RNA Folding Transitions and Cooperativity

Eivind Tøstesen,[†] Shi-Jie Chen,*[,‡] and Ken A. Dill[§]

Department of Physics 307, Technical University of Denmark, 2800 Lyngby, Denmark, Department of Physics and Astronomy and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received: August 9, 2000; In Final Form: December 11, 2000

Using a new theoretical model, we explore the conformational energy landscapes and the statistical thermodynamics of RNA secondary structure folding. The model is physical, treats the conformational entropies and excluded volume explicitly, and has previously been shown to give reasonable agreement with experimental melting curves. Here we use it to design energy landscapes that can be tested by experiments. The model predicts that even simple RNA's, less than 60 nucleotides long and having only secondary structures, can have remarkably complex and bumpy landscapes that can be altered substantially by small mutations. Moving a GC block around in a sequence of fixed composition can lead from one melting peak to two. A mutation can switch it back to one. Two melting peaks do not imply two simple stem melting events. In one system, we find 5 transitions, each of which can be either 2-state or 1-state. In some sequences we find *triple points*, where 3 macroscopically identifiable species have equal populations at the same temperature. We show designs of RNA's with multiple native states, and conformational switching between them, as functions of mutations or temperature.

1. Introduction

The central mathematical quantity for understanding the folding of an RNA or protein molecule is its *free energy landscape*, $F(\phi)$, the free energy as a function of the n degrees of freedom, $\phi = \phi_1, \phi_2, ..., \phi_n$, such as the backbone bond angles. If there is a single global minimum point on this surface, it defines the native state. Folding algorithms aim to predict that point. But if we are interested in the physical processes of folding — including transition states and intermediates — or in conformational transitions, or native fluctuations, then we need ways to explore the broader set of conformations that constitute the rest of the energy landscape. Understanding fluctuations, which are deviations away from the native conformation, may ultimately be more important than understanding the native state, since it is the fluctuations that are the key to mechanisms involved in binding and function. For example, "induced fit", ligand-induced conformational changes, and the transition states in enzyme mechanisms all involve some perturbation away from the native structure.

Developing realistic models of energy landscapes has been a challenge in computational biology. All-atom models retain atomic detail, but at the expense of sparse conformational sampling. This is like characterizing a mountain range on the basis of a few ski trails. At the opposite extreme are low-resolution models, which explore global principles of landscapes, but cannot treat realistic specific sequences. For example, probably the most extensively understood energy landscapes are for 3-dimensional 27-bead lattice models[1] and 2-dimensional 16-bead models[2] that are often based on simplified monomer alphabets.[3,4]

Recently, a third type of computational model has been developed for biomolecule energy landscapes. It is based on a polymer graph−theoretic approach.[5,6,7] We call it the nested polymer graph theory (NPGT). This approach allows both a high degree of realism in the model and yet complete coverage of the conformational space, for certain classes of chain conformations, such as those in RNA secondary structures. With this approach, we can explore not only the native structures, but also cooperative transitions,[8−11] stable intermediate states, and ensembles of partly folded structures. The NPGT method has been validated against simplified exact model studies,[5,6] and against melting experiments on simple hairpins.[7]

Our aim here is to illuminate some principles of RNA folding through the perspective of this model. Past computational modeling has aimed at rationalizing nature's existing energy landscapes; here we use the model to design new energy landscapes, and we propose how they can be experimentally tested. Our aim here is for theory to motivate experiments, rather than the reverse, and in the process we hope to learn both the strengths and limitations of the theory. We focus here on issues that are subtle or unexpected. For example, we design hairpins that can unzip in either forward or reverse directions or both; molecules that have more than one native state; RNA's that can be "conformational switches"[12] that can toggle between two different conformations as a function of temperature; and mutations that can create deep wells (stable intermediate states) on landscapes that do not otherwise have them.

Mutations that can severely distort landscapes have been found experimentally for proteins.[13] We believe such mutations will be common among RNA molecules. Recent experimental

* Corresponding author. Fax: 573-882-4195. E-mail: ChenShi@missouri.edu.
[†] Technical University of Denmark.
[‡] University of Missouri.
[§] University of California.

10.1021/jp002877q CCC: $20.00
Published on Web 02/06/2001

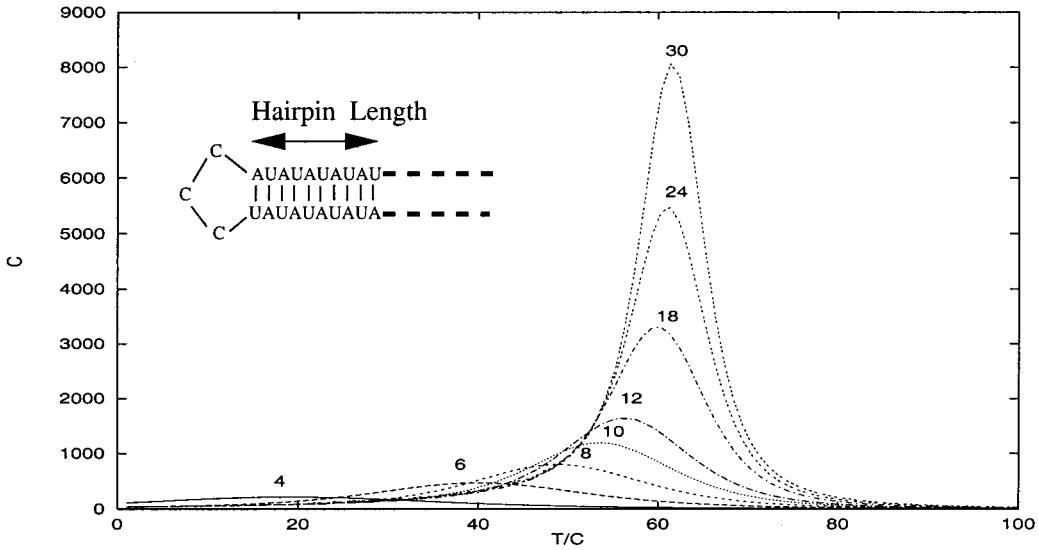

Figure 1. Predicted melting behavior (specific heat) vs chain length for AU hairpins.

advances in RNA folding are beginning to go beyond native structures to shed light more broadly on energy landscapes.[14−24]

2. The Theory

Quantitative details of the theory are given elsewhere;[5,6] here we just give an overview. Any RNA secondary structure can be described as a collection of hairpins, loops, and turns, with varying degrees to which hairpins branch off from other hairpins. The central physical idea behind the NPGT model is that the partition function for any such RNA secondary structure is a product of component partition functions for each loop and bulge and rung (base pair) of the helical hairpin "ladder" of base pairs. A component partition function is comprised of a conformational count of states, which we determine from lattice model enumerations, and a Boltzmann factor describing the energetics of the base interactions. For the latter, we use the "Serra/Turner rules".[25] The full partition function, $Q(T)$, for the whole RNA secondary structure is the sum over all the possible ways of arranging the component structures, that is, all the possible arrangements of base-pairings, branchings, and secondary structures. In NPGT model, different arrangements of the component structures are represented by polymer graphs—the diagrammatic representations of intrachain contacts,[5,6] and the full partition function $Q(T)$ is a sum over all the possible polymer graphs:

$$Q(T) = \sum_{\text{all graphs}} \Omega\, e^{-E/kT} \qquad (1)$$

where Ω is the number of chain conformations having given intrachain contacts of the polymer graph, E is the energy, k is Boltzmann's constant, and T is the temperature. The key idea of the NPGT model is that Ω for a given polymer graph is calculated as a product of the conformational count of each component structure (loops, bulges, and helical base stacks).[5,6] For a polymer graph having N component structures, Ω is given by a product of matrices:

$$U_r S^{(N)} Y S^{(N-1)} \ldots U_c$$

where $U_r = \text{row}(1,1,1,1)$, $U_c = \text{col}(1,1,1,1)$. In the NPGT model, component structures, connected through inlets and outlets, are classified into 4 types according to the inlet and outlet conformations.[5,6] $S^{(i)}$ ($i = 1,2,\ldots,N$) in the above equation is a 4×4 matrix of the ith component structure, and the matrix element $S_{mn}(m, n = 1, 2, 3, 4)$ counts conformations having a type m outlet and type n inlet.[5,6] Y is a 4×4 matrix containing 0's and 1's that ensures that one component structure has the right configuration of "outlet" (say type $m = 1, 2, 3, 4$) to couple to the "inlet" (say type $n = 1, 2, 3, 4$) of the next component structure.

All the thermodynamic properties can be obtained by standard methods from the partition function. For example, the heat capacity $C(T)$ is given by

$$C(T) = \frac{\partial}{\partial T}\left[kT^2 \frac{\partial}{\partial T} ln\, Q(T)\right]$$

The conformation counts used in the model come from lattice enumerations, so to make correspondence with real molecules, we need to make an appropriate mapping between real-chain degrees of freedom and lattice degrees of freedom. We have previously found that a single scaling constant, that is independent of chain length and sequence, is sufficient.[7] This constant multiplies the conformation count from the lattice enumeration to correct for the undercounting of the true chain conformations. We assume the Serra/Turner enthalpies and entropies for base pairing and stacking interactions,[25] we assume those values are temperature independent, and we assume all other nucleotide interaction energies are zero. We assume that E in eq 1 is the sum of the enthalpies for all the base stacks for a particular graph. Of course, these assumptions will be too simple under some conditions, hence the need to subject the model to experimental tests. Although nucleotide interactions depend on salt concentrations, the data of Serra and Turner was taken in 1 M salt, so predictions of our model may not be reliable under other salt conditions. Despite these approximations and limitations, this approach gives a reasonable first approximation to the experimental denaturation curves of several small RNA molecules.[7] On that basis, and in the absence of better methods at present, we believe the NPGT model is a good starting point for predicting RNA folding energy landscapes.

3. Cooperativity and Stability Increases with Hairpin Length

As a reference calculation, we begin with one of the simplest physical properties, but one which has not been measured experimentally yet on RNA, as far as we know. Figure 1 shows

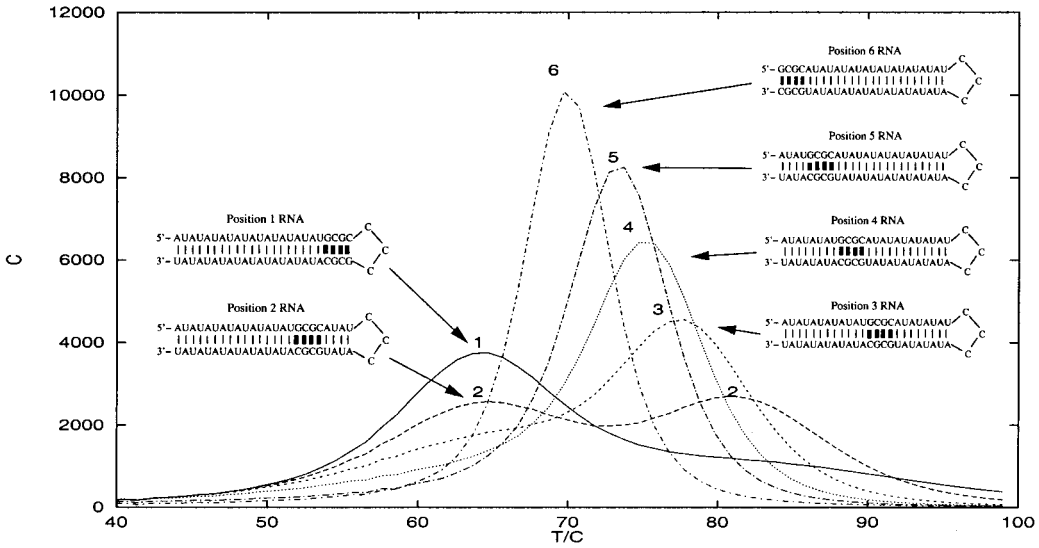

Figure 2. A series of hairpins designed to explore melting proceeding in a "forward" direction from the turn, and in the "reverse" direction from the ends.

predicted melting curves as a function of the stem length for a simple RNA sequence. In this series, all stems are AU pairs and the loops are 3 unpaired C's. We chose C's for the loops to avoid GU pairings, which are weakly stabilizing, according to the energies of Serra and Turner. The number of AU base pairs varies from 4 to 30, as indicated in the figure. The theory predicts that the melting temperature increases with length, approaching a limiting value around $T = 60-70$ °C. The cooperativity of the unzipping transition increases with length. As expected, the area under the melting curve increases linearly with length, since the base interaction free energy scales with the number of AU base pairs.

4. RNA Helix Stability and Cooperativity Depend not just on Base Composition, but also on the Monomer Sequence

Figure 2 shows a series of 6 hairpins, all having 4 GC pairs, 20 AU pairs, and a loop of 3 C's. The only difference among the 6 molecules is the placement of the GC block along the helix. If melting depended only upon the base-pair composition, i.e., the number of AU and GC pairs, then these 6 molecules should all have the same melting behavior. In contrast, the model predicts a very strong sequence dependence, a 20 °C variation in melting temperature, and a switchover from a single peak in the heat capacity curve, to two.

The least stable molecule, which we call "Position-1" RNA, has its GC block adjacent to the loop. In that case, relatively low temperatures melt the AU ends, destabilizing the GC block. The melting temperature $T_m \simeq 65$ °C for Position-1 RNA is close to that for hairpin conformations having only AU base pairs (see Figure 1).

In this series of molecules, as the GC block is moved away from the hairpin loop toward the free ends, the system is predicted to reach a point where it undergoes two transitions. The molecule that has two heat capacity peaks is labeled "Position-2 RNA" in Figure 2. We analyze its energy landscape in detail below. Position-2 RNA molecule melts in two stages: The AU chain ends melt first (i.e., at lowest temperature) at temperature $T_m \simeq 65$ °C, the same melting temperature as for Position-1 RNA. At a higher temperature, the helix in the loop region melts. The hairpin loop AU's are more stable than the chain ends because less entropy is gained when melting the loop than when melting the free ends.

Comparing all 6 sequences shows, interestingly, that moving the GC block moves toward the free ends leads not only to a systematic destabilization of the loop region (the melting temperature decreases), but also to an increasingly cooperative transition (the peak width narrows).

4.1. Population Analysis. The main value of the present theory is that it can relate such macroscopic observables as melting behavior to the microscopic conformations and ensembles that give rise to it. Here we analyze the RNA conformations that are populated at different temperatures. We focus on Position-2 RNA, the sequence that is predicted to have two peaks in its heat capacity curve. Two peaks typically implies a stable intermediate state. What is its structure? To plot the free energy landscapes, or probability distributions, as a function of *all* the degrees of freedom, would be neither possible nor illuminating. Landscapes are surfaces of high dimensionality. Instead, we show 3D plots that reduce this information to a form that is more useful for our present purposes. We take the number of *native contacts* as one useful "order parameter" that characterizes the degree of folding of an RNA molecule. The base pairs that define the native structure are referred to as the *native contacts*. All other intrachain contacts are called *nonnative contacts*. Any conformation of the chain can be characterized by its number of native contacts and its number of nonnative contacts, which might be called (i.e., *right* and *wrong* contacts, respectively, in terms of its native structure): q represents the number of native contacts in the conformation, and q_{NN} represents the number of nonnative (NN) contacts in the conformation.

Our free energy landscapes are plots of free energy F as a function of q and q_{NN}. We compute partition functions which are Boltzmann-weighted sums over all the possible conformations for each given value of q and q_{NN}:

$$Q(q,q_{NN},T) = \sum_E g(q,q_{NN},E)e^{-E/kT} \qquad (2)$$

where $g(q,q_{NN},E)$ is the number of conformations having q native and q_{NN} non-native contacts. From the partition functions we get the free energy:

$$F(q,q_{NN},T) = -kT \ln Q(q,q_{NN},T) \qquad (3)$$

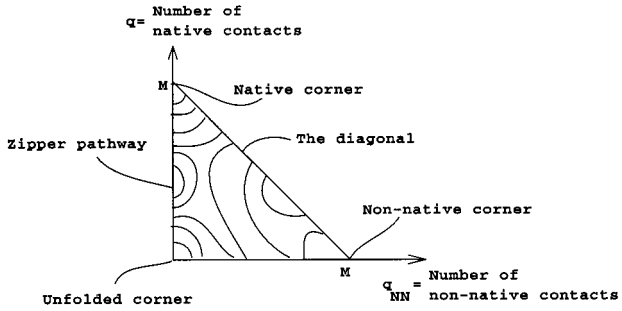

Figure 3. Orienting yourself in the landscape.

We also obtain probability distributions as

$$P(q,q_{NN},T) = \frac{1}{Q(T)} e^{-F(q,q_{NN},T)/kT} \qquad (4)$$

as a function of q and q_{NN}. The probability P carries the same information as the free energy F. But since we find that the plots of probabilities show various features more prominently and are simpler for the eye to interpret than the free energy landscapes, we show landscapes of populations rather than energies in this paper.

Our two variables (q, q_{NN}) are constrained by the requirement that the total number of contacts $q + q_{NN}$ is at most $(N - 3)/2$ where N is the number of monomers in the chain (corresponding to a fully zipped hairpin). Therefore the landscapes we show are restricted to a triangular area of the (q, q_{NN})-plane as shown in Figure 3.

In Figure 3, the unfolded coil conformations with no contacts are at the (q, q_{NN}) = (0, 0) corner of the triangle. The conformations having the highest compactness are along the opposite diagonal. If the native structure has the maximum possible number M of contacts, it would lie in the (q, q_{NN}) = (M, 0) corner of the triangle.

4.2. The Predicted Structure of the Stable Intermediate. Figure 4 shows the dependence on temperature of the conformational populations of Position-2 RNA, the molecule that has two melting peaks and a stable intermediate state.

Here we analyze Figure 4, starting at low temperatures and working toward higher temperatures. At a temperature that is well below both melting peaks (25 °C; Figure 4a), virtually all the molecules populate the native structure (the native conformation is shown in Figure 2). At 60 °C (Figure 4b), just below the lower melting temperature, the native population still dominates but fluctuations become substantial and the ends begin to fray. Only 8 native base pairs remain tightly bound together—the 4 GC pairs and the innermost 4 AU pairs, closest to the turn. At 65 °C, which is the midpoint of the lowest-temperature transition (Figure 4c), the native population is further reduced and is in equilibrium with the well-populated broad ensemble of end-frayed conformations. These end-frayed conformations are the "intermediate state". The broad distribution in Figure 4c implies that there is very little free energy barrier between native and intermediate states. Zero-barrier transitions are sometimes called "one-state" transitions in microscopic systems such as this. In macroscopic systems, zero-barrier transitions are called *critical points*. Thus the transition from native to the end-frayed state resembles a critical point.

Heating further, to 75 °C, between the two melting peaks, leads to the distribution shown in Figure 4d. Now most molecules populate the stable intermediate state, which is end-frayed. At this temperature, far more molecules are end-frayed than are either native or fully denatured. Heating further to 80 °C (Figure 4e) puts the system near the midpoint of its second transition. This one is a "two-state" transition; it has a free energy barrier. In this transition, there are equal populations of end-frayed molecules and fully denatured molecules, but no "intermediate" conformations in between. It might have been expected that the interior 4 AU pairs would melt out before the 4 GC pairs, but they do not: all 8 base pairs melt together. This transition is called two-state because the equilibrium involves two states—one is the ensemble of denatured conformations, and the other is the ensemble of end-frayed conforma-

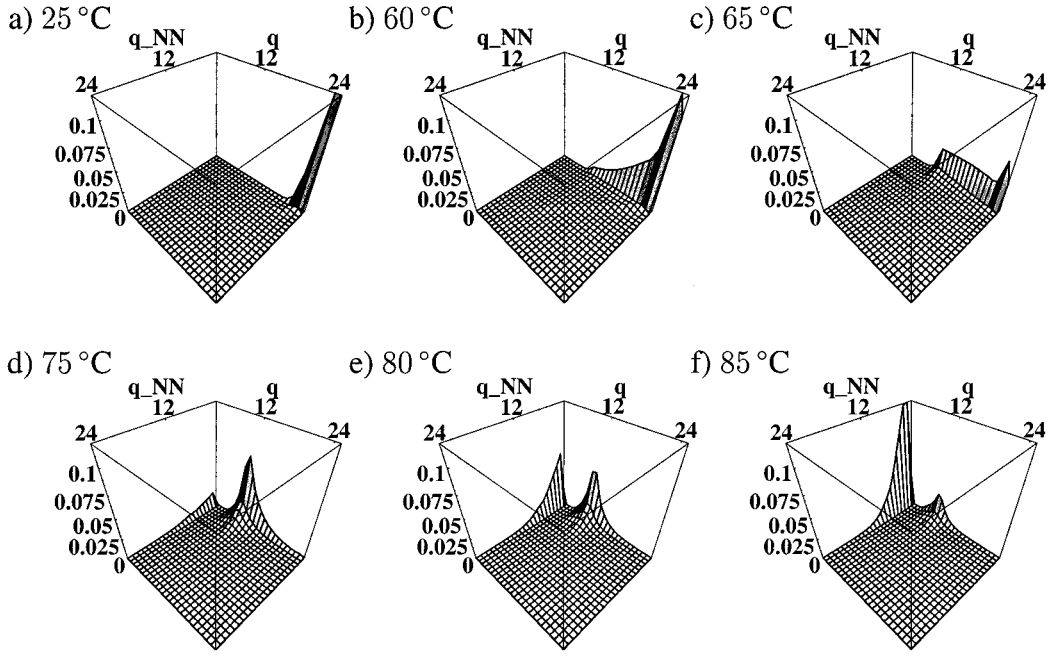

Figure 4. The probability landscapes for Position-2 RNA as a function of temperature. The predominant population is native at low temperatures, denatured at high temperature, and an end-frayed intermediate is populated at a temperature between the two melting peaks. This end-frayed intermediate has the 4 GC base pairs and the 4 innermost AU base pairs (adjacent to the turn); the long AU tails are otherwise in various degrees of unzippering.

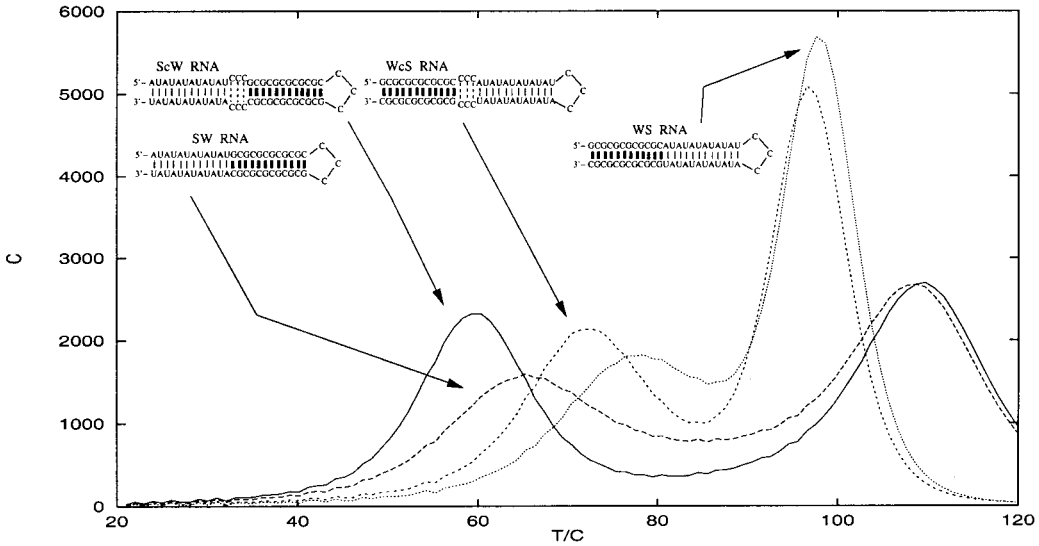

Figure 5. Melting depends on the way loops (hairpin or internal) connect strongly binded strands in one end and weak ones in the other end.

tions that are held together by the interior 8 base pairs. In macroscopic systems, an example of a two-state transition is the boiling of water at 100 °C and 1 atm pressure, i.e., well below its critical point. Under those conditions, there are two states in equilibrium: one is a gas phase, and the other is a liquid phase and none of the water is in a state of intermediate density. In macroscopic systems, this is called a "first-order" transition.

Finally, Figure 4f shows the RNA populations at a temperature above the second melting temperature: now most molecules are fully denatured.

Thus, Figure 4 illustrates how the native structure melts out with temperature in 6 different equilibrium experiments at 6 different temperatures. Note that this is not a description of kinetics, of the time sequence of populations in a single jump experiment; it is a series of equilibrium experiments at different temperatures.

Perhaps the main conclusion is that despite the simplicity of this molecule—it is an RNA molecule that only forms a simple hairpin—it has multiple transitions that are the microscopic equivalents of both a critical point and a first-order phase transition.

These results suggest the complexity that is possible in biomolecule conformational cooperativity. Much of the traditional physics of phase transitions reports on how cooperativity depends on the size of the system. For polymer conformations, the appropriate measure of system size is the number of monomers in the chain. But for biomolecule transitions, the monomer *sequence* is often a much more important determinant of cooperativity than the *chain length*. Hence much of the large literature on polymer cooperativity is inapplicable to biomolecules. Moreover, it is not just sequence effects, per se that determine the complexity of cooperativity. Helix−coil transitions are cooperative.[26,27] But helix−coil processes, because they are driven only by local interactions in the sequence, do not have multiple transitions, of multiple types, or alternative possible native structures. What makes these RNA transitions more complex and interesting, and perhaps a paradigm for folding processes in general, is that they also involve nonlocal contacts and sequence/structure relationships.

4.3. Stabilizing the Intermediates. Can we design sequences to alter the intermediate state? Figure 5 shows melting curves of four hairpins that each have two melting peaks. In this series, each stem has 12 GC pairs and 12 AU pairs, and the hairpin loop is 3 C's. Two of the molecules also have an internal loop, with 3 added C's in each strand. Our notation is SW, WS, ScW, WcS, where S indicates the strong bonds (the GC block), W indicates the weak bonds (the AU block), c indicates an internal loop, and the first letter (S in SW, for example) indicates which block is closest to the hairpin loop and most distant from the chain ends.

Some of the conformational changes that we compute for these RNA's would take place above the boiling point of water, putting them out of reach of simple experiments. We include them here because of the principles they illustrate about conformational entropies. This is a widely used strategy in chemical physics modeling; in this case it is based on extrapolating the Serra/Turner energies to other temperatures, even experimentally inaccessible ones. The main point is that such RNA's principles should also apply to shorter sequences, other compositions, or different pH or ionic strength conditions, but will just be less prominent.

Relative to the previous series of 6 molecules, this series also has pairs of melting peaks corresponding to weak and strong bonds, but the pairs of peaks are now more widely separated in melting temperature. Some of this separation in melting peaks is readily explained by conformational entropy. For example, compare the melting peaks of WcS and ScW. Here is a simple explanation. For each of the molecules the two transitions of native to intermediate, N → I, and intermediate to denatured, I → D, take place at melting temperatures

$$T_{\mathrm{NI}} = \frac{E_{\mathrm{I}} - E_{\mathrm{N}}}{S_{\mathrm{I}} - S_{\mathrm{N}}} \tag{5}$$

$$T_{\mathrm{ID}} = \frac{E_{\mathrm{D}} - E_{\mathrm{I}}}{S_{\mathrm{D}} - S_{\mathrm{I}}} \tag{6}$$

where E_j and S_j are the energies and entropies of states j = N, I, and D. Consider the structures of ScW and WcS; all quantities, except S_{I}, are roughly the same for these two molecules (see the intermediate structures of them in Figure 9), so the difference in melting peak separation between these two molecules is explained by S_{I} alone. The intermediate state of ScW has more conformational entropy than the intermediate state of WcS since there is more conformational freedom of loose ends than of a

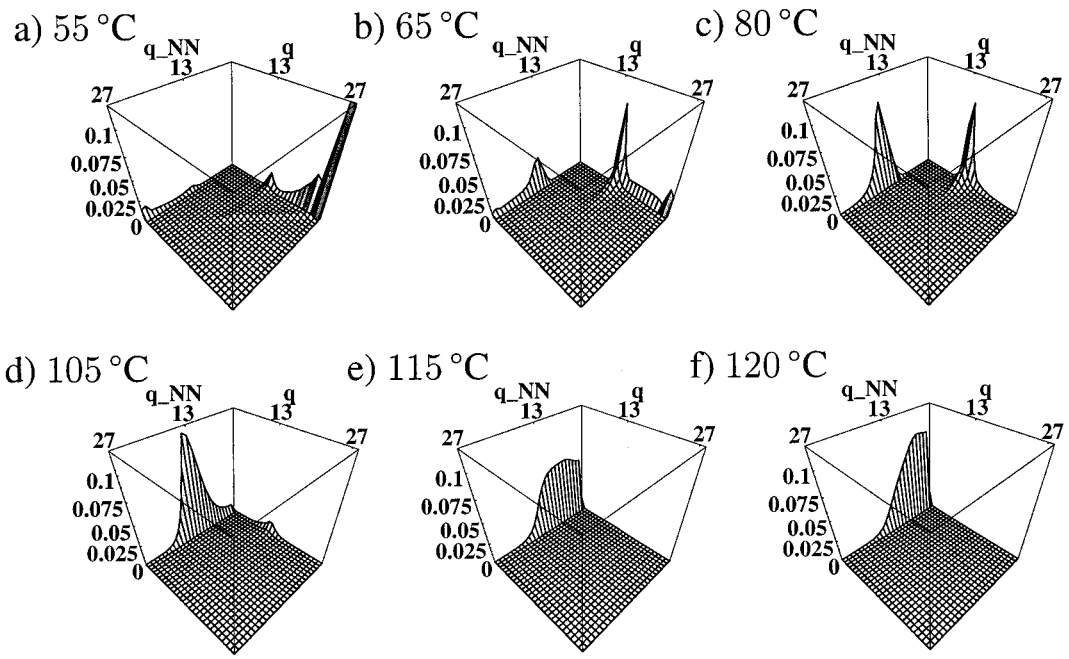

Figure 6. (a) The majority of the ensemble at 55 °C is in the native minimum at $(q, q_{NN}) = (27,0)$ with an adjacent tail in the distribution of partly unzipping down to 20 native contacts. Small fractions also populate the intermediate minimum, a $q = 12$ valley and alternative structure on the non-native axis for $q_{NN} \geq 12$. (b) At 65 °C the majority in the probability plot has shifted to the intermediate (half-zipped) minimum at $(q, q_{NN}) = (12,0)$. A small population remains in the native minimum or along the "zipper line" in between. On the non-native axis we see a "mirror image", where a small peak has developed in the molten minimum at $q_{NN} \approx 12$. (c) A slippage transition takes place around this temperature (80 °C) in which the ensemble shifts from the intermediate state to the molten state. "Slippage" refers to the helix strands sliding in opposite directions. (d) At 105 °C the majority is centered around the molten minimum at $(q, q_{NN}) = (12,0)$, with only a small peak remaining in the intermediate. Fluctuations in this molten state form a distribution that is much broader than the intermediate peak at 65 °C, hence the name "molten". (e) A one-state transition from the molten state to the denatured state takes place around this temperature. "One-state" refers to the single maximum of the probability moving from 12 to 1 non-native contacts. (f) In the denatured state most of the ensemble is random coil with a small number of non-native contacts.

constrained loop. Putting the S block at the open end increases T_{NI}, and decreases T_{ID}. The same argument applies to SW vs WS.

Next we explore the energy landscapes of ScW vs temperature. The main conclusion here is that observing two melting peaks does not always imply two simple independent stem melting events, even when the peaks are well separated. The underlying microscopic events can be more complex.

4.3.1. RNA Secondary Structures Are Predicted To Have Molten-Globule-Like States. ScW and Position-2 RNA would seem to be virtually identical sequence designs, with the same types of blocks in the same arrangements, but differing only in that ScW has a longer GC block than Position-2 RNA has. Both ScW and Position-2 RNA have two melting peaks.

However, their cooperativities are completely different. Position-2 RNA has two transitions, N → I → D. But ScW has *three* transitions, N → I → M → D. N, I ,and D correspond to the native structure, a half-zipped, end-frayed structure and random coil structures, respectively, just as in Position-2 RNA. M is an additional *molten state*[28,29] that appears in ScW but not in Position-2 RNA.

Figure 6 shows the population analysis of ScW as a function of temperature. Figure 6a,b shows that the low-temperature transition N → I is very similar to the low-temperature transition of Position-2 RNA: the intermediate structure of ScW RNA has the 12 native GC base pairs in the strong stem formed while the AU tail is frayed (see Figure 9). But in ScW, this transition is more 2-state-like, with a bigger free energy barrier, than it is in Position-2 RNA. That is, there is some partial fraying in equilibrium with the native and fully end-frayed states, but the partially frayed conformations are less populated in ScW than in Position-2 RNA.

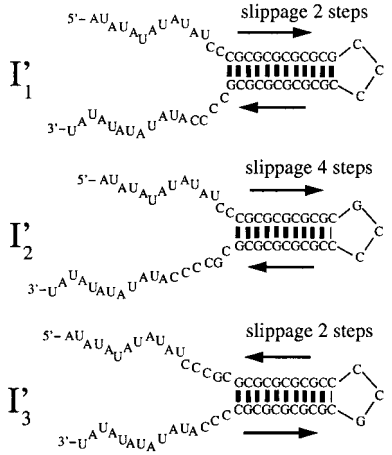

Figure 7. Members of the molten state ensemble.

Heating ScW to a higher temperature leads to a completely different structure, M, having about 12 non-native contacts, on average, and no native contacts. This does not consist of a native piece and an end-frayed piece. Rather, because the weak and the strong stems each have alternating sequences, both stems can slip to be an out-of-register zipped conformation. Such helices will have the same type of base pairs, GC's of AU's as before, but fewer of them. This might be called a shifting or slipping transition. M is more like a broken zipper, while I is more like a half-zipped zipper.

Figure 6c−f show the transitions I → M → D involving the shifted state. The strong stem is shifted and the weak stem is unzipped. The slippage transition I → M is a rearrangement of the strong helix, not an unfolding. Figure 6 shows that I → M

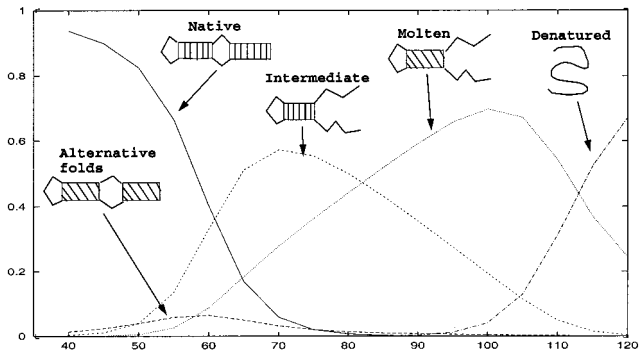

Figure 8. Fractional populations of the states of ScW RNA vs temperature. Angled bars represent slipped hairpins, and straight bars represent hairpins with native base pairs.

TABLE 1: Definitions of States (i.e., regions covered by the probability peaks) in Terms of Numbers of Native and Non-native Contacts

state	definition	
native	$18 \le q \le 27$	
alternative folds		$18 \le q_{NN} \le 27$
intermediate	$8 \le q \le 16$	$0 \le q_{NN} \le 6$
molten	$0 \le q \le 6$	$8 \le q_{NN} \le 16$
denatured	$0 \le q + q_{NN} \le 6$	

is a two-state transition (since there is a low population of states that are intermediate between I and M). The M → D transition is one-state, i.e., barrierless. Unlike the native and intermediate states, the shifted state is broad and not dominated by a single specific structure, since there are many different slipped conformations. The by a single specific structure, since there are many different slipped conformations. The- molten state ensemble M is mainly comprised of 3 equally probable slipped structures I'_1, I'_2, I'_3, shown in Figure 7.

Recently, another model has also predicted the phase diagram of such a shifted state in RNA.[30]

Thus, molecule ScW is predicted to have two 2-state transitions and one 1-state transition. A most remarkable aspect of this molecule is that at around 65 °C, there are three stable states in equilibrium, separated by two free energy barriers. The nearest correspondence in macroscopic systems is a triple point, such as in water at 0.0075 °C and 4.58 Torr, where liquid water, steam, and ice are all in equilibrium.

Figure 6a,b shows that molecule ScW also has a fifth stable state that appears at low temperature, so we call it an *alternate fold*, but is in very low population. It involves slippage in both the S and W stems, relative to the native structure. The alternate fold is highly compact but highly non-native, with up to $q_{NN} = 27$ non-native contacts.

Each population, except for the native state, represents a fairly broad ensemble. But because each ensemble is distinct from the other, we can give them labels as macroscopic states. For the ScW RNA molecule we have found five such states: *native, the alternative fold, the end-frayed intermediate, the slipped intermediate,* and *denatured*, defined by regions of the (q, q_{NN})-plane described in Table 1.

Figure 8 shows the populations of these five states vs temperature. The temperatures at which there are crossovers between populations are transition temperatures. There are three transitions: N → I → M → D. Transitions are not perfectly sharp, however, and at most temperatures, there are multiple states in equilibrium. In short, there are 3 microscopic transitions in this system, but they are manifested as only two peaks in the melting curve. From this figure it is also clear, that the system never reaches exactly a triple point: there is no point at which 3 populations are exactly equal. However, the sequence WcS, described below, has such a triple point.

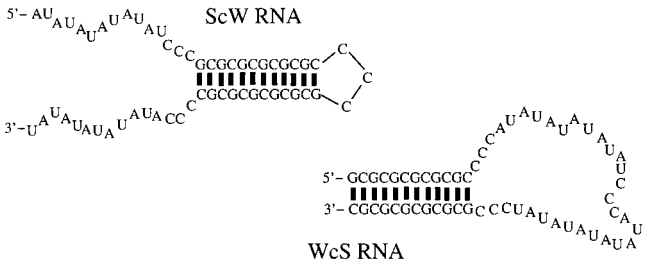

Figure 9. Intermediate structures of ScW and WcS.

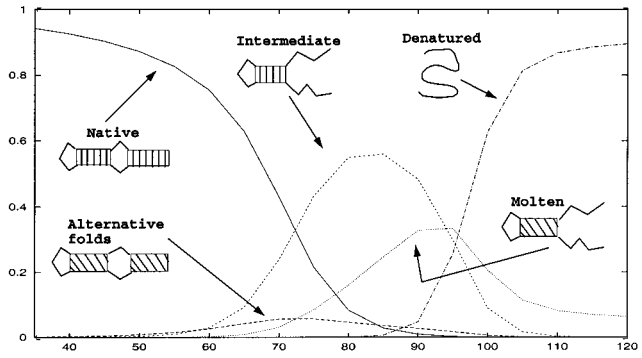

Figure 10. Fractional populations of states of WcS RNA molecule plotted against temperature. Most probable members are shown as cartoons (as in Figure 8).

4.3.2. Unzipping from the Inside. We now compare the melting of molecule WcS with ScW. The only difference between WcS and ScW is that their strong and weak stems are swapped. For WcS we expect the intermediate state might be a large floppy loop with ends clamped together, rather than the reverse in ScW, a clamped loop with frayed ends; see Figure 9.

Figure 10 shows the population analysis for WcS, indicating that this is indeed the structure of the intermediate state, and that WcS has the same 5 states as ScW. The main difference is that the slipped intermediate M is less populated in WcS than in ScW. As noted before, the intermediate state in WcS has less conformational entropy because a loop having n monomers has less conformational freedom than two free ends each having $n/2$ monomers. In WcS, only a part of the population unfolds to D via M. The other part goes directly from I to D. This transition to D happens around the higher transition temperature $T_m = 95$ °C, where the three states I, M, and D have approximately equal populations, as in a triple point.

4.4. Parallel vs Serial Stems: Designing Sequences To Have Three Melting Events with Two Stable Intermediates. Here, our aim was to design a sequence (wsm.p RNA in Figure 11) having three melting peaks. Figure 11 shows that if we put a strong stem (GC pairs) in parallel with a weak stem (AU pairs) and in parallel with a stem of medium strength (alternating AU and GC pairs), the theory predicts there will be 3 peaks in the melting curves, although they are not very distinct. If the three stems are put in series instead (smw.s RNA in Figure 11), only two peaks appear. We can increase the stability and cooperativity of the last two peaks of wsm.p RNA by increasing the length of its medium or strong stems ("modified RNA" in Figure 11).

Figure 12 shows the population analysis of the wsm.p RNA to illuminate the populations that underly the three peaks in the heat capacity. Throughout this melting, up to 9 or 10 populations appear in the landscapes, but the underlying

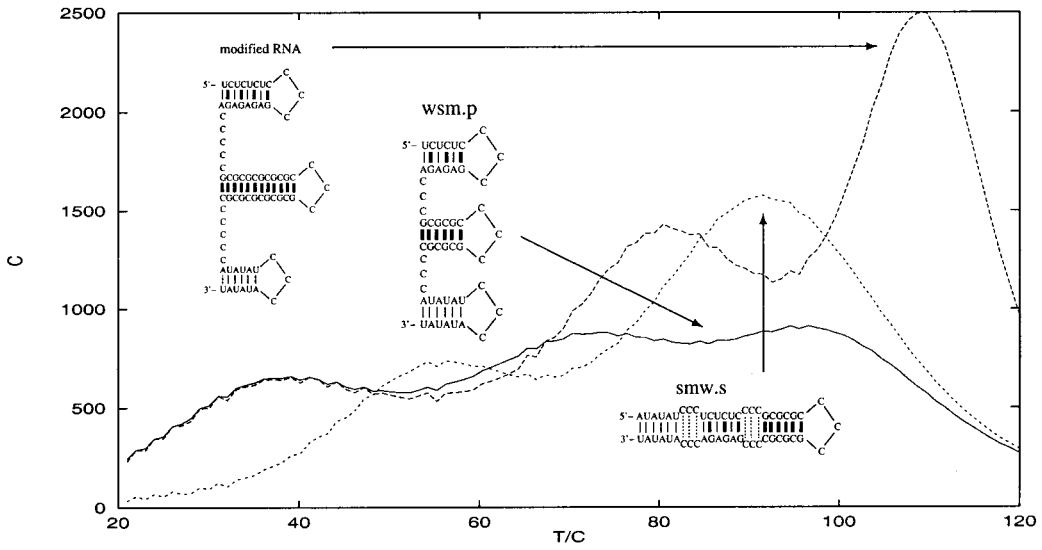

Figure 11. RNA's designed to have three melting peaks.

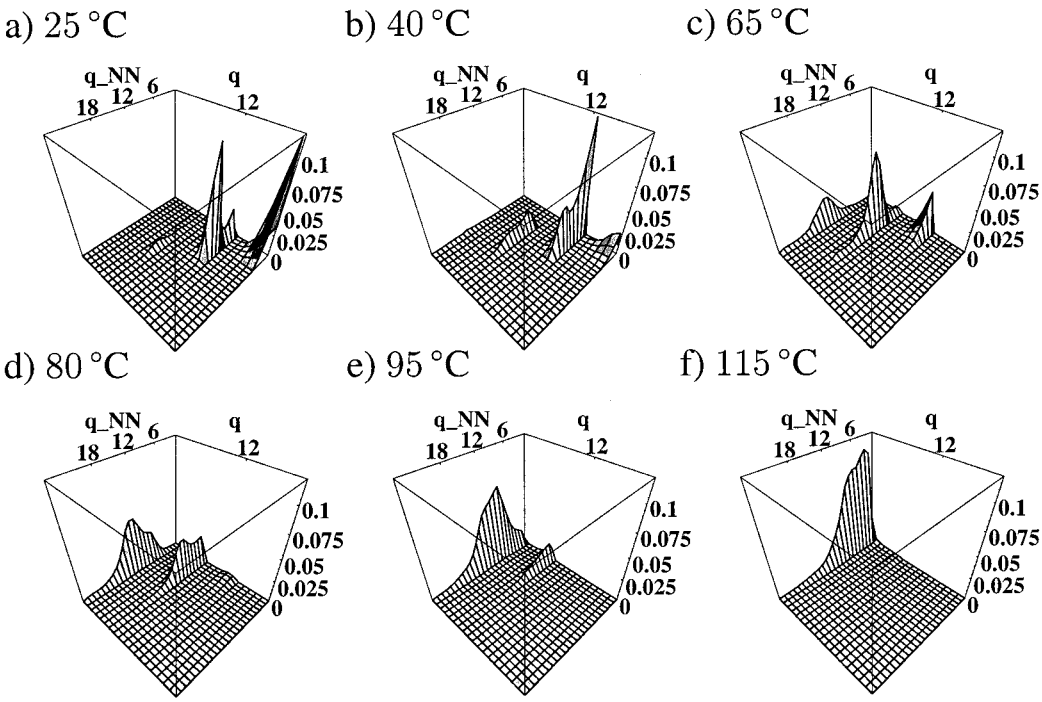

Figure 12. Probability plots for the wsm.p RNA. (a) At 25 °C the slippage transition in the weak stem is in progress, and has just begun also in the medium stem. (b) At 40 °C the weak stem is almost melted, but some part of the population remains in the molten state. (c) At 65 °C the medium stem has almost completed its slippage transition to a molten state, while the strong stem has just begun its slippage transition. (d) At 80 °C there are four peaks corresponding to two transitions: the molten-to-denatured transition of the medium stem and the slippage transition of the strong stem. (e) At 95 °C the medium stem is almost melted and the strong stem is in its molten state. (f) At 115 °C the strong stem is more than halfway through its molten-to-denatured transition.

microscopic events are simple to interpret. The three stems melt in the following order: weak, medium, and then strong, with increasing temperature. But each stem does not melt through simple transition from native to denatured. Instead the melting of each stem passes through a slipped state having non-native contacts, just as the GC region of the ScW molecule does. These several ensembles overlap considerably, so the population plot does not show distinct species.

5. Designing Mutations That Change Intermediate States

An important goal in protein folding has been to find and understand situations in which a very small change in a sequence (such as a single mutation) can cause a large change in the global shape of an energy landscape. One example has been mutations that can switch folding kinetics from two-state (single exponential) to multi-state (multiple exponentials).[31] Here we develop examples in which mutations can affect the stable intermediates of RNA secondary structures. Again we note that we are looking here at equilibrium intermediates, and not at kinetics.

Figure 13 shows that simply interchanging the C and the G at positions 32 and 33 in Position-2 RNA causes the double-peak melting of the wild-type molecule to become a single melting transition with a lower melting temperature. Figure 14 shows that the mutant RNA molecule has very simple cooperativity: a single 2-state melting transition from the native to denatured state. This is a case in which the stable intermediate

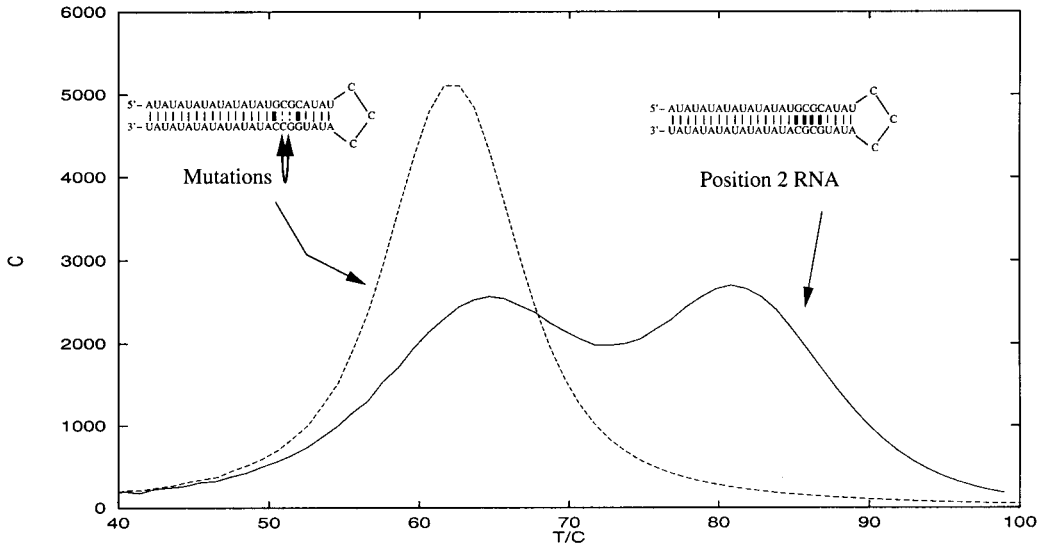

Figure 13. The two-peaked heat capacity curve of the Position-2 RNA molecule compared with its mutated version that has only one peak, a lower stability and a lower area under the curve (enthalpy change).

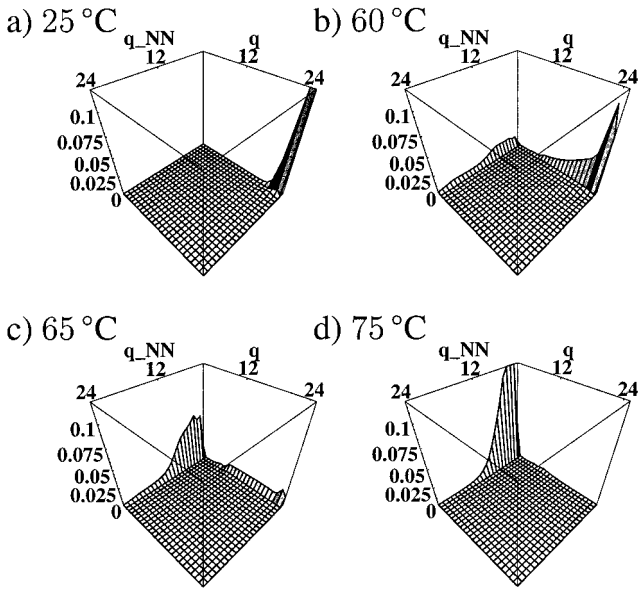

Figure 14. (a) At 25 °C almost the whole ensemble is in the native minimum. (b) 60 °C is still below the (new) melting temperature, the population peaks in the native minimum and has its end-fraying fluctuations. However, large parts of the F landscape have been altered by the mutation. A new broad population has emerged on the whole q_{NN} axis, with no native contacts. (c) 65 °C is above the new melting temperature so most of the ensemble is random coil. Here the unmutated molecule, cf. Figure 4c, was midway in its $N \rightarrow I$ transition, exhibiting a range of structures between these two states. Here we see a small fraction of the ensemble on the native axis populating the same transition structures. So this area's landscape shape is less affected by the mutation. (d) The transition structures have disappeared and all is random coil. Here the unmutated molecule was in its intermediate state, cf. Figure 4d. Actually, the F landscape of the mutant also has a shallow local minimum at $q = 9$ (not shown), but too high to be populated.

state in the wild-type has been abolished by the mutation. The "GC clamp" near the hairpin loop has been weakened in the mutant so much that it does not have a stable end-frayed intermediate state.

6. Designing Multiple Native States and Conformational Switches

In this section, we describe RNA molecules that are predicted to have *two* native states that are very different from each other: they have no base pairs in common. We call the two native states the "V fold" and the "W fold" because of their resemblance to those letters of the alphabet (see Figure 15).

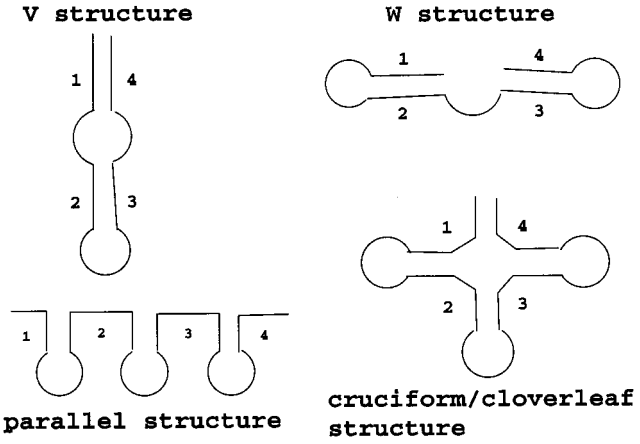

Figure 15. Definition of possible folds of the chain, the two native structures V and W, and two examples of mixed structures that have both V contacts and W contacts. Chain segments 1−4 are indicated as noncurved lines and the hinge segments as curved lines.

Each sequence has four chain segments (1, 2, 3, and 4) of equal length that can pair up to form helices. The segments are separated by three short strands that act as "hinges". Each hinge is three C's. The sequences of the four helix parts can be chosen relatively freely, but with some constraints: Segments must be pairwise complementary in the following way $1 \sim 2 \sim 3 \sim 4 \sim 1$, where $\sim$ means "matches perfectly with". This implies that sequences 1 and 3 are identical, and 2 and 4 are identical. Segments could be chosen to be all identical and self-complementary, but this would make other folds possible, for example pairing of 1 with 3.

If the four segments are long enough, there will be a possibility of forming mixed structures, such as a cruciform/cloverleaf or parallel folds (Figure 15). We describe four RNA molecules: one sequence has an "irregular" or heterogeneous sequence to minimize alternative base pairings, and one has a sequence that is more regular and homogeneous. The second pair of sequences is derived from the first pair by an identical mutation to each of the wild-types. The mutation that stabilizes the W fold destabilizes the V fold, and this creates a conformational switch.

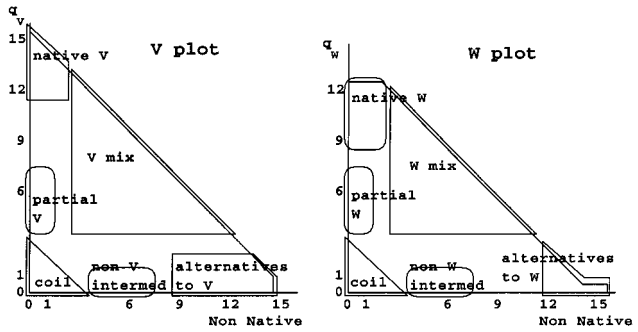

Figure 16. Regions in the landscapes covered by the defined states.

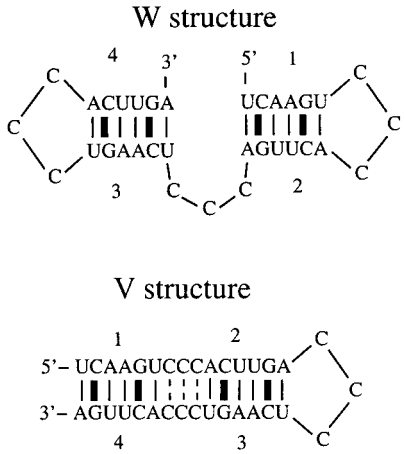

Figure 17. The sequence and native target structures of VW1 RNA.

For molecules having two native states, plots based on independent variables (q, q_{NN}) are no longer sufficient, because we must specify *which* native state defines the contacts labeled by q. Plots having more independent variables are not practical, so instead we show below two plots, one in which q counts native contacts in V and one in which q counts native contacts in W. To distinguish these cases, we use the notation q_V or q_W, respectively.

We plot the populations of eleven possible classes of conformations that are relevant for these types of molecules: native V, native W, alternatives to V, alternatives to W, partial V, partial W, V mix, W mix, non-V intermediates, non-W intermediates, and coil. These states are defined in terms of regions in the (q_V, q_{NN_V})-plane and the (q_W, q_{NN_W})-plane in Table 2, see also Figure 16. Note that there are two *equivalent* definitions of the coil state. Of course, there are ambiguities with two landscapes: At some temperatures, for example, the alternatives-to-V population will be largely equivalent to native-W, for example.

6.1. A Sequence Designed To Have Two Native States. VW1 RNA is our designation for an irregular 33-mer that was designed to have the two target native structures shown in Figure 17.

Chain segments 1 and 3 have the sequence UCAAGU, and segments 2 and 4 have the complementary sequence ACUUGA (remember the segments are antiparallel when they are paired). The melting curve (in Figure 19) has a single peak at T_m = 62 °C.

In the V- and W-landscapes (not shown) there are two free energy wells corresponding to the target native structures. They have nearly equal depth and size, but the native-V minimum is a little deeper. This small difference has a great effect on the populations; Figure 18 shows the populations of the two native structures and coil. Below the melting temperature both native structures are populated, but the V population is larger than the W population, as we expected from the energy landscape plots.

TABLE 2: Definitions of Structural States in Terms of Numbers of Contacts

state	definition	
native V	$12 \le q_V \le 15$,	$0 \le q_{NN_V} \le 2$
alternatives to W	$12 \le q_{NN_W} \le 15$,	$0 \le q_W \le 2$
partial V	$4 \le q_V \le 7$,	$0 \le q_{NN_V} \le 1$
non W intermediates	$4 \le q_{NN_W} \le 7$,	$0 \le q_W \le 1$
V mix	$4 \le q_V \le 15$,	$3 \le q_{NN_V} \le 15$
coil	$0 \le q_V + q_{NN_V} \le 3$	
native W	$9 \le q_W \le 12$,	$0 \le q_{NN_W} \le 2$
alternatives to V	$9 \le q_{NN_V} \le 15$,	$0 \le q_V \le 2$
partial W	$4 \le q_W \le 7$,	$0 \le q_{NN_W} \le 1$
non V intermediates	$4 \le q_{NN_V} \le 7$,	$0 \le q_V \le 1$
W mix	$4 \le q_W \le 12$,	$3 \le q_{NN_W} \le 15$
coil	$0 \le q_W + q_{NN_W} \le 3$	

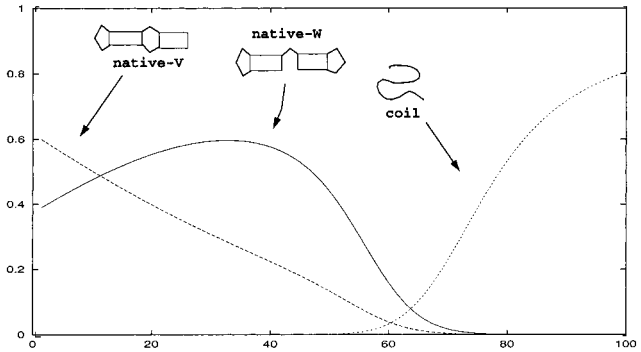

Figure 18. Populations of the structures of the VW1 RNA molecule plotted against temperature.

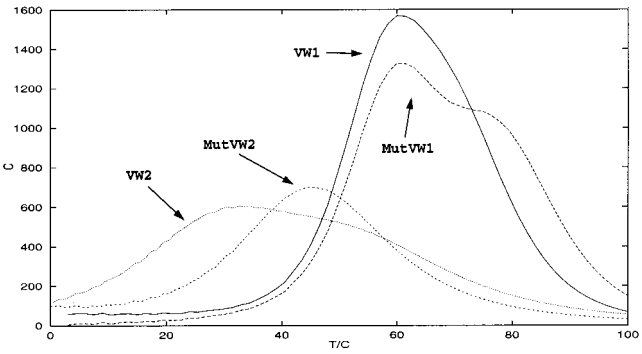

Figure 19. The melting curves of four RNA's with two native states.

In the melting transition the molecule goes from its native structures to random coil which is defined here as all conformations that do not have stable base stacks.

6.2. Switching the Native Structure and Intermediates by Mutation. Now we consider a mutant of VW1. Two bases of a pair are changed, but complementarity is retained. Chain segments 1, 2, 3, and 4 have the sequences UCACGU, ACGUGA, UCAAGU, and ACUUGA, respectively. Figure 20 shows the V and W native structures indicating the mutation sites; the mutation replaces an AU pair with a CG pair in the W structure and replaces an AU with a CU and an UA with a GA in the V structure. The mutation is designed to deepen the free energy well of W, and raise the free energy well of V to favor W at low T.

Figure 21 shows how the populations change with temperature. At low temperatures the W structure dominates the ensemble in the mutant, while in the wild-type it coexisted with the V structure, but V has been completely eradicated in the mutant (compare Figure 18).

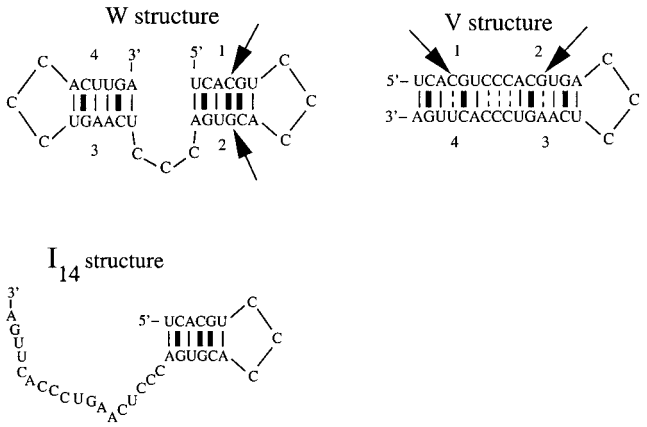

Figure 20. The MutVW1 RNA with mutation sites (arrows), native structures (top), and an intermediate structure (bottom).

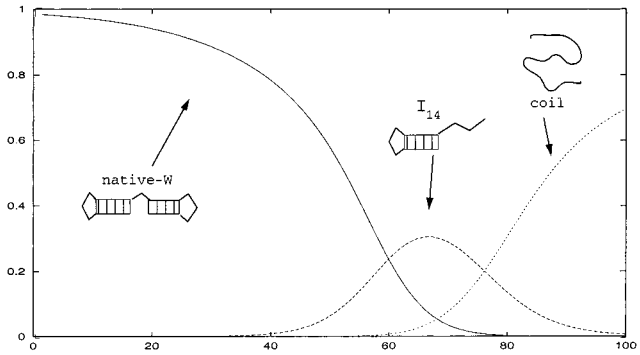

Figure 21. Populations of the structures of the MutVW1 RNA molecule plotted against temperature.

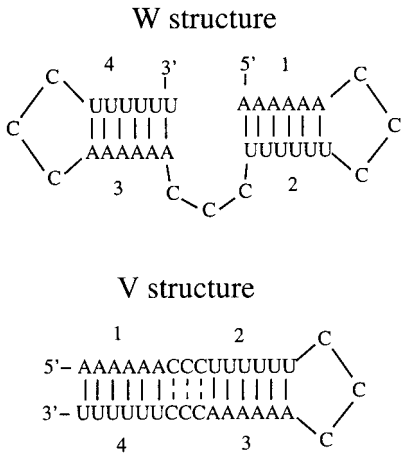

Figure 22. The sequence and native structures of VW2 RNA.

The melting temperature is $T_m \approx 62$ °C, cf. Figure 19. The molecule melts through an intermediate structure I_{14} (see Figure 20). The first transition, W $\rightarrow$ I_{14}, happens just below 60 °C corresponding to the peak of the melting curve, cf. Figure 19. The second transition at 75 °C takes the ensemble from I_{14} to coil.

6.3. Slippage in the Helices of V-W Structures. Sequence VW2 is another 33-mer. It has AAAAAA on segments 1 and 3 and the complementary UUUUUU on segments 2 and 4.

Figure 23 shows the melting curve for VW2 and its mutant. It has one peak in the range $0 \leq T \leq 100$ °C: $T_m \simeq 45$ °C. Figure 24 shows that native-W is stable at the low temperatures. The conformational ensemble is dominated by native-W at $T \leq -20$ °C (not shown in Figure 23). But at -20 °C, the system switches to native-V. The system melts at the big heat

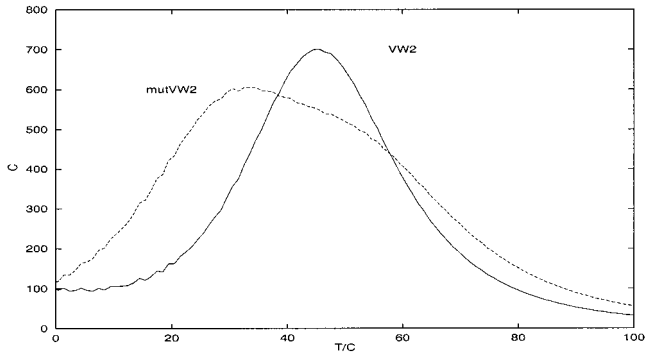

Figure 23. The melting curves of VW2 RNA and its mutated version MutVW2 RNA.

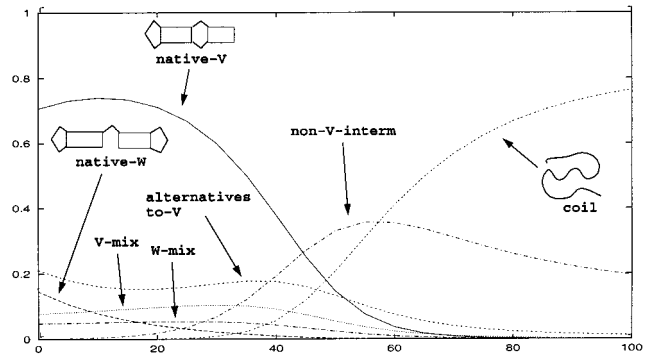

Figure 24. Populations of the states of the VW2 RNA molecule plotted against temperature.

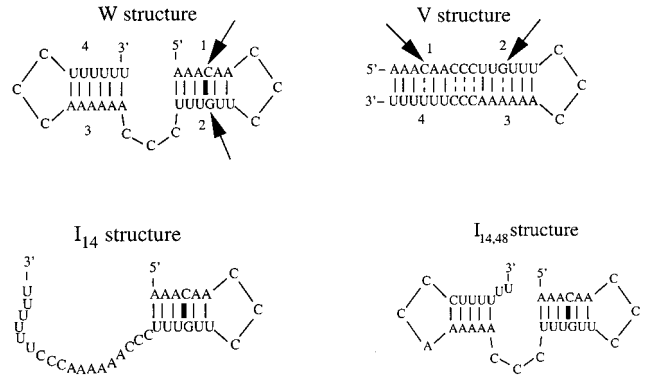

Figure 25. The MutVW2 RNA with mutation sites (arrows), its native, and stable intermediate structures.

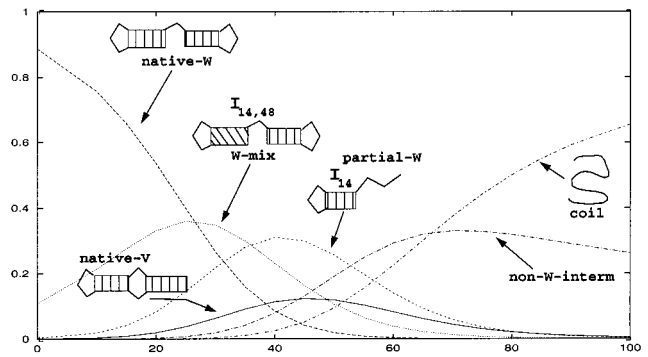

Figure 26. Fractional populations of the states of the MutVW2 RNA molecule plotted against temperature.

capacity peak at $T_m \simeq 45$ °C, where most of the ensemble goes from native-V in a transition to the coil state.

Because of its homogeneity, this sequence has many possibilities for non-native base pairing that come from sliding the

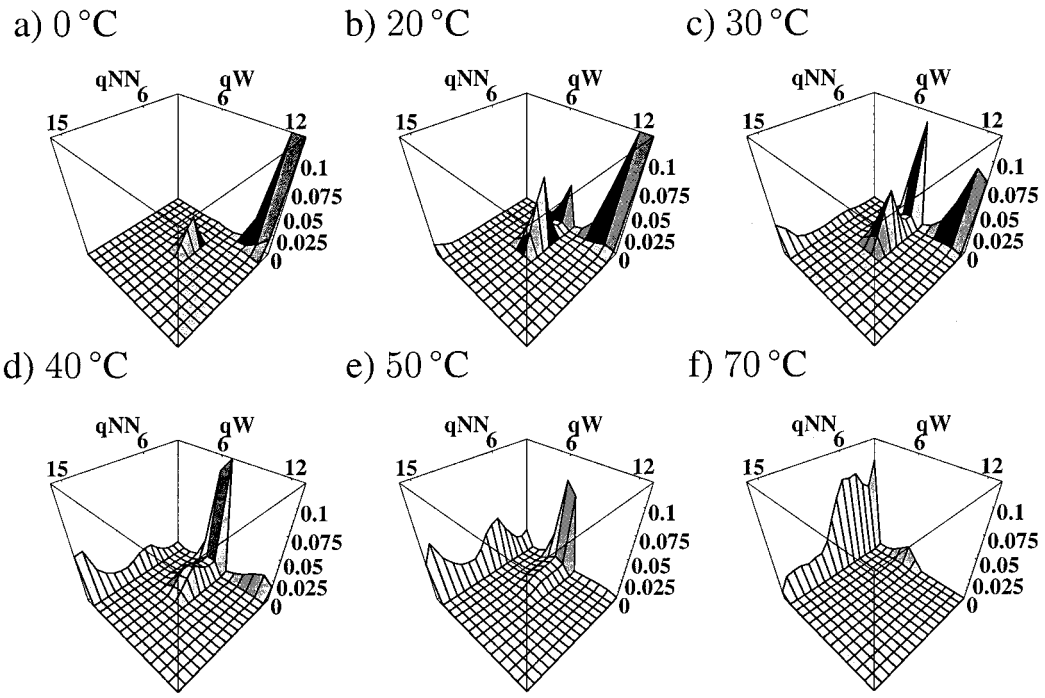

Figure 27. Temperature series of probabilities of the MutVW2 RNA. These are W plots, that show the complexity and the multiple minima. (a) At 0 °C the native-W state dominates. (b) At 20 °C the church-shaped peaks start growing. (c) At 30 °C the W-mix state reaches its maximum. (d) At 40 °C the partial-W state reaches its maximum. (e) At 50 °C the native-V state reaches its maximum. (f) At 70 °C the non-W-intermediates state reaches its maximum.

helix strands relative to each other. This gives rise to *new* wells in the energy landscapes, populated by W-mix and V-mix states. Their contacts are half native half non-native, and such structures are formed when one of the two helices in native V or W slides, while the other helix retains its native base-pairing. Although the W-mix and V-mix states only attract 5−10%, they are able to coexist with the native states all the way down to low temperatures. This is possible because the corresponding structures are energetically favorable, not entropically (many contacts). They are not just intermediate products of the melting process.

For this molecule there is also some off-pathway intermediate structure having between 4 and 11 contacts. The populations for this structure are 10−40%, and can be seen from the curves of alternatives-to-V and non-V-intermediates. These curves have maxima between 30 °C and 60 °C. These off-pathway structures are formed through slippage, and they possibly behave like the molten state described in Figure 6 for ScW RNA.

6.4. Multiple Transitions. Now we consider a mutant of VW2. We introduce the same double mutation as we did in section 6.2 to VW1, to stabilize the W fold relative to V.

The melting temperature defined by a heat capacity maximum, see Figure 23, is $T_m = 33$ °C. Segments 1, 2, 3, and 4 in MutVW2 have the sequence AAACAA, UUGUUU, AAAAAA, and UUUUUU, respectively, see Figure 25.

We expect the mutation to have a bigger effect here than in MutVW1 RNA since we introduce a strong GC contact among *only* weak AU contacts. Figure 26 shows the melting scenario in terms of population curves. Native W is the dominant state at temperatures up to around $T = 20$ °C. It coexists, however, with the W-mix state down to −20 °C. The structure $I_{14,48}$ (Figure 25) constitutes half of the W-mix population.

The system melts through a remarkable series of four states, before the coil state becomes dominant at T higher than 80 °C. These four states are the following: W-mix which reaches its maximum population at 25 °C; partial-W which is dominant at 40 °C, with structure I_{14} in Figure 25; native-V which reaches its maximum at 50 °C, with structure V in Figure 25; and finally non-W-intermediates which is populated with around 30% at 70 °C.

We see that the mutation has a strong effect in stabilizing every structure that contains the helix in the 5′-end of the W structure.

The native-V state act as an intermediate melting state together with partially unfolded structures such as I_{14}, despite the fact that it is the fully zipped V structure. So ordered structures, with little entropy, can also be produced during melting.

Figure 27 shows the W probability landscapes vs temperature, indicating the multiple populations and the various degrees to which there are intermediate states populated between them. This shows remarkable complexity for such simple molecules.

Summary

We have applied a new theoretical method to the prediction of the energy landscapes of RNA molecules having only secondary structures. The model shows that RNA folding can be exceedingly complex and sequence-dependent. Small changes can cause large changes in the shapes of landscapes. A simple short RNA hairpin can undergo 2 or 3 transitions, sometimes of different types (1-state, like critical points, or 2-states, like phase transitions), and even have triple points, where 3 different conformations have equal populations at the same temperature. There can be multiple native states, and switching between them can be driven by either mutations or temperature. These predictions are made in advance of experiments, in the hope that experiments will be developed to test them.

References and Notes

(1) Shakhnovich, E.; Gutin, A. *J. Chem. Phys.* **1990**, *93*, 5967−5971.
(2) Lau, K. F.; Dill, K. A. *Macromolecules* **1989**, *22*, 3986−3997.

(3) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(4) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phy. Chem.* **1997**, *48*, 545−600.
(5) Chen, S.-J.; Dill, K. A. *J. Chem. Phys.* **1995**, *103*, 5802−5813.
(6) Chen, S.-J.; Dill, K. A. *J. Chem. Phys.*, **1998**, *109*, 4602−4616.
(7) Chen, S.-J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646−651.
(8) Galzitskaya, O. V.; Finkelstein, A. V. *Folding Design* **1998**, *3*, 69−78.
(9) Gutin, A. M.; Galzitskaya, O. V. *Biophysics* **1993**, *38*, 77−84, and 85−89.
(10) McCaskill, J. S. *Biopolymers* **1990**, *29*, 1105.
(11) Hofacker, I. L.; et al. *Monatsh. Chem.* **1994**, *125*, 167−188.
(12) LeCuyer, K. A.; Crothers, D. M. *Biochemistry* **1993**, *32*, 5301−5311.
(13) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. *Nat. Struct. Biol.* **1994**, *1*, 149−156.
(14) Repsilber, D.; Wiese, S.; Rachen, M.; Schroder, A. W.; Riesner, D.; Steger, G. *RNA* **1999**, *5*, 574−584.
(15) Sclavi, B.; Sullivan, M.; Chance, M. R.; Brenowitz, M.; Woodson, S. A. *Science* **1998**, *279*, 1940−1943.
(16) Treiber, D. K.; Rook, M. S.; Zarrinkar, P. P.; Williamson, J. R. *Science* **1998**, *279*, 1943−1946.
(17) Batey, R. T.; Doudna, J. A. *Nat. Struct. Biol.* **1998**, *5*, 337−340.
(18) Pan, T.; Sosnick, T. R. *Nat. Struct. Biol.* **1997**, *4*, 931−938.
(19) Pan, J.; Woodson, S. A. *J. Mol. Biol.* **1998**, *280*, 597−609.
(20) Herschlag, D. *J. Biol. Chem.* **1995**, *270*, 20871−20874.
(21) Rook, M. S.; Treiber, D. K.; Williamson, J. R. *J. Mol. Biol.* **1998**, *281*, 609−620.
(22) Narlikar, G. J.; Herschlag, D. *Nat. Struct. Biol.* **1996**, *3*, 701−710.
(23) Wu, M.; Tinoco, I., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11555−11560.
(24) Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11506−11508.
(25) Serra, M. J.; Turner, D. H. *Methods Enzymol.* **1995**, *259*, 242−261.
(26) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.
(27) Poland, D. C.; Scheraga, H. A. *Theory of the helix-coil transition*; Academic Press: New York, 1970.
(28) Woodson, S. A. *Nat. Struct. Biol.* **2000**, *7*, 349−352.
(29) Russell, R.; Millett, I. S.; Doniach, S.; Herschlag, D. *Nat. Struct. Biol.* **2000**, *7*, 367−370.
(30) Bundschuureh, R.; Hwa, T. *Phy. Rev. Lett.* **1999**, *83*, 1479−1482.
(31) Callender, R. H.; Dyer, R. B.; Gilmanshin, R.; Woodruff, W. H. *Annu. Rev. Phy. Chem.* **1998**, *49*, 173−202.