

Research Article

The Globular State of the Single-Stranded RNA: Effect of the Secondary Structure Rearrangements

Zareh A. Grigoryan¹ and Armen T. Karapetian²

¹Goris State University, Avangard 4, 3204 Goris, Armenia

²Yerevan State University of Architecture and Construction, Teryan 105, 0009 Yerevan, Armenia

Correspondence should be addressed to Armen T. Karapetian; atkarapetian@gmail.com

Received 11 May 2015; Revised 6 July 2015; Accepted 6 July 2015

Academic Editor: Ben Berkhout

Copyright © 2015 Z. A. Grigoryan and A. T. Karapetian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The mutual influence of the slow rearrangements of secondary structure and fast collapse of the long single-stranded RNA (ssRNA) in approximation of coarse-grained model is studied with analytic calculations. It is assumed that the characteristic time of the secondary structure rearrangement is much longer than that for the formation of the tertiary structure. A nonequilibrium phase transition of the 2nd order has been observed.

1. Introduction

Single-stranded RNA (ssRNA) plays a central role in molecular biology. In addition to transmitting genetic information from DNA to proteins, RNA molecules participate actively in a variety of cellular processes. Examples are translation (rRNA, tRNA, and tmRNA), editing of mRNA, intracellular protein targeting, nuclear splicing of pre-mRNA, and X-chromosome inactivation. The RNA molecules involved in these processes do not code for proteins but act themselves as functional products. In addition, some RNA molecules prepared *in vitro* can bind to specific molecules such as ATP. In all these cases, the information encoded in the sequence of nucleotide bases of each RNA molecule determines its functional tertiary structure.

The forces which stabilize the secondary structure of RNA are stronger than interactions responsible for the tertiary structure and hence these two structures are characterized by two different energy scales. According to one of the currently accepted concepts of RNA folding the secondary structure elements, such as helices and loops hairpins, are formed first and then stack together to form a three-dimensional tertiary structure [1]. This is so-called hierarchical folding mechanism.

However, some experiments show that the folding rate of large RNA is lower than that predicted by the hierarchical

mechanism [2]. This might mean that two successive folding steps are not fully independent. The landscape of the energy function of large RNA is extremely rugged and contains multiple deep minima which act as kinetic traps in the folding pathways [2]. The molecule can remain trapped in the states distinct from the native structure for time periods even longer than the average lifetime of RNA in a living cell [3, 4]. It should be noted that excellent models for describing the RNA secondary structure formation have been developed as well [5–10]. What remains relatively poorly understood is the full path of formation of the tertiary structure and, in particular, the mutual interplay between the secondary and tertiary structures [11–15].

The secondary structure of RNA is determined by the base pairing pattern. It has been shown that the characteristic pattern of the secondary structure of RNA is a tree-like structure, formed by relatively short double-stranded helices. The hierarchical folding scenario has been studied, for example, in [14], where the folding of RNA with fixed secondary structure is described by the model of tree-like polymer when quenched random branching. In [15] the concept of annealed randomly branched polymer has been applied to study the equilibrium characteristics of RNA. The completely annealed branching patterns describe the ensemble of secondary structures, wherein the tertiary structure is being formed as a result of substantial rearrangements of secondary structure

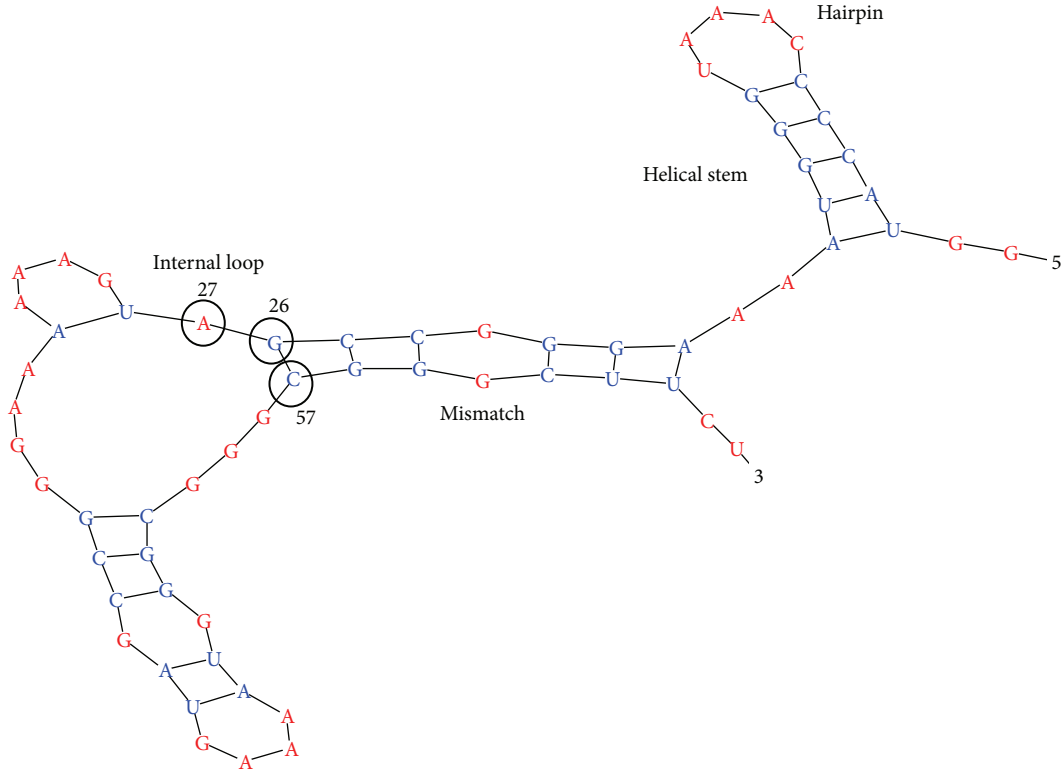


FIGURE 1: A fragment of the RNA chain with typical elements of secondary structure.

elements. This scenario is typical for large RNAs, to which the hierarchical folding mechanism is most probably not applicable. While the model of a polymer with randomly annealed branching can be applied successfully for studying the equilibrium features of RNA folding, it cannot describe the folding kinetics efficiently. At the same time, the kinetic effects in the folding process are viewed to be of great importance due to the existence of long-living intermediates, as mentioned above.

In the present study we focus on the thermodynamic behaviour of the RNA molecule in the steady nonequilibrium state that occurs in case of well-defined separation between the relaxation timescales of secondary and tertiary structures. We introduce a reasonable coarse-grained model of RNA and study its behavior through analytical equations. The obtained results provide evidence for the existence of a nonequilibrium phase transition of the second order between the glassy phase and the ensemble of freely fluctuating spatial structures.

2. The Model

Let us consider the following mental experiment. The ssRNA molecule is dissolved initially in the solvent at temperature T' which satisfies inequalities $\theta < T' < T_m$, where T_m is the melting temperature and θ is the Flory temperature. Under these conditions the RNA molecule is a random coil with a well-defined secondary structure. Next, transfer a very small amount of our RNA containing solution into the same kind of solvent but with the temperature T , such that $T < \theta$. In

the beginning, the secondary and spatial structures in this state still correspond to the temperature T' but they start to relax to the new temperature T . In the end of the process, the RNA will arrive at a compact globular state with some secondary structure pattern. The tertiary structure of RNA is stabilized by interactions between different elements of secondary structure: helical stems, hairpins, internal loops, mismatches, and so forth (Figure 1).

Interactions between helical stems can be considered as homogeneous since the nitrogen bases are located inside the double helix. At the same time, interactions between single-stranded regions are heterogeneous because of the interactions between nitrogen bases of different types. To describe the conformation of RNA in a coarse-grained approximation, we consider each nonpaired region of RNA as an effective monomer. The monomers are numbered by $i = 1, \dots, N$. The center of mass of monomer i is placed at the point with coordinates \vec{x}_i . Secondary structure is described in terms of the randomly branched polymer (Figure 2) as the matrix $\hat{B} = \|B_{ij}\|$, where $B_{ij} = 1$ if the i th and j th monomers are linked by helical stems and $B_{ij} = 0$, otherwise [14, 16]. The helices between the monomers are modelled as springs.

Then we introduce the following Hamiltonian of the model:

$$H = H_{II} + \sum_{i < j} v_{ij} \delta(\vec{x}_i - \vec{x}_j) + V_{\text{conf}}, \quad (1)$$

where the term $H_{II}(\{\vec{x}_i\}, \hat{B}) = (dT/2\ell^2) \sum_{i < j} B_{ij} (\vec{x}_i - \vec{x}_j)^2$ mimics the helical spring elasticity between the i th and j th

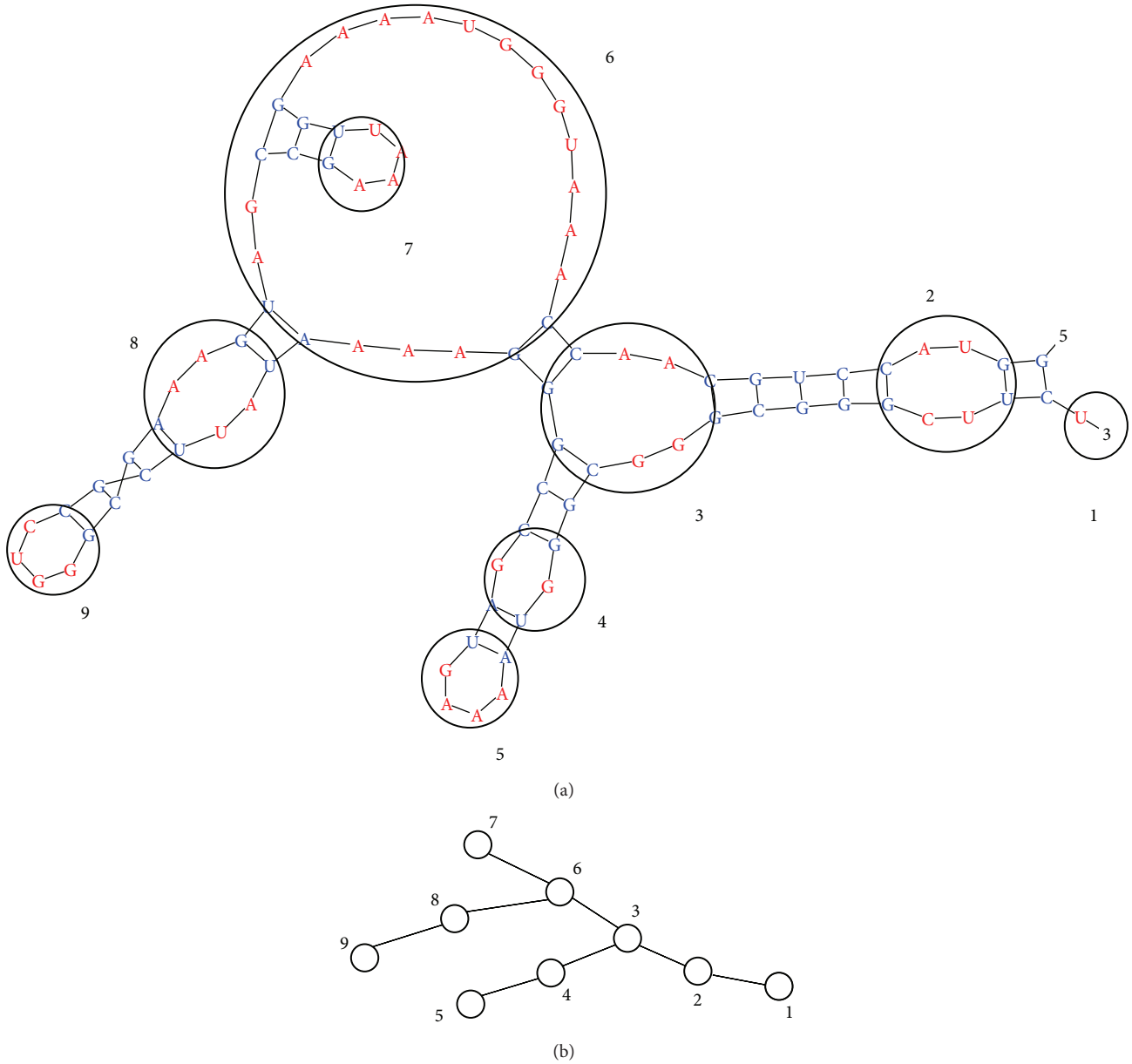


FIGURE 2: A fragment of the RNA secondary structure (a) and corresponding branched structure (b).

effective monomers and ℓ is the equilibrium distance between neighbouring monomers which here coincides with the mean length of the helical stem. Thus, the helical spring elasticity constant is assumed to be equal to $dT\ell^2$. Here v_{ij} is the second virial coefficient of interaction between the i th and j th effective monomers, which refer to the tertiary contacts between nonpaired regions (loops). $V_{\text{conf}}(\{\vec{x}_i\})$ is the confinement potential describing the homogeneous attraction between helical stems. The interactions between nonpaired regions (loops) i and j are governed by their size and nucleotide sequences (Figure 3).

Since many nucleotides contribute to the interaction of these effective monomers i and j , it is reasonable to consider coefficients v_{ij} as statistically independent random Gaussian variables with distribution

$$p(v_{ij}) \propto \exp\left(-\frac{v_{ij}^2}{2\Lambda^2}\right), \quad (2)$$

where Λ is the variance of virial coefficients v_{ij} . Collapse of the ssRNA molecule is driven mostly by electrostatic interactions and has been investigated experimentally [17, 18] and theoretically [10]. Unlike [10], here we do not take into account the counterions explicitly. Their impact is present implicitly in the confinement potential $V_{\text{conf}}(\{\vec{x}_i\}) = T(v_0 \sum_{i<j} \delta(\vec{x}_i - \vec{x}_j) + g_0 \sum_{i<j<k} \delta(\vec{x}_i - \vec{x}_j)\delta(\vec{x}_k - \vec{x}_j))$, which ensures existence of globular state of the RNA molecule. Here $v_0 < 0$ and $g_0 > 0$ are the second and third virial coefficients of interactions between helical stems, correspondingly.

Two types of conformational rearrangements in RNA are possible: rearrangement of the secondary structure with

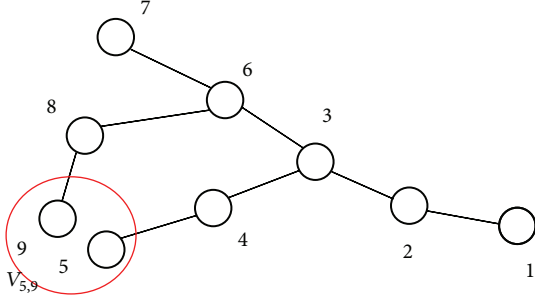


FIGURE 3: A coarse-grained model with monomers 5 and 9 interacting with the second virial coefficient $V_{5,9}$.

characteristic time scale τ_2 and tertiary structure fluctuations with characteristic time scale τ_3 . The difference between timescales τ_2 and τ_3 is well pronounced. Thus as shown in [19] the collapse of 400-nucleotide-long RNA takes about 3-4 ms while the two-order shorter 21-nucleotide sequence of RNA folds into a hairpin in about 10 ms [20]. Therefore, in further calculations it will be assumed that $\tau_3 \ll \tau_2$. The reason for this is not only the higher stability of base pairs as compared with the tertiary contacts. The kinetic factors play important role as well. Formation of the base pairs requires twisting of two single-stranded subchains into a double helix, which is kinetically hampered to unwind.

On the timescale τ , such that $\tau_3 \ll \tau \ll \tau_2$, secondary structure and spatial arrangement of the effective monomers (nonpaired regions) are not in thermal equilibrium, and this stationary nonequilibrium steady state can be described in terms of the effective partition function [21]

$$Z = \langle \langle (Z_{\hat{B}, \hat{v}})^n \rangle_{\hat{B}} \rangle_{\hat{v}}, \quad (3)$$

where $Z_{\hat{B}, \hat{v}}$ is the partition function of a branched molecule with the given branching pattern \hat{B} and interaction matrix $\hat{v} = \|v_{ij}\|$. $\langle \cdots \rangle_{\hat{B}}$ means the average over all possible branching patterns, $\langle \cdots \rangle_{\hat{v}}$ is the average over intermonomer interactions, and $n = T/T'$. T and T' are the effective temperatures of coarse-grained spatial and secondary structures correspondingly.

3. Thermodynamic Parameters

The effective partition function (3) is calculated by using the replica technique developed for systems with quenched disorder (see, e.g., [21, 22]). In our case the limit $n \rightarrow 0$ corresponds to the quenched disorder, $n = 1$ describes the completely annealed disorder, and $0 < n < 1$ for the partially annealed disorder.

Following the method described in [14] we rewrite the partition function (3) in the form

$$Z = \int D\rho e^{-\beta F\{\rho\}}, \quad (4)$$

where $\beta = 1/T$, $\rho(\vec{X}) = \sum_{i=1}^N \prod_{a=1}^n \delta(\vec{x}_i^a - \vec{x}^a)$, $\vec{X} = (\vec{x}^1, \dots, \vec{x}^n)$, and $F\{\rho\} = E\{\rho\} - TS\{\rho\}$ is the n -replica free energy with

$$E\{\rho\} = \sum_a V_{\text{conf}}(c_a) - \frac{\beta\Lambda^2}{4} \sum_{a \neq b} \int dx dy q_{ab}^2(\vec{x}, \vec{y}) \quad (5)$$

$$S\{\rho\} \cong -\frac{\ell^4}{2} \int dX \frac{(\Delta\rho(\vec{X}))^2}{\rho(\vec{X})}.$$

Here $c_a(\vec{x}) = \int dX \rho(\vec{X}) \delta(\vec{x}^a - \vec{x})$ is the one-replica density of monomers, $q_{ab}(\vec{x}, \vec{y}) = \int dX \rho(\vec{X}) \delta(\vec{x}^a - \vec{x}) \delta(\vec{x}^b - \vec{y})$ is the two-replica overlapping parameter, and Δ is the Laplace operator in the nd -dimensional space. Further in all equations we set $k_B = 1$. The energy term in (5) is obtained by averaging the n th power of the partition function over variables v_{ij} , and the entropy term includes averaging over all possible branching patterns corresponding to the rooted tree with coordination number equal to three. For function $\rho(\vec{X})$ the form $\rho(\vec{X}) = \rho_0 \exp(-(1/2) \sum k_{ab} \vec{x}^a \vec{x}^b)$ is used. The confinement potential is written as a virial expansion $V_{\text{conf}}(c) = NT((1/2)v_0c + (1/3!)g_0c^2)$. In compactly packed chain the density of monomers $c_0 \approx \ell^{-d}$ and $V_{\text{conf}}(c_0) \approx NT(v_0/4\ell^d)$. Like in [22] we parameterize the offdiagonal entries of the matrix k by function $k(u)$, where $u \in [n, 1]$, and the diagonal entries as $k_{aa} = k$. The inverse matrix $\tilde{m} = \hat{k}^{-1}$ is parameterized by $m(u)$, $u \in [n, 1]$, and $m_{aa} = \tilde{m} = \ell^2 N^{2/d}$. By introducing notations $[k](u) = uk(u) - \int_n^u dv k(v)$ and $k = \int_n^1 (du/u^2)[k](u)$, the free energy functional takes a form

$$\begin{aligned} F\{\rho\} &= nV_{\text{conf}}(c_0) \\ &+ \frac{\beta\Lambda^4}{4} nN^2 \pi^d (2\pi)^{-2d} \int_n^1 du (\tilde{m}^2 - m(u)^2)^{-d/2} \quad (6) \\ &- \frac{T\ell^4}{2} dNn \left[k^2 + \int_n^1 \frac{du}{u^2} [k](u)^2 - \bar{k}^2 \right]. \end{aligned}$$

4. Results and Discussion

Variation of the free energy (6) over \tilde{k} and $[k](u)$ as independent "variables" gives that the free energy per monomer has two branches:

$$f = \frac{F\{\rho_0\}}{Nn} = T \frac{v_0}{4\ell^d} + \begin{cases} f_{<}, & n \leq u_0 \\ f_{>}, & n > u_0 \end{cases}, \quad (7)$$

where the disordered free energy is

$$f_{<} = f_{<}^0 - \frac{n^\zeta}{\zeta} (a_1 - T\ell^4 dz_0^2), \quad (8)$$

where $n \leq u_0$, $\zeta = (3d + 4)/(4 - d)$, and $a_1(d, \beta, \Lambda, \ell) \propto \beta^{-1}(\beta\ell^{-d})^{1/\nu(d)}$. $f_{<}^0 = f_{<}^0(T, d, \Lambda, \ell)$ is the free energy of ssRNA with frozen secondary structure at $n \leq u_0$. At the same time, for $n \geq u_0$ the disordered free energy can be written as

$$f_{>} = f_{>}^0 - n \left(a_1 u_0^{\zeta-1} + T\ell^4 dz_0^2 u_0^{\zeta+1} \frac{1}{n^2} \right), \quad (9)$$

where $f_{>}^0 = f_{>}^0(T, d, \Lambda, \ell)$. Here the following notations are introduced: $z_0^2(d, \beta, \Lambda, \ell) \propto (\beta \Lambda \ell^{-d})^{1/\gamma(d)}$, $\gamma(d) = (4 - d)/8$, $\delta(d) = (d + 4)/(4 - d)$, and $u_0(d) = (3/2)((3(d/2 + 1) + (d + 1)(d/2 - 1))/(d + 2)^2) \approx 0.6$ if $d = 3$.

Thus, at $n < u_0$ a glassy phase with replica-symmetry breaking is observed while at values $n > u_0$ a replica symmetric phase is obtained.

The second derivatives of the free energy branches (8) and (9) with respect to variable n at the point of transition $n = u_0$ satisfy equations

$$\begin{aligned} \left. \frac{\partial^2 f_{<}}{\partial n^2} \right|_{n=u_0} &= -(\zeta - 1) u_0^{\zeta-2} (a_1 - T \ell^4 dz_0^2) \\ \left. \frac{\partial^2 f_{>}}{\partial n^2} \right|_{n=u_0} &= -2u_0^{\zeta-2} T \ell^4 dz_0^2 \end{aligned} \quad (10)$$

while the first derivatives are as follows: $(\partial f_{<}/\partial n)|_{n=u_0} = (\partial f_{>}/\partial n)|_{n=u_0} = -u_0^{\zeta-1} (a_1 - T \ell^4 dz_0^2)$. Thus, in the nonequilibrium state with $0 < n < 1$ the model exhibits nonequilibrium phase transition of the second order. The temperature of transition is

$$T'_c = \frac{T}{u_0}. \quad (11)$$

In the framework of the proposed model the value of the parameter $1 - n$ serves a measure of the distance from the equilibrium state, corresponding to the $n = 1$. If RNA molecule is far enough from the equilibrium ($n < u_0$), then the glassy phase is realized which is dominated by a few long-lived intermediates, which were observed experimentally in [23].

5. Conclusion

The interplay between the secondary structure formation and fast collapse of the single-stranded RNA is addressed in terms of the model with interaction between heterogeneous nonpaired and homogeneous double-stranded regions. Thus, the nucleotide sequence heterogeneity is approximately described in terms of statistical-mechanical model with disorder [22]. The memory effects in RNA compact structure formation are governed by slow rearrangements of the secondary structure with subsequent fast relaxation of the spatial degrees of freedom. Under these conditions, the mutual equilibration of fast and slow variables is hindered. Monomers rapidly attain their equilibrium at the temperature T and thus, because of the wide timescale gap, their equilibrium free energy acts as a driving force pushing the slow dynamics of the elements of secondary structure to reach a nonequilibrium stationary state at long times. This scheme is known generally as the adiabatic elimination of fast variables [24–26]. The observed experimentally [23] long-lived intermediates are obtained if RNA molecule is far enough away from the state of thermodynamic equilibrium.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] I. Tinoco Jr. and C. Bustamante, “How RNA folds,” *Journal of Molecular Biology*, vol. 293, no. 2, pp. 271–281, 1999.
- [2] D. K. Treiber and J. R. Williamson, “Exposing the kinetic traps in RNA folding,” *Current Opinion in Structural Biology*, vol. 9, no. 3, pp. 339–345, 1999.
- [3] X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu, “Correlating structural dynamics and function in single ribozyme molecules,” *Science*, vol. 296, no. 5572, pp. 1473–1476, 2002.
- [4] S. Wood, A. R. Ferré-D’Amaré, and D. Rueda, “Allosteric tertiary interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding,” *ACS Chemical Biology*, vol. 7, no. 5, pp. 920–927, 2012.
- [5] R. Bundschuh and T. Hwa, “Statistical mechanics of secondary structures formed by random RNA sequences,” *Physical Review E*, vol. 65, no. 3, Article ID 031903, 2002.
- [6] F. Krzakala, M. Mzard, and M. Mueller, “Secondary structure of RNA under tension,” *Europhysics Letters*, vol. 57, pp. 752–758, 2002.
- [7] M. Lässig and K. J. Wiese, “Freezing of Random RNA,” *Physical Review Letters*, vol. 96, no. 22, Article ID 228101, 2006.
- [8] M. Bon, G. Vernizzi, H. Orland, and A. Zee, “Topological classification of RNA structures,” *Journal of Molecular Biology*, vol. 379, no. 4, pp. 900–911, 2008.
- [9] T. R. Einert, P. Näger, H. Orland, and R. R. Netz, “Impact of loop statistics on the thermodynamics of RNA folding,” *Physical Review Letters*, vol. 101, no. 4, Article ID 048103, 2008.
- [10] M. Bon, C. Micheletti, and H. Orland, “McGenus: a Monte Carlo algorithm to predict RNA secondary structures with pseudoknots,” *Nucleic Acids Research*, vol. 41, no. 3, pp. 1895–1900, 2013.
- [11] Y. S. Mamasakhlisov, S. Hayryan, V. F. Morozov, and C.-K. Hu, “RNA folding in the presence of counterions,” *Physical Review E*, vol. 75, no. 6, Article ID 061907, 2007.
- [12] A. Šiber and R. Podgornik, “Nonspecific interactions in spontaneous assembly of empty versus functional single-stranded RNA viruses,” *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 78, no. 5, Article ID 051915, 2008.
- [13] C. E. Hajdin, F. Ding, N. V. Dokholyan, and K. M. Weeks, “On the significance of an RNA tertiary structure prediction,” *RNA*, vol. 16, no. 7, pp. 1340–1349, 2010.
- [14] A. M. Gutin, A. Yu. Grosberg, and E. I. Shakhnovich, “Conformational entropy of a branched polymer,” *Journal of Physics A: Mathematical and General*, vol. 26, pp. 1037–1049, 1993.
- [15] D. Schwab and R. F. Bruinsma, “Flory theory of the folding of designed RNA molecules,” *Journal of Physical Chemistry B*, vol. 113, no. 12, pp. 3880–3893, 2009.
- [16] P. G. de Gennes, “Statistics of branching and hairpin helices for the dAT copolymer,” *Biopolymers*, vol. 6, no. 5, pp. 715–729, 1968.
- [17] A. R. Shultz and P. J. Flory, “Phase equilibria in polymer-solvent systems,” *Journal of the American Chemical Society*, vol. 74, no. 19, pp. 4760–4767, 1952.

- [18] R. Shiman and D. E. Draper, "Stabilization of RNA tertiary structure by monovalent cations," *Journal of Molecular Biology*, vol. 302, no. 1, pp. 79–91, 2000.
- [19] R. Das, L. W. Kwok, I. S. Millett et al., "The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the tetrahymena ribozyme," *Journal of Molecular Biology*, vol. 332, no. 2, pp. 311–319, 2003.
- [20] W. Zhang and S.-J. Chen, "RNA hairpin-folding kinetics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 1931–1936, 2002.
- [21] V. Dotsenko, S. Franz, and M. Mézard, "Partial annealing and overfrustration in disordered systems," *Journal of Physics A: Mathematical and General*, vol. 27, no. 7, pp. 2351–2365, 1994.
- [22] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*, World Scientific Publishing, Singapore, 1987.
- [23] D. K. Treiber, M. S. Rook, P. P. Zarrinkar, and J. R. Williamson, "Kinetic intermediates trapped by native interactions in RNA folding," *Science*, vol. 279, no. 5358, pp. 1943–1946, 1998.
- [24] H. Haken, *Synergetics. An Introduction*, Springer, Berlin, Germany, 1983.
- [25] H. Risken, *The Fokker-Planck Equation*, Springer, Berlin, Germany, 1989.
- [26] K. Kaneko, "Adiabatic elimination by the eigenfunction expansion method," *Progress of Theoretical Physics*, vol. 66, no. 1, pp. 129–142, 1981.