

John von Neumann Institute for Computing



Loop Parameterization and RNA Secondary Structure Folding

T. R. Einert, P. Näger, H. Orland, R. R. Netz

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 201-204, 2008.

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

Loop Parameterization and RNA Secondary Structure Folding

Thomas R. Einert¹, Paul Näger¹, Henri Orland², and Roland R. Netz¹

¹ Physik Department, Technische Universität München,
85748 Garching b. München, Germany
E-mail: {einert, netz}@ph.tum.de

² Institute de Physique Théorique, CEA Saclay,
91191 Gif-sur-Yvette Cedex, France
E-mail: Henri.Orland@cea.fr

Loops are abundant in native RNA structures and proliferate close to the unfolding transition. By including a statistical weight $\sim \ell^{-c}$ for loops of length ℓ in the recursion relation for the partition function, we show that the calculated heat capacity depends sensitively on the presence and value of the exponent c , even of short tRNA. For homo-RNA we analytically calculate the critical temperature and critical exponents which exhibit a non-universal dependence on c .

We calculate the partition function of the RNA secondary structure using a formulation that allows to accurately include the statistics of terminal, internal, as well as multi-loops.¹ The statistical weight of a secondary structure depends on the free energy of base pair formation, which has been determined experimentally², but also on the entropy loss of loop formation. Polymer theory predicts the configurational weight of a loop consisting of ℓ bases to decay as ℓ^{-c} where the exponent c is universal. The loop exponent is $c_{\text{ideal}} = 3/2$ for an ideal polymer and $c_{\text{SAW}} = d\nu \simeq 1.76$ for an isolated self avoiding loop. However, helices which emerge from the loop decrease the number of configurations and consequently increase c even further.³ For instance, one obtains $c_1 = 2.06$, $c_4 = 2.16$ for the two types of loops which appear in the native structure of yeast tRNA-phe, see Fig. 1b.

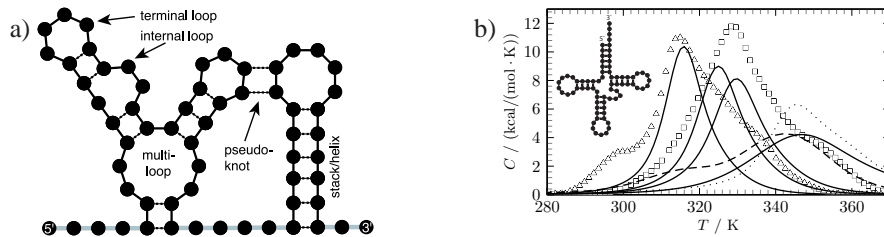


Figure 1. a) Schematic representation of a secondary RNA structure. Solid lines denote the RNA backbone, broken lines base pairs, and gray lines non-nested backbone bonds that are counted by the variable M ; here $M = 11$. b) Experimental heat capacity of the tRNA-phe of yeast for NaCl concentrations 20 mM (triangles) and 150 mM (squares).⁴ Solid lines show results using Eq. (1) with loop exponents $c = 3.0, 2.16, 1.76, 0$ (from left to right), compared with the results from the `Vienna package`⁵ which uses a linearized multi-loop entropy (dashed curve). The dotted curve is obtained with $c = 3$ and the same energy parameter set as for the solid curves, except for the loop initiation penalty which was omitted. The inset sketches the low-temperature secondary RNA structure obtained from Eq. (1) in agreement with experimental crystal-structure studies.

A valid secondary structure is a list of all base pairs, where pseudo-knots⁶ are not allowed, *i. e.* for any two base pairs (i, j) and (k, l) with $i < j$, $k < l$, and $i < k$ we have either $i < k < l < j$ or $i < j < k < l$. In our notation, the canonical partition function $Q_{i,j}^M$ of a sub-strand from base i at the 5' end through j at the 3' end depends on the non-nested backbone-length M ,^{7,8} see Fig. 1a. The recursion relations for the partition function read then

$$Q_{i,j+1}^{M+1} = Q_{i,j}^M + \sum_{k=i+M+1}^{j-N_{\text{loop}}} Q_{i,k-1}^M Q_{k,j+1}^0 \quad \text{and} \quad (1a)$$

$$Q_{k,j+1}^0 = \sum_{h=1}^{(j-k-N_{\text{loop}})/2} w_{(k+h,j+1-h)}^{(k,j+1)} \sum_{m=1}^{j-k-1-2h} \frac{Q_{k+1+h,j-h}^m}{(m+2)^c}. \quad (1b)$$

Eq. (1a) describes elongation of an RNA structure by either adding an unpaired base (first term) or by adding an arbitrary sub-strand $Q_{k,j+1}^0$ that is terminated by a helix. Eq. (1b) constructs $Q_{k,j+1}^0$ by closing structures with m non-nested bonds, summed up in $Q_{k+1+h,j-h}^m$, by a helix of length h , which is weighted with a sequence dependent Boltzmann factor w . $N_{\text{loop}} = 3$ is the minimum number of bases in a terminal loop. The unrestricted partition function of the entire RNA is given by $Z_N = \sum_M Q_{0,N}^M$. We implement the recursion relation, Eq. (1), numerically using a free energy parameter set.² In Fig. 1b we show the experimental heat capacity of the tRNA-phe of yeast compared with our predictions from Eq. (1) using $C = T\partial^2(\text{k}_B T \ln Z_N)/\partial T^2$. The heat capacity peak corresponds to the gradual melting of the secondary structure. Although the RNA consists of just 76 nucleotides and is therefore far from the thermodynamic limit where one expects asymptotic effects to be important, the loop exponent c has drastic effects. Increasing c from $c = 0$ to $c = 3$ destabilizes the structure and decreases the melting temperature by more than 30 K (solid lines). It is difficult to directly compare experimental and theoretical curves as the energy parameters were determined at 1 M NaCl concentration², while experimental data is only available at 20 mM and 150 mM. Current implementations of secondary structure prediction or partition function calculation approximate the entropy for multi-loops by an affine function $\ln(y^M M^{-c}) \approx \delta_0 + \delta_1 M$.^{5,9} This in principle corresponds to the usage of the loop exponent $c = 0$, as is corroborated by the near agreement of the results from the Vienna package⁵ (broken line) with the results from Eq. (1) using $c = 0$. Most strikingly, the melting temperature as well as the width and the height of the peaks depend dramatically on the loop exponent. Similar studies have been carried out for DNA¹⁰ where the loop exponent has a much weaker effect on denaturation curves.

In a second step, we now consider homo-polymeric RNA, which can be modeled experimentally by using a synthetic sequence like AUAU... The goal is to extract the critical asymptotic behavior embodied in Eq. (1) in the thermodynamic limit. We simply give a statistical weight $w = \exp[-\varepsilon/(\text{k}_B T)]$ to each base pair. This renders the system translationally invariant and allows to write $Q_{i,j}^M$ as Q_N^M with $N = j - i$ being the total number of backbone segments of the sub-strand ranging from i through j . This can be viewed as a coarse-graining approximation for natural or random RNA above the glass transition. To proceed, we switch to the grand canonical ensemble where we are able to study the

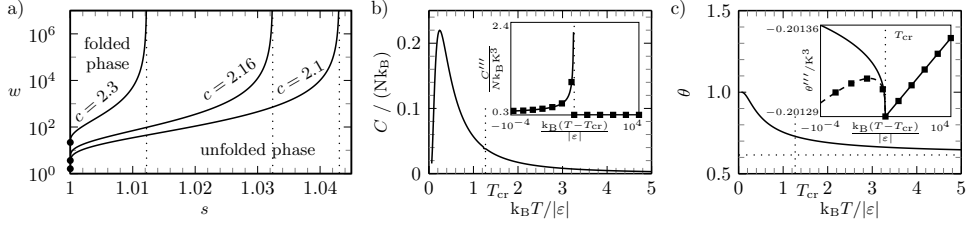


Figure 2. a) Phase diagram for three different values of the loop exponent c as a function of the base pairing weight w and force fugacity s . The dots denote the unfolding transition in the absence of external force, *i. e.* $s = 1$, which is considered in b) and c): Temperature dependence of the b) specific heat C and c) fraction of bound bases θ for $c = 2.3$. The insets show the third derivatives $C''' = d^3C/dT^3$ and $\theta''' = d^3\theta/dT^3$ which clearly exhibit singular behavior. Squares denote numerical evaluation of Eq. (2). The leading (solid lines) and next-leading (dashed line) order of the expansion around T_{cr} are shown, according to which C''' diverges with the exponent $\chi = 2/3$ for $c = 2.3$ and θ''' is characterized by the exponent $\lambda = 1/3$.

thermodynamic limit. The grand canonical partition can be calculated exactly

$$\mathcal{Z}(z, s) = \sum_{N=0}^{\infty} \sum_{M=0}^{\infty} z^N s^M Q_N^M = \frac{\kappa(z)}{1 - sz\kappa(z)}, \quad (2)$$

where $\kappa(z)$ is determined by the equation $\kappa = 1 + w/\kappa \text{Li}_c(z\kappa)$. $\text{Li}_c(x) = \sum_{n=1}^{\infty} x^n/n^c$ is the polylogarithm. The force fugacity s is $s = 1$ if no force is applied to the ends and $s > 1$ if the molecule is stretched. In Fig. 2a we show the resulting phase diagram of RNA in terms of w and s for different values of the loop exponent c . We observe a very weak phase transition only if $2 < c < c^*$ with $c^* \simeq 2.479$. For $c < 2$, the RNA is always in the folded state, whereas for $c > c^*$ RNA is always unfolded, irrespective of the temperature. The critical exponents as well as the order, which is at least four, of the phase transition depend on the loop exponent c and are calculated exactly.

The conclusion is that while the dependence of critical properties on the loop exponent c is experimentally and numerically¹¹ difficult to access and therefore largely irrelevant, the dependence of non-critical properties on c is important.

Acknowledgments

Support from the *Elitenetzwerk Bayern* within the framework of *CompInt* is gratefully acknowledged.

References

1. T. R. Einert, P. Näger, H. Orland, and R. R. Netz, *submitted to Phys. Rev. Lett.* (2008).
2. D. H. Mathews *et al.*, *J. Mol. Biol.* **288**, 911 (1999).
3. Y. Kafri, D. Mukamel, and L. Peliti, *Phys. Rev. Lett.* **85**, 4988 (2000).
4. P. L. Privalov and V. V. Filimonov, *J. Mol. Biol.* **122**, 447 (1978).
5. I. L. Hofacker *et al.*, *Mon. Chem.* **125**, 167 (1994).
6. H. Orland and A. Zee, *Nucl. Phys. B* **620**, 452, (2002).

7. R. Bundschuh and U. Gerland, *Phys. Rev. Lett.* **95**, 208104 (2005).
8. M. Müller, F. Krzakala, and M. Mézard, *Eur. Phys. J. E* **9**, 67 (2002).
9. J. S. McCaskill, *Biopolymers* **29**, 1105 (1990).
10. R. Blossey and E. Carlon, *Phys. Rev. E* **68**, 061911 (2003).
11. M. Baiesi, E. Orlandini, and A. L. Stella, *Phys. Rev. Lett.* **91**, 198102 (2003).