Published in final edited form as:

Nucleic Acids Mol Biol. 2012 ; 27: 185–212. doi:10.1007/978-3-642-25740-7_10.

Statistical mechanical modeling of RNA folding: from free energy landscape to tertiary structural prediction

Song CAO and Shi-Jie CHEN*

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

Song CAO: caos@missouri.edu; Shi-Jie CHEN: chenshi@missouri.edu

Abstract

In spite of the success of computational methods for predicting RNA secondary structure, the problem of predicting RNA tertiary structure folding remains. Low-resolution structural models show promise as they allow for rigorous statistical mechanical computation for the conformational entropies, free energies, and the coarse-grained structures of tertiary folds. Molecular dynamics refinement of coarse-grained structures leads to all-atom 3D structures. Modeling based on statistical mechanics principles also has the unique advantage of predicting the full free energy landscape, including local minima and the global free energy minimum. The energy landscapes combined with the 3D structures form the basis for quantitative predictions of RNA functions. In this chapter, we present an overview of statistical mechanical models for RNA folding and then focus on a recently developed RNA statistical mechanical model -- the **Vfold** model. The main emphasis is placed on the physics underpinning the models, the computational strategies, and the connections to RNA biology.

10.1 Introduction

RNA 3D structure, folding stability and kinetics underlie RNA functions. The 3D crystal structures of rRNAs and tRNAs have led to detailed mechanisms of protein synthesis in the ribosome machinery (Ban et al. 2000; Wimberly et al. 2000; Yusupov et al. 2001). Recent findings regarding the slow folding kinetics of self-splicing introns have revealed how their enzymatic activities arise from their global 3D folds (Hougland et al. 2005; Laederach et al. 2007; Pan and Woodson 1998; Waldsich and Pyle 2008; Woodson 2000; Zarrinkar and Williamson 1994). Theoretical and experimental analyses point to a close correlation between the efficacy of microRNA in gene regulation and the 3D structure and folding stability (Long et al. 2007; Kertesz et al. 2007) of the microRNA/target complex. RNA functions highlight the biological significance of RNA folding and the need for a predictive model for RNA folding.

Existing RNA folding theories mainly focus on secondary structures (Lu et al. 2006; Mathews et al. 2006; McCaskill 1990; SantaLucia and Turner 1997; Zuker 2003). However, RNA functions often involve structures and structural changes at the *tertiary* structural level.

* Author to whom correspondence should be addressed; chenshi@missouri.edu.

Phylogenetic modeling (Major et al. 1993; Massire et al. 1998; SantaLucia et al. 2004) as well as de novo methods (Das and Baker 2007; Das et al. 2007; Jonikas et al. 2009; Parisien and Major 2008; SantaLucia 2007; Shapiro et al. 2007; Tyagi and Mathews 2007) combined with atomic computations (Major et al. 1993; Masquida and Westhof 2006; Mathews et al. 2006) and experimental constraints (Deigan et al. 2009; Jonikas et al. 2009) have shown success in predicting RNA 3D structures. However, RNA function is determined not only by the minimum free energy state of the RNA, but also by its folding stability and the potentially large conformational changes it can undergo. Understanding RNA function requires models that predict the full free energy landscape.

Recent developments in statistical mechanical modeling of RNA folding have led to successes in predicting RNA structures, folding stabilities and folding kinetics for structures with increasing complexity. The models provide quantitative predictions and novel insights for a variety of experiments and RNA functions such as programmed ribosomal frameshifting (Cao and Chen 2009), mRNA splicing (Cao and Chen 2006a), and microRNA gene regulation (Kertesz et al. 2007; Long et al. 2007). Despite the success of this approach, several key issues remain. These issues include the computation of the entropy for RNA tertiary folds and the extraction of the energy/entropy parameters for non-canonical tertiary interactions from thermodynamic data and known structures. The primary focus of this chapter is the application of methods based on statistical mechanics to predict RNA 3D structures and folding energy landscapes and to gain quantitative understanding of RNA functions.

10.2 Overview of computational models for RNA folding

An RNA structure, defined by the nested Watson-Crick base pairs and the tertiary contacts contained in the 3D structure, can be conveniently represented by a polymer graph (Fig. 1). Such a graph (= a 2D structure) usually corresponds to many 3D conformations due to the flexible conformations of the single-stranded regions. Following Chastain and Tinoco (Chastain and Tinoco 1991), tertiary and secondary (2D) structures can be classified as the polymer graphs with and without cross-links, respectively.

Chemical and enzymatic reagents (Ehresmann et al. 1987) are highly effective structural probes for nucleic acids because the reactivity of a nucleotide can be sensitive to its local conformation and interactions, including base pairing and stacking, which are reflected in its solvent accessibility. Structure-probing experiments based on chemical and enzymatic reagents, such as the Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) analysis (Watts et al. 2009) and synchrotron-generated hydroxyl radical footprinting (Petri and Brenowitz 1997), give direct information about base pairing and local structures. The experimental data provide useful input as structural constraints for the computational modeling of complete 3D structures. In parallel with these the experimental developments, *de novo* computational modeling of RNA folding shows continuous improvements in the accuracy of the predictions of RNA structures, including the structures for long RNA sequences. Table 1 shows a list of the computational models for RNA structure prediction from single sequence input. For structure predictions from sequence

homology, see references provided in these citations: (Mathews and Turner 2002; Hofacker et al. 2002).

10.2.1 Secondary structures

RNA secondary structures contain no cross-links (non-nested interactions) in the corresponding graphs and thus permit use of efficient dynamic programming algorithms for conformational enumeration. Structural prediction algorithms based on free energy minimization (Mathews et al. 1999; Nussinov and Jacobson 1980; Williams and Tinoco 1986; Zuker 1989) can predict secondary structures, starting from a single sequence, with about 70% accuracy. Another type of approach to structural prediction is based on calculating the statistical mechanical partition function, which is an average over the conformational ensemble (Hofacker 2003; McCaskill 1990). The strategy is to determine the stable structures from the Boltzmann ensemble-averaged base pairing probabilities over all the possible base pairs. In 2003, Ding et al. developed a statistical sampling algorithm (Sfold) (Ding and Lawrence 2003) to predict RNA secondary structure. In the algorithm, 1000 structures are sampled based on the Boltzmann distribution. The cluster centroids of the 1000 structures give the predicted structures. The statistical sampling algorithm is found to give a better prediction than the free energy minimization method (Ding 2006).

While the above methods employ the same empirical thermodynamic parameters (the Turner rules) for secondary structural elements based on the nearest neighbour base pair interaction model, other models use knowledge-based scoring functions. For instance, the CONTRfold (Do et al. 2006) model uses the energy parameters derived from a training set and the MC-Fold (Parisien and Major 2008) model uses a scoring function that represents the probability of selecting a certain nucleotide cyclic motif (NCM) for the given sequence. The NCMs in the MC-Fold include the lone-pair loops and the double-stranded internal/bulge loops, which are extracted from known PDB structures. Benchmark tests show that CONTRfold and MC-Fold programs give better predictions than Mfold (Do et al. 2006, Parisien and Major 2008).

10.2.2 H-type pseudoknots --- free energy models

A general pseudoknot consists of two helical stems, S_1 and S_2 , and three loops, L_1 , L_2 and L_3 (see Fig. 2). Pseudoknots, which are neglected in most software for secondary structure prediction, play important structural and functional roles in many biochemical processes, such as viral replication (Brierley et al. 2007; Brierley et al. 2008; Draper 1990; Gesteland and Atkins 1996; Giedroc et al. 2000; Giedroc and Cornish 2009; Staple and Butcher 2005), human telomerase RNA activity (Chen and Greider 2005; Comolli et al. 2002; Qiao and Cech 2008; Shefer et al. 2007; Theimer and Feigon 2006) and metabolite-sensing riboswitches (Kang et al. 2009; Klein et al. 2009; Spitale et al. 2009).

The free energy for a pseudoknot is equal to the sum of the free energies for the helical stems and loops: $G_{pk} = G_{S_1} + G_{S_2} + G_{CS} - T S_{loops}$, where G_{S_1} and G_{S_2} are the free energies of stems S_1 and S_2 and G_{CS} is the coaxial stacking energy between stems S_1 and S_2 . While the helix free energies can be evaluated from the nearest-neighbor model based on the empirical thermodynamic parameters for base stacks, to determine the loop entropy S_{loops} requires a physical model.

From statistical mechanics, $S_{loops} = -k_B \ln \Omega_{Coil}/\Omega$, where Ω is the number of the 3D conformations of the loops and Ω_{Coil} is the number of the corresponding coil conformations. The evaluation of the conformational entropy S_{loops} is intrinsically a problem embedded in 3D space. The difficulty in the evaluation of the pseudoknot loop entropy comes from the conformational correlation between the loop and the helix. Specifically, the viability of a loop conformation is subject to the presence of the nearby helix due to the loop-helix excluded volume interaction and the end-to-end distance constraint of the loop set by the length of the helix. The presence of the nearby helix reduces the accessible space of the loop configuration, and thus decreases the number of the viable loop conformations.

The most frequently occurring pseudoknots in natural RNAs are the canonical H-type pseudoknots having a very short (single nucleotide) or completely absent inter-helix loop, L_3 . For the canonical H-type pseudoknot, the helix stems S_1 and S_2 have a strong tendency to coaxially stack on each other to form a quasi-continuous helix. Applying polymer physics theory (Fisher 1966; Jacobson and Stockmayer 1950; Poland and Scheraga 1966), Gultyaev et al. proposed the following expressions for the loop entropy (Gultyaev et al. 1999):

$$\begin{aligned}\Delta S_{L_1} &= A_{major}(S_2) + 1.75k_B \ln(1 + N - N_{\min major}(S_2)), \\ \Delta S_{L_2} &= A_{minor}(S_1) + 1.75k_B \ln(1 + N - N_{\min minor}(S_1)),\end{aligned}$$

where $N_{\min major}(S_2)$ and $N_{\min minor}(S_1)$ are the shortest allowed lengths for L_1 and L_2 , respectively. The *ad hoc* fitting of known pseudoknots with the requirement that the pseudoknot be more stable than its hairpin components yielded estimates for the entropy parameters (Gultyaev et al. 1999).

In a canonical H-type pseudoknot, loops L_1 and L_2 span the major (narrow and deep) and the minor (shallow and wide) grooves of helices S_2 and S_1 , respectively. Therefore, the two loops are highly asymmetric (Aalberts and Hodas 2005). Considering the loop asymmetry, Aalberts and Hodas (Aalberts and Hodas 2005) used the Gaussian chain approximation to derive the end-to-end distance distribution (between D and $D+d$) for an N -nt loop:

$$P_G(D, N) = 4\pi D^2 d \left(\frac{3}{2\pi N a^2} \right)^{3/2} e^{-\frac{3D^2}{2Na^2}}$$

where $d = 0.1 \text{ \AA}$ and $a = 6.2 \text{ \AA}$. The total loop entropy is

$$\Delta S = k_B \ln[P_G(D_{L_1}, L_1 + 1)P_G(D_{L_2}, L_2 + 1)]$$

where D_{L_1} and D_{L_2} are the end-to-end distance for a L_1 -nt loop L_1 and L_2 -nt loop L_2 , respectively. The end-to-end distances for L_1 and L_2 are determined by stems S_2 and S_1 , respectively.

Based on the Gaussian chain approximation, Isambert and Siggia derived the loop entropy, S , for a general three-loop (non-canonical) pseudoknot (Isambert and Siggia 2000, Isambert 2009):

$$\Delta S = k_B \ln \left(\alpha^2 \frac{e^{-A_1 S_1^2 - A_2 S_2^2} e^{2A_3 S_1 S_2} - e^{-2A_3 S_1 S_2}}{D^{3/2} 4A_3 S_1 S_2} \right)$$

where $D = L_1 L_2 + L_1 L_3 + L_2 L_3$, $A_1 = 3(L_1 + L_2)/2abD$, $A_2 = 3(L_2 + L_3)/2abD$, $A_3 = 3(L_3/2abD$ and $S_{1,2} = \sqrt{d^2 \sin^2(\pi n_{1,2}/n_p) + h^2(n_{1,2}/n_p)^2}$. In the calculation, $a = 6 \text{ \AA}$, $b = 1.5 \text{ nm}$, $n_p = 11$, $\alpha = 0.0068$, $d = 4a$, $h = 5a$ and $n_{1,2}$ are the numbers of the base pairs in stems S_1 and S_2 , respectively.

10.2.3 Pseudoknots --- structure prediction

The early computational methods for RNA pseudoknot prediction (Gultyaev et al. 1995) were based on the genetic algorithm (GA). These methods, such as the STAR (Gultyaev et al. 1995) and the MPGAfold (Shapiro and Wu 1997) models, predict the structures with the optimal kinetic accessibility instead of the ones with the lowest free energies. Other pseudoknot prediction methods based on stochastic simulations, such as the ILM (Ruan et al. 2004) and the Hotknots (Ren et al. 2005) models, can give low free energy structures. However, due to the nature of the stochastic conformational sampling, the predicted structure is not guaranteed to have the lowest free energy.

In 1999, Rivas and Eddy developed a dynamic programming method (pknotsRE) to predict pseudoknot structure (Rivas and Eddy 1999). Unlike the genetic algorithm and the other heuristic algorithms, the pknotsRE program, which uses a highly simplified energy function, is guaranteed to find the lowest energy pseudoknot. Later, with a more advanced pseudoknot energy model, Dirks and Pierce developed a partition function method (NUPACK) to predict pseudoknot structures (Dirks and Pierce 2003). In 2004, Reeder and Giegerich developed a new algorithm pknots-RG (Reeder and Giegerich 2004), which yields improved prediction than the original pknotsRE algorithm.

All the above algorithms use simplified nonphysical entropy parameters for pseudoknot loops. In 2006, Cao and Chen developed a physics-based pseudoknot prediction model based on a low-resolution structural representation, which is described below (Vfold, Cao and Chen 2006b). Benchmark tests indicate that the Vfold-based approach gives much improved predictions for pseudoknots compared to other models (Cao and Chen 2009).

10.2.4 RNA/RNA complexes

Functional RNAs often form complexes with RNA cofactors to perform catalytic and regulatory functions in a variety of RNA machineries, including ribozymes (Andronescu et al. 2005), spliceosomes (Staley and Guthrie 1998) and miRNA-argonaut complexes (Bartel 2009). Early computational models for RNA/RNA complexes, such as Hyther (Peyret et al. 1999), OligoWalk (Mathews et al. 1999), RNAhybrid (Rehmsmeier et al. 2004) and DINAMelt (Dimitrov and Zuker 2004), can give the structures and folding thermodynamics,

such as the melting curves and binding affinities, for the simple, Watson-Crick paired RNA complexes. However, these models cannot treat the formation of intra-molecular base pairs in the binding process and thus cannot treat the interplay between the intra- and inter-molecular base pairing, which are known to be critical for many RNA catalytic and regulatory reactions.

In 2005, Andronescu et al. developed the PairFold program (Andronescu et al. 2005), which can explicitly account for both intra- and the inter-molecular base pairs. Tests on seventeen experimentally validated structures show an average correct accuracy 79%. The PairFold program is based on free energy minimization, thus does not predict thermodynamic stabilities, which are determined by the properties of the complete free energy landscape. In 2006, Cao and Chen (Cao and Chen 2006a) applied the Vfold model to predict RNA/RNA complexes based on the partition function method. The Vfold model predicts in addition to the native structure, all the local minima on the free energy landscape (i.e. meta-stable states) as well as thermodynamic properties such as melting curves. By using of a physical model for conformational sampling to obtain the entropy, as well as properly including non-Watson-Crick base pairs in the conformational ensemble, the Vfold model gives improved predictions. Later, based on partition function calculations, Mückstein et al. developed the RNAup algorithm (Mückstein et al. 2006) to compute the base “un-pairing” probability. The application of fundamental theory to the analysis of RNAi-target association led to more reliable predictions for the correlation between the RNAi efficiency and the RNAi-target binding energy. In 2007, Dirks et al. (Dirks et al. 2007) established a new partition function-based theory (NUPACK). A unique feature of the theory is its ability to treat multiple (> 2) nucleic acid strands.

Most of the above mentioned folding programs are restricted to structures without pseudo-knotted folds. Although PairFold can treat pseudo-knotted complexes, but not with high accuracy (Andronescu et al. 2005). In 2006, based on a number of heuristic approaches to the energy models, Alan et al. (Alan et al. 2006) applied the minimum free energy algorithm to search for the native structure of the pseudo-knotted complexes including the kissing loop complexes. Later, Chitsaz et al. (Chitsaz et al. 2009) and Huang et al. (Huang et al. 2009) used the partition function-based algorithm to calculate the structures and folding stabilities of pseudo-knotted complexes. Test of the algorithm by Chitsaz et al. showed that the algorithm can correctly predict the thermodynamic properties for RNA/RNA complexes such as the OxyS/fhlA complex (Chitsaz et al. 2009).

10.3 RNA tertiary structural folding --- from 2D low-resolution to 3D all-atom structures

The partition function stands at the center of statistical mechanical modeling. The partition function, Q , of an RNA molecule is the Boltzmann sum over all the possible structures:

$$Q = \sum_s e^{-\Delta G_s / k_B T} \quad (10.1)$$

where s denotes an enumeration over all possible 2D structures (polymer graphs; see Fig. 1), G_s is the energy of s and $k_B = 1.99\text{cal/K}$ is the Boltzmann constant. The calculation of the partition function contains two key ingredients: sampling of all the possible conformations Σ_s and accurate evaluation of the free energy $G_s (= H - T S)$ of each 2D structure s . Here H and S are the enthalpy and entropy for the given 2D structure s .

The empirical enthalpy and entropy parameters for base stacks and loops (Turner rules) form the foundation for RNA folding free energy prediction at the secondary structural level. However, even for simple secondary structures, the answers to many biologically significant questions require information that goes beyond these parameters. For example, the stability of a hairpin or internal loop is an average over many loop conformations which may involve a variety of sequence-dependent intra-loop contacts. The intra-loop contacts dramatically reduce the loop entropy. To understand and predict the loop entropy and stability for a given sequence, we must dissect the loop entropy for different loop structures with different intra-loop contacts. Empirical thermodynamic parameters such as Turner rules cannot give such entropies. What we need is a theory to calculate the entropy.

Furthermore, most existing RNA folding prediction algorithms are unable to account for the effect of the cross-linked (i.e., tertiary) contacts. One of the challenges comes from the entropy evaluation for tertiary folds. The success of the energy and entropy parameters for the secondary structure models relies on the additive nearest-neighbor (NN) model. For the conformational entropies, the NN model assumes that the entropy for a secondary structure is equal to the sum of the entropies of the subunits (loops, base stacks). Therefore, a parameter database for the different types of subunits would suffice for the calculation of the total free energy. However, the additivity rule for secondary structures is doomed to fail for tertiary folds. This is because the tertiary contacts (cross-links) between the different secondary structural motifs (helices, loops) cause inter-dependence between the (distant) motifs. As a result, even if we knew the entropies and free energies of the individual structural subunits, we would still be unable to predict the entropy and the free energy of a tertiary structure. Therefore, a meaningful database for the tertiary energy parameters, such as a list of the entropy parameters for loops, must consider the influence of other subunits. This, in fact, makes the experimental determination of the parameters impossible. What we need rather is a first principles model. The recently developed model, called “Vfold”, is such a model.

10.3.1 The Vfold model

While predictions of the structure and full free energy landscape may not be possible at the high resolution atomic scale level, what is well within reach is to parse the complexity into two parts: To use low-resolution models to account for the complete conformational ensemble by treating the atomic details implicitly, and then to construct atomistic 3D (native and alternative) structures from the low-resolution models. Given the huge conformational space available to an RNA molecule, such a multi-scale approach has several distinctive advantages:

1. At the low-resolution level, many experimental questions are not concerned with the specific locations of the hydrogen or nitrogen atoms; instead, they are far more

concerned with the global fold, backbone flexibility, and the potential for large structural rearrangements, as investigated by small angle X-ray or neutron scattering and other low-resolution techniques (Chauhan et al. 2005; Deigan et al. 2009; Gherghe et al. 2009; Russell et al. 2002).

2. The key issue in prediction of the tertiary structural folding concerns the entropy of the global fold, which is largely determined by low-resolution properties such as the excluded volume and the chain connectivity effects. The reduced complexity of the low-resolution model allows us to maintain the rigor in physical principles when accounting for these properties. As a result, use of a low-resolution model enables first principles calculations for chain entropy, free energy, and the full free energy landscape for any given sequence.
3. The low-resolution structure provides a useful scaffold for the final all-atom folding model through structural refinements.
4. The predicted structure will provide highly needed guidance for experiments. For example, in NMR structural determinations of RNA, a severe limitation is that sequential resonance assignments rely heavily on Nuclear Overhauser Effect (NOE) data to establish connectivities (“NOE walks”), which often requires several months of data collection and analysis. The information on the nucleotide spatial proximity from the predicted (low-resolution) structure can provide useful constraints for enhancement in the efficiency and accuracy of resonance assignments.

Vfold is a recently developed low-resolution model based on the virtual bond representation of RNA conformation (Cao and Chen 2005; Cao et al. 2010; Chen 2008). As shown in Fig. 3a, the P-O₅-C₅-C₄ and the C₄-C₃-O₃-P dihedrals tend to be planar and rigid, because the torsional angles about the C₅-O₅ and C₃-O₃ bonds tend to remain in the relatively rigid *trans* (*t*) state. Therefore, the original six-torsion nucleotide backbone can be reduced to two “virtual bonds” spanning P to C₄ and C₄ to the next P in the chain (Olson and Flory 1972; Olson 1975; Olson 1980). Calculations on the nucleotide atomic structures show that each virtual bond has a length of 3.9 Å. The third virtual bond (C₄-N₁ for pyrimidine or C₄-N₉ for purine) represents the orientation of the base. In addition, a survey of known RNA structures suggests that the distance between the N₁ (in pyrimidines) or N₉ (in purines) and C₄ atoms stays close to 3.9 Å and the torsion angle between plane P_i-C₄-P_{i+1} and P_i-C₄-N₁ (N₉), close to the *g*⁻¹ isomeric state. Thus, the C₄-N₁ (N₉) virtual bond is quite rigid. The three-vector virtual bond model leads to the following Vfold model for RNA folding:

1. The virtual bond structure for a helix is constructed from the atomic coordinates of an A-form RNA helix (Arnott and Hukins 1972).
2. The ensemble of virtual bond structures for loop conformations are generated by using the usual *gauche*⁺ (*g*⁺), *trans* (*t*) and *gauche*⁻ (*g*⁻) rotational isomeric states for a polymer. A survey on the existing known structures shows that such rotational isomeric states can well represent RNA loop conformations (Cao and Chen 2005; Duarte and Pyle 1998; Duarte et al. 2003; Richardson et al. 2008). The three isomeric states can be realized in the diamond lattice. Therefore, the virtual bonds

of loop conformations are configured on the *diamond lattice*, where each lattice bond is a virtual bond (see Fig. 3b).

3. At the helix-loop junction, we fit the virtual bonds onto the diamond lattice with the minimum RMSD.

The Vfold model is fundamentally different from any of the simplified models such as the simple square or cubic lattice models used in other folding theories. Vfold is a realistic atomistic structural model because the virtual bonds are the realistic, physical P-C₄ and the C₄-P bonds in the structure, and the discretization (diamond lattice) of the virtual bond configurations for loop conformations is based on the principles of polymer physics as well as known RNA structures. So the Vfold model can directly predict experimentally measurable and biologically relevant structures and stabilities. The Vfold package is available for Windows and Unix users (to be released; URL: <http://vfold.missouri.edu/chen-software02.html>). For pseudo-knotted folds, the CPU time (t seconds) for the Vfold-based structural prediction grows with the sequence length (I nucleotides) as $\ln(t) \approx -24.3 + 7.7 \ln(I)$ on a Intel(R) Xeon(R) CPU 5150 @ 2.66GHz on Dell EM64T cluster system. The most time-consuming part of the computation is the enumeration of the different stems and loops (2D structures).

10.3.2 Pseudoknot structure and stability

The predictive power of the Vfold model is shown by its ability to compute the entropy for complex folds such as a 3-loop pseudoknot (Fig. 2). The computation involves three steps:

1. All the possible helix orientations are generated through the enumeration of (virtual bond) conformations of the loop L_3 .
2. For each helix orientation, because the probability for a loop bumping into another loop is relatively small, the loops can be treated with the independent loop approximation (Cao and Chen 2009; Chen and Dill 2000; Poland and Scheraga 1970): The total conformational count (Ω) for the 3-loop system can be estimated as the product of the conformational count ($\Omega_{loop L_i}$) for each loop.

$$\Omega = \prod_i \Omega_{loop L_i}; \quad S_{loop} = k_B \ln \Omega \quad (10.2)$$

This approach is remarkable because it reduces the 3-loop conformational enumeration into conformational enumeration for one loop at a time, resulting in a dramatic reduction in the computer time from $T(\sum_n L_n)$ to $\sum_n T(L_n)$, where L_n is the length of the n -th loop and $T(L)$ is the computer time for counting conformations for a loop of length L .

3. The volume exclusion between a loop and the helices (grooves) is the key to the evaluation of the loop entropy. In the Vfold model, this can be explicitly taken into account by disallowing overlapping virtual bonds when the loop conformations are generated in the virtual bond diamond lattice.

The Vfold model leads to loop entropy parameter tables for canonical (Cao and Chen 2006a) 2-loop H-type pseudoknots and non-canonical 3-loop H-type pseudoknot (Cao and Chen 2009). Table 1 gives the entropy parameters for the canonical (2-loop) H-type pseudoknot. For the non-canonical (3-loop) H-type pseudoknot, the entropy tables are deposited at <http://rnajournal.cshlp.org/content/15/4/696/suppl/DC1> (Cao and Chen 2009).

These loop entropy parameters allow for calculations of the folding free energy for a given pseudoknot (Cao and Chen 2006b; Cao and Chen 2009; Chen 2008). For example, for the 3-loop pseudoknot shown in Fig. 4, the free energy is $G = G_1 + G_2 - T \Delta S(S_1, S_2, L_1, L_2, L_3) = (-7.1) \text{ kcal/mol} + (-6.3) \text{ kcal/mol} + k_B T (14.2) = -4.6 \text{ kcal/mol}$, where S_1 and S_2 denote the length of stems 1 and 2 and L_1 , L_2 and L_3 are the lengths of loops 1, 2 and 3, respectively. The entropy parameter $\Delta S(S_1, S_2, L_1, L_2, L_3)$ is transcribed from Table S1 of this reference: (Cao and Chen 2009).

Application of the Vfold model to the structure prediction of a pool of biologically significant RNA molecules ranging in length from 28 to 91 nucleotides (Ren et al. 2005) indicates that the Vfold model gives significant improvements in the accuracy of the predictions as compared to other existing RNA folding models (Cao and Chen 2009). Furthermore, the successful implementation of the Vfold-predicted entropy parameters in several other software packages indicates that their use can indeed lead to significantly improved accuracy in pseudoknot structural prediction (Andronescu et al. 2010; Sperschneider and Datta 2010; Liu et al. 2010). The calculation of entropies of more complex pseudo-knotted structure is computationally demanding (Cao and Chen 2006a; Cao and Chen 2009). Monte-Carlo simulation may be a potentially useful method to generate virtual bond conformations for the evaluation of the entropy and free energy of more complex structures (Zhang et al. 2008; Zhang et al. 2009).

10.3.3 Loop-stem base triple interactions

In RNA pseudoknots, the close spatial proximity between loops and stems facilitates formation of the tertiary contacts between loop and stem nucleotides. Specifically, a nucleotide in the loop surrounding the major or minor groove is prone to form base triple interactions with a base pair in the helix stem (see Figs. 5a & b). Analysis of the known X-ray structures reveals a large number of non-canonical tertiary interactions in RNA molecules (Xin et al. 2008). These tertiary interactions include A-minor motifs first identified in ribosomal RNA (Nissen et al. 2001), the base triples seen by X-ray in the pseudoknots (Su et al. 1999) and the ribose zipper interactions originally found in group I introns (Cate et al. 1996). Non-canonical tertiary interactions can be critical for the stabilization of the structures and RNA biological functions. For instance, mutations that disrupt the base triples can dramatically reduce the biological activity of telomerase (Theimer et al. 2005) and the efficiency of ribosomal frameshifting (Cornish et al. 2005; Kim et al. 1999; Su et al. 1999).

There are eight different types of base triples in the known pseudoknot structures, A.G-C), A.(C-G), C⁺.(G-C), C⁺.(C-G), U.(A-U), U.(U-A), A.(A-U) and A.(U-A). These base triples belong to different geometric families, defined by the families to which their component basepairs belong (Almakarem et al. submitted 2011). Fig. 6 shows four types of geometric

families based on the nomenclature proposed by Leontis and Westhof (Leontis and Westhof 2001). For example, A.(G-C), C⁺.(G-C), A.(A-U) and U.(A-U) belong to *cis* WW/*trans* SW, *cis* WW/*trans* HW, *cis* WW/*cis* SW and *cis* WW/*cis* HW geometric families, respectively. According to the protonation properties, the eight base triples can be further classified into two types: the protonated [C⁺.G-C) and C⁺.(C-G)] and the unprotonated base triples. The classification is based on the fact that the protonated base triples are more stable than the unprotonated base pairs (Cornish et al. 2005) due to strong electrostatic interactions.

Computational prediction of the non-canonical tertiary interactions is extremely difficult because of the lack of accurate energy parameters for tertiary interactions and the higher conformational complexity and the larger size of the conformational space (Das and Baker 2007; Ulyanov et al. 2007; Yingling and Shapiro 2006). Yingling and Shapiro predicted the base triples in the human telomerase RNA (hTR) pseudoknot using molecular dynamics simulations (Yingling and Shapiro 2006). Though the predicted structure is not exactly consistent with the NMR structure (Kim et al. 2008), the simulation was able to provide useful insights into the interplay between bulge formation and the base triple interactions in telomerase RNAs. Molecular dynamics simulation often relies on the initial input of the 3D structure. In contrast, the Vfold model, which uses the nucleotide sequence as the only input information, does not require any additional structural information.

The Vfold-based prediction of loop-stem tertiary interactions involves the following steps:

1. Through explicit enumeration of the virtual bond conformations, the Vfold model gives the loop entropy S_{loop} as a function of loop length, helix length, and assignments of base triples (Figs. 5ab). Based on the Vfold-predicted entropy parameter set, for a given structure that contains loop-stem tertiary contacts, the free energy G can be evaluated as

$$\Delta G = \Delta G_{helix} - T \Delta S_{loop} + \sum_{i=1}^n (\Delta h^{(i)} - T \Delta s^{(i)})$$

where $h^{(i)}$ and $s^{(i)}$ are the enthalpy and entropy parameters for loop-stem contact i .

2. Theory-experiment comparisons for different systems with loop-stem contacts allow us to extract h and s parameters. As the lowest order approximation, we assume two sets of (h , s) parameters for protonated and non-protonated base triples, respectively. Fitting the melting curves for different pseudoknots that contain loop-stem base triples (Giedroc and Cornish 2009; Nixon et al. 2002; Nixon and Giedroc 2000) converged on the same set of (h , s) parameters (Fig. 5c) below:

$$(\Delta h, \Delta s) = (-7 \text{ kcal/mol}, -19 \text{ cal/mol. K}) \text{ for } (A \text{ or } U).(A- U \text{ or } U- A) \text{ and } (A).(C- G \text{ or } G- C)$$

(10.3)

$$\Delta G = (-14 \text{ kcal/mol}, -38 \text{ cal/mol/K}) (\text{pH } 7.0) \text{ for } C^+ \cdot (C-G \text{ or } G-C) \quad (10.4)$$

The above Vfold approach has led to accurate predictions for the loop-stem tertiary contacts and thermodynamic stabilities for a series of experimentally determined pseudo-knotted structures (Fig. 5c) (Cao et al. 2010).

10.3.4 All-atom 3D RNA structures

Predicting RNA 3D structure is not a solved problem (Das and Baker 2007; Ding et al. 2008; Jonikas et al. 2009; Jossinet and Westhof 2005; Parisien and Major 2008; Shapiro et al. 2007; Tan et al. 2006). Currently, challenges include adequate treatment of the problem of conformational sampling (Das and Baker 2007) and the evaluation of the energetic parameters for tertiary contacts (Ding et al. 2008; Parisien and Major 2008). The Vfold model (Cao et al. 2010) can successfully predict the 2D structures of pseudoknots, including loop-stem tertiary interactions. The Vfold model-predicted virtual bond structure provides a scaffold for the construction of all-atom models of the 3D structure. The prediction of the all-atom 3D structure from the Vfold-predicted 2D structures involves the following three steps (see Fig. 7):

1. Adding all atoms to the virtual bond structure. For nucleotides in each predicted helix, atoms are added according to the A-form helix atomic structure. The 3D conformations of the loop are generated from a combined fragment-based and diamond lattice-based method: the coordinates of the red nucleotides are adopted from the PEMV-1 fragment (PDB ID: 1KPX) and the remaining two nucleotides (blue) are generated by self-avoiding random walks on the diamond lattice. The method can effectively reduce the numbers of loop conformations to a few low-energy viable conformations. For nucleotides in the predicted loop conformations, atoms can be added using helix nucleotides as templates, by aligning the P, C₄ and N_{1,9} atoms with those in of a nucleotide in a helix. This step results in an “atomistic version” of the Vfold structure.

The product of this initial refinement step is a pre-refined atomic structure. The pre-refined structure may contain some atoms/groups that clash sterically with each other. Such steric clashes can be readily resolved by the subsequent molecular dynamics simulation in the next step.

2. Energy minimization of the whole atomistic structure using AMBER molecular dynamics simulations. With the above pre-refined structure as the initial state, molecular dynamics energy minimization with the AMBER molecular dynamics package (Case et al. 2005; Case et al. 2006; Cornell et al. 1995; Pearlman et al. 1995) yield reliable predictions for all-atom 3D structures. In the energy minimization, the negative charges on phosphates are neutralized by Na⁺ cations added to the solution. The nonbonded interactions are truncated at 12 Å. Water molecules are treated by the standard TIP3P model included in AMBER software.

As shown in Fig. 7, the above strategy gives reliable predictions for the all-atom 3D structures for simple tertiary folds such as pseudoknots.

10.4 Quantitative prediction for RNA function --- -1 programmed ribosomal frameshifting

A -1 programmed ribosomal frameshift occurs when the reading frame is shifted by one nucleotide. The change of the reading frame causes the production of multiple proteins. Ribosomal frameshifting is a mechanism used by many RNA viruses to regulate the relative number of copies of viral proteins (gag and pol). Maintaining a normal ratio of gag and gag-pol proteins is critical for viral replication. Alteration in frameshifting efficiency could lead to an abnormal ratio of gag and gag-pol proteins, resulting in reduction or elimination of viral replication (Dinman et al. 1998).

The ribosomal frameshifting machinery consists of three coupled components (see Figs. 8a and b): (a) the slippery sites (nucleotides in blue color) with the sequence XXXYYYZ, where X can be any nucleotide, Y is A or U and Z is A, U or C, (b) a spacer (black thick line) that connects the 5' slippery sites and the 3' mRNA structure, and (c) the 3' mRNA structure which can fold into a pseudoknot structure (nucleotides in red color). According to the mechanical model for frameshifting (Hansen et al. 2007; Namy et al. 2006; Plant et al. 2003), the pseudoknot blocks the entrance of the mRNA into the ribosome and causes ribosomal pausing. During ribosomal pausing, the 3'→5' movement of the tRNA in the A/T to A/A aa-tRNA accommodation process can generate a tension force in the spacer (Plant et al. 2003; Yusupova et al. 2001). The tension, if sufficiently strong, could break the codon-anticodon pair of the 0 frame, causing a 5'→3' shift by one nucleotide of the mRNA chain and the subsequent shift of the codon-anticodon pair by one nucleotide.

The three components of the ribosomal frameshifting machinery are coupled through the spacer length m : the tension force in the spacer is a function of spacer length, which varies with the folding-unfolding of the downstream mRNA. The tension force in the spacer is also dependent on the end-to-end distance X of the spacer, which increases from 3.3 nm to 4.3 nm in the 3'→5' movement of the tRNA. If the downstream mRNA structure is robust against the large tension force, the spacer would more likely be subject to a large tension to induce frameshifting before unfolding of the downstream mRNA structure occurs. Otherwise, unfolding of the downstream mRNA structure would cause the relaxation of the tension, reducing the likelihood of frameshifting.

Quantitative prediction of ribosomal frameshifting requires modeling of the folding stability and the structure of the frameshifting machinery. The partition function for the 3-component frameshifting machinery is $Q(m,X)=Q_{ds} Q_{ss} Q_{codon}$. Here Q_{ds} , Q_{ss} , and Q_{codon} are the partition functions of the 3' downstream mRNA structure ("structured" region, excluding the 5' and 3' tails), the (single-stranded) spacer of length m and end-to-end distance X , and the codon-anticodon base-pairing duplex (in the slippery region), respectively. While Q_{ss} and Q_{codon} can be evaluated using the extensible freely jointed chain model (EFJC) (Gerland et al. 2004; Hyeon and Thirumalai 2005; Liphardt et al. 2001; Smith et al. 1996; Strick et al. 2000) and the nearest neighbor model (Serra and Turner 1995), respectively, the computation of Q_{ds} requires a statistical mechanical model such as the Vfold model. The tension force predicted from the partition function is given by $F(X)=-\frac{1}{\beta} \frac{d \ln Q(m,X)}{dX}$.

The Vfold modeling for the system leads to an analytical relationship between the frameshifting efficiency and the mean tension force; see Fig. 9. It should be noted that recent single-molecule experimental data yields a highly similar relationship (Chen et al. 2009). Furthermore, because the experimental measurement (Chen et al. 2009) for human telomerase RNA and the Vfold-based theoretical predictions (Cao and Chen 2008) involve different frame-shifting systems, the fact that similar analytical relationships are derived from independent theoretical and experimental studies suggests that the quantitative results for the frame-shifting efficiency may be valid for a variety of systems.

10.5 Conclusions

For a long time, the bottleneck for RNA tertiary structural folding has been the inability to treat the free energy, especially the entropy, of structures with non-nested, long-range (tertiary) contacts between nucleotides distant in the 2D structure. Recent advances in the construction of low-resolution conformational models allow us to predict the entropy and the full free energy landscape for RNA tertiary global folds, as well as the 2D structures for the local and the global free energy minima. These 2D structures can further provide scaffolds for the construction of all-atom 3D models of the stable and metastable RNA folds through molecular dynamics calculations. Comparisons between theoretical predictions and experimental data for the 2D and 3D structures and the folding thermodynamics suggest that the statistical mechanical approach is reliable. One of the key factors that contribute to the predictive power of the statistical mechanical models is the rigorous conformational sampling/entropy.

With the rapidly growing size of the database of the experimentally measured RNA structures, fragment-based methods show promise, especially when the homologous conformations of modular components of the RNA of interest can be identified in the PDB database (Cao et al. 2011; Das and Baker 2007; Parisien and Major 2008). However, compared to the number of the deposited protein structures in the PDB database, the number of known RNA structures remains relatively small. Therefore, fragment-based methods for RNA 3D prediction may fail if no known homologous conformations can be found in the PDB database. In that case, a *de novo* construction of the (low-resolution) conformations (especially for the junctions/loops regions) is the only viable approach. The Vfold model introduced in this chapter is one such model that can build the conformations *de novo*.

The current form of the Vfold theory can successfully treat pseudo-knotted folds. However, the problem of predicting more complex pseudo-knotted folds, such as the internal ribosome entry site (IRES) of the cricket paralysis-like viruses (Filbin and Kieft 2009) and other larger tertiary folds, remains. Further development of the model should go beyond the simple loop-stem base triples by including more complex tertiary interactions, such as tetraloop-receptor interactions and kissing loop interactions (Chitsaz et al. 2009; Huang et al. 2009).

Acknowledgments

The authors thank Liang Liu for many useful discussions. This work was supported by the NSF through grants MCB0920411 and MCB0920067 and the NIH through grant R01-GM063732.

References

1. Aalberts DP, Hodas NO. Asymmetry in RNA pseudoknots: observation and theory. *Nucleic Acids Res.* 2005; 33:2210–2214. [PubMed: 15831794]
2. Alan C, Karakoc E, Nadeau JH, Sahinalp SC, Zhang KH. RNA-RNA interaction prediction and antisense RNA target search. *J Comput Biol.* 2006; 13:267–282. [PubMed: 16597239]
3. Almakarem, et al. 2010 submitted.
4. Andronescu M, Zhang Z, Condon A. Secondary structure prediction of interacting RNA molecules. *J Mol Biol.* 2005; 345:987–1001. [PubMed: 15644199]
5. Andronescu MS, Pop C, Condon A. Improved free energy parameters for RNA pseudo-knotted secondary structure prediction. *RNA.* 2010; 16:26–42. [PubMed: 19933322]
6. Arnott S, Hukins DWL. Optimised parameters for RNA double-helices. *Biochem Biophys Res Comm.* 1972; 48:1392–1399. [PubMed: 5077825]
7. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science.* 2000; 289:905–920. [PubMed: 10937989]
8. Baril M, Dulude D, Steinberg SV, Brakier-Gingras L. The frameshift stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot. *J Mol Biol.* 2003; 331:571–583. [PubMed: 12899829]
9. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. *Cell.* 2009; 2:215–233. [PubMed: 19167326]
10. Brierley I, Rolley NJ, Jenner AJ, Inglis SC. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J Mol Biol.* 1991; 220:889–902. [PubMed: 1880803]
11. Brierley I, Pennell S, Gilbert RJC. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Micro.* 2007; 5:598–610.
12. Brierley I, Gilbert RJC, Pennell S. RNA pseudoknots and the regulation of protein synthesis. *Biochem Soc Trans.* 2008; 36:684–689. [PubMed: 18631140]
13. Chen S-J, Dill KA. RNA folding energy landscapes. *Proc Natl Acad Sci USA.* 2000; 97:646–651. [PubMed: 10639133]
14. Cao S, Chen S-J. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA.* 2005; 11:1884–1897. [PubMed: 16251382]
15. Cao S, Chen S-J. Free energy landscapes of RNA/RNA complexes: with applications to snRNA complexes in spliceosomes. *J Mol Biol.* 2006a; 357:292–312. [PubMed: 16413034]
16. Cao S, Chen S-J. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.* 2006b; 34:2634–2652. [PubMed: 16709732]
17. Cao S, Chen S-J. Biphasic Folding Kinetics of RNA Pseudoknots and Telomerase RNA Activity. *J Mol Biol.* 2007; 367:909–927. [PubMed: 17276459]
18. Cao S, Chen S-J. Predicting ribosomal frameshifting efficiency. *Phys Biol.* 2008; 5:016002. [PubMed: 18367782]
19. Cao S, Chen S-J. Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA.* 2009; 15:696–706. [PubMed: 19237463]
20. Cao S, Giedroc DP, Chen S-J. Predicting loop-helix tertiary structural contacts in RNA pseudoknots. *RNA.* 2010; 16:538–552. [PubMed: 20100813]
21. Cao S, Chen S-J. Physics-Based De Novo Prediction of RNA 3D Structures. *Journal of Physical Chemistry B.* 2011; 115:4216–4226.
22. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem.* 2005; 26:1668–1688. [PubMed: 16200636]
23. Case, DA.; Darden, TA.; Cheatham, TE., III; Simmerling, J.; Wang, RE.; Duke, R.; Luo, KM. AMBER 9. University of California; San Francisco: 2006.
24. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science.* 1996; 273:1678–1685. [PubMed: 8781224]

25. Chastain M, Tinoco I Jr. Structural elements in RNA. *Prog Nucl Acid Res Mol Biol.* 1991; 41:131–177.
26. Chauhan S, Caliskan G, Briber RM, Perez-Salas U, Rangan P, Thirumalai D, Woodson SA. RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. *J Mol Biol.* 2005; 353:1199–1209. [PubMed: 16214167]
27. Chen G, Chang K-Y, Chou M-Y, Bustamante C, Tinoco I Jr. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of –1 ribosomal frameshifting. *Proc Natl Acad Sci USA.* 2009; 106:12706–12711. [PubMed: 19628688]
28. Chen JL, Greider CW. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci USA.* 2005; 102:8080–8085. [PubMed: 15849264]
29. Chen S-J. RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys.* 2008; 37:197–214. [PubMed: 18573079]
30. Chen SJ, Dill KA. RNA folding energy landscapes. *Proc Natl Acad Sci USA.* 2000; 97:646–651. [PubMed: 10639133]
31. Chitsaz H, Salari R, Sahinalp SC, Backofen R. A partition function algorithm for interacting nucleic acid strands. *Bioinformatics.* 2009; 25:i365–i373. [PubMed: 19478011]
32. Comolli LR, Smirnov I, Xu L, Blackburn EH, James TL. A molecular switch underlies a human telomerase disease. *Proc Natl Acad Sci USA.* 2002; 99:16998–17003. [PubMed: 12482936]
33. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J Am Chem Soc.* 1995; 117:5179–5197.
34. Cornish PV, Hennig M, Giedroc DP. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated –1 ribosomal frameshifting. *Proc Natl Acad Sci USA.* 2005; 102:12694–12699. [PubMed: 16123125]
35. Das R, Baker D. Automated de novo prediction of native-like RNA tertiary structures. *Proc Acad Natl Sci USA.* 2007; 104:14664–14669.
36. Das R, Kudravalli M, Jonikas M, Laederach A, Fong R, Schwans JP, Baker D, Piccirilli JA, Altman RB, Herschlag D. Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc Natl Acad Sci USA.* 2007; 105:4144–4149. [PubMed: 18322008]
37. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA.* 2009; 106:97–102. [PubMed: 19109441]
38. Draper DE. Pseudoknots and the control of protein synthesis. *Curr Opin Cell Biol.* 1990; 2:1099–1103. [PubMed: 2099803]
39. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA.* 2008; 14:1164–1173. [PubMed: 18456842]
40. Ding Y, Lawrence CE. A statistical sampling algorithm for RNA secondary structure prediction. *Nucl Acids Res.* 2003; 31:7280–7301. [PubMed: 14654704]
41. Ding Y. Statistical and Bayesian approaches to RNA secondary structure prediction. *RNA.* 2006; 12:323–331. [PubMed: 16495231]
42. Dinman JD, Ruiz-Echevarria MJ, Peltz SW. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.* 1998; 16:190–196. [PubMed: 9586242]
43. Dimitrov RA, Zuker M. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J.* 2004; 87:215–226. [PubMed: 15240459]
44. Dirks RM, Pierce NA. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput Chem.* 2003; 24:1664–1677. [PubMed: 12926009]
45. Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.* 2007; 49:65–88.
46. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics.* 2006; 22:e90–e98. [PubMed: 16873527]

47. Duarte CM, Pyle AM. Stepping through an RNA structure: a novel approach to conformational analysis. *J Mol Biol.* 1998; 284:1465–1478. [PubMed: 9878364]
48. Duarte CM, Wadley LM, Pyle AM. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucl Acids Res.* 2003; 31:4755–4761. [PubMed: 12907716]
49. Dulude D, Baril M, Brakier-Gingras L. Characterization of the frameshift stimulatory signal controlling a programmed –1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.* 2002; 30:5094–5102. [PubMed: 12466532]
50. Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B. Probing the structure of RNAs in solution. *Nucl Acids Res.* 1987; 15:9109–9128. [PubMed: 2446263]
51. Filbin ME, Kieft JS. Toward a structural understanding of IRES RNA function. *Curr Opin Struct Biol.* 2009; 19:267–276. [PubMed: 19362464]
52. Fisher ME. Effect of excluded volume on phase transitions in biopolymers. *J Chem Phys.* 1966; 45:1469–1473.
53. Gerland U, Bundschuh R, Hwa T. Translocation of structured polynucleotides through nanopores. *Phys Biol.* 2004; 1:19–26. [PubMed: 16204818]
54. Gesteland RF, Atkins JF. Recoding: Dynamic reprogramming of translation. *Annu Rev Biochem.* 1996; 65:741–768. [PubMed: 8811194]
55. Gherghe CM, Leonard CW, Ding F, Dokholyan NV, Weeks KM. Native-like RNA Tertiary Structures Using a Sequence-Encoded Cleavage Agent and Refinement by Discrete Molecular Dynamics. *J Am Chem Soc.* 2009; 131:2541–2546. [PubMed: 19193004]
56. Giedroc DP, Theimer CA, Nixon PL. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol.* 2000; 298:167–185. [PubMed: 10764589]
57. Giedroc DP, Cornish PV. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.* 2009; 139:193–208. [PubMed: 18621088]
58. Gultyaev AP, van Batenburg FHD, Pleij CWA. The computer simulation of RNA folding pathways using a genetic algorithm. *J Mol Biol.* 1995; 250:37–51. [PubMed: 7541471]
59. Gultyaev AP, Van Batenburg FHD, Pleij CWA. An approximation of loop free energy values of RNA H-pseudoknots. *RNA.* 1999; 5:609–617. [PubMed: 10334330]
60. Hansen TM, Reihani SNS, Oddershede LB, Sørensen MA. Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc Natl Acad Sci USA.* 2007; 104:5830–5835. [PubMed: 17389398]
61. Hofacker IL, Fekete M, Stadler PF. Secondary structure prediction for aligned RNA sequences. *J Mol Biol.* 2002; 319:1059–1066. [PubMed: 12079347]
62. Hofacker IL. Vienna RNA secondary structure server. *Nucl Acids Res.* 2003; 31:3429–3431. [PubMed: 12824340]
63. Hougland, J.; Piccirilli, J.; Forconi, M.; Lee, J.; Herschlag, D. How the Group I Intron Works: A Case Study of RNA Structure and Function. In: Gesteland, RF.; Cech, TR.; Atkins, JF., editors. *RNA world.* 3. Cold Spring Harbor Laboratory Press; New York: 2005. p. 133-205.
64. Huang FW, Qin J, Reidys CM, Stadler PF. Partition Function and Base Pairing Probabilities for RNA-RNA Interaction Prediction. *Bioinformatics.* 2009; 25:2646–2654. [PubMed: 19671692]
65. Hyeon C, Thirumalai D. Mechanical unfolding of RNA hairpins. *Proc Natl Acad Sci USA.* 2005; 102:6789–6794. [PubMed: 15749822]
66. Isambert H, Siggia ED. Modeling RNA folding paths with pseudoknots: Application to hepatitis delta virus ribozyme. *Proc Natl Acad Sci USA.* 2000; 97:6515–6520. [PubMed: 10823910]
67. Isambert H. The jerky and knotty dynamics of RNA. *Methods.* 2009; 49:189–196. [PubMed: 19563894]
68. Jacobson H, Stockmayer WH. Intramolecular reaction in polycondensations. I The theory of linear systems. *J Chem Phys.* 1950; 18:1600–1606.
69. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA.* 2009; 15:189–199. [PubMed: 19144906]

70. Jossinet F, Westhof E. Sequence to Structure (S2S): Display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics*. 2005; 21:3320–3321. [PubMed: 15905274]
71. Kang M, Peterson R, Feigon J. Structural Insights into Riboswitch Control of the Biosynthesis of Queuosine, a Modified Nucleotide Found in the Anticodon of tRNA. *Mol Cell*. 2009; 33:784–790. [PubMed: 19285444]
72. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007; 39:1278–1284. [PubMed: 17893677]
73. Kim N-K, Zhang Q, Zhou J, Theimer CA, Peterson RD, Feigon J. Solution Structure and Dynamics of the Wild-type Pseudoknot of Human Telomerase RNA. *J Mol Biol*. 2008; 384:1249–1261. [PubMed: 18950640]
74. Kim Y-G, Su L, Maas S, O'Neill A, Rich A. Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc Acad Natl Sci USA*. 1999; 96:14234–14239.
75. Klein DJ, Edwards TE, Ferr-D'Amar AR. Cocystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*. 2009; 16:343–344. [PubMed: 19234468]
76. Laederach A, Shcherbakova I, Jonikas MA, Altman RB, Brenowitz M. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc Natl Acad Sci USA*. 2007; 104:7045–7050. [PubMed: 17438287]
77. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA*. 2001; 7:499–512. [PubMed: 11345429]
78. Liphardt J, Onoa B, Smith SB, Tinoco I Jr, Bustamante C. Reversible unfolding of single RNA molecules by mechanical force. *Science*. 2001; 292:733–737. [PubMed: 11326101]
79. Liu B, Mathews DH, Turner DH. RNA pseudoknots: folding and finding. *F1000 Biology Reports*. 2010; 2:8. [PubMed: 20495679]
80. Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*. 2007; 14:287–294. [PubMed: 17401373]
81. Lu ZJ, Turner DH, Mathews DH. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res*. 2006; 34:4912–4924. [PubMed: 16982646]
82. Major F, Gautheret D, Cedergren R. Reproducing the 3D structure of a tRNA molecule from structural constraints. *Proc Natl Acad Sci USA*. 1993; 90:9408–9412. [PubMed: 8415714]
83. Massire C, Jaeger L, Westhof E. Derivation of the 3D architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J Mol Biol*. 1998; 279:773–793. [PubMed: 9642060]
84. Masquida, B.; Westhof, E. A Modular and Hierarchical Approach for All-Atom RNA Modeling in The RNA World. Cech, Thomas R.; Atkins, John F.; Gesteland, Raymond F., editors. 2006. (Cold Spring Harbor Monograph Series) (Cold Spring Harbor Monograph Series)
85. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*. 1999a; 288:911–940. [PubMed: 10329189]
86. Mathews DH, Burkard ME, Freier SM, Wyatt JR, Turner DH. Predicting oligonucleotide affinity to nucleic acid targets. *RNA*. 1999b; 5:1458–1469. [PubMed: 10580474]
87. Mathews DH, Turner DH. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. *J Mol Biol*. 2002; 317:191–203. [PubMed: 11902836]
88. Mathews, DH.; Schroeder, SJ.; Turner, DH.; Zuker, M. Predicting RNA Secondary Structure. In: Cech, TR.; Atkins, JF.; Gesteland, RF., editors. The RNA World. 2006. Cold Spring Harbor Monograph Series
89. McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymer*. 1990; 29:1105–1119.
90. Mükstein U, Tafer H, Hackermüller J, Bernhart SH, Stadler PF, Hofacker IL. Thermodynamics of RNA-RNA binding. *Bioinformatics*. 2006; 22:1177–1182. [PubMed: 16446276]
91. Namy O, Moran SJ, Stuart DI, Gilbert RJ, Brierley I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*. 2006; 441:244–247. [PubMed: 16688178]

92. Naphine S, Liphardt J, Bloys A, Routledge S, Brierley I. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J Mol Biol.* 1999; 288:305–320. [PubMed: 10329144]
93. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc Acad Natl Sci USA.* 2001; 98:4899–4903.
94. Nixon PL, Giedroc DP. Energetics of a strongly pH dependent RNA tertiary structure in a frameshifting pseudoknot. *J Mol Biol.* 2000; 296:659–671. [PubMed: 10669615]
95. Nixon PL, Rangan A, Kim YG, Rich A, Hoffman DW, Hennig M, Giedroc DP. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol.* 2002; 322:621–633. [PubMed: 12225754]
96. Nussinov R, Jacobson AB. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci USA.* 1980; 7:6903–6913.
97. Olson WK, Flory PJ. Spatial configuration of polynucleotide chains: I. Steric interactions in polyribonucleotides: a virtual bond model. *Biopolymers.* 1972; 11:1–23. [PubMed: 5008180]
98. Olson WK. Configuration statistical of polynucleotide chains. A single virtual bond treatment. *Macromolecules.* 1975; 8:272–275. [PubMed: 1152526]
99. Olson WK. Configurational statistics of polynucleotide chains: an updated virtual bond model to treat effects of base stacking. *Macromolecules.* 1980; 13:721–728.
100. Pan J, Woodson SA. Folding intermediates of a self-splicing RNA: Mispairing of the catalytic core. *J Mol Biol.* 1998; 280:597–609. [PubMed: 9677291]
101. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* 2008; 452:51–55. [PubMed: 18322526]
102. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P. AMBER, A package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to stimulate the structural and energetic properties of molecules. *Comp Phys Commun.* 1995; 91:1–41.
103. Petri V, Brenowitz M. Quantitative nucleic acids footprinting: Thermodynamic and kinetic approaches. *Current Opinion in Biotechnology.* 1997; 8:36–44. [PubMed: 9013649]
104. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal AA, CC, GG, and TT mismatches. *Biochemistry.* 1999; 38:3468–3477. [PubMed: 10090733]
105. Plant EP, Jacobs KL, Harger JW, Meskauskas A, Jacobs JL, Baxter JL, Petrov AN, Dinman JD. The 9 Å solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA.* 2003; 9:168–174. [PubMed: 12554858]
106. Poland DC, Scheraga HA. Occurrence of a phase transitions in nucleic acid models. *J Chem Phys.* 1966; 45:1464–1469. [PubMed: 5920201]
107. Poland, DC.; Scheraga, HA. *Theory of the helix-coil transition.* New York: Academic Press; 1970.
108. Qiao F, Cech TR. Triple-helix structure in telomerase RNA contributes to catalysis. *Nat Struct Mol Biol.* 2008; 15:634–640. [PubMed: 18500353]
109. Reeder J, Giegerich R. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinformatics.* 2004; 5:104. [PubMed: 15294028]
110. Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA.* 2004; 10:1507–1517. [PubMed: 15383676]
111. Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: heuristic prediction of RNA secondary structure including pseudoknots. *RNA.* 2005; 11:1494–1504. [PubMed: 16199760]
112. Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, Headd JJ, Richardson DC, Ham D, Hershkovits E, Williams LD, Keating KS, Pyle AM, Micallef D, Westbrook J, Berman HM. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA.* 2008; 14:465–481. [PubMed: 18192612]
113. Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol.* 1999; 285:2053–2068. [PubMed: 9925784]

114. Ruan J, Stormo GD, Zhang W. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics*. 2004; 20:58–66. [PubMed: 14693809]
115. Russell R, Zhuang XW, Babcock HP, Millett IS, Doniach S, Chu S, Herschlag D. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci USA*. 2002; 99:155–160. [PubMed: 11756689]
116. SantaLucia J, Turner DH. Measuring the thermodynamics of RNA secondary structure formation. *Biopolymers*. 1997; 44:309–319. [PubMed: 9591481]
117. SantaLucia J Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA*. 1998; 95:1460–1465. [PubMed: 9465037]
118. SantaLucia J Jr, Saro P, Aduri R, Matta V. Progress toward accurate 3D structure prediction of RNA. *Abstr Paper Am Chem Soc Natl Meet*. 2004; 227:U912–U912.
119. Serra MJ, Turner DH. Predicting thermodynamic properties of RNA. *Methods Enzymol*. 1995; 259:242–261. [PubMed: 8538457]
120. Shapiro BA, Wu JC. Predicting RNA H-type pseudoknots with the massively parallel genetic algorithm. *Comput Appl Biosci*. 1997; 13:459–471. [PubMed: 9283762]
121. Shapiro BA, Yingling YG, Kasprzak W, Bindewald E. Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol*. 2007; 17:157–165. [PubMed: 17383172]
122. Sharp PA. The centrality of RNA. *Cell*. 2009; 136:577–580. [PubMed: 19239877]
123. Shefer K, Brown Y, Gorkovoy V, Nussbaum T, Ulyanov NB, Tzfati Y. A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. 2007
124. Smith SB, Cui Y, Bustamante C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*. 1996; 271:795–799. [PubMed: 8628994]
125. Sperschneider J, Datta A. DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. *Nucl Acids Res*. 2010; 38:e103. [PubMed: 20123730]
126. Spitale RC, Torelli AT, Krucinska J, Bandarian V, Wedekind JE. The structural basis for recognition of the preQ0 metabolite by an unusually small riboswitch aptamer domain. *J Biol Chem*. 2009; 284:11012–11016. [PubMed: 19261617]
127. Staley JP, Guthrie C. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell*. 1998; 92:315–326. [PubMed: 9476892]
128. Staple DW, Butcher SE. Pseudoknots: RNA structures with diverse functions. *PLoS Biol*. 2005; 3:e213. [PubMed: 15941360]
129. Strick T, Allemand JF, Croquette V, Bensimon D. Twisting and stretching single DNA molecules. *Prog Biophys & Mol Biol*. 2000; 74:115–140. [PubMed: 11106809]
130. Su L, Chen L, Egli M, Berger JM, Rich A. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat Struct Biol*. 1999; 6:285–292. [PubMed: 10074948]
131. Tan RKZ, Petrov AS, Harvey SC. YUP: A molecular simulation program for coarse-grained and multiscaled models. *J Chem Theory Comput*. 2006; 2:529–540. [PubMed: 22844233]
132. Tan ZJ, Chen S-J. Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. *Biophys J*. 2006; 90:1175–1190. [PubMed: 16299077]
133. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell*. 2005; 17:671–682. [PubMed: 15749017]
134. Theimer CA, Feigon J. Structure and function of telomerase RNA. *Curr Opin Struct Biol*. 2006; 16:307–318. [PubMed: 16713250]
135. Tyagi R, Mathews DH. Predicting helical coaxial stacking in RNA multibranch loops. *RNA*. 2007; 13:939–951. [PubMed: 17507661]
136. Ulyanov NB, Shefer K, James TL, Tzfati Y. Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates. *Nucl Acids Res*. 2007; 18:6150–6160. [PubMed: 17827211]
137. Waldsich C, Pyle AM. A Kinetic Intermediate that Regulates Proper Folding of a Group II Intron RNA. *J Mol Biol*. 2008; 375:572–580. [PubMed: 18022197]

138. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*. 2009; 460:711–716. [PubMed: 19661910]
139. Williams AL Jr, Tinoco I Jr. A dynamic programming algorithm for finding alternative RNA secondary structures. *Nucl Acids Res*. 1986; 14:299–315. [PubMed: 3003675]
140. Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vornheln C, Hartsch T, Ramakrishnan V. Structure of the 30S ribosomal subunit. *Nature*. 2000; 407:327–339. [PubMed: 11014182]
141. Woodson SA. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci*. 2000; 57:796–808. [PubMed: 10892344]
142. Xin Y, Laing C, Leontis NB, Schlick T. Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA*. 2008; 14:2465–2477. [PubMed: 18957492]
143. Yingling YG, Shapiro BA. The prediction of the wild-type telomerase RNA pseudoknot structure and the pivotal role of the bulge in its formation. *J Mol Graph Mod*. 2006; 25:261–274.
144. Yusupova GZ, Yusupov MM, Cate JH, Noller HF. The path of messenger RNA through the ribosome. *Cell*. 2001; 106:233–241. [PubMed: 11511350]
145. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD, Noller HF. Crystal structure of the ribosome at 5.5 Å resolution. *Science*. 2001; 292:883–896. [PubMed: 11283358]
146. Zarrinkar PP, Williamson JR. Kinetic intermediates in RNA folding. *Science*. 1994; 265:918–924. [PubMed: 8052848]
147. Zhang J, Lin M, Chen R, Wang W, Liang J. Discrete state model and accurate estimation of loop entropy of RNA secondary structures. *J Chem Phys*. 2008; 128:125107. [PubMed: 18376982]
148. Zhang J, Dundas J, Lin M, Chen R, Wang W, Liang J. Prediction of geometrically feasible 3D structures of pseudo-knotted RNA through free energy estimation. *RNA*. 2009; 15:2248–2263. [PubMed: 19864433]
149. Zuker M. On findings all suboptimal foldings of an RNA molecule. *Science*. 1989; 244:48–52. [PubMed: 2468181]
150. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res*. 2003; 31:3406–3415. [PubMed: 12824337]

Figure 1.

A 2D structure can be defined by a graph, which consists of vertices (representing nucleotide monomers), connected by curved links (representing base pairing), and straight lines (representing the backbone covalent bonds). Any two base pairs on the graph can be nested, (cross-)linked, or unrelated. RNA structures are described at the secondary and tertiary structural level. Shown in the figure are the 2D structures and the corresponding graphs for (a) a secondary structure containing three helices and (b) a pseudoknot (as a simple tertiary structure). In the pseudoknot structure, the nucleotides within a loop in the secondary structure pair with the nucleotides external to the loop.

Figure 2.
A 2D structure and the 3D conformations for a pseudoknot with inter-helix loop L_3 .

Figure 3.

(a) Each nucleotide has two **backbone** virtual bonds P-C₄-P (blue) and a **sugar-base** virtual bond (red) C₄-N₁ for pyrimidine or C₄-N₉ for purine. (b) An RNA conformation can be generated through random walks of the virtual bonds in a diamond lattice. In the diamond crystal the four carbon atoms are located at the vertices and center of a tetrahedron. Four such tetrahedra connected at their vertices fit in a cube. Repetition of such cubes side by side generates a diamond lattice.

Figure 4.

The evaluation of the free energy for a 3-loop (non-canonical) H-type pseudoknot using the loop entropy parameters in Table S1 of reference (Cao and Chen 2009) and the Turner rule (Serra and Turner 1995).

Figure 5.

(a) Human telomerase RNA (hTR) pseudoknot contains 2 base triples (blue dashed lines) between loop L2 and helix stem S1 and 3 base triples between loop L1 and helix stem S2 (Theimer et al. 2005). (b) The atomic configuration for the base triple between U9 and the U22-A39 base pair. (c) Theory-experiment test for the melting thermodynamics of five experimentally measured pseudoknot molecules: PLRV and PEMV-1 (Nixon et al. 2002), BWYV and the U8 variant (Nixon and Giedroc 2000), ScYLV. The melting of these molecules show two apparent transitions. The melting at the lower temperature $T_m^{(3)}$ and the higher temperature $T_m^{(2)}$ usually correspond to the disruption of the loop-stem tertiary contacts and the secondary structure, respectively. In the theoretical predictions, the salt-dependent helix stability (in 1M NaCl) is modified according to the experimental condition (0.5 KCl, PH 7.0) by using empirical formulas (SantaLucia 1998; Tan and Chen 2006). (d) The predicted (lowest free energy) 2D structure for the ScYLV pseudoknot agrees exactly with the NMR structure (Cornish et al. 2005, PDB: 1YG3). The red lines denote the loop-stem base triples.

Figure 6. Configurations of four base triples (a) A.(G-C), (b) C⁺.(G-C), (c) A.(A-U) and (d) U.(A-U). They belong to *cis* W.W./*trans* S.W., *cis* W.W./*tran* H.W., *cis* W.W./*cis* S.W. and *cis* W.W./*cis* H.W. geometric families, respectively (Leontif and Westhof 2001; Almakarem et al. submitted 2010).

Figure 7.

(a) The Vfold-predicted Watson-Crick base pairs for the 2D structure of BWYV pseudoknot. (b) The predicted 3D low-resolution (virtual bond) structure built from the predicted 2D structure. (c) The all-atom 3D structure built from the virtual bond structure. (d) the all-atom structure refined by Amber 9. The figure is adopted from Ref. (Cao et al. 2010). The root mean square deviation (RMSD) between the experimental structure (PDB code: 473d; in sand color) and the predicted structure (in purple-blue color) is $2.7\text{\AA}$ over all-heavy atoms.

Figure 8.

A schematic diagram for the -1 ribosomal frameshifting induced by an RNA pseudoknot. The slippery sites of mRNA (blue nucleotides) is shifted in the $5' \rightarrow 3'$ direction by one nucleotide, which results in the reading frame to be shifted from the original (a) 0 frame to the new (b) -1 frame. The translated codons are (XXY,YYZ) in the 0 frame (a) and (XXX,YYY) (b) in the shifted -1 frame.

Figure 9.

The correlation between the frameshifting efficiency (η) and the mean force F_{avg} from $X=3.3$ nm to 4.3 nm. We fitted a analytical expression (the blue thick line) for the correlation: $\eta \approx 0.0083e^{0.1F_{\text{avg}}}$. The experimental data of the frameshifting efficiency is adopted from references (Brierley et al. 1991; Napthine et al. 1999) for IBV and its mutant (pKA18, 1GG, 2UU, pKA41, 3AU, 4AU, pKA49, pKA13 and pKA9), reference (Dulude et al. 2002) for HIV-1 type B, reference (Baril et al. 2003) for HIV-1 type O and reference (Cornish et al. 2005) for ScYLV.

Table 1

Computational models for RNA structure predictions.

Models	URL	References
RNA secondary structure		
Mfold	http://mfold.bioinfo.rpi.edu	Zuker 1989; Mathews et al. 1999a
Vienna software	http://rna.tbi.univie.ac.at	Rehmsmeier et al. 2004
Vfold	http://vfold.missouri.edu	Cao and Chen 2005
Sfold	http://sfold.wadsworth.org	Ding and Lawrence 2003
CONTRAFold	http://contra.stanford.edu/contrafold	Do et al. 2006
MC-Fold	http://www.major.iri.ca/MC-Fold	Parisien and Major 2008
RNA pseudoknot		
STAR	http://biology.leidenuniv.nl/~batenburg/STAR.html	Gulyaev et al. 1995
MPGAfold	http://www-lecb.ncifcrf.gov/~bshapiro/mpgaFold/mpgaFold.html	Shapiro and Wu 1997
pknotsRE	http://selab.janelia.org/software.html	Rivas and Eddy 1999
pknots-RG	http://bibiserv.techfak.uni-bielefeld.de/pknotsrg	Reeder and Giegerich 2004
NUPACK	http://nupack.org	& Dirks and Pierce 2003
ILM	http://cic.cs.wustl.edu/RNA	Ruan et al. 2004
Hotknots	http://www.cs.ubc.ca/labs/beta/Software/HotKnots	Ren et al. 2005
Vfold	http://vfold.missouri.edu	Cao and Chen 2006b
RNA/RNA complexes		
OligoWalk	http://rna.urmc.rochester.edu/software.html	Mathews et al. 1999b
DINAMelt	http://dinamelt.bioinfo.rpi.edu	Dimitrov and Zuker 2004
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid	Rehmsmeier et al. 2004
PairFold	http://www.masoft.ca/cgi-bin/RNAssoft/PairFold/pairfold.pl	Andronescu et al. 2005
Vfold	http://vfold.missouri.edu	Cao and Chen 2006a
RNAup	http://www.tbi.univie.ac.at/~ivo/RNA	Mückstein, U. et al. 2006
NUPACK	http://nupack.org	Dirks et al. 2007

Table 2

Entropy parameters for loops L_1 and L_2 . In the table, an entropy parameter S_{loop} is given in the form of $-S_{loop}/k_B$, where k_B is the Boltzmann constant. The * entries in the Table indicate the loop conformations that cannot be realized in the diamond lattice but may be viable for a realistic pseudoknot. These loop conformations usually have a long stem and short loop. For the entropies of these restricted loops, we use the values of the minimal loop for the same helix length. The table is adapted from reference (Cao and Chen 2006a).

Stem size (bp)	Loop size (nt)											
	1	2	3	4	5	6	7	8	9	10	11	12
L_1 (across the major groove of S_2)												
(S_2)												
2	-	-	-	6.2	6.4	6.4	6.6	6.8	6.9	7.1	7.1	7.2
3	-	6.4*	6.4*	6.4	6.6	6.6	6.8	6.9	7.1	7.3	7.3	7.5
4	4.5*	4.5*	4.5	5.4	5.6	6.0	6.3	6.6	6.9	7.1	7.3	7.3
5	2.3	4.4	4.6	5.7	6.0	6.5	6.9	7.2	7.5	7.8	8.0	8.0
6	2.3	4.4	4.8	5.8	6.0	6.5	6.8	7.1	7.4	7.6	7.8	7.8
7	2.3	4.4	5.0	5.9	6.2	6.8	7.0	7.3	7.6	7.8	8.0	8.0
8	-	4.4	5.2	5.7	6.4	6.7	7.1	7.3	7.5	7.7	7.9	7.9
9	-	5.5*	5.5	6.4	6.7	7.2	7.5	7.9	8.1	8.3	8.5	8.5
10	-	6.9*	6.9*	6.9	7.5	7.7	8.1	8.3	8.6	8.8	8.9	8.9
11	-	-	-	-	8.7	8.8	8.9	9.1	9.2	9.3	9.3	9.3
12	-	-	-	-	9.8	9.2	9.5	9.6	9.7	9.8	9.8	9.8
L_2 (across the major groove of S_1)												
(S_1)												
2	-	-	-	7.6	7.0	7.0	7.1	7.2	7.3	7.4	7.5	7.7
3	-	6.5*	6.5*	6.5	6.6	6.7	6.9	7.1	7.2	7.4	7.6	7.7
4	-	-	9.2*	9.2*	9.2	8.9	8.9	8.9	9.0	9.0	9.1	9.2
5	-	-	-	9.8*	9.8*	9.8	9.1	8.9	8.8	8.8	8.8	8.8
6	-	-	-	11.9*	11.9*	11.9*	11.9	11.0	10.4	10.1	9.9	9.8
7	-	-	-	-	12.4*	12.4*	12.4*	12.4	11.4	11.0	10.7	10.5
8	-	-	-	-	12.1*	12.1*	12.1*	12.1	11.6	11.4	11.2	11.1
9	-	-	-	-	-	13.7*	13.7*	13.7*	13.7	12.6	12.0	11.5
10	-	-	-	-	-	13.7*	13.7*	13.7	12.7	12.2	11.8	11.5
11	-	-	-	-	-	-	-	-	15.9	14.1	13.0	12.4
12	-	-	-	-	-	-	-	-	18.7	15.8	14.2	13.2