A two-length-scale polymer theory for RNA loop free energies and helix stacking

DANIEL P. AALBERTS and NAGARAJAN NANDAGOPAL

Department of Physics, Williams College, Williamstown, Massachusetts 01257, USA

ABSTRACT

The reliability of RNA secondary structure predictions is subject to the accuracy of the underlying free energy model. Mfold and other RNA folding algorithms are based on the Turner model, whose weakest part is its formulation of loop free energies, particularly for multibranch loops. RNA loops contain single-strand and helix-crossing segments, so we develop an enhanced two-length freely jointed chain theory and revise it for self-avoidance. Our resulting universal formula for RNA loop entropy has fewer parameters than the Turner/Mfold model, and yet simulations show that the standard errors for multibranch loop free energies are reduced by an order of magnitude. We further note that coaxial stacking decreases the effective length of multibranch loops and provides, surprisingly, an entropic stabilization of the ordered configuration in addition to the enthalpic contribution of helix stacking. Our formula is in good agreement with measured hairpin free energies. We find that it also improves the accuracy of folding predictions.

Keywords: freely jointed chain; hairpin loops; multibranch loops; RNA folding; self-avoidance

INTRODUCTION

Accurate prediction of macromolecular structure from a primary sequence is one of the grand challenges of computational biology. The secondary structure of RNA, defined by a set of canonical GC, AU, or GU base-pair interactions of the *cis* Watson–Crick/Watson–Crick-type (Leontis and Westhof 2001) contributes the great majority of the total free energy. Tertiary interactions, such as nonlocal hydrogen bonds, divalent-counterion stabilization, or helix stacking, account for the remainder.

The Turner rules (Xia et al. 1998; Mathews et al. 1999) are the basis for many RNA secondary structure prediction algorithms including Mfold (Zuker 2003), ViennaRNA (Hofacker 2003), Sfold (Ding and Lawrence 2003), and others. RNAstructure (Mathews et al. 2004) has systematically incorporated revisions to the rules (Kim et al. 1996; Diamond et al. 2001; Mathews and Turner 2002; Mathews et al. 2004; Tyagi and Mathews 2007). The Turner rules are based largely on physical–chemical measurements, with tabulated free energies for base-pair stacks and hairpin and interior loops.

Loop free energies are the most uncertain part of the Turner model, with large statistical errors for hairpin loops and systematic errors in the model of multibranch loops. The authors of Mfold confess (Zuker et al. 1999), “Because so little is known about the effects of multibranch loops on RNA stability, we assign free energies in a way that makes the computations easy.” The Mfold algorithm assumes a linear dependence of multibranch loop free energy on loop dimensions (Zuker and Sankoff 1984). Several statistical–mechanical approaches (Chen and Dill 2000; Zhang and Chen 2002) have since rejected this unphysical, albeit computationally efficient, treatment of multibranch loops.

The entropy of RNA loops has been computed recently by enumerating backbone configurations to simulate loop regions (Cao and Chen 2005; Zhang et al. 2008) based on a virtual bond representation of the backbone (Olson and Flory 1972; Olson 1975). These studies give fairly accurate numerical representations of the entropies of different loops, but their extrapolation formulas do not provide the physical insight that our theoretical framework can provide.

In this study, we describe RNA loops in terms of a two-length-scale polymer physics model with single-strand and helix-crossing segments. We derive the self-avoiding loop-closure free energy for this model, arriving at a simple functional form, which greatly reduces the number of parameters that appear in the Turner model or enumerative simulations. The standard errors for multibranch loops are 10 times

Reprint requests to: Daniel P. Aalberts, Department of Physics, Williams College, Williamstown, MA 01257, USA; e-mail: aalberts@williams.edu; fax: (413) 597-4116.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.1831710>.

smaller for our model than for Mfold's when compared with simulations. Our formula also neatly passes through the large experimental error bars for hairpin free energies.

We also introduce the surprising physical insight that the coaxial stacking of adjacent helical segments provides an entropic free energy benefit, in addition to the energetic stabilization currently incorporated in secondary structure prediction algorithms (Mathews et al. 2004). This situation marks a rare exception to the classic tradeoff between energy and entropy. This novel phenomenon may have widespread implications in determining the stability of RNA structures. Finally, in addition to simplifying the Turner model, we will show that our formula increases the accuracy of Mfold predictions.

MATERIALS AND METHODS

Polymer theory

One key observation is that RNA loops have two length scales: $a = 6.2$ Å for monomer separation in single-strand regions, and $b = 15$ Å to cross a helix (Aalberts and Hodas 2005). The properties of such a polymer can be understood by extending the freely jointed chain (FJC) model from one up to two step lengths. In our FJC2, each segment has length $r_i \in \{a, b\}$. Randomly distributed unit vectors obey $\langle \hat{n}_i \rangle = 0$, $\langle \hat{n}_i \cdot \hat{n}_j \rangle = 0$ for $i \neq j$, and $\langle \hat{n}_i \cdot \hat{n}_i \rangle = 1$. The total end-to-end separation is $\mathbf{R} = \sum_i r_i \hat{n}_i$. If there are N links of length a and M links of length b ,

$$\langle R^2 \rangle = \sum_{i,j} r_i r_j \langle \hat{n}_i \cdot \hat{n}_j \rangle = \sum_i r_i^2 = Na^2 + Mb^2, \quad (1)$$

this defines the characteristic spatial extent of the chain. If there is only one segment length type, Equation 1 reduces to the familiar FJC result (Grosberg and Khokhlov 1994).

Loops are formed when the polymer walk returns near to its starting point. If we use a characteristic volume ΔV to define “near” the origin, the loop probability scales like

$$p_{\text{near}} \sim \Delta V / \langle R^2 \rangle^{3/2} \sim (Na^2 + Mb^2)^{-3/2}.$$

For self-avoiding polymers, the classic Flory result (Grosberg and Khokhlov 1994) is that the chain extension scales with a larger power, $\langle R^2 \rangle_{\text{Flory}} \sim a^2 N^{6/5}$. The self-avoiding FJC2 analogy is

$$\langle R^2 \rangle_{\text{sa}} \sim N^{6/5} a^2 + M^{6/5} b^2.$$

The probability of end-to-end separation for a self-avoiding polymer adopts an asymptotic scaling form (Grosberg and Khokhlov 1994):

$$p_{\text{sa}}(x) \sim x^{5/18} \langle R^2 \rangle_{\text{sa}}^{-3/2},$$

when the dimensionless chain extension $x = [\mathbf{R}^2 / \langle R^2 \rangle_{\text{sa}}]^{1/2}$ is small. To form a loop, $x \sim \Delta V^{1/3} / \langle R^2 \rangle_{\text{sa}}^{1/2}$ is small, and after combining 1/2, 5/18, and 3/2 exponents,

$$p_{\text{near,sa}} \sim \langle R^2 \rangle_{\text{sa}}^{-59/36} \sim e^{-G_{\text{FJC2}}/kT}.$$

Thus, the FJC2 loop-closure free energy is

$$G_{\text{FJC2}}(N, M) = \frac{59}{36} kT \ln \left(N^{6/5} a^2 + M^{6/5} b^2 \right) + C, \quad (2)$$

with C reflecting the possibility of a different criterion for being near.

Sequence dependencies of loops are largely ignored in the Turner rules (Mathews et al. 1999, 2004); exceptions include stacking bonuses of the bases adjacent to a helix, a list of stable tetraloops (hairpin loops with $n = 4$ bases), and interior loops with up to 2×3 mismatched bases—we include these sequence dependences identically to Mfold in our testing. Implicit in neglecting sequence dependences is the approximation that the entropy of configurations dominates the free energy, particularly for long loops. This “athermal” approximation of ignoring enthalpic effects is explicit in enumerations of chain configurations (Cao and Chen 2005; Zhang et al. 2008) and in our derivation of loop free energies.

Coaxial stacking

Although coaxial stacking refers to the tertiary organization of secondary structure elements, it is included in secondary structure folding algorithms (Mathews et al. 1999, 2004; Zuker 2003). If no bases intervene between two helices, they may adopt a coaxially stacked orientation; a free energy change equal to making a conventional base-pair stack is assigned in Mfold for coaxially stacked helices (Mathews et al. 1999). If there is an unpaired base between the helices, Mfold essentially recruits a base from the loop to make an intervening mismatch (Kim et al. 1996; Tyagi and Mathews 2007). Figure 1 shows examples of each type—the acceptor stem and T Ψ CG helix stack coaxially, while the anticodon and D stems are separated by one base.

An important fact that has not been noticed is the effect of orienting the helices on the remainder of the multibranch loop. Basically, two of the long b segments are removed from the effective loop in the ordered configuration. Since the free energy cost to make

FIGURE 1. The secondary structure of tRNA includes four base-paired stems meeting at a central multibranch loop. In the multibranch loop and hairpin loops, the effective backbone is comprised of a combination of N single-strand links of length a and M helix-crossing segments of length b . Coaxial stacking orients helices, reducing the effective multibranch loop size.

a loop is less for a shorter chain, the stacked configuration is not only the low enthalpy state it is also the more stable state of the loop.

To get a sense of the magnitude of this surprising loop entropy stabilization, consider a typical tRNA (see Fig. 1) with a ($N = 12$, $M = 4$) multibranch loop. After coaxial stacking of the T Ψ CG loop and the acceptor stem, and orienting the stems leading to the D and anticodon loops, $M_{\text{eff}} = 0$. The free energy benefit is

$$G_{\text{FJC2}}(11, 0) - G_{\text{FJC2}}(12, 4) \approx -1.0 \text{ kcal/mol.} \quad (3)$$

Our M_{eff} benefit comes in addition to the conventional stacking free energy (which we treat identically to Mfold's *efn2*); its origin is the entropic benefit of shortening a loop. It is known from rubber elasticity that there are more chain configurations when the two ends are near. Stacking reduces the separation of dashed and dotted lines in Figure 1, so the stacked state has the greatest chain entropy. Terminal stacking of unpaired bases adjacent to helices could be treated in an analogous manner, but the magnitude of the effect is much smaller because it relates primarily to the shorter length scale; consequently we did not incorporate this at this stage.

Simulations

We also compute the probability of loop formation by performing FJC2 simulations. The parameters of the simulations, $a = 6.2 \text{ \AA}$ and $b = 15 \text{ \AA}$, come from analyzing the 4' carbon coordinates (Aalberts and Hodas 2005) of PDB files. To enforce self-avoidance we use hard-core beads with radius $r = 2.4 \text{ \AA}$, consistent with our observation that adjacent $C_{4'}$ -to- $C_{4'}$ unit vectors satisfy $\hat{n}_i \cdot \hat{n}_{i+1} > -0.7$ and with Turner's requirement of at least three bases ($N > 4$ segments) for a hairpin loop. Helix lengths contain interior beads (see Fig. 2). At least one single-stranded link always separates helices (see Fig. 1), so consecutive steps of length b are prohibited. Loops with $N < 3$ at $M = 0$, $N < 4$ at $M = 1$, or $N < M$ are all disallowed geometrically.

We compute the loop probability $p_{\text{loop}}(N, M)$, in simulations with 10^8 chains. After M steps of length b and $(N - 1)$ steps of length a , chosen at random, we measure the probability density of end separation in $0.1 \text{ \AA}$ bins, using the average bin density around end-to-end separation a . The Boltzmann formula,

$$G_{\text{sim}}(N, M) = -kT \ln p_{\text{loop}}(N, M), \quad (4)$$

again converts probability to free energy.

FIGURE 2. In simulations, the backbone is represented as hard-core beads of radius $r = 2.4 \text{ \AA}$. For single-stranded links the separation is $a = 6.2 \text{ \AA}$. For helix-crossing links the separation is $b = 15 \text{ \AA}$, with (hatched) intermediate beads to provide self-avoidance. A hairpin loop ($M = 1$ helical lengths) with three bases ($N = 4$ single-strand links) is depicted.

RESULTS AND DISCUSSION

Multibranch loops

We begin with the interesting case of multibranch loops with N single-stranded segments (or $n = N - M$ bases) and $M > 2$ stems. Our polymer-physics-based formula, Equation 2, correlates beautifully with simulations for a range of N , M values (see Fig. 3C).

The Mfold approach to multibranch loops is rather arbitrary and peculiar. In the initial Mfold run, a linear function is used

$$G_{\text{affine}} = (3.4 + 0n + 0.4M) \text{ kcal/mol.} \quad (5)$$

Mfold uses Equation 5 to generate a set of optimal and suboptimal folds.

Mfold then recomputes the free energies of the set of folds using the *efn2* function:

$$\begin{aligned} G_{\text{efn2}}(n \leq 6) &= (10.1 - 0.3n - 0.3M) \text{ kcal/mol} \\ G_{\text{efn2}}(n \geq 6) &= (8.3 - 0.3M) \text{ kcal/mol} + 1.75kT \ln(n/6). \end{aligned} \quad (6)$$

For the multibranch free energy, Mfold reports the *efn2* value from Equation 6, along with any free energy changes due to terminal mismatches or dangling bases.

It is clear from Figure 3 that our multibranch formula is in excellent agreement with simulations, while both Mfold approaches are poorly correlated. The standard errors per data point are: $s = 0.08 \text{ kcal/mol}$ for G_{FJC2} , $s = 0.28 \text{ kcal/mol}$ for G_{linear} , $s = 0.37 \text{ kcal/mol}$ for G_{affine} , and $s = 0.80 \text{ kcal/mol}$ for G_{efn2} . Incidentally, Zhang et al. (2008) and Cao and Chen (2005) use extrapolation formulas with multiple adjustable constants for each M . In our formula, the disparity of a and b lengths elegantly explains why chains with larger M have higher free energy and lower apparent slope with respect to changes in N .

Because our polymer theory uses exponents derived for the long chain limit, while these are relatively short chains, some systematic errors are observed. Altering the exponents could improve agreement slightly, but with errors already reduced to the same 0.1 kcal/mol range as base-pair stacking parameters, we prefer to avoid introducing additional fitting parameters for marginal improvement.

It is also not surprising that G_{FJC2} deviates the most from G_{sim} for very short hairpins ($M = 1$ and $N = 4, 5$), as self-avoidance greatly stretches these loops. Corrections to the free energies of stretched chains can be introduced (Marko and Siggia 1995), but to avoid complication we simply use G_{FJC2} rather than including corrections for the observed differences between G_{sim} and G_{FJC2} . Andronescu et al. (2007) have achieved improvements in prediction accuracy simply by reparametrizing the Turner model.

FIGURE 3. As a function of G_{sim} , for $N \leq 20$ and $M \leq 6$, we plot: (A) the Mfold initial run (Equation 5); (B) the re-evaluated free energies of Mfold's *efn2* (Equation 6); (C) our G_{FJC2} formula (Equation 2); and (D) the best linear fit to simulation data $G_{\text{linear}} = (4.3 + 0.07N + 0.2M)\text{kcal/mol}$. It is clear that G_{FJC2} best approximates G_{sim} , demonstrating that FJC2 captures the correct polymer physics lacking in the Mfold formulas. The quality of the agreement allows us to incorporate hairpin loops on the same footing as multibranch loops and to consider coaxial stacking as a change in the effective M .

Mathews and Turner (2002) proposed $(9.3 - 0.9M + 0n)\text{kcal/mol}$ on the basis of a set of experiments on $M = 3, 4$ multibranch loops. They indicated that their observation that $M = 4$ loops are more stable than $M = 3$ loops might arise from whether M is even or odd; in our framework, this can be explained as stacked helices reducing an $M = 4$ effective loop size to $M_{\text{eff}} = 0$ but $M = 3$ only to $M_{\text{eff}} = 1$. In the limit of large M , the negative slope with respect to M is unphysical because longer chains have more configurations and are thus less likely to form loops. The Mathews and Turner (2002) formula produces a standard error of $s = 2.2\text{ kcal/mol}$ against G_{sim} . An additional term related to the asymmetry of loops was also proposed in Mathews and Turner (2002); however, the functional form of this term has prohibited implementation in folding algorithms. Asymmetry is likely a proxy for the possibility of helix stacking discussed above.

Hairpin loops

Hairpin loops contain $M = 1$ helices and N single-stranded segments (or $n = N - 1$ bases). We find that our polymer theory is also consistent with the Turner values (Mathews et al. 1999). In Figure 4, our $M = 1$ theory and simulation results are compared with experiment (Mathews et al. 1999). Turner quotes values for each n , while we believe fitting

all the data to a model context is preferable, especially when extrapolating to long loops. The smoothness of Mfold's "experimental" curve for $n > 9$ is misleading as it is a Jacobson–Stockmayer extrapolation (Jacobson and Stockmayer 1950), while the actual experiments are for $3 \leq n \leq 9$. Our smooth theory and simulation curves generally fall within the large error bars of Turner.

Testing G_{FJC2} with Mfold

We test the performance of Mfold with its hairpin ($M = 1$) and multibranch ($M > 2$) loop free energies replaced by G_{FJC2} . The interior or bulge loop ($M = 2$) free energies were not altered because sequence dependencies for mismatched bases have been tabulated. For hairpins, the first mismatch and the sequence-specific tetraloop and triloop free energies were also retained. The values of C for the multibranch and hairpin forms of G_{FJC2} are set to 0.0 kcal/mol and -1.1 kcal/mol , respectively; these values ensure that G_{FJC2} hairpin and multibranch free energies are similar to their *efn2* counterparts. This similarity is important for the maintenance of relative free energy levels among the different secondary structure motifs and prevents systematic biasing toward certain types of motifs.

No Mfold default settings were changed except for the percent-suboptimality parameter, which was increased from $P = 5$ to $P = 30$ to generate a set of suboptimal folds within 30% of the minimum free energy. (For the relatively short hammerhead ribozyme sequences, $P = 100$.) This set of optimal and suboptimal folds, based on Equation 5, is rank ordered by free energy. The free energies for these Mfold affine structures are recomputed with *efn2*, Equation 6. We do the same with our G_{FJC2} formula, Equation 2. This procedure facilitates the most direct comparison of results. Because recomputing with *efn2* or G_{FJC2} does not predict any

FIGURE 4. The free energies of hairpin loops as a function of the number of bases n are given for Mfold/Turner rules (Mathews et al. 1999); for Monte Carlo simulations (Equation 4); and our theoretical expression (Equation 2) with $C = -0.8\text{ kcal/mol}$. Note that Mfold values for $n > 9$ are inferred from the Jacobson–Stockmayer theory using $n = 9$ as a reference point even though the data are not smooth for $n \leq 9$. The experimental errors for loop energies are much larger than the $\pm 0.1\text{ kcal/mol}$ for base pairs.

TABLE 1. The number of times the “most accurate” structure is predicted to have the minimum free energy (MFE) is compared between the models for five RNA classes

RNA Class	MFE = most accurate			Size of data set	Difference of proportions $p_{\text{FJC2}} - p_{\text{efn2}}$
	G_{FJC2}	<i>efn2</i>	affine		
tRNA	245	203	178	569	7.4% $\pm$ 5.7%
5S	220	221	205	309	-0.3% $\pm$ 7.1%
SRP	209	217	202	369	-2.2% $\pm$ 7.1%
Hammerhead	52	49	52	66	4.6% $\pm$ 14.4%
<i>cis</i> -reg	14	14	15	41	0.0% $\pm$ 20.5%
Net result	740	704	652	1354	2.6% $\pm$ 3.8%

The difference in proportions $p_{\text{FJC2}} - p_{\text{efn2}}$ with 95% confidence intervals (Agresti 1996) is also reported. Although only the tRNA class shows statistically significant gain, we note a net increase in performance overall.

new structures, changes in the accuracy are manifested as changes in the rank ordering of predicted folds.

We first tested using 569 tRNA sequences, previously used in a study by Gutell and coworkers (Doshi et al. 2004). This set of sequences represents a diverse collection of tRNAs corresponding to different amino acids in a wide variety of organisms. The Gutell group also makes available comparatively determined structure information for every sequence. Modified bases are represented as *N* and, as such, are not allowed to pair.

Each structure from the Mfold list is compared with Gutell’s “correct” structure. The “most accurate” structure is the one with the most correct base pairs; but while this structure is the same for G_{FJC2} , affine, and *efn2*, its rank among the list may change as the relative energies of the folds change. The number of “most accurate” structures at the optimal position shows a statistically significant increase from 203 with *efn2* to 245 with G_{FJC2} (see Table 1).

We further tested a large set of 5S ribosomal RNA sequences used in a previous evaluation of Mfold (Mathews et al. 1999), a set of SRP RNA from the SRP database (Andersen et al. 2006), and hammerhead ribozyme and *cis*-regulatory RNA elements from the Rfam database (Griffiths-Jones et al. 2005). The SRP, hammerhead, and *cis*-regulatory sequences were obtained via the RNA strand website (Andronescu et al. 2008). Performance of G_{FJC2} , measured by the percentage of times the “most accurate” structure for each sequence appears with the lowest free energy, remains similar to that of Mfold for these classes (see Table 1).

CONCLUSIONS

Our major conclusions are these: (1) RNA loops are built from two different lengths, so we introduce the FJC2 model. We encourage others to respect the *a/b* ratio in drawing two-dimensional representations. (2) The free energy of RNA loops can be derived from polymer physics principles (Equation 2). The 10 parameters for hairpins and multi-branch loops in the Turner model reduce to only two

adjustable constants in our model. (3) G_{FJC2} (*N*, *M*) is in good agreement with experiment and simulation (see Figs. 3, 4), while the Mfold formulations for multibranch loops are poor. (4) Helix stacking lowers the enthalpy as expected, but surprisingly also increases the likelihood of loop formation, because there are relatively more loop configurations for shorter effective chains. We suspect that the previously reported effect of asymmetry within loops (Mathews and Turner 2002; Zhang et al. 2008) can be understood better in terms of the entropic benefit of coaxial stacking. (5) We have shown that FJC2 produces a statistically significant increase in the accuracy of tRNA predictions while performing on par with Mfold for other RNA classes (see Table 1). (6) Our improved loop free energy function can readily be calculated if the secondary structure is known. A linear approximation of our formula as in Figure 3D, along with an M_{eff} correction, can be incorporated in all of the RNA folding algorithms based on the Turner model. (7) We believe that the FJC2 formalism is a good springboard for including stacking of single-stranded bases (Aalberts et al. 2003) in the future.

ACKNOWLEDGMENTS

We thank Dave Tucker-Smith, Dick DeVeaux, Andrea Danyluk, and Bernhard Klingenberg for discussions. This work was supported by grants from the National Science Foundation (Grant No. MCB-0641995) and the National Institutes of Health (Grant No. GM080690).

Received July 16, 2009; accepted March 22, 2010.

REFERENCES

- Aalberts DP, Hodas NO. 2005. Asymmetry in RNA pseudoknots: Observation and theory. *Nucleic Acids Res* **33**: 2210–2214.
- Aalberts DP, Parman JM, Goddard NL. 2003. Single-strand stacking free energy from DNA beacon kinetics. *Biophys J* **84**: 3212–3217.
- Agresti A. 1996. *An introduction to categorical data analysis*. Wiley, New York.
- Andersen ES, Rosenblad MA, Larsen N, Westergaard JC, Burks J, Wower IK, Wower J, Gorodkin J, Samuelsson T, Zwieb C. 2006. The tmRDB and SRPDB resources. *Nucleic Acids Res* **34**: D163–D168. doi: 10.1093/nar/gkj142.
- Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. 2007. Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics* **23**: i19–i28.
- Andronescu M, Bereg V, Hoos H, Condon A. 2008. RNA STRAND: The RNA secondary structure and statistical analysis database. *BMC Bioinformatics* **9**: 340. doi: 10.1186/1471-2105-9-340.
- Cao S, Chen SJ. 2005. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA* **11**: 1884–1897.
- Chen S-J, Dill KA. 2000. RNA folding energy landscapes. *Proc Natl Acad Sci* **97**: 646–651.

- Diamond JM, Turner DH, Mathews DH. 2001. Thermodynamics of three-way multibranch loops in RNA. *Biochemistry* **40**: 6971–6981.
- Ding Y, Lawrence CE. 2003. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* **31**: 7280–7301.
- Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR. 2004. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**: 105. doi: 10.1186/1471-2105-5-105.
- Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: Annotating noncoding RNAs in complete genomes. *Nucleic Acids Res* **33**: D121–D124.
- Grosberg A, Khokhlov A. 1994. *Statistical physics of macromolecules*. AIP, New York.
- Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**: 3429–3431.
- Jacobson H, Stockmayer WH. 1950. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J Chem Phys* **18**: 1600–1606.
- Kim J, Walter AE, Turner DH. 1996. Thermodynamics of coaxially stacked helices with GA and CC mismatches. *Biochemistry* **35**: 13753–13761.
- Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512.
- Marko JF, Siggia ED. 1995. Stretching DNA. *Macromolecules* **28**: 8759–8770.
- Mathews DH, Turner DH. 2002. Experimentally derived nearest-neighbor parameters for the stability of RNA three- and four-way multibranch loops. *Biochemistry* **41**: 869–880.
- Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911–940.
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* **101**: 7287–7292.
- Olson WK. 1975. Configurational statistics of polynucleotide chains. A single virtual bond treatment. *Macromolecules* **8**: 272–275.
- Olson WK, Flory PJ. 1972. Spatial configurations of polynucleotide chains. I. Steric interactions in polyribonucleotides: A virtual bond model. *Biopolymers* **11**: 1–23.
- Tyagi R, Mathews DH. 2007. Predicting helical coaxial stacking in RNA multibranch loops. *RNA* **13**: 939–951.
- Xia T, Santalucia J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37**: 14719–14735.
- Zhang W, Chen SJ. 2002. RNA hairpin-folding kinetics. *Proc Natl Acad Sci* **99**: 1931–1936.
- Zhang J, Lin M, Chen R, Wang W, Liang J. 2008. Discrete state model and accurate estimation of loop entropy of RNA secondary structures. *J Chem Phys* **128**: 125107. doi: 10.1063/1.2895050.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res* **31**: 3406–3415.
- Zuker M, Sankoff D. 1984. RNA secondary structures and their prediction. *Bull Math Biol* **46**: 591–621.
- Zuker M, Mathews D, Turner D. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In *RNA biochemistry and biotechnology* (ed. J Barciszewski, BFC Clark), pp. 11–43. Kluwer Academic, Dordrecht, The Netherlands.

RNA

A PUBLICATION OF THE RNA SOCIETY

A two-length-scale polymer theory for RNA loop free energies and helix stacking

Daniel P. Aalberts and Nagarajan Nandagopal

RNA 2010 16: 1350-1355 originally published online May 26, 2010

Access the most recent version at doi:[10.1261/rna.1831710](https://doi.org/10.1261/rna.1831710)

References

This article cites 27 articles, 6 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/16/7/1350.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to RNA go to:
<http://rnajournal.cshlp.org/subscriptions>