doi:10.1016/j.jmb.2005.12.014 J. Mol. Biol. (2006) 357, 292–312

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Free Energy Landscapes of RNA/RNA Complexes: With Applications to snRNA Complexes in Spliceosomes

Song Cao and Shi-Jie Chen*

Department of Biochemistry and Department of Physics University of Missouri-Columbia, Columbia MO 65211, USA

We develop a statistical mechanical model for RNA/RNA complexes with both intramolecular and intermolecular interactions. As an application of the model, we compute the free energy landscapes, which give the full distribution for all the possible conformations, for U4/U6 and U2/U6 in major spliceosome and U4atac/U6atac and U12/U6atac in minor spliceosome. Different snRNA experiments found contrasting structures, our free energy landscape theory shows why these structures emerge and how they compete with each other. For yeast U2/U6, the model predicts that the two distinct experimental structures, the four-helix junction structure and the helix Ib-containing structure, can actually coexist and specifically compete with each other. In addition, the energy landscapes suggest possible mechanisms for the conformational switches in splicing. For instance, our calculation shows that coaxial stacking is essential for stabilizing the four-helix junction in yeast U2/U6. Therefore, inhibition of the coaxial stacking possibly by protein-binding may activate the conformational switch from the four-helix junction to the helix Ib-containing structure. Moreover, the change of the energy landscape shape gives information about the conformational changes. We find multiple (native-like and misfolded) intermediates formed through base-pairing rearrangements in snRNA complexes. For example, the unfolding of the U2/U6 undergoes a transition to a misfolded state which is functional, while in the unfolding of U12/U6atac, the functional helix Ib is found to be the last one to unfold and is thus the most stable structural component. Furthermore, the energy landscape gives the stabilities of all the possible (functional) intermediates and such information is directly related to splicing efficiency.

Keywords: free energy landscape; folding pathways; misfolded state; major and minor spliceosomes; RNA/RNA complex

*Corresponding author

Introduction

The coding sequence (exons) in the pre-messenger RNA (pre-mRNA) is separated by the intervening non-coding sequences (introns).[1,2] In order to produce the mature and functional mRNA, the introns must be accurately removed from the pre-mRNA (splicing). In most eukaryotes, the process is accomplished by the spliceosome, which is composed of five small nuclear RNAs (snRNAs) (U1,U2,U4,U5,U6) and over 50 proteins.[3–6] Two sequential transesterification reactions are found in splicing the pre-mRNA. In the first step, the 2′ hydroxyl of a bulged adenosine near the 3′ end of intron attacks the 5′ splice site, generating a branched lariat intermediate and the 5′ exon containing a 3′ hydroxyl. In the second step, the 3′ hydroxyl of the free 5′ exon attacks the 3′ splice site, ligating the exons and excising the intron to create the mature mRNA.

The U4/U6 and U2/U6 complexes are functionally important for splicing in the major introns. In yeast, the U4/U6 complex is found to assemble the spliceosome before the two steps of splicing. Subsequently, the U2/U6 complex is formed, which requires unwinding the U4/U6 complex.[7–11] The dynamic process may be found in other species such as human[12–15] and *Nosema locustae*

Abbreviations used: pre-mRNA, pre-messenger RNA; snRNAs, small nuclear RNAs; snRNPs, small nuclear ribonucleoprotein particles; ss, splice site.

E-mail address of the corresponding author: chenshi@missouri.edu

(*N. locustae*).[16,17] Distinct structures have been proposed for the U2/U6 complex in yeast and human.[8,12] The main difference between these structures is whether an intramolecular stem in U2 (stem I) is unzipped to form an intermolecular stem (helix Ib).

The formation of helix Ib was first proposed by Madhani & Guthrie.[8] Subsequent studies for yeast and human confirmed the central role of the intermolecular helix Ib.[11,18–20] Furthermore, it was found that strengthening the intramolecular U2 stem I can destabilize the helix Ib and consequently block splicing.[18,21] Also, it was found that destabilizing the extended U6 helix can enhance the splicing in yeast.[11] These experimental findings clearly show the functional importance of the intermolecular helix Ib and the interplay between intermolecular and intramolecular interactions.

On the other hand, it was proposed that U2/U6 complex may contain the intramolecular U2 stem I in both yeast[9,10] and human.[12] Recently, an NMR structural measurement showed that a protein-free U2/U6 complex from yeast is a four-helix junction,[22] which consists of the intramolecular U2 stem I. The intermolecular helix Ib was absent in the NMR structure of the U2/U6 complex. The four-helix junction and the intramolecular stem I in U2 can be relevant to the first step of splicing by helping position U6 intramolecular stem–loop adjacent to the 5′ splice site, while the other structure (with the intermolecular helix Ib) may be protein-dependent and occur in the second step of splicing. The conformational switch may occur in the active site between the two steps. In addition, Moore & Sharp proposed that the conformational switch is from a group II-like structure to a group I-like structure.[23]

A new non-consensus splice was first recognized by Jackson.[24] Later, Hall & Padgett proposed that there was a distinct minor-class of introns (AT-AC).[25] Further studies to unveil the splicing machinery for this type of introns lead to the discovery of a new-abundance spliceosome (minor-class spliceosome). The components of this spliceosome include five identified small nuclear ribonucleoprotein particles (snRNPs). They are, respectively, U11, U12, U5, U4atac and U6atac.[26–28] The U12 snRNA is the analog of U2 snRNA in the major spliceosome and the U4atac and U6atac are the analogs of U4 and U6 snRNA in the major spliceosome. The functional similarities between these two sets of snRNAs are supported by the conservation of RNA–RNA interactions among the pre-mRNA splice sites, U12 snRNA, U6atac snRNA, and U11 snRNA.[29–37] The proposed structure of U12/U6atac complex shows high conservation from vertebrates to higher plants.[34] For example, the U12/U6atac complex contains the intermolecular helix Ib which agrees with the novel U2/U6 base-pairing interaction in yeast.[8,18] Moreover, U4 snRNA is found to function in both major and minor spliceosomes,[37] which gives further strong support for the possible high structural conservation for the complexes in the major and minor classes of spliceosomes.

To understand the physical mechanism for snRNA functions in spliceosome requires detailed quantitative analysis on the folding stabilities and conformational changes for snRNA complexes. For DNA complexes, DNA Software Inc. has successfully developed a numerical software to predict DNA hybridization by computing the equilibrium distribution for hundreds of conformational species as a function of temperature.[38] For RNA complexes, several computational algorithms have been developed to successfully predict the native structures and thermal stabilities for RNA/RNA complexes.[39–42] These models are not without limitations. For example, the OligoWalk algorithm[39] does not treat suboptimal structures or the full conformational distribution and other thermodynamic models[40–42] can treat conformational ensemble but fail to consider the intramolecular interactions. Recently, another successful algorithm (PairFold) has been developed to predict the optimal and suboptimal structures for RNA/RNA complexes.[43] The algorithm can account for both the intermolecular and intramolecular base-pairs. However, the model does not consider the non-canonical mismatched base-pairs and the model focused on the native and suboptimal structures rather than thermal stability and full conformational distributions, which requires a rigorous treatment for the thermodynamic conformational sampling and entropy. A previous lattice conformation-based model[44] can predict the free energy landscapes and the conformational changes for RNA/RNA complexes. But the model was based on simplified lattice conformations, so it cannot treat realistic RNA–RNA complexes.

We present a new statistical mechanical theory to compute the free energy landscapes for the RNA/RNA complexes. In the model, we fully take account of both the intermolecular and intramolecular interactions. We also account for both canonical and non-canonical (mismatched) base-pairs. The free energy landscape can give the full distribution for all the possible conformations. Such information is particularly useful for snRNA complexes, whose functions involve multiple stable structures and structural transitions. Based on the free energy landscape, we investigate the folding stability and (equilibrium) pathways for the spliceosome snRNA complex. The motivation for the application of the model to snRNA complexes is fourfold. First, we can validate the model through extensive theory–experiment comparisons for the stable (low-free energy) structures. Second, snRNA experiments found contrasting structures, our free energy landscape theory shows why these distinct structures emerge and how they may compete with each other. Third, the predicted folding stability and pathways suggest possible mechanisms for conformational transitions in splicing. Fourth, the free energy landscape can give stabilities for all the possible (functional) intermediates, and such

information is important for understanding splicing efficiency.

Results and Discussion

Prediction of the ribozyme–RNA complexes

In Table 1, we compare the accuracy for the native structure prediction between our model and the PairFold program for 17 ribozyme–substrate complexes.[43] The difference between the present model and the model of Andronescu *et al.* is that: (i) the model of Andronescu *et al.* is for optimal and suboptimal structure prediction only, while our model gives the full energy landscape with the native structure as the global minimum point; (ii) our model is based on the first principle calculation (with a virtual bond model) for the conformational entropy and folding thermodynamics;[45] (iii) our model accounts for the non-canonical mismatched base-pairs. To measure the accuracy of the predicted native structures, we follow Andronescu *et al.* and use sensitivity (SE) and specificity (SP) as two criteria, where sensitivity is defined as the ratio of the number of the correctly predicted base-pairs to the number of the base-pairs in the experimentally determined structure, and the specificity is defined as the ratio of the number of the correctly predicted base-pairs to the total number of the canonical base-pairs in the predicted structure. From Table 1, we find that the present model can give improved predictions for the native structures. For example, as shown in Figure 1, the present model gives about 15% improvement in both SP and SE for a modified hairpin ribozyme and substrate complex.

The low free energy structures of snRNA complexes and the spliceosome functions

Low free energy structures of snRNA complexes

Truncated sequences. In our calculations, we truncate the snRNA sequences (see Table 2) and focus only on the sequences that are involved in the snRNA/snRNA binding, except for human U2/U6, for which we also remove the intermolecular helix III.[12] The removed sequences are located in the 5′ or 3′ terminal regions. These sequences contain (conserved) snRNA sequences that bind to pre-mRNA to form snRNA/pre-mRNA complexes in the respective 5′ or 3′ regions. For example, in the U2/U6 complex, the removed U6 5′ nucleotides (…ACAGAGA…, conserved in yeast and human) pair with the 5′-

Table 1. The comparison for the algorithm's accuracy between our model and PairFold[43]

			PairFold		This model	
ID	Description (reference)	S_1, S_2	SE	SP	SE	SP
	Hammerhead and hammerhead-like ribozymes					
1	Hammerhead ribozyme R32 and substrate (Figure 1a[83])	32, 11	0.82	1.00	0.88	1.00
2	AUG-cleaving hammerhead-like ribozyme and substrate (Figure 2[84])	36, 16	0.95	0.90	1.00	1.00
3	Hammerhead ribozyme and no-tail gene target mRNA in zebrafish (Figure 3[84])	38, 17	0.95	1.00	1.00	1.00
4	Rz12×12 hammerhead ribozyme and HIV-1 target sequence (Figure 1[85])	46, 25	0.88	1.00	0.94	1.00
5	αYRz60 hammerhead ribozyme and HIV-1 target sequence (Figure 1[85])	100, 70	0.90	0.88	0.93	0.93
	Hairpin and hairpin-like ribozymes					
6	Reverse-joined hairpin ribozyme HP-RJTL and substrate (Figure 1(c)[86])	65, 14	1.00	0.76	1.00	0.79
7	Conventional hairpin ribozyme with sequence variation HP-WTSV1 and substrate (Figure 1(d)[86])	55, 14	0.83	0.75	0.80, 0.95	0.73, 0.83 (2)
8	Hairpin-derived twin ribozyme HP-DS1 and substrate (Figure 1(e[86])	120, 34	0.93	0.78	0.93, 1.00	0.80, 0.84 (2)
9	Hairpin ribozyme and substrate (Figure 1(b)[50])	92, 21	0.74	0.85	0.83	0.90
10	Modified hairpin ribozyme and substrate (Figure 1[82])	143, 18	0.84	0.75	1.00	0.89
11	Hairpin ribozyme RzG101 and substrate (Figure 4(a)[87])	50, 14	0.78	0.74	0.83, 1.00	0.71, 0.82 (2)
12	Minimal two-way helical junction 2WJ-SV5 hairpin ribozyme–substrate complex (Figure 1[88])	40, 32	0.92	0.81	0.88	0.84
	Other ribozymes					
13	X-motif ribozyme model 43X and S21 RNA (Figure 1(a)[89])	43, 21	0.65	0.58	0.78	0.90
14	MR8-1 ribozyme(derived from X-motif ribozyme and S21 substrate RNA (Figure 5[89])	49, 21	0.71	0.48	0.72	0.62
15	ATP-sensitive allosteric ribozyme construct IV-up and substrate (Figure 1(a)[90])	59, 14	1.00	0.96	1.00	1.00
16	ATP-insensitve ribozyme construct IV-down and substrate (Figure 4(a)[90])	59, 14	0.83	0.70	0.83	0.76
17	5′CYbUT RNA crosslinked to gCYb-558 (Figure 7(b)[91])	59, 28	1.00	0.94	1.00	1.00

We use the predicted ribozyme–substrate complexes by Andronescu *et al.* as the references. S_1 and S_2 are the lengths of the two sequences. In the calculation, the ion concentration is 1 M NaCl, and the temperature is 37 °C.

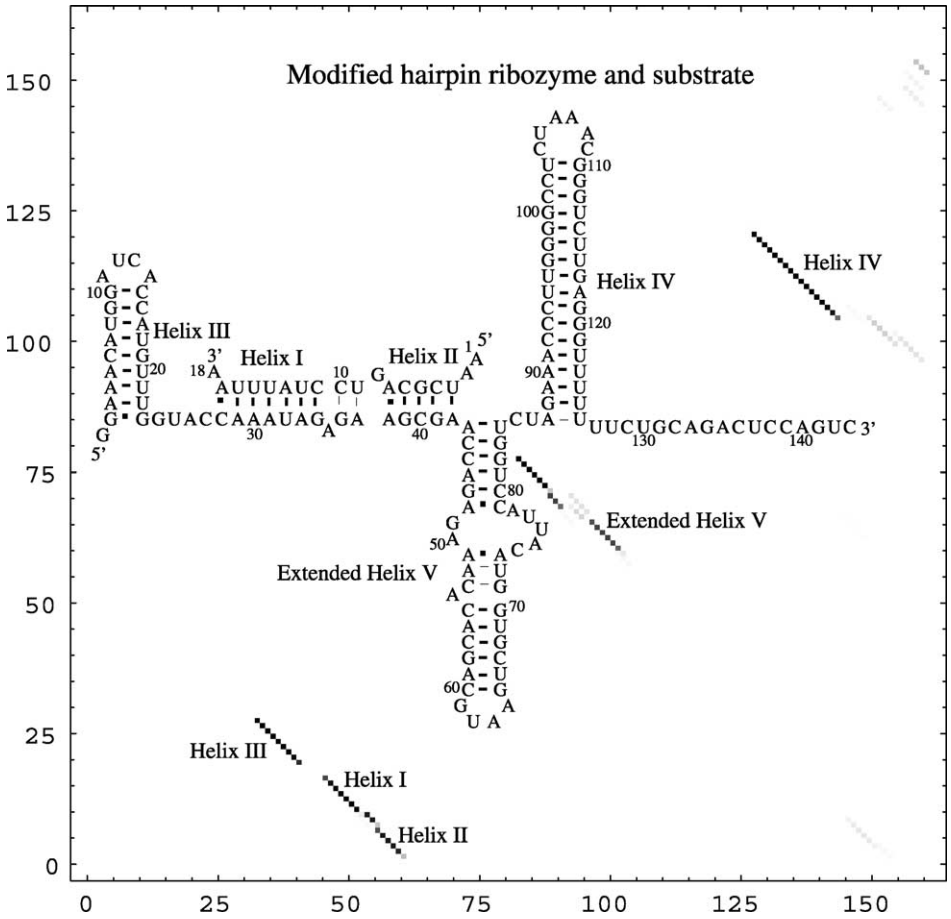

Figure 1. The density plots for base-pairing probabilities and the predicted stable structures of modified hairpin ribozyme and substrate. The short thick lines (for base-pairs) indicate the correct base-pairs as determined by the experiments.[82] The row and column indexes in the density plots denote the nucleotides along the two sequences. The two sequences are aligned such that the short sequence is followed by the long sequence.

splice site (ss) of the pre-mRNA to form a U6/5′ ss helix, and the removed U2 3′ nucleotides (…AGUGUAGUA…, also conserved) pair with the branch point of the intron to form a U2/intron branch point helix. The present form of the model does not treat the triple-chain (pre-mRNA/snRNA/snRNA) complexes. The use of the truncated sequences allows us to focus on the binary snRNA/snRNA complex, which forms the catalytic core. The study of the conformational changes for snRNA/snRNA complexes is helpful for the understanding of the splicing process.

In addition, for U2/U6, the removed U6 3′ sequence and U2 5′ sequences, which are separated from the central part of the complex by the long helix II (see Figure 2(a)–(c)), are expected to have minimal effects on the structure and the stability of the central part of the U2/U6 complex. For U4/U6 and U4atac/U6atac complexes, the termini are stabilized by the longer helix I and helix II (see Figure 5), so the removed sequences in the 5′ and the 3′ termini may have negligible effects on the U4/U6 and U4atac/U6atac complexes.

General results. We apply our statistical mechanical model (see Theory and Methods) to systematically predict the low free energy structures for snRNA complexes in major and minor spliceosomes. We summarize our results in Table 2 and show the predicted structures in Figures 2, 4 and 5.

For yeast and human U2/U6 complexes (Figure 2(a) and (b)), because the experiments were performed around room temperature,[14,15,22] we show the low free energy structures at 25 °C in order to make close comparisons with the experiments. For U2/U6 complex in *N. locustae*[16] (Figure 2(c)), U12/U6atac[34] (Figure 4), U4/U6[46] (Figure 5), and U4atac/U6atac[34] (Figure 5), however, the experimental studies for the structures were mostly performed at 37 °C, therefore we present the 37 °C results.

A notable feature in the low-free energy structures in Figures 2, 4 and 5 is the stabilization of the multiple competing alternative structures of snRNA/snRNA complexes. Also shown in the Figures are the calculated fractional populations of the states. Our model shows that some structures can coexist with equal or similar stability.

Table 2. The accuracy of the model predictions on the snRNA complexes in major and minor spliceosomes

ID	Description (reference)	Truncated sequences		SE	SP
	U2/U6 snRNA complexes in major spliceosome	*U2*	*U6*		
18	U2/U6 snRNA complex of yeast (Figure 5(c[22])	G3-A30	U54-U102	0.91	1.00 (I)
19	U2/U6 snRNA complex of human (Figure 8,[12] Figure 1[14,15])	C3-A29	A48-G95	0.96	1.00 (I)
20	U2/U6 snRNA complex of *N. locustae* (Figure 5[16])	A1-A27	A39-U84	0.88	0.77 (I)
	U4/U6 snRNA complexes in major spliceosome	U4	U6		
21	U4/U6 snRNA complex of yeast (Figure 1[8])	A1-U64	G55-U80	0.85	0.77 (I)
				1.00	0.92 (II)
22	U4/U6 snRNA complex of human (Figure 1[46])	C3-U63	G49-G72	1.00	0.94 (I)
	U12/U6atc snRNA complexes in minor spliceosome	*U12*	*U6atac*		
23	U12/U6atac snRNA complex of Arobidopsis (Figure 7[34])	G1-A16	A20-U53	0.59	0.67 (I)
				0.93	1.00 (II)
24	U12/U6atac snRNA complex of human (Figure 8[34])	A1-A17	A20-G53	0.75	1.00 (I)
				0.88	0.93 (II)
	U4atac/U6atc snRNA complexes in minor spliceosome	*U4atac*	*U6atac*		
25	U4atac/U6atac snRNA complex of human (Figure 1(c)[35])	C3-C65	G23-G48	1.00	0.89 (I)
				0.72	0.78 (II)
26	U4atac/U6atac snRNA complex of Drosophila (Figure 1(d)[35])	C2-C63	G23-G48	0.97	0.88 (I)

The (truncated) sequences of U2/U6, U4/U6, U12/U6atac and U4atac/U6atac are shown in the Table. For U2/U6 complex, the calculation is performed at $T=25$ °C in order to compare with the NMR determined structure for yeast in Sashital *et al.*[22] and the structure for human *in vitro*[14] and protein-free structure for human,[15] which are performed at about room temperature. For other structures, the calculation is performed at standard condition (1 M NaCl and $T=37$ °C). See also Figure 2 for the predicted low-free energy structures (with and without the coaxial stacking). In the results listed in this Table, we consider the coaxial stacking for the four-helix junction. For the structure with multiple coexisting states, we select the first two most populous structures of populations higher than 10%.

In general, the predicted structures agree with the experimental results very well. For instance, for yeast U2/U6 complex, experimental studies give two distinct structures: a four-helix junction structure (Figure 2(a)(I)) that resembles the hairpin ribozyme.[22] Moreover, the extended U6 intramolecular stem–loop in the four-helix junction closely resembles domain 5 of Group II intron with the conserved AGC triad. Similarly, the intermolecular structure (Figure 2(a)(III)) resembles domains 5 and 6 of Group IIA intron with the helix Ib-like stem for the conserved AGC triad.[8,18] Our theory predicts both structures as possible coexisting/competing alternative states of yeast U2/U6.

Given the possible errors in the entropy calculation[45] and in the experimental measurement for the stacking energy parameters,[47,48] our predicted results cannot be error-free. In fact, the relative distribution between the competing states can be quite sensitive to the entropy and enthalpy parameters used in the model. In the statistical mechanical framework, the fractional population of state i is given by:

$$p_i = \frac{e^{-\Delta G_i/k_B T}}{\sum_j e^{-\Delta G_j/k_B T}}$$

where ΔG_j is the free energy of state j (relative to the fully unfolded state) and $\sum_j$ is the sum over all the possible states in the complete conformational ensemble.

We test the sensitivity on the entropy ΔS and the stacking enthalpy ΔH_{stack} separately. We use yeast U2/U6 complex (Figure 2(a)) as an example. A fluctuation (increase) in ΔH_{stack} of 5′CU3′-5′AG3′ is found to cause a small shift in the populational distribution by $(\Delta p_I, \Delta p_{II}, \Delta p_{III}) = (-10\%, -5\%, \text{and } 15\%)$ for structures I, II, and III in Figure 2(a), respectively. Such a change in ΔH_{stack} would stabilize helix Ib in structure III and thus cause an increase in p_{III}. As another test, an increase in ΔH_{stack} of 5′*CC*3′-5′*GG*3′ would cause a more significant change in stability: $(\Delta p_I, \Delta p_{II}, \Delta p_{III}) = (35\%, -30\%, \text{and } -5\%)$, corresponding to the stabilization of structure I.

To test the error-tolerance for the entropy ΔS calculation, we vary (increase) the loop entropy that is computed from the virtual bond model[45] by 10%. We find $(\Delta p_I, \Delta p_{II}, \Delta p_{III}) = (-15\%, 15\%, \text{and } 0\%)$, indicating the stabilization of other unfolded states (of large entropies).

From the above test results, we find that the snRNA/snRNA stability, which involves multiple loops and many base-pairs, can be sensitive to both the base stacking parameters and the loop entropy parameters. This clearly shows the importance for accurate experiments on ΔH_{stack} and rigorous (physical) models for entropy calculation.

Helix Ib in U2/U6 and U12/U6atac complex

U2/U6 (yeast). (a) Experiments suggest two distinct structures. From the genetic and biochemical studies, Madhani & Guthrie[8] proposed a yeast U2/U6 complex structure that contains a novel intermolecular base-pairing interaction, helix stem helix Ib in Figure 2(a)(III). Mutational studies[8,11,18] demonstrated that this inter-molecular helix, formed through the base-pairing between U2 and the conserved U6 triad sequence AGC, is critical for splicing. More specifically, it was suggested that the

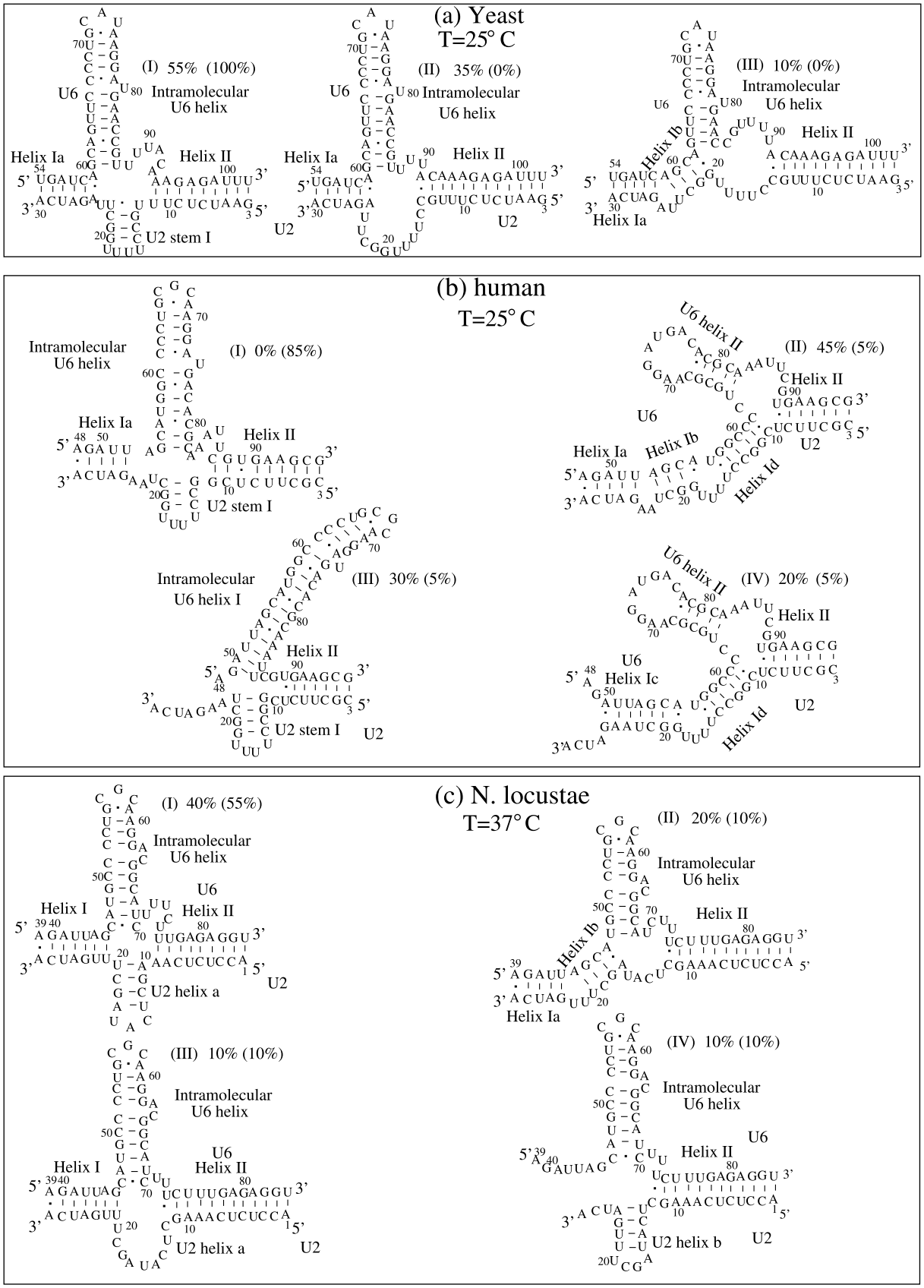

Figure 2. The predicted stable structures for U2/U6 complex in (a) yeast, (b) human and (c) *N. locustae*. Also shown in the Figure are the fractional populations with the possible coaxial stacking (numbers in the parenthesis) and without the coaxial stacking. In order to make close experimental comparisons, we show the results at relevant experimental conditions: 25 °C for yeast and human[1,4,15,22] and 37 °C for *N. locustae*.[16]

formation of helix Ib may be required for exon ligation.[18] Furthermore, structural similarity between the proposed U2/U6 complex and the domains 5 and 6 of Group IIA intron[49] further supports the proposed intermolecular interaction.

Recently, Sashital, Cornilescu & Butcher performed NMR studies on the protein-free yeast U2/U6 complex structure.[22] The NMR study suggested a four-helix junction structure (see Figure 2(a)(I)), which is quite different from that proposed by Madhani & Guthrie.[8] The most notable feature in the NMR structure is the absence of the critical inter-molecular helix, helix Ib. In the four-helix junction, the AGC triad is unzipped from the inter-molecular helix, helix Ib, to form (intramolecular) base-pairs in the elongated U6 intramolecular helix. Furthermore, by making analogy to the hairpin ribozyme,[50] Butcher *et al.*[22,51] proposed that the helix stems of the four-helix junction can form coaxial stacking.

(b) Theory predicts the experimental structures as coexisting/competing states. For a multi-helix junction structure, the adjacent helices may form coaxial stacks. However, with the presence of other molecules such as proteins, such coaxial stacking may be inhibited. Therefore, in our calculation, we perform the calculations with and without coaxial stacking separately.

(i) With coaxial stacking. By including the possible coaxial stacking in the conformational ensemble, we predict a single dominant structure–structure I in Figure 2(a) for 25 °C. In fact, we find that structure I is still the dominant structure even at a higher temperature (37 °C). This is consistent with the experiment.[22] In the calculation, the statistical weights of coaxial stacked structures are evaluated using the coaxial stacking parameters in Walter & Turner.[52] We found that the coaxial stacking can significantly stabilize the four-helix junction structure in Figure 2(a)(I). The most stable structure is found to be formed through the coaxial stacking between U2 stem I and helix II and between helix Ia and the intramolecular U6 helix. It occupies nearly 100% fractional population. From the good SE and SP values of Table 2 (#18), we find that our predicted structure agrees well with the NMR result.[22]

There exist two possible coaxial stacking patterns: U2 stem I–helix II and helix Ia–intramolecular U6 helix stacking, or U2 stem I–helix Ia and helix II–intramolecular U6 helix stacking. Our thermodynamic model shows that the first arrangement is thermodynamically much more stable than the second one. This is because the helix II–intramolecular U6 helix stacking in the second coaxial pattern requires the juxtaposition of the two ends of the loop from U88 to A93, resulting in a large unfavorable entropic loss. In the predicted stable (first) coaxial pattern, the intron-binding site A26 (U2) and U52 (U6) are brought into close proximity, which can facilitate the formation of the tertiary contact between them.[51] This is consistent with the experimental finding.[14,53,54] The second coaxial stacking arrangement, though thermodynamically unstable, supports the juxtaposition between the metal-binding site U80 in the intramolecular U6 helix and the intron-binding region of U2. Our calculation suggests that the possible stabilization of the second coaxial pattern requires assistance from protein or the two coaxial stacking structures exist kinetically before reaching thermal equilibration.

(ii) Without coaxial stacking. Without considering coaxial stacking, the theory predicts three coexisting structures for the U2/U6 complex from yeast at about $T = 25$ °C (see Figure 2(a)). These three structures correspond to three minima on the free energy landscape in Figure 3(a). Structure II is a partially unfolded form of structure I and can thus be classified as the same structural category of structure I. So effectively we can classify the three predicted structures into two structural categories as represented by structures I and III, respectively.

The predicted structures agree with the experimental structures derived from the functional and biochemical studies. For example, both structures I and III contain inter-molecular helix Ia and II and intramolecular U6 helix. Genetic and biochemical studies show that these inter/intra-molecular interactions are important for splicing. In addition, helix Ia has been proposed to play a role in splicing activation by juxtaposing the removed U6 5′ nucleotides (…ACAGAGA…, conserved in yeast and human), which are base-paired with the 5′ ss of the pre-mRNA, and the removed U2 3′ nucleotides (…AGUGUAGUA…, also conserved), which are base-paired with the branch point.

A key difference between structures I and III is whether it contains the intermolecular helix Ib. Previous experimental studies suggested that structure I competes specifically with helix Ib.[11,18,21] Our first principle calculation provides unambiguous support for the experimental finding. From the perfect SE and SP scores for structure III (1.00, 1.00), we find that our predicted structure III, which contains helix Ib, agrees exactly with the experimentally derived structure.[8,18]

For structure I, there is a slight difference between the predicted structure and the NMR structure. The key difference is for the base-pairing for A59 in the AGC triad. A59 in the predicted intramolecular U6 helix in Figure 2(a)(I) is unpaired while it is paired with U88 in the NMR structure. This is because the A59-U88 pair (5′*AG*3′-5′*UU*3′ stack; $\Delta G = -0.6$ kcal/mol)[47] is less stable than the alternative pair A59-A25 (5′*AG*3′-5′*CA*3′ mismatched stack; $\Delta G = -1.8$ kcal/mol). However, the coaxial stacking between helix Ia and the intramolecular U6 helix would stabilize the A59-U88 pair. This result further supports the formation of the possible coaxial stacking between Ia and the intramolecular U6 helix.

Both coaxial stacking and coaxial stacking-free calculations predict that the four-helix junction structure I is more stable than the helix Ib-containing structure III at 25 °C. This is consistent

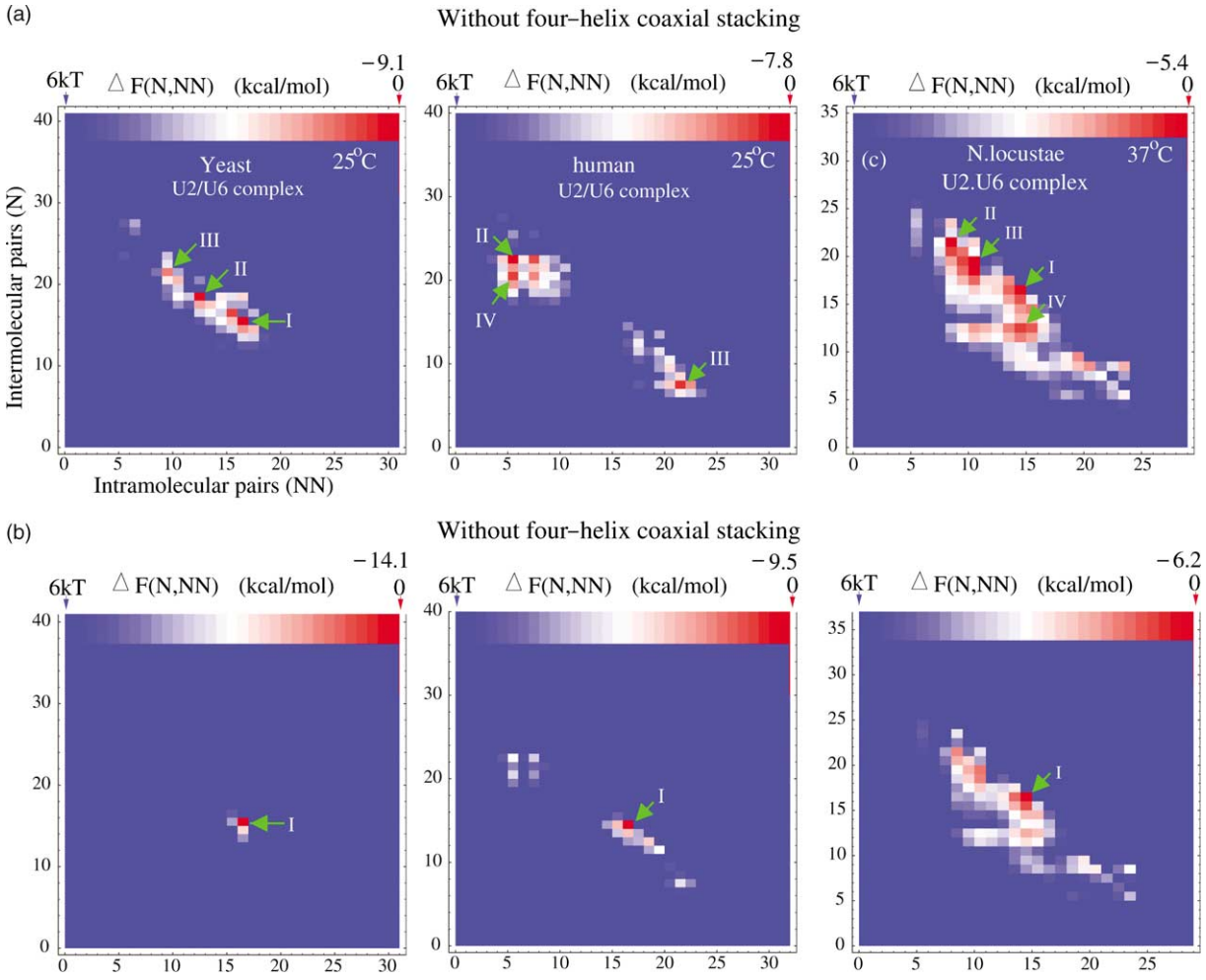

Figure 3. The free energy landscapes for U2/U6 complex in yeast, human and *N. locustae* without (a) and with (b) four-helix coaxial stacking. Only conformations with free energies within $6k_BT$ above the global minimum are shown. For U2/U6 complex in human with four-helix coaxial stacking, structure IV is consistent with the experimental determined structure in Sun & Manley.[12]

with the NMR study, which shows that the four-helix junction is dominant at least for temperatures at 283 K, 298 K, and 303 K.[22] Furthermore, as we will show in the later section on the folding pathway, an increase in temperature would destabilize the four-helix junction structure, resulting in a I→III conformational change. Such conformational switch has been hypothesized to occur during the two transesterification steps of splicing,[22,51,55] and our model supports the experimental hypothesis.

In summary, the statistical mechanical model predicts both experimentally suggested yeast U2/U6 structures (I and III in Figure 2(a)). The two structures are found to coexist and compete with each other. The formation of coaxial stacking can significantly stabilize the four-helix junction structure I. Comparison with the NMR structure supports the proposed coaxial stacking in the four-helix junction. In the section on U2/U6 folding pathways, we will further investigate the conformational change of the structures and their implications on splicing.

U2/U6 (human). Based on the biochemical experiments, Sun & Manley[12] and Valadkhan & Manley[14,15] proposed that human U2/U6 structure contains a long U6 intramolecular helix extended beyond the A–C mismatch (at room temperature). The extension of the U6 intramolecular helix would disrupt the intermolecular helix Ib that is critical for splicing in yeast as demonstrated in biochemical and genetic studies. Furthermore, Sun & Manley[12] suggested that, in analogy to hairpin ribozyme structure, human U2/U6 may form a four-helix junction; see Figure 2(b)(I). Such a structure has been further supported by several other studies.[13–15]

Shown in Figure 2(b) are our predicted low free energy structures for human U2/U6 complexes and the fractional populations with and without the coaxial stacking, respectively. Similar to yeast U2/U6 complex, human U2/U6 also has multiple competing structures.

By including all the possible coaxial stacking in the conformational ensemble, we find that at 25 °C, the four-helix junction (I in Figure 2(b)) is the most

stable structure. It occupies 85% of the total population, which means that 85% of the molecules would have this conformation, or, equivalently, a molecule has a probability of 85% to have this conformation. In fact, we find that the four-helix junction structure I is stable at higher temperature of 37 °C with a fractional population of 35% as compared to the 20% population for structures II, III, and IV (each). This is consistent with the experiment.[12] The predicted structure agrees almost exactly with the experimental structure except for a lone base-pair between A53 (in U6) and U19 (in U2), which is missing in the predicted stable structure. Such lone pairs are treated as unstable in RNA secondary structure. In the four-helix junction structure, the adjacent helix II and U2 stem I are predicted to stack through the 5′*GG*3′-5′*CC*3′ base stacking, which contributes about −4.5 kcal/mol to the total stability.[52]

Moreover, the tertiary interaction between G46 in U6 with A24 in U2[14,15] will further stabilize the four-helix junction structure. In conclusion, our human U2/U6 model supports the proposed hairpin ribozyme-like structure reported in experiments.[12–15]

Without the coaxial stacking, structure II in Figure 2(b) is the most stable structure and structure I becomes unstable ($\Delta G_I - \Delta G_{II} = +2.7$ kcal/mol). This shows that the experimentally derived four-helix junction structure is stable only in the presence of coaxial stacking. This clearly shows the importance of the coaxial stacking in human U2/U6 folding stability.

Further experimental studies suggested the possible role of other alternative structures for human U2/U6. For example, mutational study suggested that the formation of helix Ib (shown in Figure 2(b)(II)) may be important for the first step of splicing in mammalian.[20] In addition, strengthening the helix I in the four-helix junction (in Figure 2(b)(I)) was found to block splicing.[21] This implies that in the splicing process, the coaxial-stacked four-helix junction may be destabilized so that other catalytically important structures can form. Structures II, III, and IV in Figure 2(b) are the competing stable structures if the coaxial stacking is disrupted. However, unlike the U2/U6 structures in yeast, due to the lack of direct structural measurements for human U2/U6, it is unknown whether these predicted structures (II, III, and IV) would correspond to the functional structures of human U2/U6.

For human U2/U6, the truncated sequences in U2 and U6 are highly modified, including the 2′-O-methylations, pseudouridines and base methylations.[56–58] These modifications can affect the stabilities of the predicted structures in Figure 2(b). Qualitatively, 2′-O-methylation (m) and pseudouridine (Ψ) modifications can increase the stability of helices and the base methylations would likely destabilize the base-pairs. As a result, the m25-U2 modification can stabilize structures I, II and IV, which contain m25-U2 in helix Ia or Ic. Structure III is not stabilized because m25-U2 is unpaired in the structure. In contrast, modified bases Ψ6-U2, Ψ7-U2, m11-U2, m12-U2, m19-U2 as well as m53-U6 and m54-U6 are all paired (in different helices) in the four structures (I, II, III and IV). Therefore, the changes in the relative stabilities between the four structures caused by these modifications may not be significant. Due to the lack of the detailed thermodynamical parameters for the modified sequences, however, we are presently unable to give quantitative results for the stability changes caused by the above and other sequence modifications in U2/U6.

U2/U6 (N. locustae). For *N. locustae*, we obtain four low free energy structures. Since structure III is the partially unfolded form of structures I, we can effectively classify the four predicted structures into three structural categories as represented by structures I, II and IV, respectively. Structures I and II are similar to that of yeast. Structure II, which consists of a helix Ib, is supported by experiments.[16] In fact, structure I can be further stabilized by the coaxial stacking between helix II and U2 helix a (5′AA3′-5′UU3′).

By comparing Figure 2(a) (yeast) and (c) (*N. locustae*), we find clear parallel analogy: I (yeast) ↔ I (*N. locustae*) and III (yeast) ↔ II (*N. locustae*). We again find that the helix Ib-containing structure II competes with the structure I.

In fact, even without the coaxial stacking, as shown in Figure 2(a) and (c), the four-helix junction structure I is the most stable structure in yeast and *N. locustae* U2/U6. In contrast, for human U2/U6, the four-helix junction is stable only with the presence of coaxial stacking. This is because the U2 stem I and the intramolecular U6 helix are much more stable in yeast and *N. locustae* than in human, The coaxial stacking further enhances the stability of the four-helix junction structure I and destabilizes the helix Ib-containing structure II. Inhibition of the coaxial stacking may activate the formation of helix Ib, which is functionally important for splicing.[8,18]

U12/U6atac (Arabidopsis and human). The U12/U6atac base-pairing interaction is crucial for the splicing of a minor AT-AC intron.[27,59,60] A structural conservation for the U12/U6atac complex has been demonstrated in Arabidopsis and human.[34] In addition, comparing the U12/U6atac complex (see Figure 4) with the U2/U6 complex (see Figure 2), we find that helix II in U2/U6 in Figure 2 is absent in the U12/U6atac complex. This seems to support the functional redundancy of helix II in the U2/U6 complex (see Figure 2) in the presence of helix Ib in yeast.[61] The structural similarities between the minor and major spliceosome may be explained as an evolutionary consequence.

We calculated the base-pair probability and free energy landscapes of U12/U6atac complex for Arabidopsis and human. The calculated results

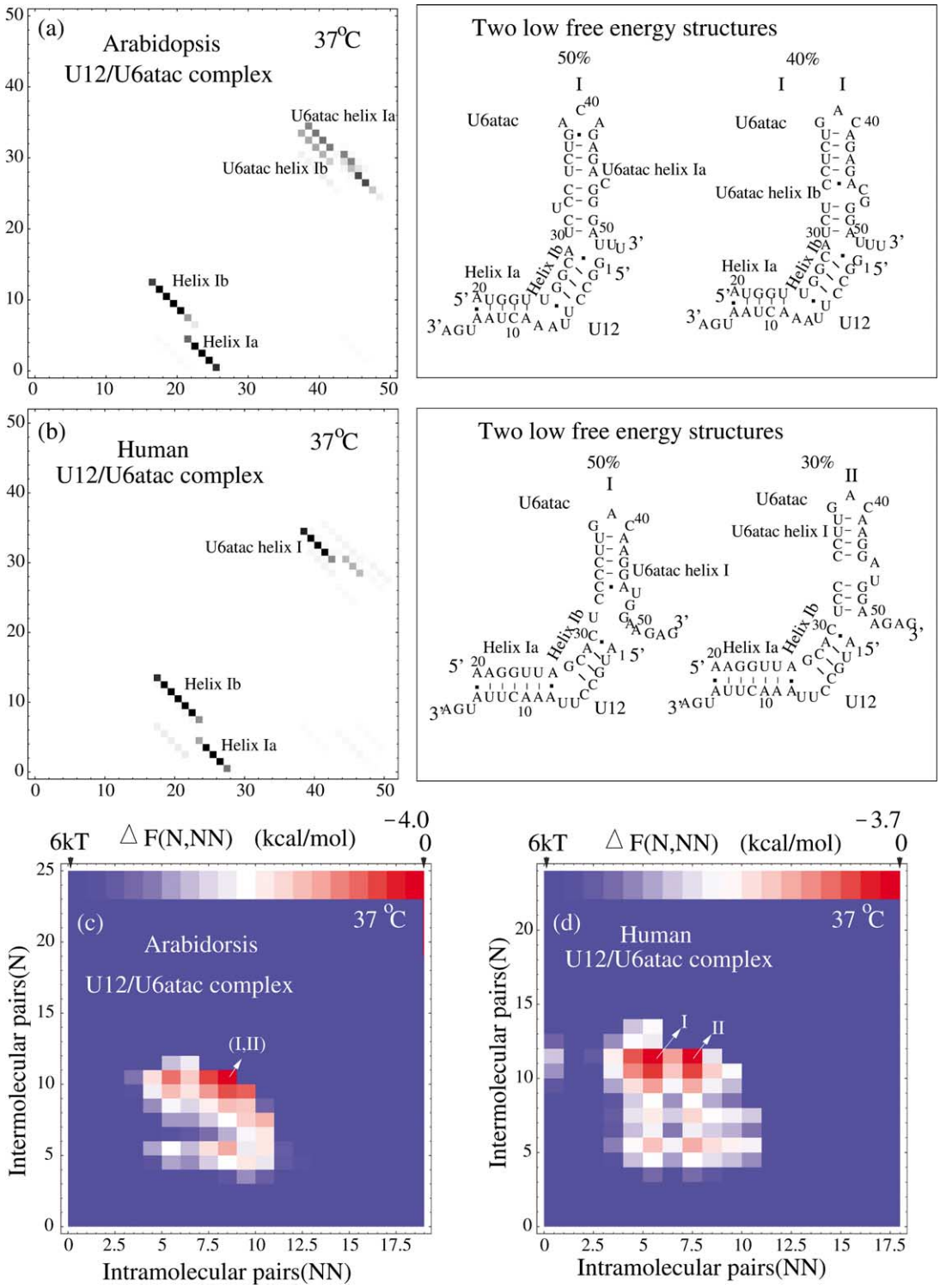

Figure 4. The density plots for base-pairing probabilities and the predicted stable structure of U12/U6atac complex in minor spliceosomes: (a) Arabidopsis and (b) human. The row and column indexes in the density plots denote the nucleotides along the two sequences. The corresponding free energy landscapes are shown as (c) and (d), respectively.

are presented in Figure 4. The predicted structures consist of three important parts: helix Ia, helix Ib and U6atac intramolecular stem–loop structure. Mutational and biochemical studies support the novel base-pairing interaction of U12/U6atac in the predicted structure II in Figure 4.[27,34] Moreover, we find an alternative structure (I in Figure 4) in the U12/U6atac complex for both species. The U6atac intramolecular stem–loop structure is rearranged as compared with structure II in Figure 4.

Structures and structural conservation for the U4/U6 and U4atac/U6atac complexes

The U4 and U6 base-pair interaction in yeast and human consists of two intermolecular helices and one U4 intramolecular stem–loop structure.[46,62] One of the primary functions of U4 is to assemble the U6 chain in an inert conformation.[63] The U4 snRNA does not necessarily participate in the chemical step of splicing since a spliceosome intermediate that lacks U4 has been shown to be functional.[64]

U4/U6 (yeast, human). We predict the low free energy structures from the base-pairing probability. Figure 5(a) and (b) show the predicted stable structures for U4/U6 in yeast and in human, respectively. We find two competing alternative structures (I and II in Figure 5(a)) in yeast. Structure II agrees with the proposed structure in experimental study[62] and the alternative structure I differs from structure II in a single-base shift in the stem II helix (see Figure 5(a)). In human, a single low free energy structure (in Figure 5(b)) is predicted. The structure is strongly supported by the experiments.[35,46] From the predicted U4/U6 structures, we find high conservation for structures of the U4/U6 complex in yeast and human, both consist of two intermolecular base-pairing helices, separated by an intramolecular U4 stem–loop component.

(C.2) U4atac/U6atac (human, Drosophila). The novel U4atac and U6atac base-pairing interaction has been proposed by Tarn & Steitz.[27] Later, the model for U4atac/U6atac base-pairing interactions was revised by Padgett & Shukla[35] to account for a 2 nt error in U4atac snRNA. We predicted the low free energy states using the corrected sequences for both human and Drosophila. The predicted structures are shown in Figure 5(c) and (d). The structure I in both human and Drosophila are in good agreement with that proposed in Padgett & Shukla.[35] An alternative structure (II in Figure 5(c)) is found in the U4atac/U6atac complex from human.

The high structural similarity between U4/U6 and U4atac/U6atac and between U2/U6 and U12/U6atac support the structural conservation in the evolution. Experiments *in vivo* have demonstrated that the U6 snRNA and U6atac stem–loop are functionally analogous, and U6 snRNA can functionally replace the U6atac snRNA stem–loop.[65] Furthermore, it was found that the U4 snRNA can function in both major and minor spliceosomes.[37] This evidence gives additional support for the structural conservation in the snRNA complexes between major and minor spliceosomes.

Folding pathways and implications on splicing functions

Our motivation of the present study is not only on the native structures but also, more importantly, on the structural stability and the conformational changes away from the native structure. The structural transitions can be vividly revealed by the free energy landscape that gives the full distribution of the complete conformational ensemble. Mathematically, we compute the free energy landscape $F(N,NN)$ as a function of structural parameters (N,NN), where N and NN denote the numbers of intermolecular and intramolecular base-pairs, respectively.

The minima of the landscape indicate the stable states and the global minimum corresponds to the most stable state. From the change of the shape of the landscape, we can obtain the information about the conformational change. For example, from the energy landscape, we found that two folding intermediates existed in the *E. coli* 23 S rRNA,[66] one is native-like and the other is misfolded.

In addition, from the energy landscape, we can investigate whether the folding/unfolding is two-state (cooperative) or multi-state with the formation of intermediates, and whether the folding intermediates are "on-pathway" or "off-pathway" in the RNA–RNA complexes. Here we define an on-pathway intermediate as a stable state containing native base-pairs only and an off-pathway intermediate as a state containing non-native base-pairs. Here a base-pair is called "native" if it exists in the native structure and "non-native" otherwise.

Yeast U2/U6

From the detailed experimental comparisons and thermodynamic analysis for the structures at 25 °C shown in Figure 2(a), we find the following conclusions about the structure of yeast U2/U6 complex:

(1) Without coaxial stacking, there exist two competing structures: the four-helix junction structure I and the three-helix junction III structure that contains intermolecular helix Ib.
(2) With coaxial stacking, the four-helix junction structure I becomes the only stable structure at 25 °C. The thermodynamically favorable coaxial stacking is formed between U2 stem I and helix II and between helix Ia and the intramolecular U6 helix.
(3) The predicted four-helix junction structure I agrees very well with that derived from the NMR study.[22] The four-helix junction structure is consistent with the hypothesized structural role to promote the formation of the catalytic core.
(4) The predicted three-junction structure III agrees exactly with the one derived from the biochemical and genetic studies.[8,18] The three-helix junction structure contains the intermolecular helix Ib which has been shown to be critical for splicing.[8,18]

From the NMR study by Sashital *et al.*,[22] it was proposed that the four-helix junction structure I

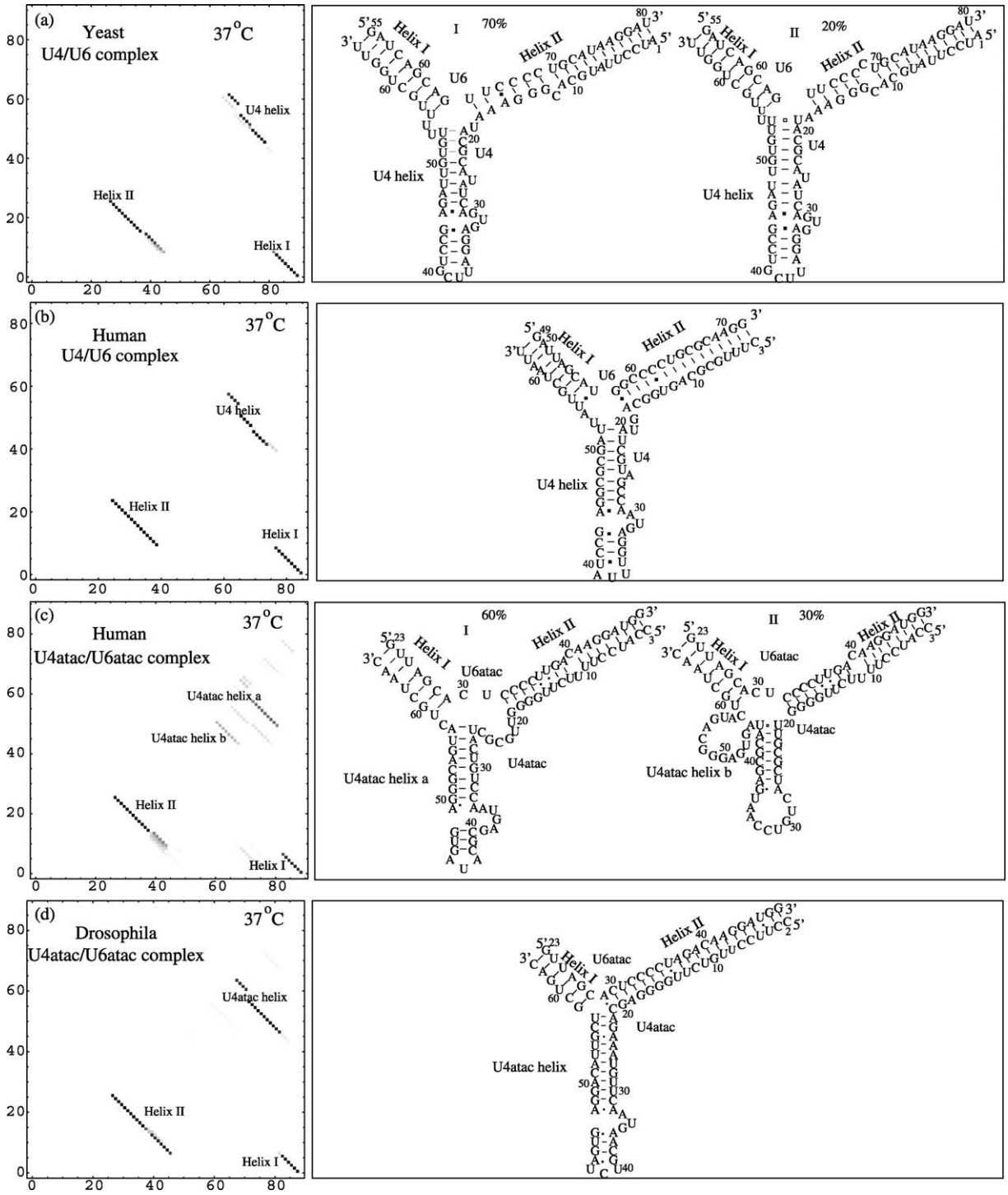

Figure 5. The density plots for base-pairing probabilities and the predicted stable structures in U4/U6 complex in (a) yeast and (b) human and U4atac/U6atac complex in (c) human and (d) Drosophila. The row and column indexes in the density plots denote the nucleotides along the two sequences.

may be important for the first step of splicing and the helix Ib-containing structure III may be important for the second step of splicing. Therefore, a conformational switch between the two structures may occur in the splicing process. How to understand the formation of the functionally important helix Ib-containing three-helix junction III from the (most) stable four-helix junction structure I? It was proposed that protein-binding may play certain unknown roles to promote the formation of helix Ib.[8,18] From our quantitative analysis for the role of coaxial stacking in stabilizing the four-helix junction, we further hypothesize that protein may intrude the four-helix junction structure to disrupt the coaxial stacking. With the coaxial stacking disrupted, the four-helix junction structure is

significantly destabilized and the competing helix Ib-containing three-helix junction structure is stabilized and formed.

We use the free energy landscape to investigate the equilibrium pathways for the thermal unfolding and conformational switch of U2/U6. The existence of the alternative structures for the U2/U6 complexes at 25 °C implies that the folding/unfolding of U2/U6 may involve multiple pathways with on-pathway and off-pathway intermediates. In Figure 6, we show the folding pathways for yeast U2/U6 complex. First, the proposed protein binding would disrupt the coaxial stacking in the coaxial stacked structure N, causing a structural change from N to its non-coaxial-stacking form (structure I in the Figure). We then investigate the change in the free energy landscape and structural distribution in the thermal unfolding process.

At low temperature ($T=0$ °C), we find a single native structure I, as shown as the single global minimum on the free energy landscape. As temperature increases, the helix Ib-containing

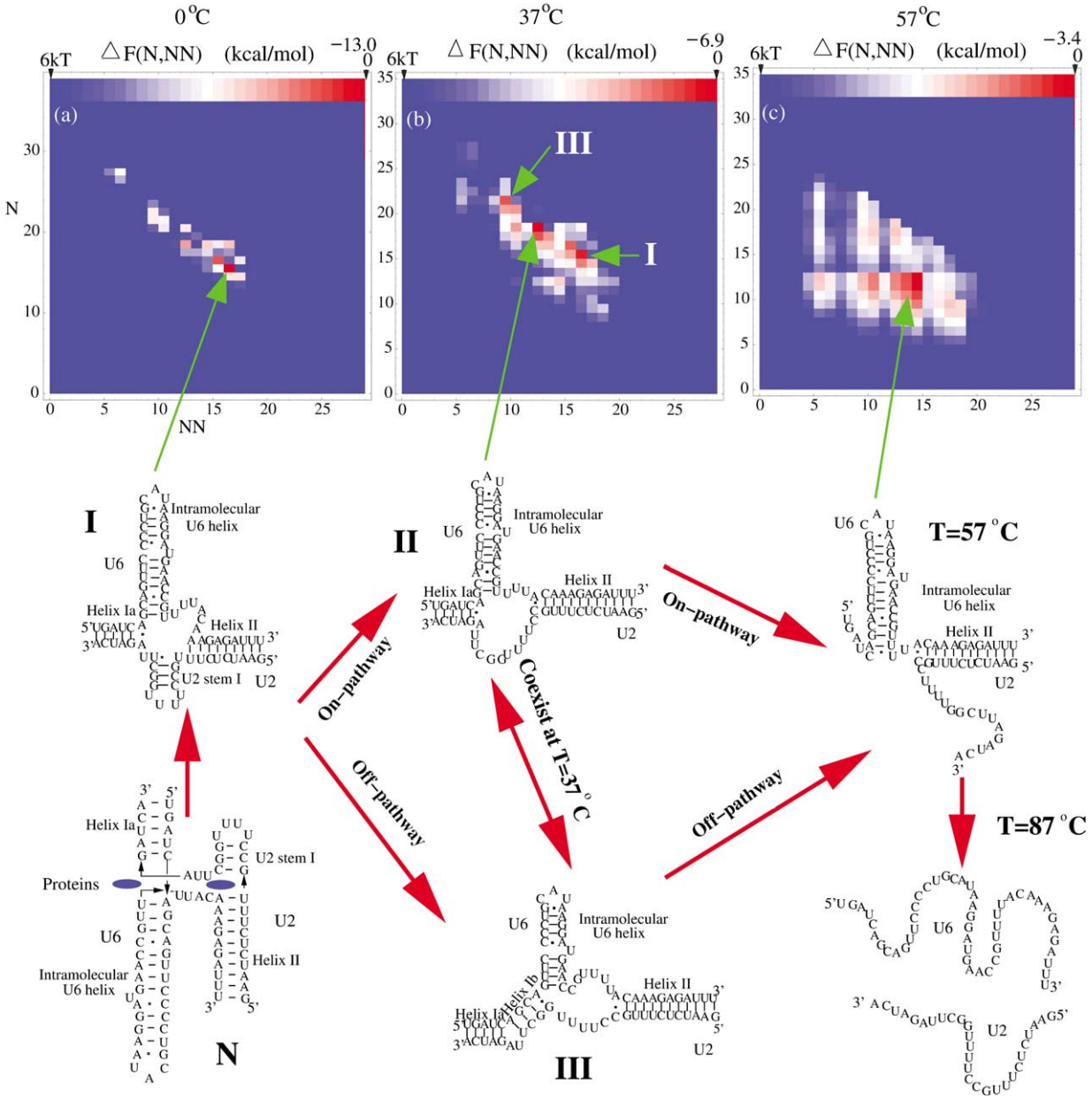

Figure 6. The stable structures of the U2/U6 complex from yeast at different temperatures. *N* is the four-helix junction structure with the formation of the coaxial stacking. With the postulated inhibition of the coaxial stacking (by protein binding), $N \rightarrow$ I, where I is the four-helix junction without the coaxial stacking. As shown in the Figure, I is stable at low temperature. The misfolded structure (III) emerges at $T=37$ °C, which requires the unwinding of U2 stem I in structure I. Structure III is believed to be functionally active for splicing. Though structure I is the sole stable structure at $T=0$ °C, as shown in Figure 2(a), the three structures I, II, and III coexist at $T=37$ °C. Conformation II is a partially unfolded form of conformation I.

structure emerges as an off-pathway intermediate III. The formation of helix Ib requires a structural change to unwind U2 stem I (see Figure 6). Such conformational change from intramolecular U2 stem I to intermolecular helix Ib is consistent with mutational experiments.[18,21] In the following steps, helix Ia is unzipped before helix II is unfolded, i.e. helix II is more stable than the helix Ia. Finally, at about 87 °C, the double-stranded complex dissociates into two single-stranded chains.

U12/U6atac

Multiple intermediates are also found in the U12/U6atac complex for Arabidopsis (see Figure 7 for the corresponding free energy landscapes). We note that structures I and II have the same (*N*,*NN*) (*N* and *NN* are the numbers of intermolecular and intramolecular base-pairs, respectively). Therefore, they are embedded in the same point on the free energy landscape. To distinguish these two structures, we further plot the base-pair probability (data not shown). From the base-pairing probability plot, we find a single stable native structure I at $T=0$ °C. As the temperature is increased, the native structure is partially unfolded to form a new structure II. At $T=37$ °C, structures I and II are nearly equally stable. Structure II is found to play a crucial role in splicing.[34] As the temperature is further increased, helix Ia and U6atac intramolecular pairs start to be disrupted. At about $T=87$ °C, helix Ib, the most stable helix stem in the complex, is disrupted and the complex is dissociated.

For both U2/U6 and U12/U6atac complexes, folding pathways show complicated on-pathway (with native-like intermediates) and off-pathway (with misfolded intermediates) folding processes. Moreover, the free energy landscapes (in Figures 3 and 4), for the snRNA/snRNA complexes are quite bumpy with multiple local minima, corresponding to alternative stable structures. For RNA/RNA complexes, depending on the sequence, the interplay between the intermolecular and intramolecular interactions can cause more profound structural

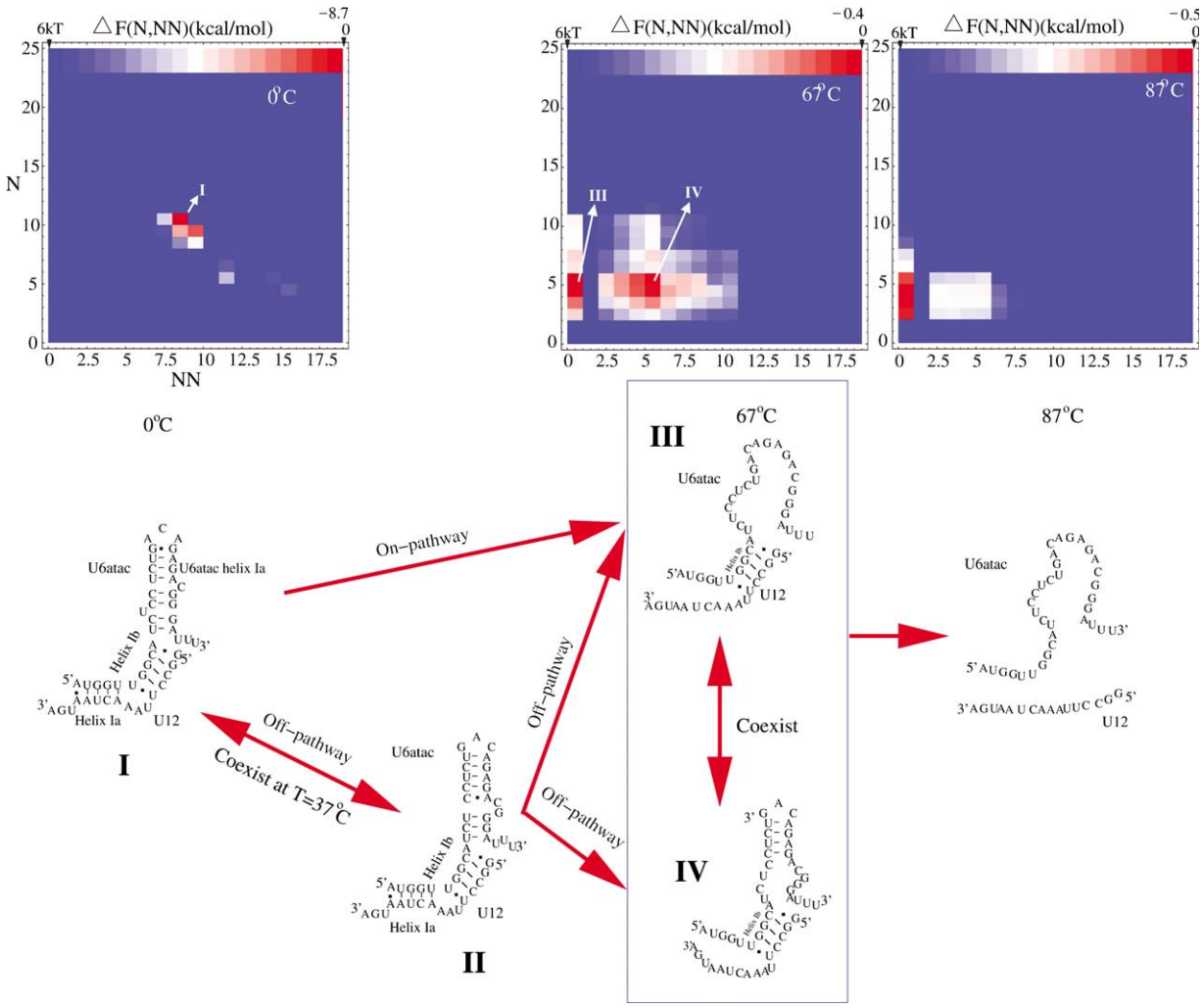

Figure 7. The stable structures of the U12/U6atac complex from Arabidopsis at different temperatures. The misfolded structure (II) at $T=37$ °C, which involves the refolding of U6atac intramolecular helix stem, is believed to be functionally active for splicing.

rearrangements than a single-stranded RNA,[66] resulting in a more bumpy energy landscape. For most of the proteins, the thermodynamic and mutational stability are optimized in the evolution.[67] For snRNA complex, however, the molecule may be optimized in such a way that multiple structural transitions can occur in order to perform different functions. This observation is in agreement with extensive experimental findings.[68–70]

For the snRNA/snRNA complexes studied here, the dissociated states, corresponding to the positions with (the number of intermolecular base-pairs) $N=0$, are shown as hills (=high free energy states) on the landscapes. This means that the dissociated states are unstable and snRNAs have a strong affinity to form complexes.

Conclusion

We develop a statistical mechanical model to predict the folding free energy landscapes and conformational changes for RNA/RNA complexes. In the model, we explicitly take account of: (i) both the intermolecular and the intramolecular interactions and (ii) the canonical and the non-canonical (mismatched) base-pairs. The interplay between the intermolecular and the intramolecular interactions leads to a variety of distinctive folding behaviors of RNA complexes. More specifically, our studies on the energy landscapes and folding pathways for snRNA complexes lead to the following conclusions:

(1) snRNA complexes have rugged free energy landscapes, corresponding to multiple coexisting/competing structures. For example, different experiments suggested two contrasting yeast U2/U6 structures (four-helix junction and helix Ib-containing structure), our model shows that these distinct structures actually coexist: they emerge as two free energy minima on the free energy landscape for the full conformational ensemble. Furthermore, our calculation supports the experimental hypothesis that helix Ib competes specifically with the four-helix junction.
(2) In analogy to the hairpin ribozyme structure, coaxial stacking between adjacent helix stems can form in snRNA complexes. The coaxial stacking can significantly stabilize the helix junction structures in some snRNA complexes. For example, in yeast U2/U6, the coaxial stacking can significantly stabilize one of the two competing structures, the four-helix junction structure. Therefore, we hypothesize that the conformational switch from the four-helix junction to the helix Ib-containing structure may be activated by the disruption of the coaxial stacking, presumably by protein, in the splicing process.
(3) The folding pathways for snRNA complexes involve multiple native-like and misfolded intermediates. For example, at low temperature, the functional helix Ib structure is unstable. Upon heating, at $T=37\,^{\circ}\mathrm{C}$, the intramolecular U2 stem I is unzipped and the structure is rearranged to form a misfolded state, which contains the functionally important intermolecular helix Ib. The model can quantitatively predict the stability of all the possible intermediates. Such quantitative analysis for the formation of the intermediate states is essential for the understanding of splicing efficiency.

Although the application of our model is focused on the folding of the snRNA complexes, the model is general and can be widely used on predicting the free energy landscapes and folding pathways of other RNA/RNA complexes. The model may also be applied to treat the DNA/DNA and DNA/RNA complexes[38] if we use the corresponding enthalpy/entropy parameters.[71]

Theory and Methods

Free energy landscape of RNA/RNA complex

As shown in Figure 8, we consider a thermodynamic system consisting of single-stranded RNA molecules S_1, S_2 and double-stranded RNA/RNA complex S_{12}. The folding thermodynamics of the RNA/RNA complex is determined by the partition functions Z_s for S_s ($s=1,2,12$). From the partition functions, we can compute the free energies F_s for the respective single-stranded chains and the double-stranded complex from: $F_s=-k_BT\ln Z_s$.

To relate the free energy to the structure, we define free energy landscape.[66,72–74] We define the free energy landscape $\Delta F(\Gamma)$ (Γ=parameter set used to describe the structure) for the folding (formation) of

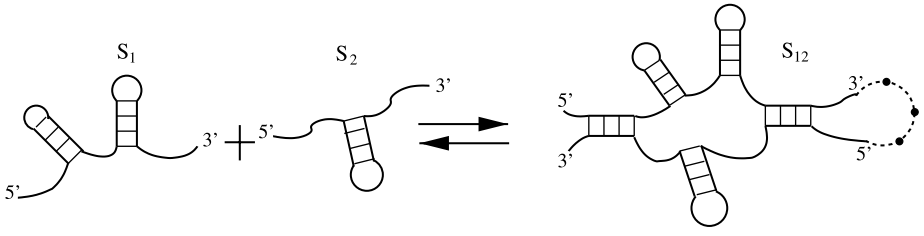

Figure 8. Formation and dissociation of a double-stranded complex S_{12} from two single-stranded chain molecules S_1 and S_2. A three-nucleotide phantom link (broken line) is added between the 3′ end of S_1 and the 5′ end of S_2.

the complex as:

$$\Delta F(\Gamma) = F_{12} - (F_1 + F_2) + \Delta G_{\text{init}}$$
$$= -k_B T \ln\left(\frac{Z_{12}}{Z_1 Z_2}\right) + \Delta G_{\text{init}} \quad (1)$$

where ΔG_{init} is the free energy change associated with the nucleation of the two strands. Experimental measurement gives $\Delta G_{\text{init}} = 4.1$ kcal/mol for RNAs at 37 °C.[40] As an approximation, we can assume that ΔG_{init} is independent of structure Γ and treat ΔG_{init} as a constant, because it does not cause any change in the shape of the free energy landscape.

The calculation of the total partition function for RNA/RNA complex

The partition function calculation for single-stranded molecules has been well studied.[75–78] Recently, we developed a statistical mechanical thermodynamic theory for single-stranded RNA secondary structure folding.[45] The theory is based on the virtual bond chain representation of realistic RNA structures. Therefore, unlike previous lattice models, the model can directly treat realistic RNA conformations. In addition, the model has two distinctive features. First, it explicitly accounts for non-canonical base-pairs. Second, it gives rigorous chain entropy (including the multi-branched loop entropy) from polymer principle calculation. In the entropy calculation, the theory can treat the steric hindrance between different nucleotide groups in the sequence and the correlations between different structural units (=loops, base stacks). Therefore, the model accounts for the non-additivity effect (i.e. total entropy is unequal to the addition of entropy of each structural unit). The theory can successfully predict folding thermodynamics for RNAs tested (against experiments).

Despite the several previous successful efforts on the modeling of RNA/RNA complexes,[39–43,79–81] accurate prediction for double-stranded RNA/RNA complex folding thermodynamics by accounting for both intermolecular and intramolecular interactions, non-canonical base-pairs, thermal fluctuations and full conformational distributions remains a great challenge. In the present work, we develop a statistical mechanical model to compute the folding thermodynamics for RNA/RNA complexes.

We consider RNA secondary structures only. According to the nearest neighbor model for RNA secondary structure, we describe the chain conformations in terms of base stacks (formed by neighboring base-pairs). Especially, we consider not only the canonical (Waston–Crick) base-pairs but also the non-canonical (mismatched) base-pairs as well. We use the recently developed virtual bond model to compute the chain entropy.[45] For the enthalpy ΔH_{stack} for the base stacks, we use experimentally measured empirical parameters. Specifically, we use Table I in Serra & Turner[47] for the base stacking, including mismatched stacking, parameters. Our model can also include the coaxial stacking in the calculation. We use Table II in Walter & Turner[52] for the coaxial stacking parameters.

The basic idea in the model is to transform the double-stranded complex into an equivalent single-stranded chain. As shown in Figure 8, we connect the two molecules in the complex by a three-nucleotide phantom linker. With the phantom linker, the two-stranded complex appears as a "virtual" single-stranded chain. However, since the phantom linker is not real, thermodynamic parameters associated with the phantom linker are set to be zero. Because of this, the previous theory for single-stranded RNA folding cannot be used to treat the double-stranded complex (with the phantom linker). We need a new theory.

We note that a previous model[40] also introduced a phantom linker. However, that model cannot treat intramolecular base-pairs, which are believed to play crucial structural and functional roles in many RNA/RNA complexes. As we show in the following, the model developed in this study can account for both intermolecular and intramolecular base-pairs.

With the phantom linker as a divider, the chain $a \to b$ in Figure 9(a) and (b) can have two types of conformations: closed (Figure 9(a)) and open (Figure 9(b)) according to whether the chain is closed by a base stack or not. Furthermore, in order to keep track of the position of the phantom linker so that the contribution from the phantom linker can be conveniently excluded, we further distinguish two types of "closed" chain conformations ($C^0(a,b)$ and $C^1(a,b)$) (e.g., $a \to b$ in Figure 9(a)):

$$\begin{array}{ll} \text{type}-1 & \text{if the phantom linker resides} \\ & \text{inside the closed chain } a \to b \\ \text{type}-0 & \text{otherwise} \end{array}$$

From the above definition for the closed conformations, we can classify two types of "open" conformations $a \to b$ (see Figure 9(b)): "type-0" if the phantom linker resides in the single-stranded chain segments $a \to a_1$, $b_1 \to a_i$, $b_i \to a_n$, $b_n \to b$, $5' \to a$ or $b \to 3'$, and "type-1" if the phantom linker resides inside one of the (smaller) closed conformations $a_1 \to b_1, \ldots a_i \to b_i, \ldots$, or $a_n \to b_n$.

RNA secondary structure shows a recursive hierarchy: the closing base stack of a closed conformation (e.g., stack $(a-1,a,b,b+1)$ Figure 9(a)) can be connected to another smaller closed conformation through either a real loop (C1) or the open conformations (C2) or through a closed conformation (C3).

In order to account for the correlation between the closing base stack (with (a,b) base-pair; see Figure 9(c)(C2)) and the other connected structural units (from $a+1$ to $b-1$ in Figure 9(c)(C2)), we distinguish the following four types for the (open) chain from a to b (see Figure 9(b) and (c)(C2)):

type − LR	if a_1 is adjacent to a (i.e. $a_1 = a + 1$) and b_n is adjacent to b (i.e. $b_n = b - 1$)
type − L	if only a_1 is adjacent to a
type − R	if only b_n is adjacent to b
type − M	is neither a_1 nor b_n is adjacent to a or b.

In Figure 9(d), we note that a is the (left) 5′-terminal nucleotide and b is the (right) 3′-terminal nucleotide. Types-L, R, LR and M correspond to a bulge on the strand close to the 3′ end, on the strand close to the 5′ end, on both the 5′ and the 3′ strands, and an internal unstacked loop, respectively.

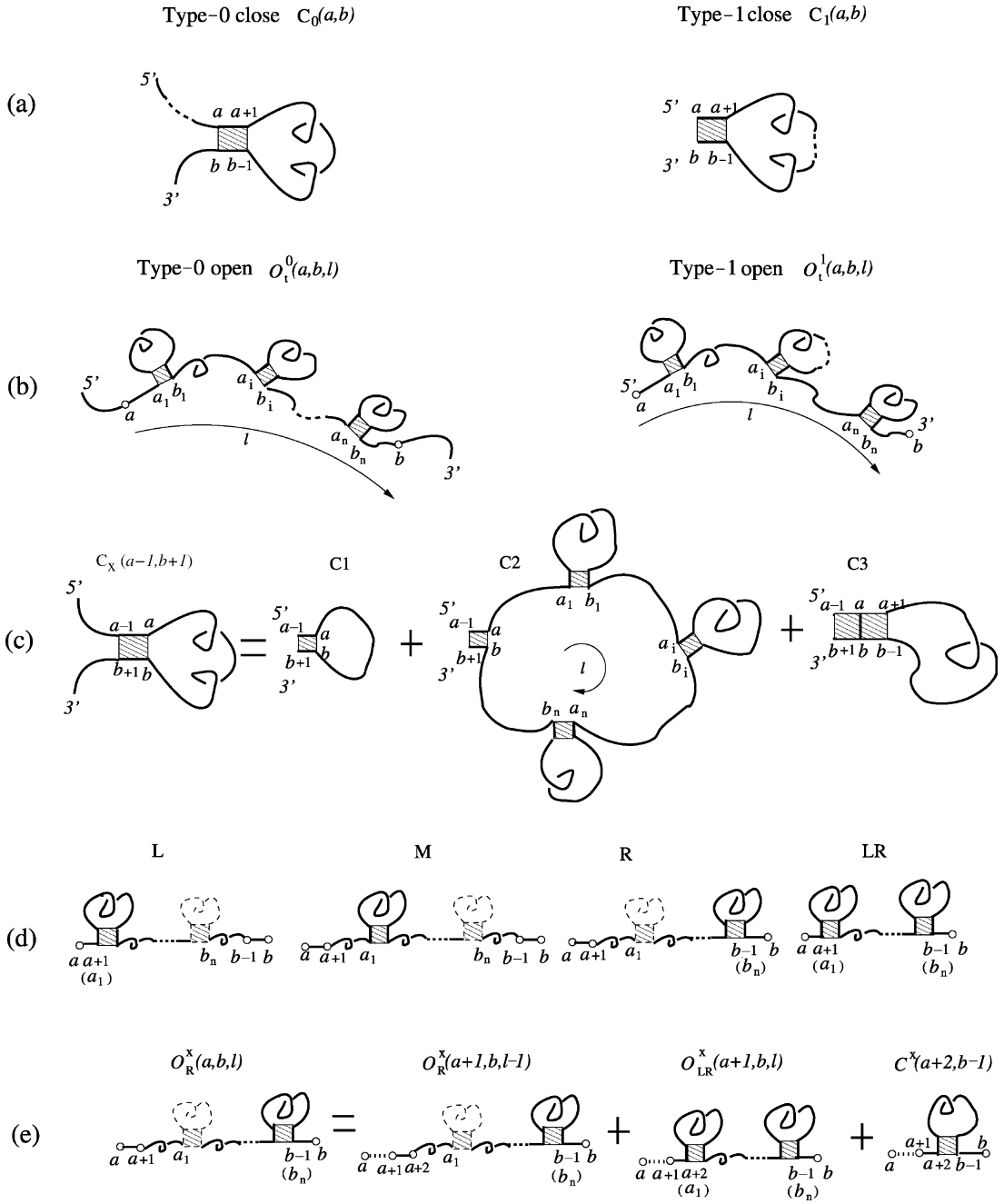

Figure 9. (a) Two types of closed conformations for the chain $a \to b$ with the phantom link outside and inside the closed conformation. The respective partition functions are denoted by $C_0(a,b)$ and $C_1(a,b)$. (b) Two types of open conformations from a to b: $O_t^0(a,b,l)$ and $O_t^1(a,b,l)$. (c) The closed conformation $C_x(a-1,b+1)$ can have three types of possible conformations for the chain $a \to b$: a loop (C1), an open conformation (C2), or a closed conformation (C3). (d) The four types of open conformation: t=L,R,M,LR for $O_t^0(a,b,l)$ and $O_t^1(a,b,l)$. (e) The partition function for the R-type conformations for a chain from a to b can be computed as the sum of the partition functions for shorter chains ($(a+1) \to b$ and $(a+2) \to (b-1)$).

For a chain $a \rightarrow b$, we use $C(a,b)$ and $O(a,b,l)$ to denote the partition functions for the closed and the open conformations, respectively. Here l is the loop length as illustrated in Figure 9(b). We also use the superscript (0 or 1) and the subscript (t=L,R,LR,M) to denote the conformational types. The hierarchical relationship of RNA secondary structure results in the following recursive relations for the partition functions:

$$\begin{aligned} C^0(a-1,b+1) = (e^{-\Delta G_{\text{stack}}/k_B T})\{&C^0(a,b) \\ &+ e^{\Delta S_{\text{unstacked}}(b-a-1)/k_B} \\ &+ \sum_{t,l} e^{\Delta S_{\text{unstacked}}(l)/k_B} O_t^0(a,b,l)\} \end{aligned} \tag{2}$$

$$\begin{aligned} &C^1(a-1,b+1) \\ &= (e^{-\Delta G_{\text{stack}}/k_B T})\left\{ C^1(a,b) + 1 + \sum_{t,l,x=0,1} f^x O_t^x(a,b,l) \right\} \end{aligned} \tag{3}$$

$$f^x = \begin{cases} 1 & \text{if } x = 0 \\ e^{\Delta S_{\text{unstacked}}(l)/k_B} & \text{if } x = 1 \end{cases} \tag{4}$$

where ΔG_{stack} is the free energy of the closing stack formed by base-pairs (a,b) and $(a-1,\ b+1)$ in Figure 9(c) and $\Delta S_{\text{unstacked}}(l)$ is the loop entropy for the unstacked loop of length l (see Figure 9(c)(C2)) of type t (=LR,L,R,M). The f^x factor accounts for the contribution from the unstacked loop of length l: f^x = 1 if the phantom linker is embedded in the unstacked loop (and so the loop is not a real loop), and $f^x = e^{\Delta S_{\text{unstacked}}(l)/k_B}$ if the phantom linker is not in the unstacked loop (and so the loop is a real closed loop).

Figure 9(c) gives an illustration for how to calculate the partition function $C_x(a-1,b+1)$ for closed conformation from $a-1$ to $b+1$. There are three types of possible conformations from a to b: a single (real) loop (C1), an open conformation (C2), and a closed conformation (C3) connected a smaller closed conformation.

The key for the partition function calculation is to obtain $O_t^x(a,b,l)$ (x=0,1; t=M,L,R,LR) for different a and b. From the recursive relations in equations (2) and (3), we can calculate the partition function of the chain from the partition functions of shorter chain segments using the following recursive relationships:

$$\begin{aligned} O_{\text{L}}^x(a,b,l) = O_{\text{L}}^x(a,b-1,l-1) + O_{\text{LR}}^x(a,b-1,l) \\ + C^x(a+1,b-2) \end{aligned}$$

$$O_{\text{M}}^x(a,b,l) = O_{\text{M}}^x(a,b-1,l-1) + O_{\text{R}}^x(a,b-1,l)$$

$$\begin{aligned} O_{\text{R}}^x(a,b,l) = O_{\text{R}}^x(a+1,b,l-1) + O_{\text{LR}}^x(a+1,b,l) \\ + C^x(a+2,b-1) \end{aligned}$$

$$\begin{aligned} O_{\text{LR}}^0(a,b,l) = \sum_{a<y<b} C^0(y,b-1)\{&O_{\text{L}}^0(a,y,l-2) \\ &+ O_{\text{LR}}^0(a,y,l-1) + C^0(a+1,y-1)\} \end{aligned}$$

$$\begin{aligned} O_{\text{LR}}^1(a,b,l) = \sum_{\substack{a<y<b \\ x1+x2=1}} C^{x1}(y,b-1)\{&O_{\text{L}}^{x2}(a,y,l-2) \\ &+ O_{\text{LR}}^{x2}(a,y,l-1) + C^{x2}(a+1,y-1)\} \end{aligned}$$

Figure 9(e) gives an illustration for the recursive relation for the calculation of $O_{\text{R}}^x(a,b,l)$. The recursive relations for the other types can be understood through similar diagrammatic illustrations.

The total partition function $Q_{\text{tot}}(a,b)$ for a chain from a to b is given by the sum of the partition functions for all the different types of conformations:

$$Q_{\text{tot}}(a,b) = 1 + \sum_{x=0,1} \left\{ C^x(a,b) + \sum_{l,t} O_t^x(a-1,b+1,l) \right\} \tag{5}$$

The first term is the contribution from the unfolded coil state. The computational time scales with the chain length N as $O(N^4)$ and the memory scales as $O(N^2)$.

From the total partition function $Q_{\text{tot}}(a,b) = Z_{12} + Z_1 Z_2$, we obtain the partition function for the complex Z_{12} from the following equation:

$$Z_{12} = Q_{\text{tot}}(a,b) - Z_1 Z_2 \tag{6}$$

where $Q_{\text{tot}}(a,b)$ is obtained from equation (5) and Z_1 and Z_2 are obtained from the statistical mechanical model that we developed for single-stranded RNA secondary structures.[45]

The calculation of the base-pairing probability and the free energy landscape

To obtain the detailed structural information from the free energy landscape, we compute the base-pairing probability $p_s(x,y)$ for each and every base-pair (x,y) for both the double-stranded complex (s=12) and the single-stranded free molecules (s=1 or 2):

$$p_{12}(x,y) = \frac{Z_{12}(x,y)}{(Z_{12} + \alpha Z_1 Z_2)} \quad \text{for the complex } S_{12}$$

$$p_1(x,y) = \frac{\alpha Z_1(x,y) Z_2}{(Z_{12} + \alpha Z_1 Z_2)} \quad \text{for } S_1$$

$$p_2(x,y) = \frac{\alpha Z_2(x,y) Z_1}{(Z_{12} + \alpha Z_1 Z_2)} \quad \text{for } S_2$$

where $\alpha = e^{\Delta G_{\text{init}}/k_B T} \simeq 735$ for T=37 °C and $Z_s(x,y)$ (s=1,2,12) are the conditional partition functions of all the conformations with base-pair (x,y) formed in state S_s. From the base-pairing probability, we can obtain the information about the most probable structures.

In contrast to the previous models,[39–43] the advantages of the present model are that: (i) it

accounts for not only the intermolecular interactions, but also the intramolecular interactions, (ii) it explicitly treats the non-canonical (mismatched) base-pairs, and (iii) it gives not only the native state, but also the full energy landscape and the full conformational distribution, (iv) it is based on a new statistical mechanical theory for RNA folding and RNA/RNA complexes, and the chain entropy is calculated from a polymer physics theory using the virtual bond chain representation.

We compute the free energy landscape $F(\Gamma)$ from the partition functions Z_s ($s=1,2,12$) (see equation (1)). The minima on the free energy landscape $F(\Gamma)$ represent the stable states. From the change of the landscape shape, we can vividly observe the conformational change and the change of the conformational distribution. To directly "observe' the association and dissociation of the complex from the free energy landscape, we choose a two-component structural parameter set $\Gamma=(N,NN)$ to describe the structure of the complex. Here N and NN are the numbers of intermolecular and intramolecular base-pairs, respectively. The free energy landscape $\Delta F(N,NN)$ can be plotted as a visualizable three-dimensional landscape. $N=0$ corresponds to the state that the double-stranded complex S_{12} is completely dissociated into two free single-stranded molecules S_1 and S_2. The detailed (stable) structure at the minima on the free energy landscape can be confirmed/identified from the base-pairing probabilities.

Acknowledgements

We thank Drs Kathleen Hall, Brenda Peculis and Jonathan Staley for valuable discussions. Funding for this project was provided by NIH (GM063732, to S.-J.C.) and MU life science fellowship (to S.C.).

References

1. Berget, S. M., Moore, C. & Sharp, P. A. (1977). Spliced segments at 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA*, **74**, 3171–3175.
2. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. (1977). An amazing sequence arrangement at 5′ ends of adenovirus 2 messenger-RNA. *Cell*, **12**, 1–8.
3. Will, C. L. & Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**, 320–328.
4. Staley, J. P. & Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**, 315–326.
5. Collins, C. A. & Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? *Nature Struct. Biol.* **7**, 850–854.
6. Jurica, M. S. & Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell.* **12**, 5–14.
7. Madhani, H. D., Bordonné, R. & Guthrie, C. (1990). Multiple roles for U6 snRNA in the splicing pathway. *Genes Dev.* **4**, 2264–2277.
8. Madhani, H. D. & Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803–817.
9. McPheeters, D. S. & Abelson, J. (1992). Mutational analysis of the yeast U2 snRNA suggests a structural similarity to the catalytic core of group-I introns. *Cell*, **71**, 819–831.
10. Fortner, D. M., Troy, R. G. & Brow, D. A. (1994). A stem/loop in U6 RNA defines a conformational switch for pre-messenger RNA splicing. *Genes Dev.* **8**, 221–233.
11. Ryan, D.E. & Abelson, J. (2002). The conserved central domain of yeast U6 snRNA: importance of U2-U6 helix I-a in spliceosome assembly, *RNA*, **8**, 997–1010.
12. Sun, J. S. & Manley, J. L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian messenger-RNA splicing. *Genes Dev.* **9**, 843–854.
13. Sun, J. S. & Manley, J. L. (1997). The human U6 snRNA intramolecular helix: structural constraints and lack of sequence specificity. *RNA*, **3**, 514–526.
14. Valadkhan, S. & Manley, J. L. (2000). A tertiary interaction detected in a human U2-U6 snRNA complex assembled *in vitro* resembles a genetically proven interaction in yeast. *RNA*, **6**, 206–219.
15. Valadkhan, S. & Manley, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature*, **413**, 701–707.
16. Fast, N. M., Roger, A. J., Richardson, C. A. & Doolittle, W. F. (1998). U2 and U6 snRNA genes in the microsporidian *Nosema locustae*: evidence for a functional spliceosome. *Nucl. Acids Res.* **26**, 3202–3207.
17. Gultyaev, A. P., Franch, T. & Gerdes, K. (2000). Coupled nucleotide covariations reveal dynamic RNA interaction patterns. *RNA*, **6**, 1483–1491.
18. Hilliker, A. K. & Staley, J. P. (2004). Multiple functions for the invariant AGC triad of U6 snRNA. *RNA*, **10**, 921–928.
19. Luukkonen, B. G. M. & Séraphin, B. (1998). A role for U2/U6 helix Ib in 5′ splice site selection. *RNA*, **4**, 915–927.
20. Wolff, T., Menssen, R., Hammel, J. & Bindereif, A. (1994). Splicing function of mammalian U6 small nuclear-RNA: conserved positions in central domain and helix-I are essential during the 1st and 2nd step of pre-messenger RNA splicing. *Proc. Natl Acad. Sci. USA*, **91**, 903–907.
21. Wu, J. A. & Manley, J. L. (1992). Multiple functional domains of human U2 small nuclear-RNA: strengthening conserved stem I can block splicing. *Mol. Cell. Biol.* **12**, 5464–5473.
22. Sashital, D. G., Cornilescu, G. & Butcher, S. E. (2004). U2–U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nature Struct. Mol. Biol.* **11**, 1237–1242.
23. Moore, M. J. & Sharp, P. A. (1993). Evidence for 2 active-sites in the spliceosome provided by stereochemistry of pre-messenger RNA splicing. *Nature*, **365**, 364–368.
24. Jackson, I. J. (1991). A reappraisal of non-consensus messenger-RNA splice sites. *Nucl. Acids Res.* **19**, 3795–3798.
25. Hall, S. L. & Padgett, R. A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with nonconsensus splice sites. *J. Mol. Biol.* **239**, 357–365.
26. Tarn, W. Y. & Steitz, J. A. (1996). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell*, **84**, 801–811.
27. Tarn, W. Y. & Steitz, J. A. (1996). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science*, **273**, 1824–1832.

28. Patel, A. A. & Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. *Nature Rev. (Mol. Cell. Biol.)*, **4**, 960–970.
29. Tarn, W. Y. & Steitz, J. A. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**, 132–137.
30. Nilsen, T. W. (1998). RNA–RNA interactions in nuclear pre-mRNA splicing. In *RNA Structure and Function* (Simons, R. W. & Grunberg-Manago, M., eds), p. 279, Cold Spring Harbor Laboratory Press, New York.
31. Yu, Y. T., Scharl, E. C., Christine, M. S. & Steitz, J. A. (1999). The growing world of small nuclear ribonucleoproteins. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds) 2nd edit., p. 487, Cold Spring Harbor Laboratory Press, New York.
32. Burge, C. B., Tuschl, T. & Sharp, P. A. (1999). Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds) 2nd edit., p. 525, Cold Spring Harbor Laboratory Press, New York.
33. Wu, Q. & Krainer, A. R. (1999). AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol. Cell. Biol.* **19**, 3225–3236.
34. Shukla, G. C. & Padgett, R. A. (1999). Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA*, **5**, 525–538.
35. Padgett, R. A. & Shukla, G. C. (2002). A revised model for U4atac/U6atac snRNA base pairing. *RNA*, **8**, 125–128.
36. Shukla, G. C., Cole, A. J., Dietrich, R. C. & Padgett, R. A. (2002). Domains of human U4atac snRNA required for U12-dependent splicing *in vivo*. *Nucl. Acids Res.* **30**, 4650–4657.
37. Shukla, G. C. & Padgett, R. A. (2002). A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol. Cell.* **9**, 145–150.
38. SantaLucia, J., Jr & Hicks, D. (2004). The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415–440.
39. Mathews, D. H., Burkard, M. E., Freier, S. M., Wyatt, J. R. & Turner, D. H. (1999). Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, **5**, 1458–1469.
40. Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31**, 3406–3415.
41. Dimitrov, R. A. & Zuker, M. (2004). Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.* **87**, 215–226.
42. Rehmsmeier, M., Steffen, P., Höchsmann, M. & Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
43. Andronescu, M., Zhang, Z. & Condon, A. (2005). Secondary structure prediction of interacting RNA molecules. *J. Mol. Biol.* **345**, 987–1001.
44. Zhang, W. B. & Chen, S.-J. (2001). Predicting free energy landscapes for complexes of double-stranded chain molecules. *J. Chem. Phys.* **114**, 4253–4266.
45. Cao, S. & Chen, S.-J. (2005). Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA*, **11**, 1884–1897.
46. Wersig, C. & Bindereif, A. (1990). Conserved domains of human U4 snRNA required for snRNP and spliceosome assembly. *Nucl. Acids Res.* **18**, 6223–6229.
47. Serra, M. J. & Turner, D. H. (1995). Predicting thermodynamic properties of RNA. *Methods Enzymol.* **259**, 242–261.
48. Kierzek, R., Burkard, M. E. & Turner, D. H. (1999). Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, **38**, 14214–14223.
49. Michel, F., Umesono, R. & Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns—a review. *Gene*, **82**, 5–30.
50. Rupert, P. B. & Ferre-D'Amare, A. R. (2001). Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780–786.
51. Butcher, S. E. & Brow, D. A. (2005). Towards understanding the catalytic core structure of the spliceosome. *Biochem. Soc. Trans.* **33**, 447–449.
52. Walter, A. E. & Turner, D. H. (1994). Sequence dependence of stability for coaxial stacking of RNA helices with Watson–Crick base paired interfaces. *Biochemistry*, **33**, 12715–12719.
53. Madhani, H. D. & Guthrie, C. (1994). Randomization-selection analysis of snRNAs *in vivo*: evidence for a tertiary interaction in the spliceosome. *Genes Dev.* **8**, 1071–1086.
54. Ryan, D. E., Kim, C. H., Murray, J. B., Adams, C. J., Stockley, P. G. & Abelson, J. (2004). New tertiary constraints between the RNA components of active yeast spliceosomes: a photo-crosslinking. *RNA*, **10**, 1251–1265.
55. Konarska, M. M. & Query, C. C. (2005). Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev.* **19**, 2255–2260.
56. Donmez, G., Hartmuth, K. & Luhrmann, R. (2004). Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA*, **10**, 1925–1933.
57. Gu, J., Patton, J. R., Shimba, S. & Reddy, R. (1996). Localization of modified nucleotides in *Schizosaccharomyces pombe* spliceosomal small nuclear RNAs: modified nucleotides are clustered in functionally important regions. *RNA*, **2**, 909–918.
58. Patton, J. R., Jacobson, M. R. & Pederson, T. (1994). Pseudouridine formation in U2 small nuclear-RNA. *Proc. Natl Acad. Sci. USA*, **91**, 3324–3328.
59. Frilander, M. J. & Steitz, J. A. (2001). Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol. Cell.* **7**, 217–226.
60. Otake, L. R., Scamborova, P., Hashimoto, C. & Steitz, J. A. (2002). The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol. Cell.* **9**, 439–446.
61. Field, D. J. & Friesen, J. D. (1996). Functionally redundant interactions between U2 and U6 spliceosomal snRNAs. *Genes Dev.* **10**, 489–501.
62. Brow, D. A. & Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**, 213–218.
63. Guthrie, C. & Patterson, B. (1988). Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387–419.
64. Yean, S. L. & Lin, R.-J. (1991). U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction. *Mol. Cell. Biol.* **11**, 5571–5577.
65. Shukla, G. C. & Padgett, R. A. (2001). The intramolecular stem-loop structure of U6 snRNA can functionally replace the U6atac snRNA stem–loop. *RNA*, **7**, 94–105.
66. Chen, S.-J. & Dill, K. A. (2000). RNA folding energy landscapes. *Proc. Natl Acad. Sci. USA*, **97**, 646–651.
67. Bornberg-Bauer, E. & Chan, H. S. (1999). Modeling evolutionary landscapes: mutational stability, top-

ology, and superfunnels in sequence space. *Proc. Natl Acad. Sci. USA*, **96**, 10689–10694.
68. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. (1998). Kinetic intermediates trapped by native interactions in RNA folding. *Science*, **279**, 1943–1946.
69. Pan, J., Thirumalai, D. & Woodson, S. A. (1997). Folding of RNA involves parallel pathways. *J. Mol. Biol.* **273**, 7–13.
70. Tinoco, I., Jr, & Bustamante, C. (1999). How RNA folds *J. Mol. Biol.* **293**, 271–281.
71. SantaLucia, J., Jr (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
72. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, **267**, 1619–1620.
73. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
74. Dill, K. A. (1999). Polymer principles and protein folding. *Protein Sci.* **8**, 1166–1180.
75. McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
76. Chen, S.-J. & Dill, K. A. (1995). Statistical thermodynamics of double-stranded polymer molecules. *J. Chem. Phys.* **103**, 5802–5813.
77. Chen, S.-J. & Dill, K. A. (1998). Theory for the conformational changes of double-stranded chain molecules. *J. Chem. Phys.* **109**, 4602–4616.
78. Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucl. Acids Res.* **31**, 3429–3431.
79. Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C. & Marks, D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1.
80. Lewis, B. P., Snih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
81. Rajewsky, N. & Socci, N. D. (2004). Computational identification of microRNA targets. *Dev. Biol.* **267**, 529–535.
82. Vitiello, D., Pecchia, D. B. & Burke, J. M. (2000). Intracellular ribozyme-catalyzed *trans*-cleavage of RNA monitored by fluorescence resonance energy transfer. *RNA*, **6**, 628–637.
83. Kasai, Y., Shizuku, H., Takagi, Y., Warashina, M. & Taira, K. (2002). Measurements of weak interactions between truncated substrates and a hammerhead ribozyme by competitive kinetic analyses: implications for the design of new and efficient ribozymes with high sequence specificity. *Nucl. Acids Res.* **30**, 2383–2389.
84. Vaish, N. K., Kore, A. R. & Eckstein, F. (1998). Recent developments in the hammerhead ribozyme field. *Nucl. Acids Res.* **26**, 5237–5242.
85. Hormes, R., Homann, M., Oelze, I., Marschall, P., Tabler, M., Eckstein, F. & Sczakiel, G. (1997). The subcellular localization and length of hammerhead ribozymes determine efficacy in human cells. *Nucl. Acids Res.* **25**, 769–775.
86. Schmidt, C., Welz, R. & Muller, S. (2000). RNA double cleavage by a hairpin-derived twin ribozyme. *Nucl. Acids Res.* **28**, 886–894.
87. Barroso-delJesus, A. & Berzal-Herranz, A. (2001). Selection of targets and the most efficient hairpin ribozymes for inactivation of mRNAs using a self-cleaving RNA library. *EMBO Rep.* **2**, 1112–1118.
88. Hampel, K. J. & Burke, J. M. (2001). A conformational change in the "loop E-like" motif of the hairpin ribozyme is coincidental with domain docking and is essential for catalysis. *Biochemistry*, **40**, 3723–3729.
89. Lazarev, D., Puskarz, I. & Breaker, R. R. (2003). Substrate specificity and reaction kinetics of an X-motif ribozyme. *RNA*, **9**, 688–697.
90. Tang, J. & Breaker, R. R. (1998). Mechanism for allosteric inhibition of an ATP-sensitive ribozyme. *Nucl. Acids Res.* **26**, 4214–4221.
91. Leung, S. S. & Koslowsky, D. J. (2001). Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to its cognate gRNA. *RNA*, **7**, 1803–1816.

Edited by D. E. Draper

(Received 13 July 2005; received in revised form 2 December 2005; accepted 3 December 2005)
Available online 21 December 2005