Published in final edited form as:

Methods Mol Biol. 2016 ; 1490: 63–72. doi:10.1007/978-1-4939-6433-8_5.

A method to predict the structure and stability of RNA/RNA complexes

Xiaojun Xu and Shi-Jie Chen

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

Summary

RNA/RNA interactions are essential for genomic RNA dimerization and regulation of gene expression. Intermolecular loop-loop base pairing is a widespread and functionally important tertiary structure motif in RNA machinery. However, computational prediction of intermolecular loop-loop base pairing is challenged by the entropy and free energy calculation due to the conformational constraint and the intermolecular interactions. In this chapter, we describe a recently developed statistical mechanics-based method for the prediction of RNA/RNA complex structures and stabilities. The method is based on the virtual bond RNA folding model (Vfold). The main emphasis in the method is placed on the evaluation of the entropy and free energy for the loops, especially tertiary kissing loops. The method also uses recursive partition function calculations and two-step screening algorithm for large, complicated structures of RNA/RNA complexes. As case studies, we use the HIV-1 Mal dimer and the siRNA/HIV-1 mutant (T4) to illustrate the method.

Keywords

RNA/RNA complex; tertiary motif entropy; partition function; two-step screening

1. Introduction

RNA-RNA interactions play widespread roles in RNA biological functions from mRNA splicing^{1–3} and microRNA target recognition^{4–7} to RNA/RNA dimerization.^{8–10} RNA folding induced by RNA-RNA binding can be important for RNA-related cellular processes. For example, during the mRNA splicing process, RNA/RNA complexes formed by small nuclear RNAs undergo multiple structural rearrangements in the different steps of splicing,¹¹ microRNAs regulate gene expression by binding to gene targets (at 3' untranslated regions of target mRNA transcripts),¹² and RNA/RNA dimerization plays essential role in viral replication.^{13, 14} Many of the processes of RNA-RNA binding are facilitated by the intermolecular loop-loop interactions between RNA molecules.

Accurate prediction of RNA/RNA complex formation requires a reliable method to treat (1) the change of the conformational statistics in the folding process, (2) interplay between intermolecular and intramolecular base pairing, (3) tertiary interactions, such as kissing-loop

interactions and pseudoknotted interactions between two RNAs. Most of the currently available folding algorithms for RNA/RNA complexes can treat intermolecular and intramolecular competitions at the secondary structure level.^{11, 15–19} However, many biologically important RNA/RNA complexes involve tertiary (cross-linked) intermolecular contacts. As a result of the cross-linkage between the different loops and between the different loops and helices, the folding free energy of a structure becomes nonadditive, i.e., the total stability of a structure is not the simple additive sum of the stability of each structure subunit. Studies by us and other groups show that a physical entropy model can lead to improvements in the predictions of RNA secondary structures and thermodynamic stabilities.²⁰ For example, the recently developed models,^{21, 22} based on the partition function calculations and simplified thermodynamic models for simple kissing interactions can account for complex loop kissing interactions. However, to explicitly account for the nonadditive free energy, especially the entropy for the different loop kissing motifs, we need a new physics-based model.

The recently developed Vfold model is a statistical mechanics-based RNA folding model. The model relies on a coarse-grained (virtual bond) representation for RNA conformation.^{23–25} Compared with other free energy-based RNA/RNA complex structure prediction models, the Vfold model computes loop entropy parameters from explicit conformational sampling in three-dimensional space. Furthermore, by enumerating all the possible structures, including tertiary structures containing cross-linked loops, in the partition function calculation, the Vfold model computes the free energy landscape for secondary and simple tertiary structures. The results led to several predictions for RNA mechanisms, such as pseudoknot-involved conformational switch between bistable secondary structures,²⁶ microRNA-gene target interactions,¹² and RNA/RNA kissing dimerization in viral replication.^{13, 14}

2. Tertiary loop entropy calculation

With two virtual bonds per nucleotide to represent the backbone conformation, the Vfold model samples fluctuations of loops/junction conformations in three-dimensional space through conformational enumeration (see Fig. 1). By calculating the probability of loop formation, the Vfold model gives the conformational entropy parameters for the formation of the different types of loops including pseudoknot loops^{27, 28} and hairpin kissing loops.¹³ The model is based on the complete conformational ensemble and accounts for the chain connectivity and exclude volume effects. Here we use the a hairpin loop stretched by a bound RNA (Fig. 1) to show the steps of entropy calculation for a kissing loop.

1. We first use A-form helix (see^{Note 1}) to configure the 4-nt stem in Fig. 1.
2. We then enumerate all the possible virtual bond backbone conformations for a given chain length (see^{Note 2}) and count the total number Ω_{coil} of the conformations (see^{Note 3}).

¹The coordinates (x, θ, z) for atoms P and C4' in the A-form helix are $(8.71 \text{ \AA}, 70.5 + 32.7 i, -3.75 + 2.81 i)$ and $(9.68 \text{ \AA}, 46.9 + 32.7 i, -3.10 + 2.81 i)$ ($i = 0, 1, 2, \dots$)²⁹

²A survey of the known structures suggests that the virtual bonds (P-C4' and C4'-P) have bond length of $\sim 3.9 \text{ \AA}$ and bond angle in the range of $90^\circ - 120^\circ$.

3. From the conformational ensemble above, we identify loop conformations according to the loop closure condition. For the structure in Fig. 1, we fit the two ends of a loop to a base pair configuration in an A-form helix. We generate the virtual bond conformations through self-avoiding random walks and count the total number Ω_{loop} of viable loop conformations.
4. We calculate the loop entropy $S_{\text{loop}} = k_B \ln(\Omega_{\text{loop}}/\Omega_{\text{coil}})$, where k_B is the Boltzmann constant.
5. The above Vfold-based computation leads to pre-tabulated entropy parameters for hairpin loops,²³ internal/bulge loops,²³ H-type pseudoknots with/without inter-helix junction,^{27, 28} hairpin-hairpin kissing motifs¹³ and other tertiary motifs.³⁰

3. Methods

Partition function is a Boltzmann-weighted sum over all the possible structures. From the partition function, we can compute (a) the base pairing probability for each nucleotide pair, from which we can predict the probable structures, (b) the melting curve and (c) the other folding thermodynamic properties such as folding free energies. Furthermore, the method can predict not only the global minimum free energy structure but also all the possible alternative structures (local minima on the free energy landscape).

One of the key ingredients in partition function calculation is the enthalpy and entropy parameters for a given structure. In our calculation, the enthalpy and entropy for the canonical and the mismatched base stacks are given by Turners experimental data.³¹ The loop entropies are calculated from the Vfold pre-tabulated parameters.

3.1 RNA/RNA complex structure prediction based on complete structure enumeration

1. For given sequences of two interacting RNAs, we insert a three-nucleotide phantom link (*see* Note 4) between the RNAs. The three-nucleotide linker transforms the original two-RNA system into an effective one-RNA system (*see* Fig. 2).
2. For each RNA sequence, we first generate the structure ensemble by enumerating all the possible base pair arrangements, including H-type pseudoknots with/without inter-helix loop and other tertiary motifs shown in Fig. 2 (*see* Note 5).
3. For each sequence, we compute the partition function by summing over all the structures in the ensemble generated above. We denote the partition functions as Z_1 and Z_2 for the two RNAs, respectively.

³An acceptable loop conformation (with virtual bond backbone) is restricted by the chain connectivity, excluded volume effects, and the loop closure condition.

⁴The three nucleotides in the phantom linker cannot form base pairs with any other (physical) nucleotides. Moreover, for the effective one-RNA system, no loop parameters will be assigned to any (nonphysical) loops that contain phantom nucleotides.

⁵To consider the compatibility for the connection between the different structural subunits, we classify the partition functions into different types (e.g., different types of loops) in our recursive partition function algorithm.^{11, 13, 23, 27}

4. We repeat the above procedure to calculate the partition function Z_3 for the effective one-RNA system above. Note that Z_3 includes structures with and without inter-molecular base pairs.
5. We calculate the partition function for the complex (structures with intermolecular base pairs) as $Z_{12} = Z_3 - Z_1 \cdot Z_2$.
6. We compute the total partition function $Z(c_s, T)$ for the system: $Z(c_s, T) = Z_1 \cdot Z_2 + \alpha e^{-(G_{\text{init}}/k_B T)} Z_{12}$. Here c_s is the RNA strand concentration, T is the temperature, and $G_{\text{init}} = 3.61 + 0.75 k_B T$ (kcal/mol) (see Note 6). The parameter α is equal to $c_s/4$ for non-self-complementary strands and c_s for self-complementary strands.
7. By following the similar procedure as above, we compute the conditional partition function $Z_{ij}(c_s, T)$ for all the structures that contain the (i, j) base pair (see Note 7).
8. From the above partition functions, we calculate the probability of finding the (i, j) base pair: $p_{ij}(c_s, T) = Z_{ij}(c_s, T)/Z(c_s, T)$.
9. From the base pairing probability for all the possible (i, j) pairs, we predict the most probable (see Note 8) as well as alternative structures (see Note 9).

We use the HIV-1 Mal complex as an example to show how Vfold predicts the RNA/RNA complex structure. Given the 23-nt RNA sequence of the HIV-1 Mal (see Note 10), Vfold calculates the base pairing probabilities p_{ij} for all the possible base pairs (see Note 7) for the given temperatures T and RNA strand concentration (see Fig. 3). Our prediction indicates that thermal heating can induce the conformational switch from the kissing complex to the extended-duplex dimer for the HIV-1 Mal dimer. The result is consistent with the experimental observation.³³

3.2 Two-step screening method for RNA/RNA complex prediction

The partition function method with complete structure enumeration is computationally expensive, especially when the tertiary motifs are included in the conformational ensemble. To treat large and complicate RNA/RNA complexes, we use a two-step screening method to efficiently predict the structure and the folding stability of RNA/RNA complexes. The essence of the model is to parse the structure prediction for the whole system into two steps: we first identify the most probable binding sites (binding mode), we then calculate the base pairing probabilities based on the most probable binding site (binding mode).

1. For the given RNA sequences, we enumerate all the possible binding regions (see Note 11) between two RNAs (see the red base pairs in Fig. 4A). We call RNA-RNA binding at a given region m as a binding mode.

⁶The physical origin of an additional G_{init} is due to the entropy loss associated with the conversion from two single-stranded RNAs to a single RNA complex.

⁷Vfold only calculates conditional partition function $Z_{ij}(c_s, T)$ ($|i-j| > 3$) for A-U, G-C or G-U base pairs.

⁸The predicted most probable structure is formed by the base pairs of the largest base pairing probabilities.

⁹We recommend users to also predict the possible alternative structures from the base pairing probability pattern, since many biological functions involve the suboptimal (alternative) structures.

¹⁰The sequence of the HIV-1 Mal is: 5'CUUGCUGAGGUGCACACAGCAAG3'.

2. For each binding mode m , we use the nearest-neighbor model to calculate the free energy $\Delta G_{\text{binding-helix}}^{(m)}$ for the intermolecular base pairs:

$\Delta G_{\text{binding-helix}}^{(m)} = \sum_{\text{stack}} \Delta G_{\text{stack}}$, where G_{stack} is the free energy of a constituent base stack in the binding region.

3. For a given binding mode, we enumerate all the possible intramolecular base pairs (see Note 12), with the condition that the nucleotides in the binding region are not allowed to base pair with other nucleotides (see the blue base pairs in Figs. 4BC).

4. We assign the free energy for each structure in the conformational ensemble:

$\Delta G_s^{(m)} = \Delta G_{\text{helix}}^{(m)} + \Delta G_{\text{loop}}^{(m)}$. Here, $\Delta G_{\text{helix}}^{(m)}$ includes the intermolecular helix $\Delta G_{\text{binding-helix}}^{(m)}$ in the binding region and $\Delta G_{\text{loop}}^{(m)}$ accounts for the free energy (loop entropic) changes caused by the conformational constraint due to the formation of the loop-kissing interactions (Fig. 2) (see Note 13).

5. We calculate the total partition function for mode m as the sum over all the possible structures:

$$Z_{\text{tot}}^{(m)} = \sum_{\text{structures}} e^{-\Delta G_s^{(m)} / k_B T}$$

6. By following the similar procedure, We enumerate all the possible binding modes and calculate the partition function $Z_{\text{tot}}^{(m)}$ for each binding mode. The most probable binding mode (M) is the one with the maximum total partition function Z_{tot}^M (see Note 14).

7. In the next step, we compute the RNA/RNA complex structure for the most probable binding mode. For the most probable mode M , following the above procedure, we compute the partition function Z_{ij}^M for all the conformations with base pair (i, j) .

8. From the partition functions, we calculate the probability of finding intramolecular base pair (i, j) : $p_{ij}^M = Z_{ij}^M / Z_{\text{tot}}^M$.

9. For the most probable binding site (M), we predict the intermolecular base pairs by minimizing the free energy for the binding region. We assume that the binding region is a helix stem that can possibly contain at most a bulge loop (see Note 11).

¹¹In the binding region, we assume that the helix has at least two base pairs. We allow the formation of a bulge loop in the intermolecular helix stem at the binding site.

¹²For each binding mode, with the assumption that the intermolecular interactions occur only at the predicted binding site, the enumeration of intramolecular base pairs can be performed independently for the two RNAs. This assumption significantly enhances the computational efficiency since the two RNAs are treated independently.

¹³We have pre-tabulated parameters for the different types of the loops.^{13, 23, 27, 28, 30}

¹⁴The two-step screening method for RNA/RNA complex structure prediction calculates base pairing probabilities only for the most stable binding modes.

10. From the base pairing probabilities for all the (i, j) pairs, we predicted the most probable (see^{Note 8}) as well as the alternative complex structures (see^{Note 9}).

We use the *siRNA/HIV-1 mutant (T4)* complex (see^{Note 15}) to illustrate the above two-step screening method. As shown in Fig. 4(D), both RNAs can fold into a stem-loop structure. We found¹² a single dominant binding site for the two-RNA system. The stem of T4 is completely disrupted upon siRNA binding at the dominant binding site. Meanwhile, the nucleotides in the 3' tail of T4 refolds into a new hairpin-like structure after siRNA binding. Since the two RNAs in this case are short (19 nt and 47 nt, respectively), we can also predicted its complex structure using the exact structure enumeration as described in the previous section. The two methods give consistent results.¹²

It should be noted that depends on the sequences of the two RNAs, we might find multiple dominant binding sites with comparable binding affinities, as the case of the HIV-1 Mal dimer shown in Fig. 3. If there exist multiple most probable binding modes, we need to follow the above procedure for each mode to predict the respective complex structures.

Acknowledgments

This research was supported by NIH grant GM063732.

References

1. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem.* 2003; 72:291–336. [PubMed: 12626338]
2. Roy SW, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet.* 2006; 7:211–221. [PubMed: 16485020]
3. Taggart AJ, DeSimone AM, Shih JS, Filloux ME, Fairbrother WG. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat Struct Mol Biol.* 2012; 19:719–721. [PubMed: 22705790]
4. 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]
5. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009; 136:215–233. [PubMed: 19167326]
6. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem.* 2010; 79:351–379. [PubMed: 20533884]
7. Westerhout EM, Berkhout B. A systematic analysis of the effect of target RNA structure on RNA interference. *Nucleic Acids Res.* 2007; 35:4322–4330. [PubMed: 17576691]
8. Russell RS, Liang C, Wainberg MA. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology.* 2004; 1:23. [PubMed: 15345057]
9. Lorenz C, Piganeau N, Schroeder R. Stabilities of HIV-1 DIS type RNA loop-loop interactions in vitro and in vivo. *Nucleic Acids Res.* 2006; 34:334–342. [PubMed: 16410613]
10. Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol.* 2004; 2:461–472. [PubMed: 15152202]
11. Cao S, Chen S-J. Free energy landscapes of RNA/RNA complexes: with applications to snRNA complexes in spliceosomes. *J Mol Biol.* 2006; 357:292–312. [PubMed: 16413034]
12. Cao S, Chen S-J. Predicting kissing interactions in microRNA-target complex and assessment of microRNA activity. *Nucleic Acids Res.* 2012; 40:4681–4690. [PubMed: 22307238]

¹⁵The sequence of the siRNA is: 5' CUGUAUCAUCUGCUCCUGU3'. The sequence of the HIV-1 mutant (T4) is: 5' ACAGGAGCAGAUGAUACAGUUCAGAGAAUGUAUAAUCUGCUUAUGU3'.

13. Cao S, Chen S-J. Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. *RNA*. 2011; 17:2130–2143. [PubMed: 22028361]
14. Cao S, Xu X, Chen S-J. Predicting structure and stability for RNA complexes with intermolecular loop-loop base-pairing. *RNA*. 2014; 20:835–845. [PubMed: 24751648]
15. Andronescu MS, Zhang ZC, Condon AE. Secondary structure prediction of interacting RNA molecules. *J Mol Biol*. 2005; 345:987–1001. [PubMed: 15644199]
16. Bernhart SH, Tafer H, Flamm C, Stadler PF, Hofacker IL. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol*. 2006; 1:3. [PubMed: 16722605]
17. Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA. Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Rev*. 2007; 49:65–88.
18. Dimitrov RA, Zuker M. Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids. *Biophys J*. 2004; 87:215–226. [PubMed: 15240459]
19. Mathews DH. Using OligoWalk to identify efficient siRNA sequences. *Methods in Molecular Biology*. 2010; 629:109–121. [PubMed: 20387146]
20. Andronescu MS, Pop C, Condon AE. Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *RNA*. 2010; 16:26–42. [PubMed: 19933322]
21. Huang FWD, Qin J, Reidys CM, Stadler PF. Partition function and base pairing probabilities for RNA/RNA interaction prediction. *Bioinformatics*. 2009; 25:2646–2654. [PubMed: 19671692]
22. Chitsaz H, Salari R, Sahinalp SC, Backofen R. A partition function algorithm for interacting nucleic acid strands. 2009
23. Cao S, Chen S-J. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA*. 2005; 11:1884–1897. [PubMed: 16251382]
24. Chen S-J. RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. *Annual Review of Biophysics*. 2008; 37:197–214.
25. Xu X, Zhao P, Chen S-J. Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One*. 2014; 9:e107504. [PubMed: 25215508]
26. Xu X, Chen S-J. Kinetic mechanism of conformational switch between bistable RNA hairpins. *J Am Chem Soc*. 2012; 134:12499–12507. [PubMed: 22765263]
27. Cao S, Chen S-J. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res*. 2006; 34:2634–2652. [PubMed: 16709732]
28. Cao S, Chen S-J. Predicting structures and stabilities for H-type pseudoknots with inter-helix loop. *RNA*. 2009; 15:696–706. [PubMed: 19237463]
29. Arnott S, Hukins DW, Dover SD. Optimised parameters for RNA double-helices. *Biochem Biophys Res Commun*. 1972; 48:1392–1399. [PubMed: 5077825]
30. Cao S, Chen S-J. A domain-based model for predicting large and complex pseudoknotted structures. *RNA Biol*. 2012; 9:200–211. [PubMed: 22418848]
31. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res*. 2010; 38:D280–D282. [PubMed: 19880381]
32. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol*. 2001; 8:1064–1068. [PubMed: 11702070]
33. Muriaux D, De Rocquigny H, Roques BP, Paoletti J. NCp7 activates HIV-1 RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem*. 1996; 271:33686–33692. [PubMed: 8969239]

Figure 1.

The Vfold model uses two bonds (P-C4' and C4'-P) to represent each nucleotide and computes loop entropies by sampling virtual bond conformations in three-dimensional space. (a) A tertiary motif which shows a hairpin loop stretched through base pairing with a bound RNA. (b) A virtual bond (in cyan) backbone conformation of the three-dimensional tertiary structure corresponding to the two-dimensional structure shown in (a). In the three-dimensional structure, the two ends of the loops L_1 and L_2 are fitted to the base pairs of A-form helices (in red). Starting from an A-form helix with the given number of base pairs, Vfold enumerates loop backbone conformations on a diamond lattice with bond length of 3.9 Å, bond angle of $\sim 109.5^\circ$ and three equiprobable torsional angles (60° , 180° , 300°). From the probability of loop closure, Vfold calculates RNA motif based loop entropies.

Figure 2.

Examples of intermolecular interactions (base pairing) between two RNAs that can be treated by the current Vfold model. The thick curved links in the polymer graph (right panels) denote helices. The straight lines represent RNA main chain. The dashed lines denote the phantom link, which is used to convert the two-RNA system into an effective one-RNA system. The blue lines in the left panels are intramolecular base pairs and the red lines are the intermolecular base pairs. (A) The effective one-RNA system containing secondary structures only. (B) and (C) Tertiary kissing motifs with cross-linked base pairs.

Figure 3. Density plots for the predicted base pairing probabilities at different temperatures for the HIV-1 Mal dimer. The RNA strand concentration C_s is $150 \mu\text{M}$, which is adopted from the experiment.³² At room temperature, the kissing-loop and the extended-duplex dimer (structures shown in inset) coexist. As the temperature increases, the extended-duplex becomes dominant.

Figure 4.

(A–C) Enumeration of RNA/RNA complex structures. For a given bindingmode (base pairs in red), the intramolecular base pairs (shown in blue) are enumerated independently for the two RNAs. The final RNA/RNA complex structures are generated by the combination of the different intramolecular base pairs. (D) The conformational change caused by the siRNA binding to a HIV-1 mutant (T4). siRNA can induce the complete unzipping of T4, which refolds into a new structure after siRNA binding.