. Since simulation with the G<sub>0</sub>-potential involves a bias towards the native/target structure, it is often used to explore large conformational changes during transitions between various states. RMSD per residue was compared between MD and hybrid MD-G<sub>0</sub> simulations, and a linear relationship was observed with high confidence for both RNA ( $R^2 = 0.71$ ) and protein ( $R^2 = 0.66$ ) (Figure 3). Particularly, it has been shown that N-terminal domain of S4 is highly disordered in solution,<sup>36</sup> and the same behavior is seen in the hybrid MD-G<sub>0</sub> simulations (red dots in Figure 3, top panel). Thus, the consistency of nucleotide flexibility between simulations and SHAPE, as well as between two kinds of simulations suggests that structure-based simulation with G<sub>0</sub>-potential is able to recapitulate dynamics of molecules at atomic resolution.

### Multiple conformations of the five-way junction

The dynamics of individual molecules observed in real time complements the structural information from footprinting assays like SHAPE, and thus provides a more detailed picture of molecular interaction. We directly measured the dynamics of the five-way junction by single molecule FRET (smFRET).<sup>57</sup> Fluctuations in the distance between h16 and h3 were measured by hybridizing Cy3 and Cy5 labeled primers to extensions in the RNA sequence (Figure 1). These extensions do not change the RNA secondary structure nor S4 binding affinity (S.A. and M.M., unpublished). Figure 4 A shows the Mg<sup>2+</sup> dependence of the RNA dynamics obtained from smFRET observations. The complex behavior of the unliganded RNA is demonstrated by multiple FRET levels emerging at different Mg<sup>2+</sup> concentrations. At 1 mM Mg<sup>2+</sup>, the majority of the molecules fluctuate rapidly between low FRET states (0.1 ~ 0.35). The fluctuation appears in the histogram as a broad peak instead of two well-defined separated FRET levels. The fluctuations slow down and resolve into low and medium-low states at higher Mg<sup>2+</sup>, as seen in the example trace for 5 mM Mg<sup>2+</sup>. At the same time, a fraction of the molecules temporarily exhibits a medium-high FRET state (0.55), implying that the RNA transiently folds into a structure with smaller distance between h16 and h3. At 20 mM Mg<sup>2+</sup>, the population in this mid-FRET state increases, and excursions to higher FRET levels become visible, as illustrated in the sample trace in Figure 4 A. At 100 mM Mg<sup>2+</sup>, this transition becomes more frequent and thus the high-FRET population increases.

The behaviors described above are not uniformly exhibited by all molecules under the same conditions. There is a high degree of molecule-to-molecule variation. As an example, Figure S2 in the Supporting Material depicts additional traces at 20 mM Mg<sup>2+</sup> showing that some molecules stay mostly in the low-FRET state, a small fraction resided in the medium-high FRET state, and yet others fluctuated quickly between the states. On rare occasions, we find molecules switching between different behaviors. These observations suggest a complex folding energy landscape of the RNA, which bears similarity to the rugged multidimensional folding energy landscape with two distinct conformational states previously reported for the *Tetrahymena* ribozyme.<sup>58,59</sup> In the five-way junction system, we were able to detect a larger number of conformational states describing more complex folding behavior of the ribosomal RNA. Even then, the observed states might still be degenerate due to different configurations of the bound Mg<sup>2+</sup> ions or different conformations of the whole RNA molecule giving rise to similar distances by chance.

All-atom MD simulations of the five-way junction without S4 protein can help to identify meta-stable states corresponding to high and low FRET signals. It was shown in our previous study that the five helices in the five-way junction start to separate after 40 ns of simulation in NaCl buffer, while maintaining their secondary structure (except h3).<sup>36</sup> It follows that conformational states of the unfolding are best described by the relative orientations among the constituent helices, which give rise to a ten-member angle vector (see Methods).

Accordingly, four conformational states of the five-way junction have been identified from the MD unfolding trajectory: the native structure (N), the intermediate state (I), the extended state (II) and the misfolded state (III) (Figure 4 B & C). State I is closest to the native structure, though h17 is stretched and separated from h4. Contacts between h16 and h18 fluctuate, but the two helices remain close to each other. h3 has the largest fluctuations in secondary structure but still maintains most of its tertiary contacts with h18. State II is the most extended conformation; however, the five helices of the molecule keep moving within a plane so that inter-conversion between states I and II may occur occasionally. State III distinguishes itself from the other states by the loss of the planar conformation adopted in the native structure. Particularly, h16 and h3 are moving out of the plane while h17 and h18 move into the plane in such a way that a right angle between h18 and the co-axially stacked h16 & h17 is made. Interestingly, the relative positioning of h16 and h18 in state III is very similar to the conformation adopted in the eukaryotic ribosome structure,<sup>60,61</sup> in which the S4 N-terminus is nonhomologous to that in the bacterial ribosome.<sup>62</sup> The main difference among various conformational states lies in the relative positioning of h16 to h18. It is shown in the Supporting Material that the distribution of angle between h16 and h18 shows three peaks, corresponding to state I (and the native structure N), II and III, respectively.

To explore the stability of each identified conformation, further MD simulations were carried out. Representative conformations of states II, III, as well as the native structure were taken as the starting conformations, and Mg<sup>2+</sup> ions were placed using protocol developed in Eargle *et al.*<sup>43</sup> The same clustering analysis was applied to these magnesium quenching simulations, and the results showed that both states N and III would stay in their local minima within the 100 nanoseconds of simulation. State II, however, was shown to be inter-changeable with state I during the simulation period. Importantly, clustering using either the ten-dimensional angle vector or the distribution of the angles between h16 and h18 showed similar results. This consistency granted us the simplicity to use angle between h16 and h18 as the single criterion in determining conformational states of the five-way junction (see Supporting Material).

Having clarified all major conformers of the five-way junction in MD simulations, we examined their correlations with the FRET states observed in experiments (Figure 4 A). Due to the uncertainty of both FRET labels on the five-way junction posed by large extension in RNA structure (especially for the position of Cy3, see Figure 1), relative distance fluctuations are considered more critical in correlating with the FRET states rather than the absolute distance values. In the unfolding simulation of the five-way junction, distance between the FRET labels on h16 and h3 oscillated over a large range (Figure 4 D), which would correspond to both high and low FRET efficiency values. However, once  $Mg^{2+}$  ions were added, the near-native structure and state III are both locked in the small-distance states (Figure 4 E). Distance between labeled positions on h3 and h16 occasionally reach lower than 50 Å, which could be consistent to the medium-high FRET signals seen in traces obtained under 20mM and 100mM of  $Mg^{2+}$  concentrations. The fact that distance between FRET labels fluctuated in the same range in simulations starting from both the native structure and the misfolded conformation, indicates that the medium-high FRET population, though very small, could still consist of a mixture of conformational states. On the other hand, distance between FRET labels in the simulation starting with the extended state II has fluctuations as much as 40 Å, and the labeled positions seldom get closer than 70 Å. The data demonstrates flexibility of the extended state and is consistent with the observation in experiment that different RNA molecules would oscillate between different low-FRET states. Since the average behavior of hundreds of molecules shows only one peak in the FRET efficiency histogram, the various FRET signal levels should be best described as sub-states within the extended state II of the five-way junction. Further experiments where the S4 protein is included are required to precisely correlate FRET efficiencies with conformations identified in simulations.

### Simultaneous folding and binding of the five-way junction and S4 with Gō-potential

We employed the structure-based  $G_0$  model to study the simultaneous folding and binding of the five-way junction and S4. Two conformations of the five-way junction, state II (extended) and III (mis-folded or non-planar) from MD simulation, were considered as initial conformations in the simulations. These two distinct starting conformations share some, though definitely not all, of the changes in structure upon S4 binding. We present here the shared features of the binding process starting from these two different RNA conformations, differences will be discussed in the next section. Results of simulations starting from state II were used to illustrate the common theme in the five-way junction:S4 binding process. Additional details of the choice of initial conformations are provided in the Methods section.

Within the designed number of steps, the five-way junction and S4 bound in 61 out of the 100 replicates starting from state II and in 72 simulations starting from conformation III. To describe the binding process, we first monitored the center of mass (COM) distances between different domains of S4 (see Methods for domain definition) and their binding sites on the five-way junction. The results for all bound replicates from simulations starting from state II are shown in Figure 5 A. From the histogram, we see two phases of the binding process — a diffusing phase and a binding phase. In the diffusing phase, the two domains of S4 randomly explore the configuration space, each trying to reach into contacts with the five-way junction independently. However, when the COM distance gets below 25 Å between S4N and h16 or below 30 Å between S4 C2-domain and the five-way junction, the binding phase begins. The binding phase consists of two major pathways. In one pathway, the N-terminal domain of S4 binds first, followed by the binding of the C2-domain. In the other, the order is reversed. While binding initiated from the C2-domain seems dominant from the histogram, its diffusion might be relatively free compared to the S4N. As long as

the S4N remains close to the five-way junction, the C2-domain can diffuse as far away as 45 Å and the whole protein may still be able to bind.

To take a closer look at the “binding phase”, we calculated the fraction of native interface contacts (interface Q) between the N- or C2-domains of S4 protein and the five-way junction. When starting from state II of the five-way junction with the two domains of S4 put at equivalent distances from their binding sites, the first contact to be made is, for the majority of the bound replicates (56 out of 61), between the S4 N-terminus and h16. However, this first contact between S4N and h16 does not always lead to complete binding. Starting from state II, for example, it takes another  $5.5 \times 10^5$  time steps (one fourth of total simulation time) on average before the S4 C2-domain makes contact with the five-way junction. After both domains of S4 establish their initial interactions with the five-way junction, the binding phase starts from the left bottom corner of the histogram (Figure 5 B) where both domains have interface Q values close to zero. One dominant binding pathway can be seen clearly from the histogram, in which the C2-domain binds first, and S4 N-terminus follows with two distinct barriers (yellow). The barrier located at interface Q  $\sim$  0.15 represents a binding intermediate where the disordered coil in the S4 N-terminus interacts with the 5' strand of h16 (A408 – A412). This region faces the protein interface in the extended conformation state II, and it is close to the junction where the C2-domain binds. In order for the binding to continue, a slight rotation in h16 is needed then in order for the S4 N-terminus to make further contacts with both strands of the helix in the bulge region. A previous study<sup>36</sup> has shown that the h16 internal loop is relatively flexible compared to the helical regions. Therefore, the requirement for simultaneous rearrangements in the internal loop of h16 and the S4 N-terminal coil gives rise to the second barrier at Q  $\sim$  0.3.

In addition to the dominant one, two other pathways exist in which the N-terminus of S4 initiates binding. In one path, S4 N-terminus binds completely before the C-terminal domain starts to make contacts (green). In the other path, only 30% of the native contacts between the S4 N-terminus and the five-way junction are made before the C-terminus begins to bind (pink). The ratio of the number of molecules going through each of the three pathways is about 3:1:1 in simulations starting from state II. The exact partition changes according to initial conformation (data not shown) while the positions of the three pathways as well as the energetic barriers remain the same regardless of the starting conformations.

In summary, the two domains of S4 behave differently. The C-terminal domain of S4 binds quickly and tightly once the right contacts are made. Conversely, complete binding of S4 N-terminus is the rate-limiting step in the binding phase due to two energetic barriers arising from structural rearrangements in both h16 and S4N. The reason for their different functions in binding lies in the distinct structural properties of both the protein and the RNA binding sites. The C-terminal domain of S4 binds to the central junction of the five helices, while the S4 N-terminus binds to both h16 and h18, which may separate widely (as in extended state II) when S4 is not bound. Figure 5C shows that the S4 C-terminal domain and its binding site are both stable, fluctuating independently through a relatively small range of conformations. However, conformations of the S4 N-terminus and its binding site are clearly correlated. In the diffusing phase, both of them are flexible, with RMSD of S4N fluctuating between 4 to 8 Å and  $R_G$  of h16 and h18 fluctuating between 17 to 21 Å. Only after binding occurs are both constrained close to their native conformations. This suggests that the protein-RNA interactions have co-evolved such that their structural features match to facilitate binding.

Since the S4 C-terminal domain is stable throughout the diffusing and binding phases, and no apparent energy barrier is present for its binding to the five-way junction, we suggest that

the binding of S4 C-terminus is mainly determined by the search in entropic space. Once it is oriented correctly with respect to its binding site, all the contacts are made at once and binding occurs within a short time. The binding of N-terminus needs more investigation. It has been suggested in<sup>36</sup> that the main purpose of the S4 N-terminus is to bring the separated h16 and h18 back together into their native configuration. Interactions between h16 and h18 can be characterized by interface Q between the two helices. There is intrinsic fluctuation between h16 and h18, and they may make and loose contacts constantly when the S4 N-terminus is not bound (Figure 5 D). This is consistent with the observation from MD simulations that states I and II interchange spontaneously. Once the S4 N-terminus is tightly bound, fluctuations between the two helices are suppressed, and their interactions remain until the end of the simulations (Figure 5 E). In general, the S4 N-terminus initiates the interaction between h16 and h18. However, there are also cases where it binds to one of the states while h16 and h18 are fluctuating (Figure 5 F). Supposedly, if the N-terminus of S4 catches the five-way junction in state I in which h16 and h18 are close together, the binding barriers should be eliminated. However, this circumstance is rarely seen in the simulations.

### Binding mechanism of the S4 N-terminus and the five-way junction

We consider two possible scenarios for the mechanism of S4 N-terminus binding — population shift or induced fit. Results in the previous section suggest that conformations of the S4 N-terminus and its binding site on the five-way junction are closely related, and it is the interactions between them that correlate with the conformational changes. In order to determine the binding mechanism, we plot the RMSD of S4 N-terminus against its interface Q with the five-way junction. A population shift scheme (Figure 6A) results in two separate conformational populations existing regardless of whether the interactions between the S4N and the five-way junction are present or not. However, the increase in interface interactions should result in a change in the relative amount of molecules stay in each of the two conformational states. In contrast, binding achieved through induced fit (Figure 6B) predicts two distinct conformations populated under different conditions (no interaction, or bound).

Our simulations show that the S4 N-terminus clearly adopts an induced fit mechanism as it remains flexible before any interaction with the five-way junction is established (Figure 6C). Only upon binding to the five-way junction (interface  $Q > 0.5$ ) will the N-terminal coil fold into a stable conformation even when a bias toward the native state is present in the G - potential model.

The binding mechanism is more complicated for the five-way junction, and the interpretation strongly depends on the starting conformation. In simulations starting from the extended conformation of the five-way junction, state II, conformational states were measured using radius of gyration ( $R_G$ ) of h16 and h18. Two populations existed in the absence of S4N interaction (Figure 7A), one with  $R_G \sim 17 \text{ \AA}$  close to that in the native structure (state N or I), the other with a higher and fluctuating  $R_G$  (state II). This is consistent with the occasional interconversions between state I and II seen in the MD simulations of the five-way junction without S4, and with the hybrid MD-G<sub>-</sub> simulations where h16 and h18 fluctuate between states with or without interactions between each other. After S4 binds, density for the conformation with large  $R_G$  disappears completely. Binding of S4N is able to begin from either state, as shown with representative traces (yellow from state I, and orange from state II) on the histogram (Figure 7A). Interestingly, binding from state II occurs in a step-wise manner, alternately establishing interface interactions and increasing compactness of the h16-h18 entity. In this case, conformations of the five-way junction (particularly, h16 and h18) are distributed in two populations intrinsically, and the interactions from S4 eliminate one of them in an extreme example of population shift.

On the other hand, folding of the five-way junction starting from conformation III is a clear example of the induced fit mechanism (Figure 7B). Due to stability of conformation III seen in the MD simulations, it is expected that h16 and h18 should stay at a right angle without S4 bound. Therefore, we see only one population centered at  $65^\circ$  when  $Q < 0.05$ . Folding of h16 and h18 back to the native structure will not happen before a substantial fraction of the S4 N-terminus interactions are established ( $Q > 0.4$ ). This fact supports our previous finding that state III is a relatively stable conformation of the five-way junction that will only fold correctly with the help of S4.

## Discussion

A useful way to understand the folding of biological molecules is by describing the folding energy landscape, which shows the positions of alternative conformations (including the native state, unfolded state, and all intermediates) as well as the energetic connections between them.<sup>58</sup> In addition, structural features that define the intermediate states and the energy barriers between them will further assist to depict the folding behaviors along the pathways. Folding of the RNA molecules, in particular, is known to be hierarchical,<sup>63,64</sup> such that secondary and tertiary structures fold sequentially due to the large difference in their energetic contributions. Folding of the secondary structure is characterized by large negative  $\Delta G$  and is sensitive to kinetic traps,<sup>63</sup> while forming tertiary contacts is highly cooperative. However, recent studies on the folding of riboswitches and ribozymes showed the non-hierarchical nature of RNA folding and the existence of multiple collapsed and stable non-native structures in the native basin on the energy landscape.<sup>59,65,66</sup>

In this paper, we assumed that the secondary structure of the five-way junction forms during the co-transcriptional self-assembly of the rRNA. Based on this assumption, we tried to obtain information about the topology of its folding landscape in the vicinity of the native basin. Since formation of RNA tertiary structure is sensitive to the amount of Mg<sup>2+</sup> in solution, we first depleted all Mg<sup>2+</sup> ions in the system and ran MD simulation so that the five-way junction would unfold in a way that tertiary contacts were disrupted while secondary structure was maintained. By doing this, we pushed the RNA molecule out of the most stable native structure at the bottom of its folding funnel (Figure 8, top left). Then the unfolding trajectory was analyzed, and conformations in the trajectory were grouped into discrete states according to the angles between helices (a measure of tertiary interactions with respect to the native structure). In order to check whether these discrete clusters of the five-way junction states stand for stable intermediates during the final folding process, Mg<sup>2+</sup> ions were put back into the simulations starting from representative conformations chosen from each of these groups (Figure 8, bottom left).

We found, similar to the presence of multiple native states of the *Tetrahymena* group I ribozyme,<sup>59</sup> that three stable states existed for the five-way junction alone in solution in addition to its native state. First, two conformations (states I and II) are separated by a relatively small barrier, inferred from their rapid interconversion. Second, a third intermediate (state III) with a free energy even lower than that of the native state, is probably separated from the native state by a large barrier because of its apparent stability and the difficulty folding back to the native structure on its own. These observations from MD simulations, which were also seen in the FRET experiments, gave rise to the proposed energy landscape depicted in Figure 8 (bottom left). Furthermore, visualizing atomic details in MD simulations provided us with a straight forward way to recognize structural characteristics of the different states. Interestingly, identification of the multiple stable conformations can be simply reduced to a low dimensional metric that involves only the angle between h16 and h18 (see Results).

*In vitro* binding of S4 to the five-way junction and rearrangement to the final stable complex takes a few seconds to minutes to complete,<sup>11</sup> which is a time scale currently inaccessible to MD simulation. To understand what role the protein plays in the folding of the five-way junction, simultaneous binding and folding of the two molecules were performed using G - potentials developed in the framework of the NAMD package (see Methods). States II and III were chosen as the initial unfolded conformations of the five-way junction. Considering that states I and II were observed to interconvert spontaneously, we viewed the protein-assisted folding process to start from all possible stable points on the energy landscape (Figure 8 middle column). By adding in the r-protein S4, the folding landscape was altered in terms of relative stability among the different states and the barrier height between them. Apparently, the binding of S4 greatly stabilizes the native structure of the five-way junction, since no fluctuations back to the three non-native states were ever observed after S4 was correctly bound. Folding from state III was now possible with the help of S4, showed by a decreased barrier between it and the native structure. Most likely, this barrier is even smaller than that of the other folding pathway, due to the higher probability of successful binding events observed in multiple replicates of the binding simulations. The folding pathway that went through states I and II downhill to the native state had been smoothed, however, the shape of the landscape was preserved which is consistent with the presence of two barriers seen in the S4N binding process starting from state II.

S4N was shown to be intrinsically disordered by many experiments as well as MD simulations, and consequently it was suggested that the flexibility of S4N is utilized by the protein to speed up the molecular recognition between the five-way junction and S4.<sup>36</sup> Here, we examined this possibility by visualizing the binding process using hybrid MD-G simulations. The flexibility of S4N facilitates the binding process in two ways. First, by closely working with the globular domain of S4C which serves as an anchoring point, it decreases the five-way junction folding barriers at the cost of its own entropy loss upon binding. Second, because of its flexibility, it is capable of binding to all possible conformers of the five-way junction, and, thus, may increase the probability of successful binding. After the initial binding, S4N alters its own conformation accordingly, and induces both the folding of five-way junction and the complete docking between the two molecules (Figure 8 right column). This mechanism is in agreement with the extended view of biomolecular recognition, in which binding of the biomolecules is initiated by selection among a dynamical conformational ensemble followed by subsequent adjustments by induced fit.<sup>67,68</sup>

## Conclusion

Simultaneous folding and binding of the five-way junction in the 5'-domain of 16S rRNA and the primary binding protein S4 is essential for the assembly of the bacterial ribosomal SSU. This process is studied in this paper using the hybrid MD-G simulations developed within the framework of all-atom molecular dynamics CHARMM force field. With the support from all-atom MD simulations, single-molecule FRET experiment and SHAPE chemistry, we have identified multiple metastable states of the five-way junction near its native basin and estimated the shape of its folding landscape. Analysis of the five-way junction-S4 interactions from hundreds of hybrid MD-G replicate simulations shows that the intrinsic disorder of the N-terminus of S4 controls the rate of binding in two ways. First, it is capable of binding to all different conformations of the five-way junction during the self-assembly process of the RNA molecule. Second, it induces the folding of the five-way junction to the native state. Our simulations capture the cooperativity in the simultaneous folding and binding of the five-way junction and S4, and shed further light on the general principle of protein:RNA interactions and evolution.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

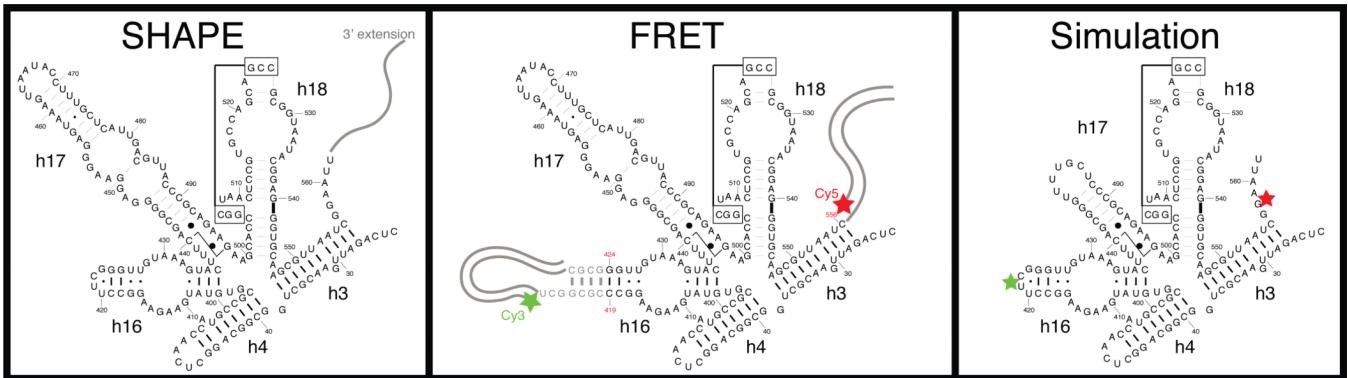
The authors would like to thank the NSF Center for the Physics of Living Cells at the University of Illinois at Urbana-Champaign for partial support of this work. The work was supported by grants from the National Science Foundation (MCB-0844670, MCB-0613643, PHY08-22613, EF-0526747, 0822613 and 0646550), and National Institutes of Health (GM-065367). T.H. is an investigator with the Howard Hughes Medical Institute. Supercomputer and local computing time was provided by NCSA LRAC (MCA03T027) and NSF CRIF (0541659).

## References

1. Held W, Ballou B, Mizushima S, Nomura M. *J. Biol. Chem.* 1974; 249:3103–3111. [PubMed: 4598121]
2. Nowotny V, Nierhaus K. *Biochemistry*. 1988; 27:7051–7055. [PubMed: 2461734]
3. Talkington M, Siuzdak G, Williamson J. *Nature*. 2005; 438:628–632. [PubMed: 16319883]
4. Sykes M, Williamson J. *Ann. Rev. Biochem.* 2009; 38:197–215.
5. Mulder A, Yoshioka C, Beck A, Bunner A, Milligan R, Potter C, Carragher B, Williamson J. *Science*. 2010; 330:673–677. [PubMed: 21030658]
6. Ha T, Zhuang X, Kim H, Orr J, Williamson J, Chu S. *Proc. Natl. Acad. Sci. USA*. 1999; 96:9077. [PubMed: 10430898]
7. Kim H, Nienhaus G, Ha T, Orr J, Williamson J, Chu S. *Proc. Natl. Acad. Sci. USA*. 2002; 99:4284. [PubMed: 11929999]
8. Klostermeier D, Sears P, Wong C, Millar D, Williamson J. *Nucl. Acids Res.* 2004; 32:2707. [PubMed: 15148358]
9. Adilakshmi T, Ramaswamy P, Woodson S. *J. Mol. Biol.* 2005; 351:508–519. [PubMed: 16023137]
10. Adilakshmi T, Bellur D, Woodson S. *Nature*. 2008; 455:1268–1272. [PubMed: 18784650]
11. Mayerle M, Bellur D, Woodson S. *J. Mol. Biol.* 2011; 412:453–465. [PubMed: 21821049]
12. Poland D, Scheraga H. *J. Chem. Phys.* 1965; 43:2071. [PubMed: 5845857]
13. Momany F, McGuire R, Burgess A, Scheraga H. *J. Phys. Chem.* 1975; 79:2361–2381.
14. Scheraga H, Vila J, Ripoll D. *Biophys. Chem.* 2002; 101:255–265. [PubMed: 12488006]
15. Stagg S, Mears A, SC H. *J. Mol. Biol.* 2003; 328:49–61. [PubMed: 12683996]
16. Trylska J, McCammon J, Brooks C III. *J. Am. Chem. Soc.* 2005; 127:11125–11133. [PubMed: 16076220]
17. Hamacher K, Trylska J, McCammon J. *PLoS Comput. Biol.* 2006; 2:e10. [PubMed: 16485038]
18. Onuchic J, Luthey-Schulten Z, Wolynes P. *Ann. Rev. Phys. Chem.* 1997; 48:545–600. [PubMed: 9348663]
19. Shoemaker B, Wang J, Wolynes P. *Proc. Natl. Acad. Sci. USA*. 1997; 94:777–782. [PubMed: 9023333]
20. Clementi C, Nymeyer H, Onuchic J. *J. Mol. Biol.* 2000; 298:937–953. [PubMed: 10801360]
21. Koga N, Takada S. *J. Mol. Biol.* 2001; 313:171–180. [PubMed: 11601854]
22. Pogorelov TV, Luthey-Schulten Z. *Biophys. J.* 2004; 87:207–214. [PubMed: 15240458]
23. Levy Y, Wolynes P, Onuchic J. *Proc. Natl. Acad. Sci. USA*. 2004; 101:511–516. [PubMed: 14694192]
24. Levy Y, Cho S, Onuchic J, Wolynes P. *J. Mol. Biol.* 2005; 346:1121–1145. [PubMed: 15701522]
25. Schug A, Onuchic J. *Current Opinion in Pharmacology*. 2010; 10:709–714. [PubMed: 20951644]
26. Cho S, Pincus D, Thirumalai D. *Proc. Natl. Acad. Sci. USA*. 2009; 106:17349–17354. [PubMed: 19805055]
27. Clementi C, Garcia A, Onuchic J. *J. Mol. Biol.* 2003; 326:933–954. [PubMed: 12581651]
28. Wu L, Zhang J, Qin M, Liu F, Wang W. *J. Chem. Phys.* 2008; 128:235103.

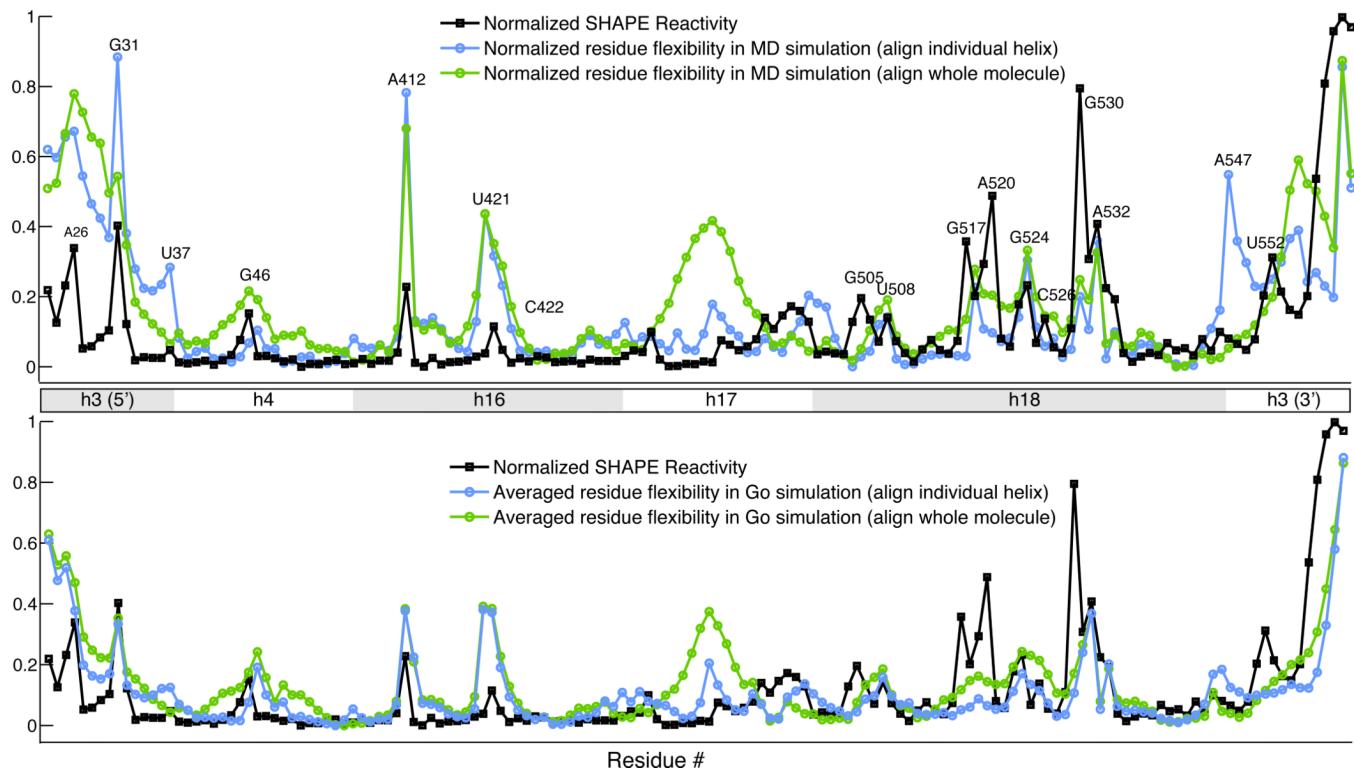
29. Whitford P, Noel J, Gosavi S, Schug A, Sanbonmatsu K, Onuchic J. PROTEINS: Struc. Func. and Bioinformatics. 2009; 75:430–441.
30. Whitford P, Schug A, Saunders J, Hennelly S, Onuchic J, Sanbonmatsu K. Biophys. J. 2009; 96:L7–L9. [PubMed: 19167285]
31. Whitford P, Geggier P, Altman R, Blanchard S, Onuchic J, Sanbonmatsu K. RNA. 2010; 16:1196–1204. [PubMed: 20427512]
32. Bellur D, Woodson S. Nucl. Acids Res. 2009; 37:1886–1896. [PubMed: 19190093]
33. Woese CR, Kandler O, Wheelis M. L. Proc. Natl. Acad. Sci. USA. 1990; 87:4576–4579. [PubMed: 2112744]
34. Winker S, Woese C. R. Syst. Appl. Microbiol. 1991; 14:305–310.
35. Roberts E, Sethi A, Montoya J, Woese C, Luthey-Schulten Z. Proc. Natl. Acad. Sci. USA. 2008; 105:13953–13958. [PubMed: 18768810]
36. Chen K, Eargle J, Sarkar K, Gruebele M, Luthey-Schulten Z. Biophys. J. 2010; 99:3930–3940. [PubMed: 21156135]
37. Hohng S, Wilson T, Tan E, Clegg R, Lilley D, Ha T. J. Mol. Biol. 2004; 336:69–79. [PubMed: 14741204]
38. Balaeff, A.; Eargle, J.; Roberts, E. Ionize v 1.6. Urbana, Illinois: University of Illinois – Urbana-Champaign; 2005. Website: <http://www.ks.uiuc.edu/Development/MDTools/ionize/>
39. Humphrey W, Dalke A, Schulten K. J. Mol. Graphics. 1996; 14:33–38.
40. Grubmueller, H.; Groll, V. Solvate v 1.0. Göttingen, Germany: Max Planck Institute for Biophysical Chemistry; 1996. Website: <http://www.mpiibpc.mpg.de/home/grubmueller/downloads/solvate/index.html>
41. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J. Comp. Chem. 2005; 26:1781–1802. [PubMed: 16222654]
42. Foloppe N, MacKerell AD Jr. J. Comp. Chem. 2000; 21:86–104.
43. Eargle J, Black AA, Sethi A, Trabuco LG, Luthey-Schulten Z. J. Mol. Biol. 2008; 377:1382–1405. [PubMed: 18336835]
44. Serdyuk I, Gogia Z, Venyaminov S, Khechinashvili N, Bushuev V, Spirin A. JMB. 1980; 137:93–107.
45. Sayers E, Gerstner R, Draper D, Torchia D. BCH. 2000; 39:13602–13613.
46. Frey B, Dueck D. Science. 2007; 315:972–976. [PubMed: 17218491]
47. Cox, TF.; Cox, MA. Multidimensional scaling. 2nd ed.. Chapman & Hall/CRC; 2001.
48. Markus M, Gerstner R, Draper D, Torchia D. EMBO J. 1998; 17:4559–4571. [PubMed: 9707416]
49. Berk V, Zhang W, Pai RD, Cate JHD. Proc. Natl. Acad. Sci. USA. 2006; 103:15830–15834. [PubMed: 17038497]
50. Merino E, Wilkinson K, Coughlan J, Weeks K. J. Am. Chem. Soc. 2005; 127:4223–4231. [PubMed: 15783204]
51. Deigan K, Li T, Mathews D, Weeks K. Proc. Natl. Acad. Sci. USA. 2009; 106:97. [PubMed: 19109441]
52. Vicens Q, Gooding A, Laederach A, Cech T. RNA. 2007; 13:536–548. [PubMed: 17299128]
53. Lucks J, Mortimer S, Trapnell C, Luo S, Aviran S, Schroth G, Pachter L, Doudna J, Arkin A. Proc. Natl. Acad. Sci. USA. 2011; 108:11063–11068. [PubMed: 21642531]
54. Gerstner R, Pak Y, Draper D. Biochemistry. 2001; 40:7165–7173. [PubMed: 11401563]
55. Gherghe C, Shajani Z, Wilkinson K, Varani G, Weeks K. J. Am. Chem. Soc. 2008; 130:12244–12245. [PubMed: 18710236]
56. Gherghe C, Mortimer S, Krahn J, Thompson N, Weeks K. J. Am. Chem. Soc. 2008; 130:8884–8885. [PubMed: 18558680]
57. Ha T, Enderle T, Ogletree D, Chemla D, Selvin P, Weiss S. Proc. Natl. Acad. Sci. USA. 1996; 93:6264. [PubMed: 8692803]
58. Russell R, Zhuang X, Babcock H, Millett I, Doniach S, Chu S, Herschlag D. Proc. Natl. Acad. Sci. USA. 2002; 99:155–160. [PubMed: 11756689]
59. Solomatin S, Greenfeld M, Chu S, Herschlag D. Nature. 2010; 463:681–684. [PubMed: 20130651]

60. Ben-Shem A, Jenner L, Yusupova G, Yusupov M. *Science*. 2010; 330:1203–1209. [PubMed: 21109664]
61. Rabl J, Leibundgut M, Ataide S, Haag A, Ban N. *Science*. 2011; 331:730–736. [PubMed: 21205638]
62. Chen K, Roberts E, Luthey-Schulten Z. *BMC Evol. Biol.* 2009; 9:179–195. [PubMed: 19640295]
63. Brion P, Westhof E. *Ann. Rev. Biophys. Biomol. Struct.* 1997; 26:113–137. [PubMed: 9241415]
64. Tinoco I Jr, Bustamante C. *J. Mol. Biol.* 1999; 293:271–281. [PubMed: 10550208]
65. Buchmueller K, Weeks K. *Biochemistry*. 2003; 42:13869–13878. [PubMed: 14636054]
66. Duncan C, Weeks K. *Biochemistry*. 2010; 49:5418–5425. [PubMed: 20533823]
67. Boehr D, Nussinov R, Wright P. *Nature Chem. Biol.* 2009; 5:789–796. [PubMed: 19841628]
68. Csermely P, Palotai R, Nussinov R. *Trends Biochem. Sci.* 2010; 35:539–546. [PubMed: 20541943]



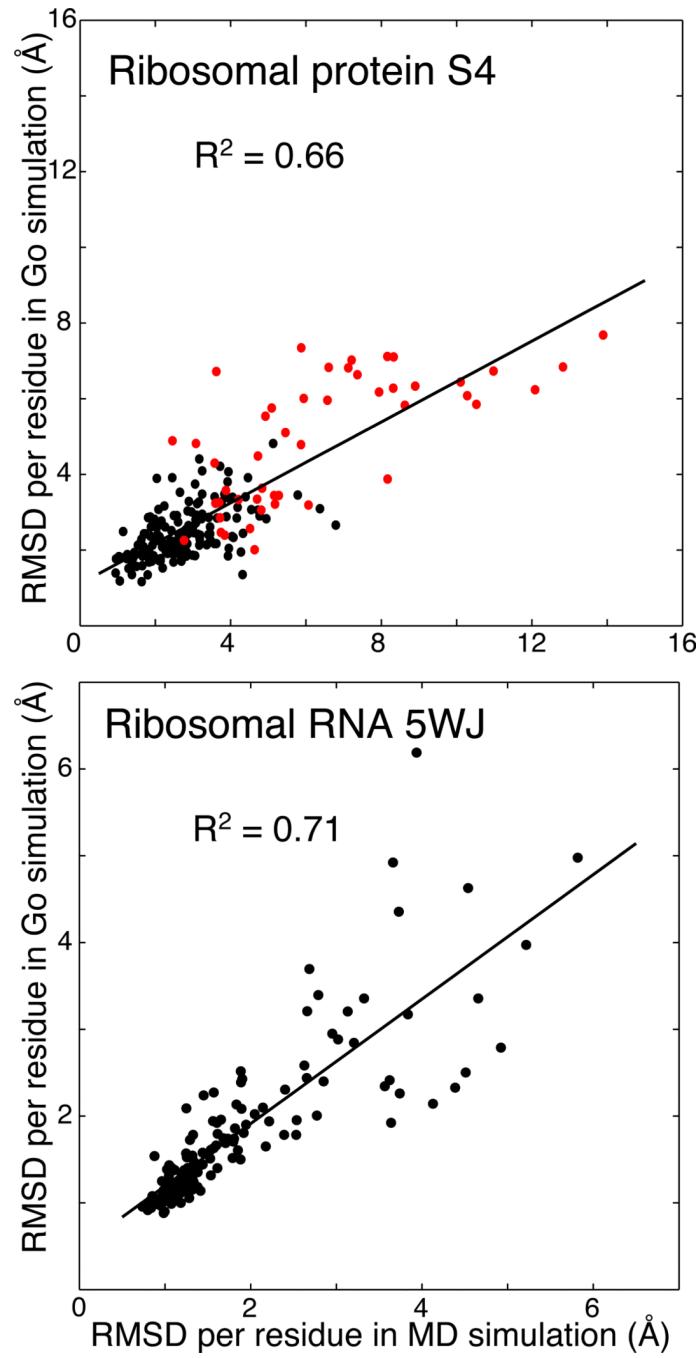
**Figure 1. Constructs used in this work**

Secondary structures of the five-way junction in the SHAPE and FRET experiments, and in simulations. Stars on the FRET system indicate fluorophore locations, and stars in the simulation system show corresponding positions used to measure distances for comparison to the FRET results.



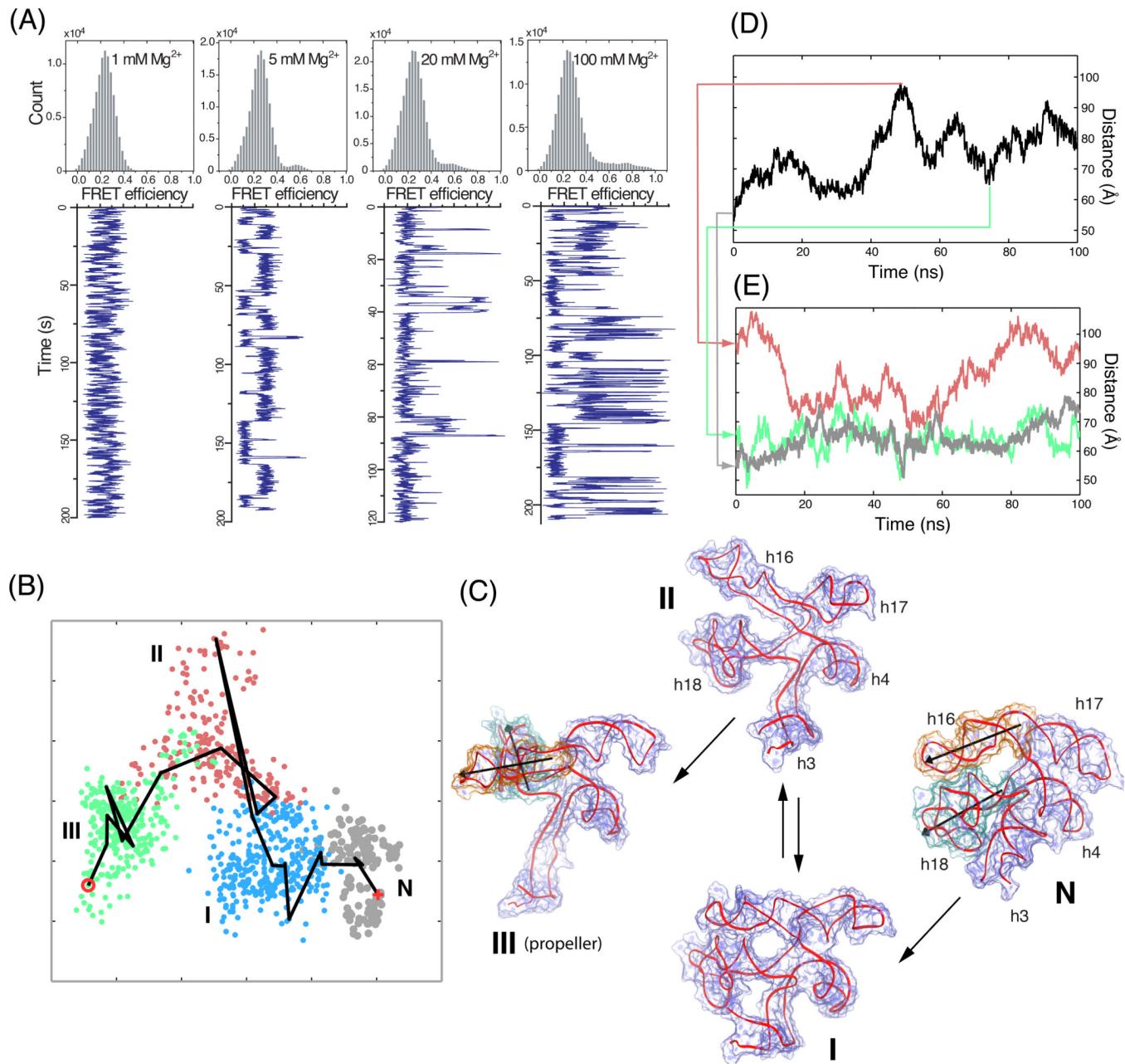
**Figure 2. Comparison of simulated and experimental SHAPE data**

Normalized SHAPE reactivity (black squares) compared to flexibility measured for each residue in the five-way junction using MD (top) and hybrid MD-G<sub>o</sub> simulations (bottom). Fluctuations in the tertiary structure are shown by alignment of the whole molecule to the average position (green circles), and in the secondary structure by alignment of individual helices to their average positions (blue circles). SHAPE data is taken from Mayerle *et al.*<sup>11</sup>



**Figure 3. Correlations between MD and hybrid MD-G simulations of native state fluctuations in the five-way junction:S4 complex**

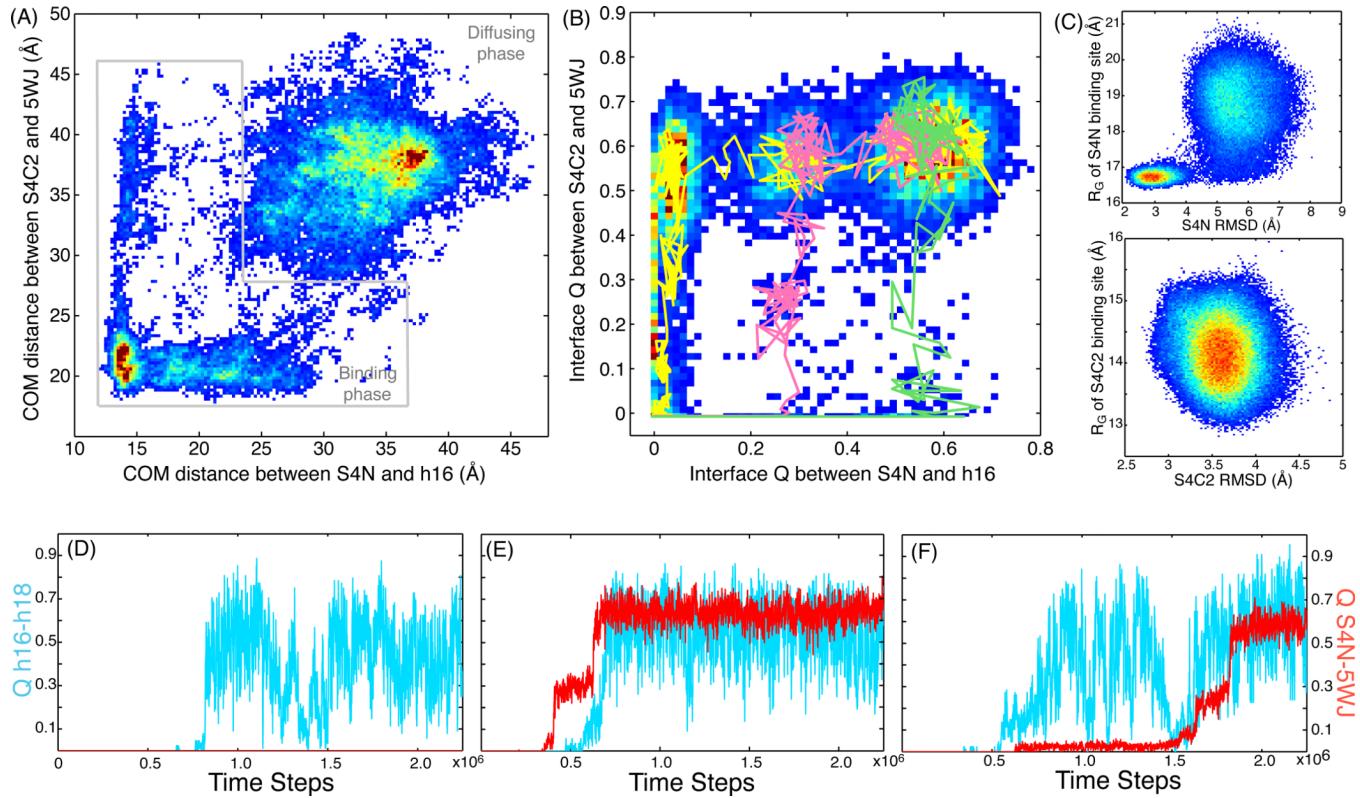
RMSD per residue is linearly correlated between MD and hybrid MD-G simulations for both the protein (top) and the RNA (bottom). Red dots in the top panel are RMSD of the S4 N-terminal residues which show higher flexibility in both MD and hybrid MD-G simulations.



**Figure 4. Multiple states are observed in sm-FRET experiment and MD simulations of the five-way junction**

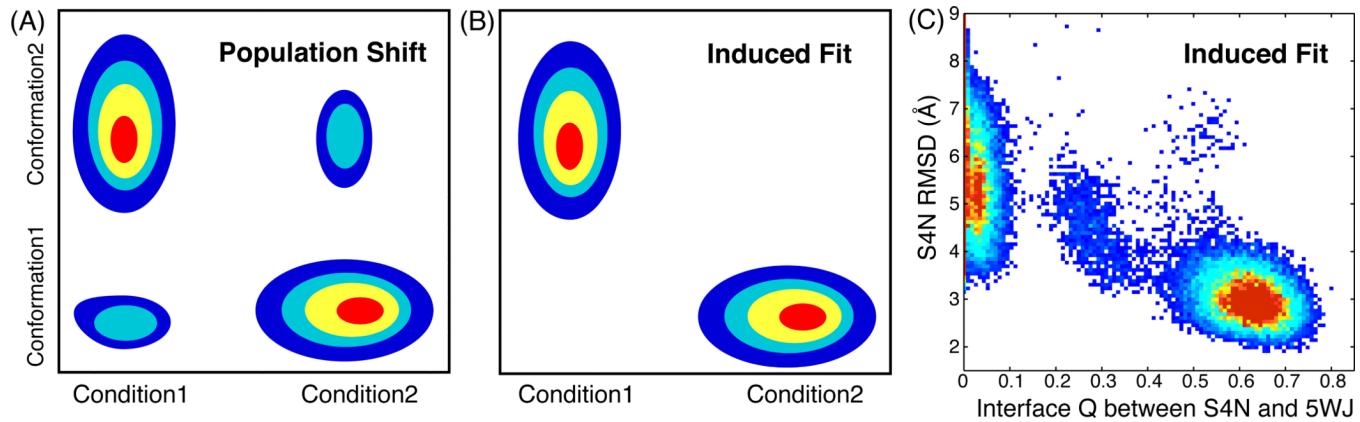
(A) FRET signal distributions in rRNA of the five-way junction under varying  $Mg^{2+}$  concentration. Representative time traces are shown for each concentration to illustrate signal fluctuations over time. (B) Multi-dimensional scaling (MDS) reduction of the ten-dimensional angle vector into a two-dimensional space for visualization of the unfolding MD trajectory simulated in NaCl solution. The solid line shows the time trace with a low sampling rate. On the path, the red (+) marks the starting frame and the red (○) marks the ending frame. (C) Representative snapshots of the five-way junction are shown together with a proposed transition map among different states: N (native structure), I (intermediate state), II (unfolded state), and III (misfolded state). (D) FRET distance measured over time between U422 on the tetraloop of h16 and A559 of h3 for the five-way junction in the

unfolding simulation in NaCl solution. (E) FRET distances measured over time in the Mg<sup>2+</sup> quenching simulations starting from corresponding conformational states (color-coded the same as in B).



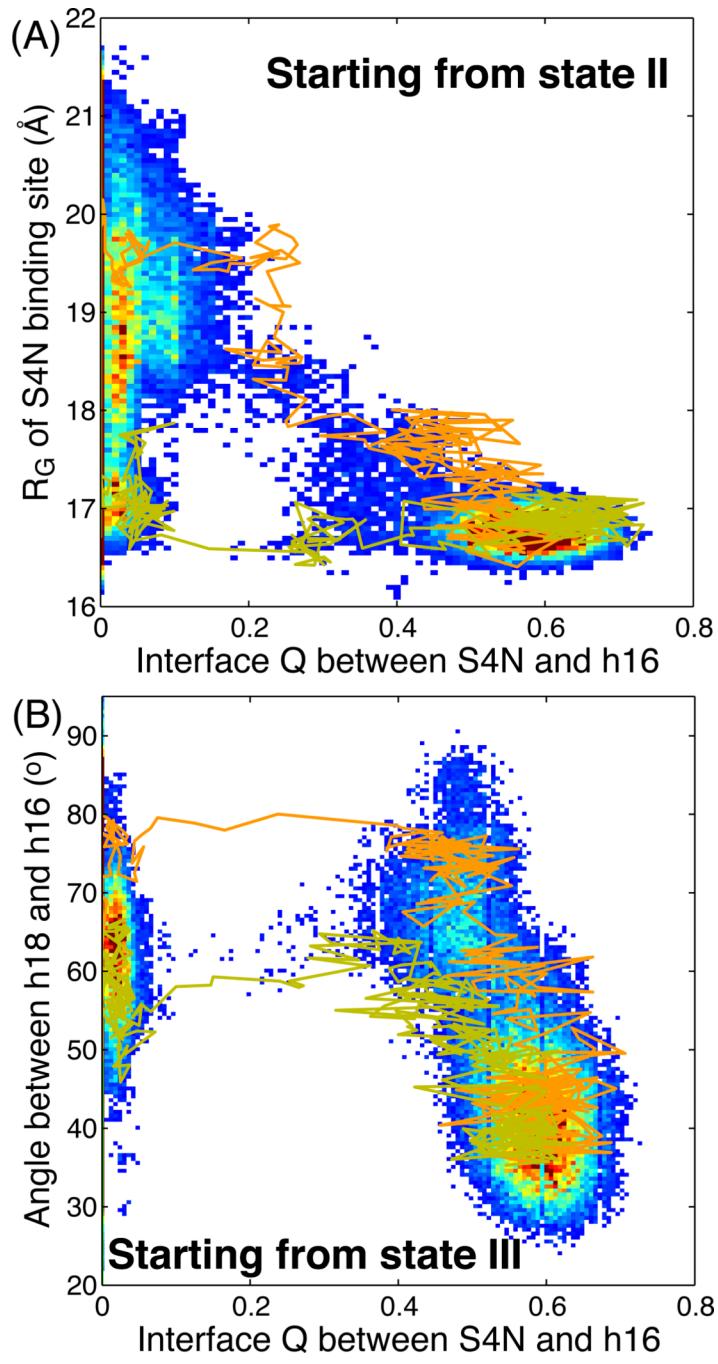
**Figure 5. S4 binds to the five-way junction in multiple steps**

(A) Histogram of the center-of-mass distance between S4N and h16 versus that between S4-C2 domain and the five-way junction shows two distinctive phases in the binding process: the diffusing phase and the binding phase. (B) Histogram of the fraction of native interface interactions between S4N and h16 versus that between S4-C2 domain and the five-way junction. Three representative pathways are shown in yellow, red and green from left to right. (C) Correlation between the conformations of protein and RNA is visualized by histograms as a function of RMSD of the protein N-terminus (top) or C2-domain (bottom) and  $R_G$  of the corresponding binding site on the RNA. (D)–(F) Representative time traces of the fraction of native contacts between h16 and h18 (shown in light blue), and the fraction of native interface contact between S4N and the five-way junction (shown in red).



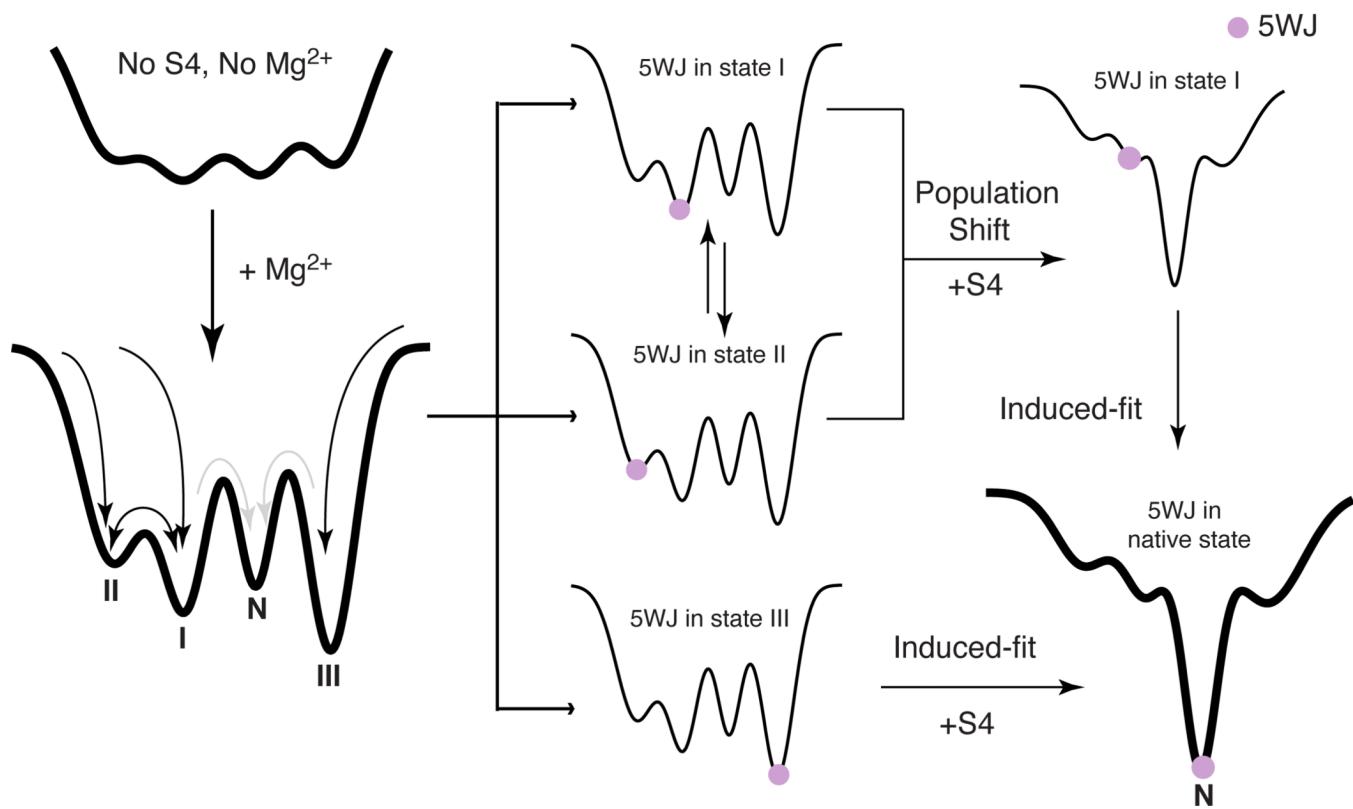
**Figure 6. S4N binding proceeds through induced fit**

Schematic representations of changes in conformational densities results in (A) population shift or (B) induced fit mechanisms. (C) 2D Histogram of the folding of S4N (RMSD) as a function of its docking to the five-way junction as measured by the fraction of native interface contacts between them. The histogram resembles the induced fit schematic in that the minor component in the unfolded population is absent.



**Figure 7. Five-way junction binding mechanism depends on the starting conformation**  
 Histograms showing conformational changes in the five-way junction upon docking with S4N. Conformations of the five-way junction are measured using radius of gyration ( $R_G$ ) of h16 and h18 when simulations are started from state II (A) and using angles between h16 and h18 when initiated from the state III (B). In panel (A), a population shift mechanism is seen since one of the two existing populations is eliminated by the binding of S4N. In panel (B), conformational change of the five-way junction occurs only after the binding of S4N which induces a fit to the native conformation.

# Energy Landscape for Five-Way Junction (5WJ) Folding



**Figure 8. Folding energy landscape of the five-way junction**

Three forms of the landscapes are highlighted: without S4 and no  $Mg^{2+}$  (top left), with  $Mg^{2+}$  (bottom left), and with S4 (bottom right). The S4 binding mechanism are depicted through separate pathways shown by duplication of the five-way junction folding landscapes.