



## Simulation of RNA Folding on the Simple Cubic Lattice

S. Hayryan, K. G. Sargsyan, C.-K. Hu

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. 101-104, 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>

# Simulation of RNA Folding on the Simple Cubic Lattice

Shura Hayryan<sup>1</sup>, Karen G. Sargsyan<sup>2</sup>, and Chin-Kun Hu<sup>1,3</sup>

<sup>1</sup> Institute of Physics, Academia Sinica, Taipei 11529, Taiwan  
*E-mail: {shura, huck}@phys.sinica.edu.tw*

<sup>2</sup> Yerevan Physics Institute, Alikhanian Brothers 2, 375036 Yerevan, Armenia

<sup>3</sup> Center for Nonlinear and Complex Systems and Department of Physics,  
Chung Yuan Christian University, Chungli 32023, Taiwan

We consider the RNA molecule as a self-avoiding walk on the simple cubic lattice and use a modified Wang-Landau method to calculate its specific heat  $C_v$  and mean square end to end distance  $R^2$  as a function of temperature. The energy function includes hydrogen bond energy  $\epsilon$ , stacking energy  $\bar{\epsilon}$ , or chain rigidity energy  $J$ . We find that inclusion of pseudoknots or  $\bar{\epsilon}$  greatly changes the behavior of  $C_v$  and inclusion of stacking energy greatly changes the behavior of  $R^2$ .

## 1 Introduction

RNA molecule is a highly charged heteropolymer built from 4 types of nucleotide bases. Like the proteins, a RNA molecule can attain secondary and tertiary structure with the latter being function-related. It is widely believed that the folding of a RNA molecule occurs in two stages. First the elements of the secondary structure are formed, and then such elements are folded to form the tertiary structure with the energy change much smaller than that in the first stage. This mechanism is known as hierarchical folding<sup>1</sup>. The typical physical interactions in RNA molecules include complementary hydrogen bonds between A-U and G-C pairs of nucleotide bases, strong electrostatic interactions and stacking interaction between neighbor bounded pairs<sup>2,3</sup>. The important feature of hydrogen bonding between bases is that this bonding is of saturated nature meaning that once two bases are bonded, there can be no other bonds with other bases.

Most theoretical models for predicting the secondary structures of RNA sequences neglect the formation pseudoknots (see e.g. Fig. 1 in Ref. 3) and assume the hierarchical mechanism of folding. Among other theoretical approaches, the coarse grain lattice models have been applied to study the statistical behavior of RNA folding, e.g. Leoni and Vanderzande<sup>4</sup> applied so called self avoiding two-tolerant trail model to study the RNA folding on a square lattice, in which the saturated nature of the hydrogen bonds is taken into account by allowing the chain to visit each bond at most twice. The twice visited bond corresponds to the H-bonded bases. They consider the simple Hamiltonian which includes only the energy of hydrogen bonds. With this model they obtained the phase diagram of the system which includes native, coiled and branched polymer phases. The similar two-tolerant trail model is applied for simulations on face centered cubic lattices<sup>5</sup> and Husimi lattice<sup>6</sup>.

In this paper we model a RNA molecule as a self avoiding walk on the simple cubic lattice and use Monte Carlo method to study its thermodynamic behavior.

## 2 The RNA Model and Monte Carlo Moves

We consider a RNA molecule of  $N$  bases as a self-avoiding random walk of length  $N$  on a 3D simple cubic lattice. Each lattice site corresponds to one nucleotide base and can be visited only once. The lattice bonds mimic chemical links between monomers. We consider following interactions: (i) *Hydrogen bonds between two bases*. The homopolymeric approximation is assumed wherein each base can be H-bonded to any other if their distance equals to 1 lattice constant. This approximation is not principal, and the procedure described below can be easily extended to the heteropolymeric case. However, here the saturated nature of H-bonding is maintained meaning that each base is allowed to be bonded only with one other. The energy value  $\varepsilon$  is associated with each bounded base-pair. Due to the homopolymeric approximation, here the value of  $\varepsilon$  is taken as an average of the energies of A-U and G-C bonds. We use the value<sup>2</sup>  $\varepsilon = 15kcal/mol$ . (ii) *Stacking interactions between two successively H-bonded pairs*. When two neighbor base-pairs are H-bonded then an energy  $\bar{\varepsilon}$  is added to the Hamiltonian. We take<sup>2</sup>  $\bar{\varepsilon} = 0.5\varepsilon$  or  $\varepsilon$ . (iii) *Bending energy or the rigidity of the chain*. Bending of the chain is penalized by some energy  $J$ . Some unit vectors  $\vec{s}_i$  are assigned to each lattice bond to describe this term. With these basic interactions included, we have the Hamiltonian

$$H = - \sum_{(i,j) \in S} \varepsilon - \sum_{(i,j) \in S} \bar{\varepsilon} g_{i,i+1,j-1,j} - J \sum_{i=1}^{N-1} \vec{s}_i \vec{s}_{i+1}. \quad (1)$$

Here the three sums correspond to the three types of the interactions describe above. In the first and the second terms the summation goes over all base pairs  $(i, j)$ . The function  $g_{i,i+1,j-1,j}$  in the stacking energy term takes value 1 if the base pairs  $(i, j)$  and  $(i+1, j-1)$  are H-bonded simultaneously, and 0, otherwise. The sum in the third term goes over all chemical bonds. At this point we purposely don't include the electrostatic interactions which will be discussed later.

To investigate the thermodynamics of our system we use the Monte Carlo technique to sample the phase space. Starting from some initial configuration of the chain we apply the pivot algorithm<sup>7</sup> as a proposal move to the next configuration. In the initial configuration the hydrogen bonds are assigned starting from some random site first in one direction, and then in the opposite one. The basic procedure is as this: if the chosen site is not H-bonded then a random search is made between its nearest neighbors (not chemically connected to the given site). When the first non-bonded site is found then a bond is placed between it and the chosen one. If no free neighbors are found then the next site is considered. Two strategies are applied for assigning the hydrogen bonds during the Monte Carlo move: (i) When the center for pivot rotation is chosen the hydrogen bond of the given site as well as of the next and the previous sties are considered broken. After the pivot move these bonds are restored again using the procedure described above. After restoring the bonds the energy of the new configuration is calculated. (ii) All hydrogen bonds along the chain are considered broken and reconstructed anew after the pivot move. The Wang-Landau<sup>8</sup> algorithm is used for calculating the density of states. Then the average thermodynamic or the structural quantities are being calculated with the general formula

$$\langle A \rangle_T = \frac{\sum_i \langle A \rangle_{E_i} g(E_i) e^{-E_i/kT}}{\sum_i g(E_i) e^{-E_i/kT}}. \quad (2)$$

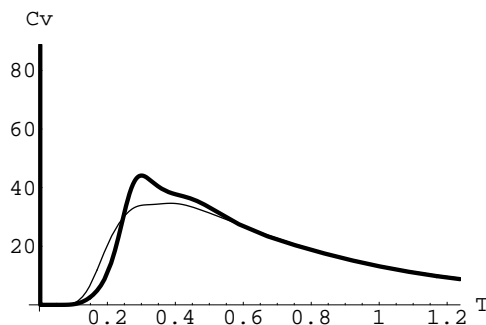


Figure 1. The specific heat  $C_v$  as a function of the temperature  $T$  (in unit  $\varepsilon/k$ ) for the system with only hydrogen bond energy and including pseudoknots (bold line) or excluding pseudoknots (light line),  $N = 100$ .

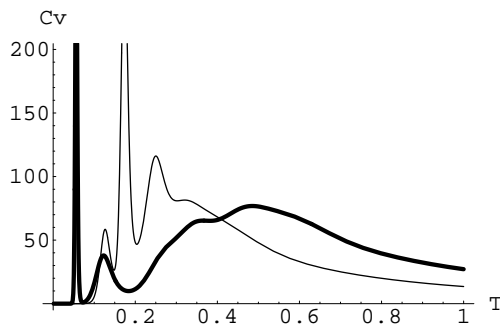


Figure 2.  $C_v$  as a function of  $T$  (in unit  $\varepsilon/k$ ) for the system with hydrogen bonding & stacking interactions of  $\bar{\varepsilon} = 0.5\varepsilon$  (bold line) or  $\bar{\varepsilon} = \varepsilon$  (light line), and without pseudoknots,  $N = 100$ .

Here  $T$  is the temperature,  $k$  is the Boltzmann constant,  $\langle A \rangle_T$  is the quantity under consideration,  $E_i$  is the energy of the state  $i$ , and  $g(E_i)$  is the density of state. In the simulations, we allow formation of the pseudoknots<sup>3</sup>.

### 3 Results

Before starting simulations with the full Hamiltonian (1), we have studied the behavior of the systems with simplified energy functions in order to get more detailed insight. Due to the limited space we bring here the results for only these simplified systems.

*The system with the hydrogen bond energy only.* This corresponds to the first sum in (1). The simulated results of specific heat  $C_v$  as a function of the temperature  $T$  (in unit  $\varepsilon/k$ ) for the system with pseudoknots (bold line) or without pseudoknots (light line) are plotted in Fig. 1, which shows a well defined compaction transition; the transition is sharper in the system with pseudoknots. The same transition is seen also on the graph of the mean squared end-to-end distance (not shown here).

*Hydrogen bonds plus stacking energy.* This corresponds to the sum of the first two terms in (1). In this case (Fig. 2) the specific heat is not single-peaked. While the small

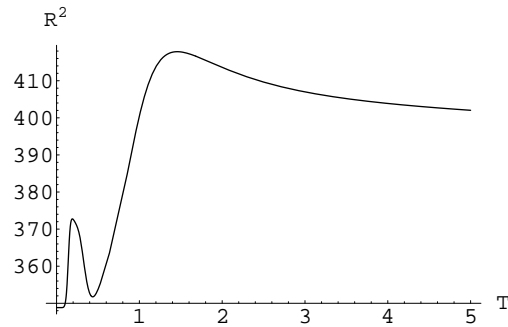


Figure 3. The mean square end to end distance  $R^2$  vs.  $T$  (in unit  $J/k$ ) with  $\bar{\epsilon} = 0$ ,  $J = 2kJ/mol$  and  $N = 100$ .

peak at the left side of the main one might be non-essential (within the sampling error) the one at the right side can not be overlooked; it might be the manifestation of two-step RNA folding in the hierarchical folding mechanism. Figure 2 shows that changing the stacking energy affects considerably the position of the peak.

*Hydrogen bonds plus chain rigidity.* This corresponds to the sum of the first and third terms in (1), where  $J$  may be estimated by investigating the behavior of the correlation length for the Hamiltonian at room temperature. Some estimates show that the correlation length is not larger than several thousands of monomers which gives the value between  $0.5kJ/mol$  and  $4kJ/mol$ . For this case we plot the mean squared end-to-end distance as a function of  $T$  in Fig. 3 with  $J = 2kJ/mol$ . One can see that even such a small rigidity brings a strong frustration into the system; the small peak on the graph manifests the existence of glassy structure at low temperatures.

## Acknowledgments

This work was supported by Grants NSC 96-2911-M 001-003-MY3 & AS-95-TP-A07, and National Center for Theoretical Sciences in Taiwan.

## References

1. I. Tinoco and C. Bustamante, *J Mol Biol* **263**, 271, (1999).
2. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
3. Y. S. Mamasakhlisov, S. Hayryan, V. F. Morozov, and C.-K. Hu, *Phys. Rev. E* **75**, 061907 (2007).
4. P. Leoni and C. Vanderzande, *Phys Rev E* **68**, 051904 (2003).
5. M. Baiesi, E. Orlandini and A. L. Stella, arXiv:cond-mat/0303226v1.
6. R. A. Zara and M. Pretti, *Physica A* **371**, 88 (2006).
7. N. Madras, G. Slade, *The self-avoiding walk* (Birkhäuser, Boston, 1993).
8. F. G. Wang and D. P. Landau, *Phys Rev Lett* **86**, 2050 (2001); *Phys Rev E* **64**, 056101 (2001).