

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

Daniel P. Aalberts and Nagarajan Nandagopal
Physics Department, Williams College,
Williamstown, MA 01267
RNA, in press (2010).

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.

I. INTRODUCTION

Accurate prediction of macromolecular structure from 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* WatsonCrick/WatsonCrick-type [17], contributes the great majority of the total free energy. Tertiary interactions such as non-local hydrogen bonds, divalent-counterion stabilization, or helix stacking account for the remainder.

The Turner rules [21, 26] are the basis for many RNA secondary structure prediction algorithms including MFOLD [29], ViennaRNA [14], Sfold [10], and others. RNAstructure [19] has systematically incorporated revisions to the rules [9, 16, 20, 22, 25]. 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 [30], “Because so little is known about the effects of multi-branch 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 [31]. Several statistical-mechanical approaches [8, 28] have since rejected this unphysical, albeit computationally efficient, treatment of multibranch loops.

The entropy of RNA loops have been computed recently by enumerating backbone configurations to simulate loop regions [7, 27] based on a virtual bond representation of the backbone [23, 24]. 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 paper, 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 which appear in the Turner model or enumerative simulations. The standard errors for multibranch loops are ten times smaller for our model than for MFOLD’s when compared to 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 [19]. 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.

II. METHODS

A. Polymer Theory

One key observation is that RNA loops have two length scales: $a = 6.2 \text{ \AA}$ for monomer separation in single-strand regions, and $b = 15 \text{ \AA}$ to cross a helix [1]. 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 $\langle \hat{\mathbf{n}}_i \rangle = 0$, $\langle \hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j \rangle = 0$ for $i \neq j$, and $\langle \hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_i \rangle = 1$. The total end-to-end separation is $\mathbf{R} = \sum_i r_i \hat{\mathbf{n}}_i$. If there are N links of length a and M links of length b ,

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

this defines the characteristic spatial extent of the chain. If there is only one segment length type, Eq. (1) reduces to the familiar FJC result [13].

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 [13] 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

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

The probability of end-to-end separation for a self-avoiding polymer adopts an asymptotic scaling form [13]:

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

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 $\frac{1}{2}$, $\frac{5}{18}$ and $\frac{3}{2}$ 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

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

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

Sequence dependencies of loops are largely ignored in the Turner rules [20, 21]; 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 [7, 27] and in our derivation of loop free energies.

B. Coaxial stacking

Although coaxial stacking refers to the tertiary organization of secondary structure elements, it is included in secondary structure folding algorithms [19, 21, 29]. If no bases intervene between two helices, they may adopt a coaxially-stacked orientation; a free energy change equal to making a conventional basepair stack is assigned in MFOLD for coaxially stacked helices [21]. If there is an unpaired base between the helices, MFOLD essentially recruits a base from the loop to make an intervening mismatch [16, 25]. Figure 1 shows examples of each type —

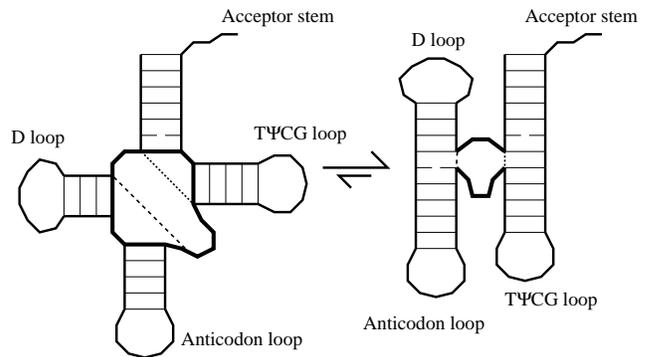


FIG. 1: The secondary structure of tRNA includes four base-paired stems meeting at a central multibranch loop. In the multibranch loop and hairpin loops, the effective backbone is comprised of a combination of N single-strand links of length a and M helix-crossing segments of length b . Coaxial stacking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

the acceptor stem and TΨCG helix stack coaxially, while the anticodon and D stems are separated by one base.

An important fact which 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 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 Figure 1) with an $(N = 12, M = 4)$ multibranch loop. After coaxial stacking of TΨCG 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.

C. Simulations

We also compute the probability of loop formation by performing FJC2 simulations. The parameters of the

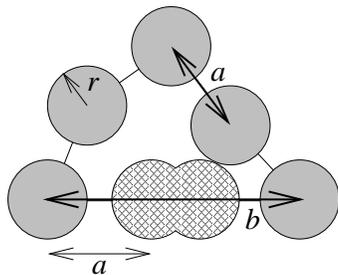


FIG. 2: In simulations, the backbone is represented as hard core beads of radius $r = 2.4$ Å. For single-stranded links the separation is $a = 6.2$ Å. For helix-crossing links the separation is $b = 15$ Å, with (hatched) intermediate beads to provide self-avoidance. A hairpin loop ($M = 1$ helical lengths)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

simulations, $a = 6.2$ Å and $b = 15$ Å, come from analyzing the $4'$ carbon coordinates [1] of PDB files. To enforce self-avoidance we use hard-core beads with radius $r = 2.4$ Å, consistent with our observation that adjacent $C_{4'}$ to $C_{4'}$ unit vectors satisfy $\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_{i+1} > -0.7$ and with Turner’s requirement of at least three bases ($N \geq 4$ segments) for a hairpin loop. Helix lengths contain interior beads, see Figure 2. At least one single-stranded link always separates helices (see Figure 1), so consecutive steps of length b are prohibited. Loops with $N < 3$ at $M = 0$, or $N < 4$ 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 Å 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.

III. RESULTS AND DISCUSSION

A. 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 Eq. (2) correlates beautifully with simulations for a range of N, M values, see Figure 3(c).

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 Eq. (5) to generate a set of optimal and suboptimal folds.

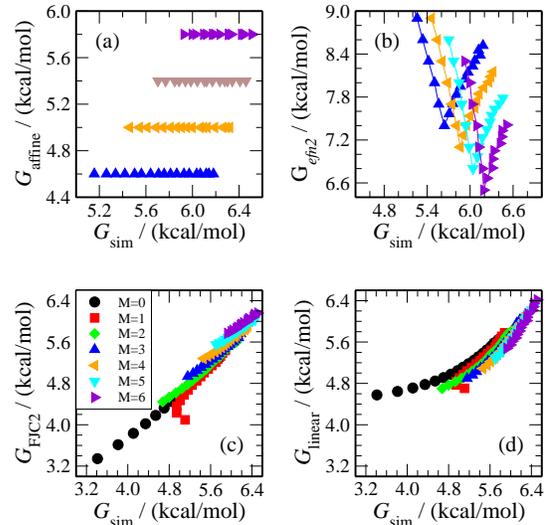


FIG. 3: As a function of G_{sim} , for $N \leq 20$ and $M \leq 6$, we plot: (a) the MFOLD initial run [Eq. (5)], (b) the reevaluated free energies of MFOLD’s *efn2* [Eq. (6)], (c) our G_{FJC2} formula [Eq. (2)], and (d) the best linear fit to simulation data $G_{\text{linear}} = (4.3 + 0.07N + 0.2M)$ 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 .

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

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

For the multibranch free energy, MFOLD reports the *efn2* value from Eq. (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 error per data point is: $s = 0.08$ kcal/mol for G_{FJC2} , $s = 0.28$ kcal/mol for G_{linear} , $s = 0.37$ kcal/mol for G_{affine} , and $s = 0.80$ kcal/mol for G_{efn2} . Incidentally, Zhang et al. [27] and Cao and Chen [7] 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 [18], 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.* [6] have achieved improvements in prediction accuracy simply by re-parametrizing the Turner model.

Mathews and Turner [22] proposed $(9.3 - 0.9M + 0n)$ 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 [22] formula produces a standard error of $s = 2.2$ kcal/mol against G_{sim} . An additional term related to the asymmetry of loops was also proposed in Mathews and Turner [22]; 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.

B. 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 [21]. In Figure 4, our $M = 1$ theory and simulation results are compared with experiment [21]. 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 [15], 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.

C. 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 dependences for mismatched bases have been tabulated. For hairpins, the first mismatch and the sequence-specific tetraloop and triloop free energies were also retained. The val-

ues 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 towards 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 percent of the minimum free energy. (For the relatively short hammerhead ribozyme sequences, $P = 100$). This set of optimal and suboptimal folds, based on Eq (5), is rank ordered by free energy. The free energies for these MFOLD affine structures are recomputed with *efn2*, Eq (6). We do the same with our G_{FJC2} formula Eq (2). This procedure facilitates the most direct comparison of results. Because recomputing with *efn2* or G_{FJC2} does not predict any 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 *et al* [11]. 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 to 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

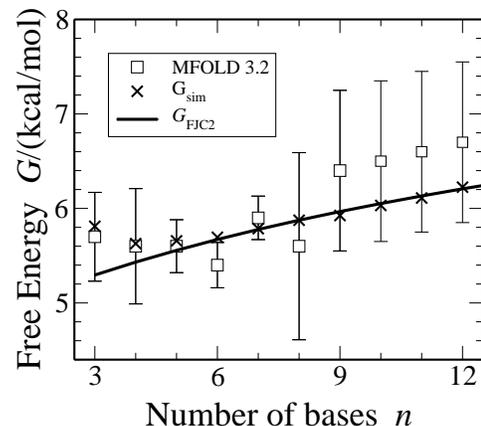


FIG. 4: The free energies of hairpin loops as a function of the number of bases is given for MFOLD /Turner rules [21], for Monte Carlo simulations Eq. (4), and our theoretical expression Eq. (2) with $C = -0.8$ kcal/mol. Note that MFOLD values for $n > 9$ are inferred from Jacobson-Stockmayer theory using $n = 9$ as a reference point even though the data is not smooth for $n < 9$. The experimental errors for loop energies are much larger than the ± 0.1 kcal/mol for base pairs.

RNA Class	MFE = most accurate			Size of Dataset	Diff. of Props. $p_{\text{FJC2}} - p_{\text{efn2}}$
	G_{FJC2}	efn2	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	740	704	652	1354	$2.6\% \pm 3.8\%$

TABLE I: 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. The difference in proportions $p_{\text{FJC2}} - p_{\text{efn2}}$ with 95% confidence intervals [3] is also reported. Although only the tRNA class shows statistically significant gain, we note a net increase in performance overall.

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 on a large set of 5S ribosomal RNA sequences used in a previous evaluation of MFOLD [21], a set of SRP RNA from the SRP database [4], and hammerhead ribozyme and *cis*-regulatory RNA elements from the Rfam database [12]. The SRP, hammerhead and *cis*-regulatory sequences were obtained via the RNA STRAND website [5]. 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).

IV. 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 (Eq 2). The ten parameters for hairpins and multibranch loops in Turner model reduce to only two adjustable constants in our model. (3) $G_{\text{FJC2}}(N, M)$ is in good agreement with experiment and simulation (see Figures 3 and 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 [22, 27] 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 3(d), along with an M_{eff} correction, can be incorporated in all of the RNA folding algorithms based on the Turner model. (7) We feel that the FJC2 formalism is a good springboard for including stacking of single-stranded bases [2] in the future.

V. FUNDING

This work was supported by grants from the National Science Foundation [grant number MCB-0641995] and the National Institutes of Health [grant number GM080690].

VI. ACKNOWLEDGEMENTS

The authors thank Dave Tucker-Smith, Dick DeVeaux, Andrea Danyluk, and Bernhard Klingenberg for discussions.

-
- [1] Aalberts, D. P. and Hodas, N. O. (2005). Asymmetry in RNA pseudoknots: observation and theory. *Nucl. Acids Res.*, 33(7):2210–2214.
 - [2] Aalberts, D. P., Parman, J. M., and Goddard, N. L. (2003). Single-strand stacking free energy from dna beacon kinetics. *Biophys. J.*, 84:3212–3217.
 - [3] Agresti, A. (1996). *A Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc., New York.
 - [4] Andersen, E. S., Rosenblad, M. A., Larsen, N., Westergaard, J. C., Burks, J., Wower, I. K., Wower, J., Gorodkin, J., Samuelsson, T., and Zwieb, C. (2006). The tmrdb and srpdb resources. *Nucl. Acids Res.*, 34(suppl.1):D163–168.
 - [5] Andronescu, M., Bereg, V., Hoos, H., and Condon, A. (2008). RNA STRAND: The RNA secondary structure and statistical analysis database. *BMC Bioinformatics*, 9(1):340.
 - [6] Andronescu, M., Condon, A., Hoos, H. H., Mathews, D. H., and Murphy, K. P. (2007). Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics*, 23(13):i19–28.
 - [7] Cao, S. and Chen, S. J. (2005). Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA*, 11(12):1884–1897.
 - [8] Chen, S.-J. and Dill, K. A. (2000). RNA folding energy landscapes. *Proc. Natl. Acad. Sci.*, 97(2):646–651.
 - [9] Diamond, J. M., Turner, D. H., and Mathews, D. H. (2001). Thermodynamics of three-way multibranch loops in RNA. *Biochemistry*, 40(23):6971–6981.
 - [10] Ding, Y. and Lawrence, C. E. (2003). A statistical sampling algorithm for RNA secondary structure prediction. *Nucl. Acids Res.*, 31(24):7280–7301.

- [11] Doshi, K. J., Cannone, J. J., Cobaugh, C. W., and Gutell, R. R. (2004). Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics*, 5.
- [12] Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucl. Acids Res.*, 33(suppl.1):D121–124.
- [13] Grosberg, A. and Khokhlov, A. (1994). *Statistical Physics of Macromolecules*. AIP Press, New York.
- [14] Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucl. Acids Res.*, 31(13):3429–3431.
- [15] Jacobson, H. and Stockmayer, W. H. (1950). Intramolecular reaction in polycondensations. i. the theory of linear systems. *J. Chem. Phys.*, 18(12):1600–1606.
- [16] Kim, J., Walter, A. E., and Turner, D. H. (1996). Thermodynamics of Coaxially Stacked Helices with GA and CC Mismatches *Biochemistry*, 35: 13753–13761.
- [17] Leontis, N. B. and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, 7(4):499–512.
- [18] Marko, J. F. and Siggia, E. D. (1995). Stretching DNA. *Macromolecules*, 28:8759–8770.
- [19] Mathews, D. H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization *RNA*, 10:1173–1190.
- [20] Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci.*, 101(19):7287–7292.
- [21] Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288(5):911 – 940.
- [22] Mathews, D. H. and Turner, D. H. (2002). Experimentally derived nearest-neighbor parameters for the stability of RNA three- and four-way multibranch loops. *Biochemistry*, 41(3):869–880.
- [23] Olson, W. K. (1975). Configurational statistics of polynucleotide chains. a single virtual bond treatment. *Macromolecules*, 8(3):272–275.
- [24] Olson, W. K. and Flory, P. J. (1972). Spatial configurations of polynucleotide chains. i. steric interactions in polyribonucleotides: a virtual bond model. *Biopolymers*, 11(1):1–23.
- [25] Tyagi, R. and Mathews, D. H. (2007). Predicting helical coaxial stacking in RNA multibranch loops. *RNA*, 13(7):939–951.
- [26] Xia, T., Santalucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998). Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with watson crick base pairs. *Biochemistry*, 37(42):14719–14735.
- [27] Zhang, J., Lin, M., Chen, R., Wang, W., and Liang, J. (2008). Discrete state model and accurate estimation of loop entropy of RNA secondary structures. *J. Chem. Phys.*, 128(12).
- [28] Zhang, W. and Chen, S. J. (2002). RNA hairpin-folding kinetics. *Proc. Natl. Acad. Sci.*, 99(4):1931–1936.
- [29] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.*, 31(13):3406–3415.
- [30] Zuker, M., Mathews, D., and Turner, D. (1999). *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide In RNA Biochemistry and Biotechnology*. NATO ASI Series. Kluwer Academic Publishers.
- [31] Zuker, M. and Sankoff, D. (1984). RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46:591-621.