

Development and Applications of Coarse-Grained Models for RNA

Changbong Hyeon,^{*[a]} Natalia A. Denesyuk,^[b] and D. Thirumalai^[b]

Abstract: In contrast to proteins, much less attention has been focused on the development of computational models for describing RNA molecules, which are being recognized as playing key roles in many cellular functions. Current atomically detailed force fields are not accurate enough to capture the properties of even simple nucleic acid constructs. In this article, we review our efforts to develop coarse-grained (CG) models that capture the underlying physics for the particular length scale of interest. Two models are discussed. One of them is the three interaction site (TIS) model, in which each nucleotide is represented by three beads corresponding to sugar, phosphate, and base. The other is the self-organized polymer (SOP) model, in which each nucleotide is represented as a single interaction

center. Applications of the TIS model to study the complexity of hairpin formation and the effects of crowding in a shifting equilibrium between two conformations in human telomerase pseudoknot are described. The work on crowding illustrates a direct link to the activity of telomerase. We use the SOP model to describe the response of the *Tetrahymena* ribozyme to force. The simulated unfolding pathways agree well with single molecule pulling experiments. We also review predictions for the unfolding pathways for the *Azoarcus* ribozyme. The success of the CG applications to describe dynamics in RNA gives hope that more complex processes involving RNA-protein interactions can be tackled using variants of the proposed models.

Keywords: coarse-grained models · nucleotides · protein folding · ribozymes · RNA

1. Introduction

The use of models in obtaining insights into the organization of nucleic acids can be traced back to the publication in 1969 of a seminal paper^[1] outlining a detailed structure for transfer nucleic acid (tRNA) by Levitt, one of the recipients of the 2013 Nobel prize in chemistry. By assuming that conformations of different tRNA molecules are similar (homologous), and that the association of hydrophobic moieties, stabilized by hydrogen bond interactions and hydration of charged and other solvent exposed polar groups, are free-energetically favorable, Levitt predicted, with stunning accuracy, a detailed structure of tRNA. Although there has not been a great deal of attention paid to RNA structures until recently, several major discoveries have made it abundantly clear that RNA is not merely a passive carrier of information, but is involved in many cellular functions (many are most likely in the category of “unknown unknowns”). Thus, it is important not only to determine RNA structures, but also to determine how they fold (the RNA folding problem)^[2] and function in crowded cellular environments, and respond to the binding of metabolites, as occurs in the process of gene expression in bacteria. The growing importance of RNA has ushered this important class of molecules to the center stage of biology, raising questions that were once familiar only in the study of proteins. From a computational per-

spective of RNA folding, the dream would be to generate movies using atomically detailed molecular dynamics simulations, starting from different initial conditions, until RNA reaches the folded structure. For RNA, this is not yet a reality because force fields are not accurate^[3] and the folding time scales are extremely long (for large ribozymes it can exceed seconds). This forces us to use concepts rooted in polyelectrolyte theory and coarse-grained (CG) models to make much needed progress, and in the process, discover general principles of RNA folding. Here, we describe some developments along these lines that we have made over the past decade or more.

The use of CG models, based on theoretical ideas to describe essential phenomena in a number of areas, has a rich history in science. Insights into many problems in

[a] C. Hyeon
School of Computational Sciences
Korea Institute for Advanced Study
Seoul 130-722 (Korea)
e-mail: hyeoncb@kias.re.kr

[b] N. A. Denesyuk, D. Thirumalai
Biophysics Program
Institute for Physical Science and Technology
University of Maryland
College Park, MD 20742 (USA)

condensed matter physics have come from the use of minimal CG models that capture the essence of complex phenomena. Such models can sometimes be solved by purely analytical methods. However, complex problems, such as spin glasses,^[4] structural glasses,^[5–7] and a host of problems in biology, such as RNA folding^[8] and enzyme functions^[9] require computational methods covering a broad spectrum of time and length scales. With increasing computational power and continued focus on developing increasingly accurate force fields, in the spirit of pioneering works in the context of protein folding and dynamics^[10–12] and enzyme reactions, it is likely that in the not too distant future, one will be able to represent the folding of proteins in atomic detail. In fact, rapid progress has been made in obtaining folding trajectories of a few single domain proteins.^[13–15] At present, the prospects of performing atomically detailed and accurate molecular dynamics (MD) simulations for nucleic acids seems daunt-

ing. Given the paucity of predictions for RNA folding using MD simulations, there should be interest in developing CG models for nucleic acids, with a view to describing many aspects of self-assembly of RNA folding, folding of chromatin, and protein-RNA (and DNA) interactions. Although there are several examples of using simulations of coarse-grained models in condensed matter and the study of proteins, there is still a good deal of resistance to the use of CG models to make testable predictions in biology. The purpose of this article is to document a few case studies from our own research to illustrate the predictive power of CG models in the difficult area of RNA folding.

Modeling reality requires a level of abstraction, depending on the phenomenon of interest. For example, near a critical point of a fluid (or a magnet), exponents that describe the vanishing of the order parameter (density or magnetization) or divergence of the correlation length (describing density-density or spin-spin correlation function) are universal; they depend only on the spatial dimensionality (d), and are impervious to atomic details. These findings, which are rooted in the concepts of universality and renormalization groups,^[16] are also applicable to the properties of polymers.^[17] In nucleic acids, at short length scales ($l \sim 5 \text{ \AA}$), the detailed chemical environment determines the basic forces (hydrogen bonds and dispersion forces) between two nucleotides. If $l \sim 1\text{--}1.5 \text{ nm}$, interactions between the two bases, base stacks and grooves of the nucleic acids become relevant. Understanding how RNA folds ($l \sim 1\text{--}3 \text{ nm}$) requires energy functions that provide at least a CG description of nucleotides and of interactions between them in the native state and excitations around the folded structure. On the persistence length scale, $l_p \approx 150 \text{ bp} \approx 50 \text{ nm}$,^[18] and beyond, it suffices to treat dsDNA as a stiff elastic filament without explicitly capturing the base pairs (bps). If $l \sim O(1) \mu\text{m}$, dsDNA behaves like a self-avoiding polymer.^[19] On the scale of chromosomes ($l \sim \text{mm}$), a much coarser description suffices. Thus, models for DNA and RNA vary because the scale of structural organization changes from nearly millimeters in chromatin to a few nanometers in the folded states of RNA. To describe RNA folding and the response of these molecules to mechanical forces, we have introduced two classes of CG models. The first represents each nucleotide by three interaction sites, referred to as the TIS model. Here, we employ the TIS model to illustrate the complexity of RNA hairpin formation and the effects of crowding on the equilibrium shift between the pseudoknot and hairpin formations in human telomerase RNA. In the second model, each nucleotide is represented using a single interaction center. The efficacy of the resulting model, referred to as the self-organized polymer (SOP) model, is illustrated using simulations to predict the outcomes of single molecule pulling experiments.

Changbong Hyeon is a professor at the School of Computational Sciences of the Korea Institute for Advanced Study. He did his PhD with Prof. D. Thirumalai at the University of Maryland, College Park, and his post-doctoral studies with Prof. Jose N. Onuchic at the UC San Diego. His current interests are in RNA and protein dynamics, single molecule biophysics, molecular motors, and many other subcellular processes.



Natalia Denesyuk is a Research Associate with Prof. D. Thirumalai at the University of Maryland, College Park. She did her PhD with Prof. Jean-Pierre Hansen at the University of Cambridge. Her current interests are in developing coarse-grained and all-atom models to study the self-organization of large RNA and RNA-protein complexes, and in modeling crowding effects in a cellular environment.



D. Thirumalai is a professor at the Institute of Physical Science and Technology of the University of Maryland. He did his PhD with Prof. Donald Truhlar at the University of Minnesota, and his post-doctoral studies with Prof. Bruce Berne at Columbia University. He is interested in theoretical chemistry, soft matter science, and biophysics. His current interests range from protein and RNA folding to functions of molecular machines.



2. General Considerations in Modeling RNA

In general, as a natural consequence of a rugged folding landscape, the folding of RNA is partitioned into several pathways, the effect of which can be detected in the form of multi-exponential functions in kinetics experiments.^[8] For several reasons, the folding landscape of RNA is rugged, consisting of multiple basins of attraction separated by substantial free energy barriers: (i) Phosphate groups are negatively charged, which implies that polyelectrolyte (PE) effects oppose the collapse and folding of RNA. Valence, size and shape of counterions, necessary to induce compaction and folding,^[20] can dramatically alter the thermodynamics and kinetics of RNA folding. Thus, PE effects, which are controlled by a number of factors, play a key role in determining RNA folding thermodynamics and kinetics (see below). (ii) The purine and pyrimidine bases of nucleotides have different sizes but are chemically similar. (iii) Only ~54% of bases form canonical Watson-Crick base pairs, while the remaining nucleotides are in non-pairing regions^[21] (discussed further below). (iv) The lack of chemical diversity in the bases results in RNA easily adopting alternate stable misfolded conformations, resulting in a relatively small stability gap between the folded and misfolded structures.^[8] Thus, the homopolymer nature of the RNA monomers, the critical role of counterions in shaping the folding landscape, and the presence of multiple low-energy excitations around the folded state, make RNA folding a challenging problem.

2.1 Polyelectrolyte Effects

To fold, RNA must overcome the large electrostatic repulsion between the negatively charged phosphate groups. PE-based theory (see Appendix 1) shows that multivalent cations ($Z > 1$) are more efficient in neutralizing the backbone charges than monovalent ions – a prediction that is borne out in experiments. For example, the midpoint of the folding transition C_m , the ion-concentration at which the populations of the folded and unfolded states are equal, for *Tetrahymena* ribozyme is $\sim 3 \times 10^6$ -fold greater in Na^+ than in cobalt-hexamine ($Z = 3$)! The nature of the compact structures depends on Z , with the radius of gyration scaling as $R_G \propto 1/Z^2$,^[22] implying compact intermediates have lower free energy as Z increases. Thus, folding rates should decrease as Z increases, which also accords well with experiments.^[23] Polyelectrolyte theory also shows that counterion charge density, $\zeta = Z e/V$, (V is the volume occupied by the ion) should control RNA stability. As ζ increases, RNA stability should increase – a prediction that was validated using a combination of PE-based simulations and experiments. For instance, it was shown that the changes in stability of *Tetrahymena* ribozyme in various Group II metal ions (Mg^{2+} , Ca^{2+} , Ba^{2+} , and Sr^{2+}) showed a remarkable linear varia-

tion with ζ .^[24,25] The extent of stability is greatest for ions with largest ζ (smallest V corresponding to Mg^{2+}). We further showed using Brownian dynamics simulations that this effect could be captured solely by non-specific ion-RNA interactions.^[24] These findings, and similar variations of stability in different sized diamines, show that: (i) the bulk of the stability arises from non-specific association of ions with RNA; and (ii) stability can be greatly altered by the valence, shape, and size of the counterions.

2.2 Statistics of Base Pairs in RNA Structures

Because of the systematic experimental studies pioneered by Tinoco, Uhlenbeck, Crothers, and Pörschke in 1970s,^[26–28] and Turner and Zuker in the 1980s and 1990s,^[29–31] good estimates of the free energies of RNA at the level of secondary structure are available. However, it is still difficult to accurately predict RNA secondary structures (especially if the number of nucleotides is big) because a substantial portion of the nucleotides do not participate in the “canonical” GC, AU, and GU type of base pairing, with base stacks whose free energy of association can be accurately measured. To quantify the fraction of nucleotides forming base pairs, we searched through RNA structures for pairs of nucleotides (A with U, G with C or U) in which at least one of their possible pairs of donor-acceptor heavy atoms is within a cut-off distance of 4 Å. For example, a Watson–Crick (WC) AU pair is identified if N6 from A and O4 from U, or N1 from A and N3 from U, is within 4 Å from each other. The 4 Å cut-off, larger than the typical 2.8–3.0 Å in an ideal hydrogen-bonded base pair, is chosen to account for possible non-ideal bonds and imperfections in the RNA structures, due to the X-ray resolution or NMR NOESY signals.^[21] Hence, the list of hydrogen-bonded pairs includes all the WC and reverse WC, Hoogsteen and reverse Hoogsteen, wobble and reverse wobble pairs for AU, GC, and GU; thus, in our statistics, base pairs found in pseudoknots are also included. (Parenthetically, using a more liberal definition of base pairs, Westhof and coworkers^[32] have shown that nearly 70% of nucleotides are involved in base pairs; however, they included non-canonical base pairs, such as G-A, A-C, C-C, A-A, and U-U, whose free energy of stabilization is not known or measured in their analysis. Even with this higher estimate, it is clear that a large fraction of nucleotides are not accounted for in canonical bp formation.) Based on our criterion for bp formation, we counted the number of the bps, N_{bp} , for a given structure in the PDB (protein data bank), and compared this number with the length, L , of the sequence. We found that N_{bp} varies linearly with L as expected, and the linear growth of N_{bp} with L is satisfied with great accuracy (Figure 1(a)). If all the nucleotides are engaged in WC-type canonical base pairing, the expected value for the ratio between N_{bp} and L is 0.5. However, the slope is only 0.27, i.e., $N_{bp} = 0.27 \times L$. The ratio

$N_{bp}/L=0.27$ implies that only 54% ($f_{bp}=0.54$) of nucleotides in an RNA chain form base pairs. The remaining 46%, which is a large percentage of nucleotides, remain as single strands, forming bulges, loops, or dangling ends, the free energy of which varies significantly, depending on their size. Even with the more liberal definition used by Westhof, it is clear that a large fraction of nucleotides do not participate in WC-type base pairing.

Using the calculated relationship between N_{bp} versus L , the total length of a given RNA structure can be decomposed into the base-paired and unpaired regions. Thus,

$$L = N_{ss} + 2 \times N_{bp} \quad (1)$$

or

$$f_{bp} = \frac{2 \times N_{bp}}{N_{ss} + 2 \times N_{bp}} = \frac{2 \langle n_{bp} \rangle}{\langle n_{ss} \rangle + 2 \langle n_{bp} \rangle} \quad (2)$$

where f_{bp} ($=0.54$) is the proportion of nucleotides making base pairs in the structure, and N_{ss} is the number of single-stranded nucleotides and nucleotides making non-canonical base pairs. From the statistics of the length of stem surveyed from ribosomal RNA (Figure 1(b)), the average length of stem is $\langle n_{bp} \rangle \approx 4.1$, and hence from Eq. 2, with $f_{bp} \approx 0.54$, the average length of a single strand is $\langle n_{ss} \rangle \approx 7.0$.^[33]

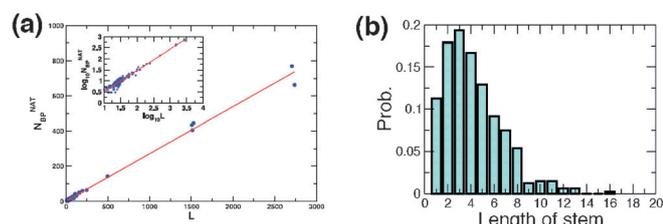


Figure 1. Statistics of base pairs in RNAs. (a) The dependence of the number of base pairs in a given native structure as a function of sequence length, L . The figure was adapted from Dima *et al.*^[21] (b) Distribution of the number of consecutive base pairs composing stems, calculated from 16S, 23S *E. coli* and *T. thermophilus* ribosomes.

The free energies associated with C-C, A-A, and G-A, as well as loops, bulges, internal loops, and internal multi-loops, are difficult to determine experimentally. In addition, the PE effects are hard to take into account accurately. These considerations explain the difficulties in devising reliable models for simulations. Nevertheless, we show here, using examples from our studies, that the two classes of CG models not only provide insights into a number of RNA problems, but also make specific testable predictions.

3. CG Models for RNA

Although CG models have been applied to a number of problems involving proteins, their use in the context of RNA is relatively recent.^[34–38] To our knowledge, we were the first to introduce CG models for RNA and use them to predict the outcome of single molecule pulling experiments and RNA folding,^[34,39–44] and to describe the effects of molecular crowding on the stability of folded states of RNA.^[45,46] We introduced two levels of coarse graining for nucleic acids. The first genre of models utilized the three site interaction model, which has been subsequently used in interesting applications to problems in DNA biology. In the original applications, we used only semi-realistic parameters in the TIS force field. In more recent developments, we parameterized the energy function to ensure that experimental melting curves for a number of oligonucleotides are quantitatively reproduced,^[47] resulting in a transferable TIS force field. To cope with larger systems, including complexes with proteins, we also developed the self-organized polymer model.

3.1 Three Interaction Site Representation of RNA

In the TIS model,^[34,46] each nucleotide is replaced by three spherical beads P, S, and B, representing a phosphate, a sugar, and a base (the center panel in Figure 11a, Appendix 2), located at the center of mass of the chemical groups. The energy function in the TIS model, U_{TIS} , is given by

$$U_{TIS} = U_{BL} + U_{BA} + U_{EV} + U_{ST} + U_{HB} + U_{EL}, \quad (3)$$

corresponding to bond length and angle constraints, excluded volume repulsions, single-strand base stacking, inter-strand hydrogen bonding and electrostatic interactions, respectively. Interactions resulting from chain connectivity, such as bond lengths, ρ , and angles, α , are constrained by harmonic potentials, $U_{BL}(\rho) = k_{\rho}(\rho - \rho_0)^2$ and $U_{BA}(\alpha) = k_{\alpha}(\alpha - \alpha_0)^2$, where the equilibrium values ρ_0 and α_0 are obtained by coarse-graining an ideal A-form RNA helix. The values of k_{ρ} and k_{α} are given in Ref. [47].

3.1.1 Excluded Volume Interaction

Excluded volume interaction between the interacting sites is modeled by a Weeks-Chandler-Andersen (WCA) potential,^[48]

$$U_{EV}(r) = \varepsilon_0 \left[\left(\frac{D_0}{r} \right)^{12} - 2 \left(\frac{D_0}{r} \right)^6 + 1 \right], \quad r \leq D_0, \quad (4)$$

$$U_{EV}(r) = 0, \quad r > D_0,$$

where $\varepsilon_0 = 1 \text{ kcal mol}^{-1}$ and $D_0 = 3.2 \text{ \AA}$ for all interacting sites.

3.1.2 Stacking Interactions

We assign stacking interactions, U_{ST} , between two consecutive nucleotides along the chain,

$$U_{\text{ST}} = \frac{U_{\text{ST}}^0}{1 + k_r(r - r_0)^2 + k_\phi(\phi_1 - \phi_{1,0})^2 + k_\phi(\phi_2 - \phi_{2,0})^2}, \quad (5)$$

where r (Å) and ϕ_1 , ϕ_2 (rad) are defined in Figure 11b, Appendix 2. The equilibrium values of r_0 , $\phi_{1,0}$ and $\phi_{2,0}$ are obtained by coarse-graining an A-form RNA helix. The parameters k_r , k_ϕ and U_{ST}^0 are derived from available thermodynamic measurements of single-stranded and double-stranded RNA,^[21,27,31] as described in detail in Ref. [47].

3.1.3 Hydrogen Bond Interactions

The TIS description of an RNA molecule includes coarse-grained hydrogen bond interactions, U_{HB} , which mimic the atomistic hydrogen bonds present in the molecule's PDB structure. Each hydrogen bond found in the PDB structure is modeled by a coarse-grained interaction potential,

$$U_{\text{HB}} = U_{\text{HB}}^0 \times [1 + 5(r - r_0)^2 + 1.5(\theta_1 - \theta_{1,0})^2 + 1.5(\theta_2 - \theta_{2,0})^2 + 0.15(\psi - \psi_0)^2 + 0.15(\psi_1 - \psi_{1,0})^2 + 0.15(\psi_2 - \psi_{2,0})^2]^{-1}, \quad (6)$$

where the geometrical definition of r (Å) and θ_1 , θ_2 , ψ , ψ_1 , ψ_2 (rad) is illustrated in Figure 11c, Appendix 2. For canonical Watson-Crick base pairs, the equilibrium values r_0 , $\theta_{1,0}$, $\theta_{2,0}$, ψ_0 , $\psi_{1,0}$ and $\psi_{2,0}$ are adopted from the coarse-grained structure of an ideal A-form RNA helix. For all other hydrogen bonds, the equilibrium parameters are obtained by coarse-graining the PDB structure of the RNA molecule. The parameter U_{HB}^0 has been calibrated to reproduce experimental melting curves of RNA hairpins and pseudoknots (see Ref. [47] for details). Although the terms $\theta_{1,0}$, $\theta_{2,0}$, ψ_0 , $\psi_{1,0}$, $\psi_{2,0}$ in Eq. 6 are defined based on those of PDB structure, and Eq. 6 is for the native hydrogen bonds, it is also possible to consider generic angles, and extend U_{HB} for non-native hydrogen bond interactions.

3.1.4 Electrostatic interactions

To model electrostatic interactions, we employ the Debye-Hückel approximation,^[49] combined with the concept of counterion condensation.^[50] Using this approach, the electrostatic energy of an RNA conformation is computed as

$$U_{\text{EL}} = \frac{Q^2 e^2}{2\epsilon} \sum_{ij} \frac{\exp(-|\mathbf{r}_i - \mathbf{r}_j|/\lambda)}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (7)$$

where Q is the reduced charge of the phosphate groups, e is the proton charge, $|\mathbf{r}_i - \mathbf{r}_j|$ is the distance between two phosphates i and j , ϵ is the dielectric constant of water and λ is the Debye-Hückel screening length. In monovalent salt solutions, Manning's theory of counterion condensation predicts^[50]

$$Q = \frac{b}{l_B}, \quad (8)$$

where b is the length per unit charge in the polyelectrolyte in the absence of counterion condensation and l_B is the Bjerrum length, $l_B = e^2/\epsilon k_B T$. We previously established that using $b = 4.4$ Å gives good agreement between the simulation and experiment for the RNA thermodynamics in monovalent salt solutions.^[47]

3.2 Self-Organized Polymer Model

The self-organized polymer model, a new class of versatile coarse-grained structure-based models, is well suited to understanding dynamics at the spatial resolution that single-molecule spectroscopy of large RNA provides. We refer to the model as the SOP model because it only uses the polymeric nature of the biomolecules and the crucial topological constraints that arise from the specific fold. We introduced the SOP model to study the response of proteins and RNA to mechanical force. The reason for using the SOP model in force spectroscopy applications is the following: (i) Forced unfolding and force-quench refolding lead to large conformational changes of the order of 10–100 nm. Currently, single molecule experiments (laser optical tweezers or atomic force microscopy) cannot easily resolve structural changes below 1 nm. As a result, details of the rupture of hydrogen bonds or local contacts between specific residues cannot be discerned from force-extension curves or the dynamics of the end-to-end distance, R , alone. Because only large changes in R are monitored, it is not crucial to model details due to local interactions, such as bond-angle and various dihedral angle potentials. (ii) In the context of mechanical unfolding, as well as the folding of RNAs, many of the details of the unfolding and folding pathways can be accurately computed by taking into account only the interactions that stabilize the native fold. Previous studies also suggested that it is crucial to take into account chain connectivity and attractive interactions that faithfully reproduce the contact map of a fold. The basic idea of the SOP model is to use the simplest possible Hamiltonian to simulate the low-resolution global dynamics for biopolymers of arbitrary size. The potential function for biopolymers in the SOP representation is

$$\begin{aligned}
H_{SOP} = & - \sum_{i=1}^{N-1} \frac{k}{2} R_0^2 \log \left[1 - \frac{(r_{i,i+1} - r_{i,i+1}^0)^2}{R_0^2} \right] + \sum_{i=1}^{N-2} \varepsilon_l \left(\frac{\sigma}{r_{i,i+2}} \right)^6 \\
& + \sum_{i=1}^{N-3} \sum_{j=i+3}^N \varepsilon_h \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \Delta_{ij} + \sum_{i=1}^{N-3} \sum_{j=i+3}^N \varepsilon_l \left(\frac{\sigma}{r_{ij}} \right)^6 (1 - \Delta_{ij})
\end{aligned} \quad (9)$$

The first term in Eq. 9 is the finite extensible nonlinear elastic (FENE) potential for chain connectivity with parameters, $k=20 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, $R_0=0.2 \text{ nm}$. The distance between neighboring beads at i and $i+1$ is $r_{i,i+1}$, and $r_{i,i+1}^0$ is the distance in the native structure. The use of the FENE potential is more advantageous than the standard harmonic potential, especially for the purpose of simulating force-induced stretching of biopolymer, because the fluctuations of $r_{i,i+1}$ are strictly restricted around $r_{i,i+1}^0$ with variations of $\pm R_0$, which enable us to produce worm-like chain behavior, without including an additional bond-angle potential term to the energy potential. The second term ensures non-crossing of the chain by making $i, i+2$ pairs interact repulsively with $\sigma=3.8 \text{ \AA}$. The Lennard-Jones potential, with a special condition for the native pairs, is used to account for interactions that stabilize the native fold. The native pairs are defined for beads, between i and j with $|i-j|>2$, whose distance in the native state is less than 8 \AA . We use $\varepsilon_h=1-2 \text{ kcal mol}^{-1}$ for native pairs, and $\varepsilon_l=1 \text{ kcal mol}^{-1}$ for nonnative pairs. Although potential attraction between nonnative pairs are neglected in the SOP model, this should not qualitatively affect the results for forced unfolding of RNA. Especially when simulating force-induced unfolding dynamics of native structure, such interactions do not make a significant contribution to the dynamics. To probe forced unfolding of RNA, it is sufficient to only include attractive interactions between contacts that stabilize the native state.

There are five parameters in the SOP force field. In principle, the ratio of ε_h and ε_l can be adjusted to obtain agreement with experiments. For simplicity, a uniform value of ε_h is assumed for proteins or RNA. But, if such modification is critical for understanding the dynamic behavior of the molecular system of interest, ε_h can be made sequence-dependent and ion-implicit. The time spent in calculating the Lennard-Jones potential scales as $\sim O(N^2)$. From an algorithmic point of view, drastic savings in computational time can be achieved by truncating the forces due to the Lennard-Jones potential for interacting pairs with $r_{ij}>3r_{ij}^0$ or 3σ , and by taking into account the neighboring list for numerical enumeration. Variation of the original SOP model has been used with remarkable success to predict the outcomes of single molecule pulling experiments. These predictions were subsequently validated.

4. Applications

4.1 Complex Kinetics of Hairpin Formation

When viewed on length scales that span several base pairs and time scales of the order of μs , folding of an RNA hairpin is simple. Hairpin formation can be described using a two-state model. However, when probed on short times (ns– μs range), the formation of a small hairpin, involving a loop and base-stacking, is complex. Temperature jump experiments have shown that the kinetics of hairpin formation in RNA deviates from the classical two-state kinetics, and is best described as a multi-step process.^[51] Additional facets of hairpin formation have been revealed in single molecule experiments that use mechanical force (f). Inspired by these experiments, we performed simulations using the TIS model, with the energy function given in Ref. [34], by varying T and f .^[52] The equilibrium phase diagram showed two basins of attraction (folded and unfolded) at the locus of critical points (T_m, f_m), which are the transition midpoints separating the unstructured and hairpin states. At T_m and f_m , the probability of being unfolded and folded is the same. The free energy surface obtained from simulations explained the sharp bimodal transition between the folded and unfolded state when the RNA hairpin is subject to f .^[34,52] characteristic of a two-state system.

However, upon temperature quench, a hairpin forms by multiple steps,^[52] as observed in the kinetic experiments.^[51] Folding pathways between T -quench and f -quench refolding are markedly different (Figure 2b), reflecting the differing initial conditions. The initial conformations generated by forced unfolding are fully extended and structurally homogeneous. The first event in folding, upon f -quench, is loop formation, which is a slow nucleation process (see Figure 2b). Zipping of the remaining base pairs leads to rapid hairpin formation. Refolding upon T -quench commences from a structurally broad ensemble of unfolded conformations. Therefore, nucleation can originate from many regions in the molecule (see Figure 2b). Our simulations showed that the complexity of the folding kinetics, observed in ribozyme experiments,^[53] is reflected in the formation of simple RNA hairpin.^[34,54]

4.2 Mechanical Unfolding of RNA

The stability of the RNA molecule in the native state can roughly (neglecting entropy) be decomposed into the secondary and tertiary interactions, $e^{tot} = \sum_i \varepsilon_i^{sec} + \sum_k \varepsilon_k^{ter}$, where i and k refer to the number of secondary and tertiary structural elements, respectively. The difference in the free energy between secondary and tertiary interactions, $\sum_i \varepsilon_i^{sec} \gg \sum_k \varepsilon_k^{ter}$, ensures that force disrupts tertiary contacts prior to secondary structures; thus, rips in force-extension curves (FECs) are due to disruptions of domains

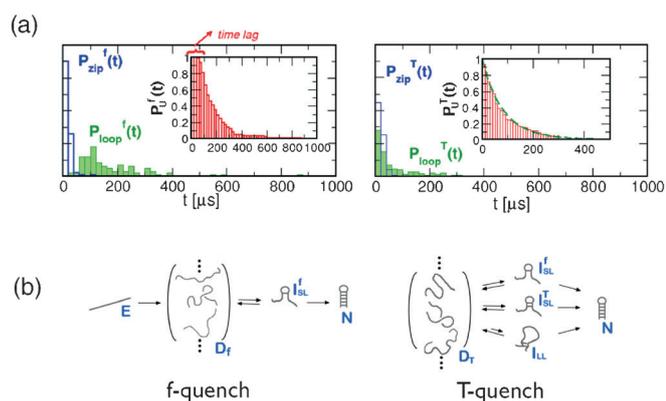


Figure 2. Folding of an RNA hairpin. (a) Kinetics of hairpin formation under *f*-quench and *T*-quench conditions. In each molecule the hairpin formation time can be decomposed into looping and zipping time as $\tau_f = \tau_{loop} + \tau_{zip}$. The dynamics of looping are more time-consuming in the *f*-quench condition than in the *T*-quench condition (compare $P_{loop}^f(t)$ and $P_{loop}^T(t)$), but the dynamics of zipping occurs on a similar time scale (compare $P_{zip}^f(t)$ and $P_{zip}^T(t)$). The probability of molecules remaining unfolded at time, t , ($P_U(t)$), is plotted in the inset. Under *f*-quench and *T*-quench conditions, the probabilities are fit to $P_U^f(t) = e^{-(t-50)/138}$ for $t > 50 \mu s$ and $P_U^T(t) = 0.44 e^{-t/63} + 0.56 e^{-t/104}$, respectively, where the time t is in the unit of μs . In particular, the lag phase of $P_U^f(t)$ at $0 < t < 50 \mu s$ suggests an obligatory intermediate (I_{SL} , where *SL* stands for small loops). (b) Schematic of refolding pathways of an RNA hairpin upon quenching the force from a high to low value (left) and from a temperature quench (right) using the TIS model. Upon *f*-quench, folding commences from an extended (*E*) state by forming the turn, which nucleated the hairpin formation (left). However, folding occurs by multiple pathways upon *T*-quench. The figures were adapted from Hyeon and Thirumalai^[52] (reproduced with copyright permission from the American Chemical Society).

defined by secondary structural elements. The separation in the energy scales ($\sum_i \epsilon_i^{sec} \gg \sum_k \epsilon_k^{ter}$) enables analysis of FEC for RNA to be made domain by domain. The hairpin stacks, which can vary in the length and sequence, are the simplest structural motifs, and the presence of hairpin loops, bulges, internal loops, and internal multiloops, increases the complexity of RNA structures. The effect of force on RNA structures with increasing complexity from simple hairpins, hairpins with bulges and internal loops to three-way junctions is quantitatively described elsewhere.^[40] We showed that many aspects of the physics of the mechanical unfolding of RNA, such as a force-dependent free energy profile as a function of the end-to-end distance, loading rate dependent movement of transition state, and pulling speed dependent unfolding pathways, can be dissected using structural motifs.

4.2.1 The L-21 T. Ribozyme

The *Tetrahymena* ribozyme and its independently folding subdomains (P4–P6 and P5abc) have been extensively studied since the discovery of self-splicing enzymatic activity in 1980s.^[20,55,56] By probing the unfolding character-

istics of increasingly larger constructs of the *T.* ribozyme using LOT (laser optical tweezers) experiments, Onoa *et al.*^[57] were able to associate the force peaks in the FECs to rupture specific substructures, providing an outline of the forced unfolding pathway of RNA. They assumed that extension by a certain length corresponds to the unraveling of the entire helical substructures, and inferred the unfolding pathway of ribozymes from FECs alone. However, it is difficult to unambiguously assign the specific paired helices for a given FEC. The number of rips in the FECs also varies depending on the specific molecule that is being stretched. In addition, there are multiple unfolding routes,^[57] indicative of heterogeneity in force-induced unfolding.

To provide molecular details of forced unfolding, we used the SOP representation of the Westhof model,^[58a,b] and computed the FEC^[58a] using Brownian dynamics simulations. In agreement with LOT experiments,^[57] in the majority of the cases, the FECs have about eight peaks. The number of peaks varies from molecule to molecule, which is increasingly being recognized as a characteristic of heterogeneity at the single-molecule level.^[59–61] Explicit comparison of the FECs with the dynamics of the contact disruption (Figure 3b) enables us to read off the molecular events associated with each rip in the FEC. There are two major unfolding pathways (Figure 3): (i) $[N] \rightarrow [P9.2] \rightarrow [P9.1, P9, P9.1a] \rightarrow [P2] \rightarrow [P2.1] \rightarrow [P3, P7, P8] \rightarrow [P6] \rightarrow [P4, P5] \rightarrow [P5a, P5b, P5c]$. (ii) $[N] \rightarrow [P2] \rightarrow [P2.1] \rightarrow [P9.2] \rightarrow [P9, P9.1, P9.1a] \rightarrow [P3, P7, P8] \rightarrow [P6] \rightarrow [P4, P5] \rightarrow [P5a, P5b, P5c]$. The difference between the two pathways is the switch in the order of unfolding of the peripheral domains (P2 and P9). The experimentally inferred, most probable pathway is $[N] \rightarrow [P9.2] \rightarrow [P9.1] \rightarrow [P9, P9.1a] \rightarrow [P2, P2.1] \rightarrow [P3, P7, P8] \rightarrow [P6, P4] \rightarrow [P5] \rightarrow [P5a, P5b, P5c]$.

Snapshots from simulations in Figure 3 show the conformational changes that occur in the force-induced unfolding transition. Although the predicted FECs do not quantitatively agree with the measurements, due to the differences between the loading rates and the spring constant used in the simulations and experiments, the order of the unfolding of the helices and the heterogeneous nature of the unfolding pathways are consistent with experiments. A few additional comments about our results from SOP simulations are worth making: (i) Both simulations and experiments^[57] find that the peripheral domains unravel before disruption of the tertiary interactions involving the catalytic core. Complete rupture occurs when helices P6, P4, and P5abc unfold. (ii) The major and minor unfolding pathways are not trivially related to each other. In one pathway, unfolding starts from P2, while in the other, unraveling starts from the P9 end. From a structural perspective, P2 forms tertiary interactions with P5c, whereas the P9 helix is in contact with P5. The free energies of the tertiary interactions involving the P2 and P9 domains are also different. Thus, from both the energetic

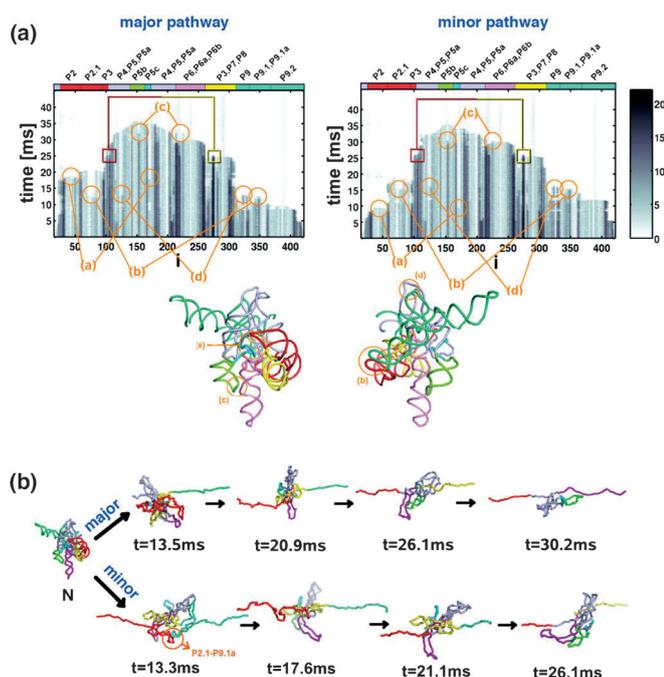


Figure 3. Unfolding of *T.* ribozyme. (a) The dynamics of disruption of individual contacts ($Q_i(t)$) for one of the trajectories in the major unfolding pathway. The scale in differing shades represents the number of contacts that survive at t . The tertiary contacts indicated with the circles (a–d) are shown in the *T.* ribozyme structure. The yellow squares show interactions that stabilize the P3 pseudoknot. The figure on the right shows that unfolding can also occur by an alternate pathway in which the initial event is the opening of P2. (b) Snapshots of *T.* ribozyme structures are shown along the major and minor pathways. The figures adapted from Hyeon *et al.*^[62] (reproduced with copyright permission from Elsevier).

and structural considerations, the differences in the unfolding pathways are significant. (iii) The rips corresponding to the peripheral domains P9 in the simulations are $[P9.2] \rightarrow [P9.1, P9, P9.1a]$, whereas in the experiments three rips corresponding to $[P9.2] \rightarrow [P9.1] \rightarrow [P9, P9.1a]$ are identified. The two rips corresponding to $[P2] \rightarrow [P2.1]$ also differ from the single rip $[P2, P2.1]$ in the experiment. The minor differences may be due to the slight variations in the constructs used in experiments (390-nt) versus simulations (407-nt).

4.2.2 Stretching the Azoarcus Ribozyme

Building on the good agreement between unfolding pathways in simulations and single molecule experiments on the *Tetrahymena* ribozyme, we performed simulations for a smaller, but structurally related, ribozyme. Mechanical unfolding trajectories of the 195-nt *Azoarcus* ribozyme (PDB code: 1u6b), generated at three different loading rates using the SOP model, reveal distinct unfolding pathways (Figure 4). At the highest loading rate, the FEC has six conspicuous rips,^[62] whereas at the lower loading rate, the number of peaks is reduced to between two and four.

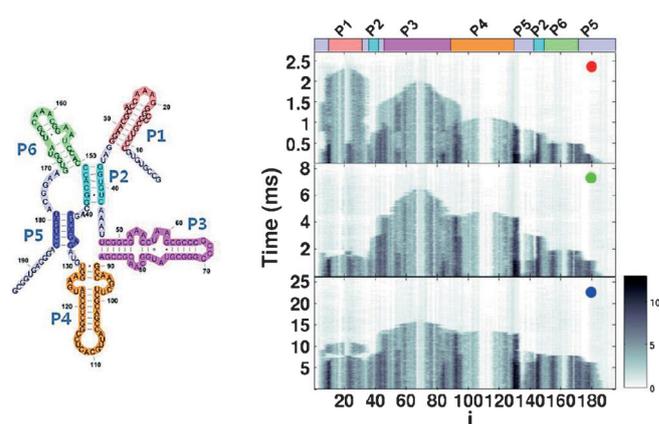


Figure 4. Loading rate dependent variation of unfolding pathway of *Azoarcus* ribozyme. The secondary structure of the ribozyme is shown on the left. Time evolutions of contact disruption at three different loading rates $r_f = 1.2 \times 10^3$ pN s⁻¹, 3.6×10^2 pN s⁻¹, and 1.9×10^1 pN s⁻¹ are shown from the top to bottom. The figure adapted from Hyeon *et al.*^[62] (reproduced with copyright permission from Elsevier).

These simulations showed that the unfolding pathways can be altered by changing the loading rate.

At the highest loading rate, the dominant unfolding pathway is $N \rightarrow [P5] \rightarrow [P6] \rightarrow [P2] \rightarrow [P4] \rightarrow [P3] \rightarrow [P1]$. At medium loading rates, the ribozyme unfolds via $N \rightarrow [P1, P5, P6] \rightarrow [P2] \rightarrow [P4] \rightarrow [P3]$, which produces four rips in the FECs. Multiple helices in the square bracket mean that they unravel nearly simultaneously. At the lowest loading rate, the ribozyme unfolds by two steps via $N \rightarrow [P1, P2, P5, P6] \rightarrow [P3, P4]$. The simulations using the SOP model indicate that unfolding pathways depend on the loading rate, a result that does not seem to be well appreciated. The physical origin of the change in the unfolding pathways as the loading rate is varied is explained elsewhere.^[62]

These results showed that the unfolding pathways can drastically change, depending on the loading rate, r_f . The dominant unfolding rate depends on r_f , suggesting that the outcomes of unfolding by LOT and AFM experiments can be dramatically different. Our predictions for the response of *Azoarcus* ribozyme can be tested readily using pulling experiments.

4.3 Towards Folding under Cellular Conditions

4.3.1 Background

Because the cytosol is crowded, replete with macromolecules such as ribosomes, lipids, proteins, and RNA, whose collective volume fraction (ϕ) can exceed 0.2, it is expected that RNA folding in the crowded environment is different from *in vitro* experiments that are conducted under infinite dilution conditions.^[63–66] Describing the transition between folded and unfolded states of RNA under crowd-

ed conditions is complicated because the nature of the interactions between the crowding agents and RNA is not fully understood. However, to a first approximation, the dominant effect of crowding agents is to exclude the molecule of interest from the volume occupied by crowders (see Appendix 3 for depletion interaction arising from excluded volume interactions). If excluded volume interactions dominate (an assumption that has to be tested before the theory can be applied to analyze experiments), then the stability of the folded state of the RNA is enhanced, compared with the infinite dilution limit. In this case, the loss in entropy of the folded state due to crowding is much less than that of the unfolded state, resulting in the stabilization of the folded state. The resulting entropic stabilization of the folded state^[64,67] has been recently affirmed in a number of studies.^[45,46,68]

We showed, using theoretical arguments and coarse-grained simulations, that crowding can modestly stabilize RNA secondary structures.^[45] However, RNA requires counterions (Mg^{2+} , for example) for tertiary folding. Thus, the effect of macromolecular crowding on tertiary structures of RNA may be complicated depending on the interplay between electrostatic and excluded volume interactions. Using small-angle X-ray scattering measurements, it has been shown that, in the presence of polyethylene glycol (PEG), the 195 nucleotide *Azoarcus* ribozyme is more compact relative to $\phi = 0$.^[68] It was concluded that excluded volume effects play a dominant role in the compaction of RNA in low molecular weight PEG. Interestingly, the transition to the folded state occurred at a lower Mg^{2+} concentration in the presence of PEG.^[68] Even if excluded volume interactions largely determine the stability of the folded states of RNA, a number of variables besides ϕ , such as size and shape of crowding agents, also contribute to the stability of RNA in the presence of inert crowding agents. Thus, a systematic study of the influence of macromolecular crowding on RNA is required. To date there are only a handful of works that have considered this problem.

4.3.2 Crowding Shifts Conformational Equilibrium between Pseudoknot (PK) and Hairpin (HP) in Human Telomerase RNA

As a biologically relevant example, we simulated crowding effects on the transition between the hairpin and pseudoknot conformations (Figure 5) in the pseudoknot domain of human telomerase RNA (hTR).^[69] The activity of the pseudoknot domain, which is conserved in different organisms, is closely linked to chromosome stability.^[70,71] Mutations that either increase or decrease the stability of the PK conformation result in a reduction in telomerase activity.^[72,73] Therefore, it is important to compare the impact of physical factors, such as macromolecular crowding, with that of naturally occurring chemical mutations.

To provide quantitative estimates of crowding-induced changes in the stability of RNA, we augmented the TIS

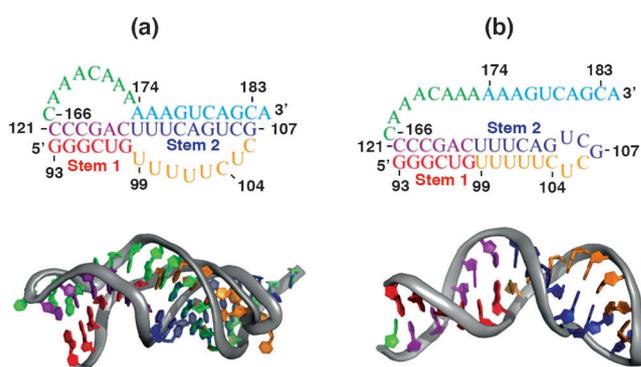


Figure 5. Secondary and tertiary structures of the pseudoknot and hairpin conformations for ΔU177 . (a) PK secondary and tertiary structures. (b) HP secondary and tertiary structures. Nuclear magnetic resonance (NMR) structure of the HP includes residues G93 to C166 only (PDB code 1NA2). To quantify the effect of crowders on the PK–HP equilibrium, we added an unstructured tail A167–A184 to the NMR structure in (a). This inclusion ensures that identical RNA sequences are used in simulations of the PK and HP conformations.

model for RNA with interactions between the crowders and the nucleotides. We modified the Lennard-Jones potential to model interactions of RNA with the spherical crowders of arbitrary size,

$$U_{\text{L}}(r) = \varepsilon \frac{2R_i}{D_0} \left[\left(\frac{D_0}{r + D_0 - D} \right)^{12} - 2 \left(\frac{D_0}{r + D_0 - D} \right)^6 + 1 \right],$$

$$r \leq D, U_{\text{L}}(r) = 0, \quad r > D,$$
(10)

where r is the distance between the centers of mass of two interacting particles, D_0 is the effective penetration depth of the interaction, R_i is the radius of an RNA coarse-grained bead, r_c is the radius of a crowder, and $D = R_i + r_c$. The ratio $2R_i/D_0$ in Eq. 10 is used to scale the interaction strength, ε , in proportion to the surface contact area. This potential accounts for nonspecific surface interactions between spherical crowders representing large macromolecules and individual segments of the coarse-grained RNA. The global effects of hard sphere crowders on RNA can be qualitatively explained using the concept of depletion forces, first discussed by Asakura and Oosawa (see Appendix 3).

4.3.3 Crowding Effect is Negligible for Large Crowders

The high density of macromolecules in the cell (volume fractions $\phi \approx 0.2$ – 0.4) reduces the space available for conformational fluctuations. Therefore, macromolecular crowding should shift the thermodynamic equilibrium between the HP and PK states of the hTR pseudoknot domain towards the more compact PK. To assess the extent to which PK is favored at $\phi \neq 0$, we first discuss

the simulation results^[46] for the HP and PK states of the modified pseudoknot domain, $\Delta U177$ (Figure 5). The molecular construct $\Delta U177$ has been examined experimentally *in vitro* at $\phi=0$.^[73] The atomistic structures of the HP and PK conformations of $\Delta U177$ are available from the protein data bank, codes 1NA2 and 2K96, respectively.

Here, we only consider spherical crowders with radius, r_c . For monodisperse particles, the volume fraction is $\phi=4\pi r_c^3 \rho/3$, where ρ is the number density. Thus, ϕ can be changed by increasing or decreasing ρ , or by altering the size of the crowding particles. For clarity of presentation, we fix $\phi=0.3$ and examine the consequences of changing r_c . Based on general theoretical considerations^[74,75] (see Appendix 3), it can be shown that, in the colloid limit $r_c > R_G^0$, the crowding agents would have negligible effects on RNA stability. Here, R_G^0 is the size of the RNA in the absence of the crowding agent. It is only in the opposite polymer limit, $r_c < R_G^0$, that the crowding particles would affect RNA stability. We therefore expect that the magnitude of the crowding effect should depend only on the ratio r_c/R_G^0 . Although this scaling type result follows from theory it has not been adequately tested in experiments involving RNA.

The HP melting profile, taken to be the negative derivative of the number of intact base pairs N_{BP} with respect to $k_B T$, is plotted in Figure 6a for the crowder radius, $r_c=26 \text{ \AA}$.^[46] Such crowders are larger than the radius of gyration of strand G93–C121 in the unfolded state, $R_G^0=20 \text{ \AA}$. As discussed above, large crowders should have minimal effects on the melting of the HP even at $\phi=0.3$. For a fixed ϕ , the average distance between two spherical crowders will increase with the crowder size. If the unfolded hairpin can easily fit in the interstitial space, the folding/unfolding transition will not be affected significantly by the presence of crowders. For $\phi=0.3$ and the crowder radius $r_c=26 \text{ \AA}$, which is only slightly larger than R_G^0 , the increase ΔT in the melting temperature is 1.5°C for stem 1 of the HP and is negligible for stem 2 (Figure 6a). A further increase in r_c results in $\Delta T \approx 0$ for both stems (data not shown).

Figure 6a also shows the melting profile of the HP in a ternary mixture of crowders, containing volume fractions $\phi=0.11, 0.11,$ and 0.08 of particles with $r_c=104 \text{ \AA}, 52 \text{ \AA},$ and 26 \AA , respectively.^[46] The sizes and volume fractions of individual components in the model mixture correspond to the ribosome, large enzymatic complexes and relatively small individual proteins, found in *E. coli*. Because all the values of r_c in the *E. coli* mixture are larger than R_G^0 , we expect only small changes in the melting profile of the HP (Figure 6a). For the total volume fraction of 0.3, the melting temperature of the HP stem 1 increases only by 2°C with respect to $\phi=0$ (Figure 6a). Interestingly, the effect of the *E. coli* mixture is similar in magnitude to that of a monodisperse suspension with $r_c=26 \text{ \AA}$ and $\phi=0.3$. In contrast, a monodisperse suspension

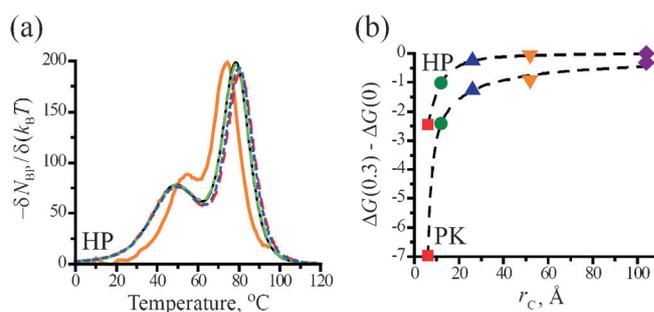


Figure 6. (a) Melting profiles of the HP in various crowding environments at $\geq 1 \text{ M}$ monovalent salt concentration. Black solid curve: without crowders. Red dashed curve (the model *Escherichia coli* mixture): $\phi=0.11$ of crowders with $r_c=104 \text{ \AA}$, $\phi=0.11$ of crowders with $r_c=52 \text{ \AA}$ and $\phi=0.08$ of crowders with $r_c=26 \text{ \AA}$. Green dashed-dotted curve: $\phi=0.08$ of crowders with $r_c=26 \text{ \AA}$. Blue dotted curve: $\phi=0.3$ of crowders with $r_c=26 \text{ \AA}$. Thick orange curve is the experimental UV data at 200 mM KCl , from Figure 2b in Theimer *et al.* (2003),^[69] divided by 5.53×10^{-5} . The two peaks indicate melting of stems 1 and 2 of the HP. The peak positions (melting temperatures) and the overall width of the melting curve (melting range) serve as a measure of agreement between theory and spectroscopic data. (b) Changes in stability (kcal mol^{-1}) of the HP and PK at 37°C due to crowders at $\phi=0.3$, as a function of the crowder radius r_c . The figure adapted from Denesyuk and Thirumalai^[76] (reproduced with copyright permission from Springer).

with $r_c=26 \text{ \AA}$ and $\phi=0.08$, which is equivalent to the smallest particle component in the mixture, has negligible effect on the melting of the HP (Figure 6a).

In summary, the crowding effect of polydisperse mixtures is largely the effect of the smallest particle component, but taken at the total volume fraction of the mixture. The excess stability of the folded state due to crowding decreases nonlinearly with the radius of the crowding particle r_c (see the next subsection). We therefore propose that, for crowding in the cellular environment, the main role of large macromolecules is to increase the effective volume fraction of the relatively small macromolecules.

4.3.4 Role of Crowder Size in the PK-HP Equilibrium

The change in stability of the HP and PK at 37°C , in the presence of monodisperse crowders for different crowder radii r_c ($\phi=0.3$), shows that the magnitude of the excess stability $\Delta G(0.3) - \Delta G(0)$ is small if $r_c/R_G^0 > 1$, and increases sharply for $r_c/R_G^0 < 1$. The crowding effect is larger for the PK for all values of r_c (Figure 6b), indicating an equilibrium shift towards this conformation. Macromolecular crowding promotes the binding of strand C166–A184 to the remainder of the structure in the PK (Figure 5), because fluctuations associated with this strand are restricted in a crowded environment. The crowder radius, $r_c=12 \text{ \AA}$, corresponds to the size of an average protein *in vivo*. For $\phi=0.3$ and $r_c=12 \text{ \AA}$, $\Delta G_{PK}(0.3) - \Delta G_{PK}(0) = -2.4 \text{ kcal mol}^{-1}$ and

$\Delta G_{\text{HP}}(0.3) - \Delta G_{\text{HP}}(0) = -1.0 \text{ kcal mol}^{-1}$, which results in the relative stabilization of the PK conformation by $-1.4 \text{ kcal mol}^{-1}$ (Figure 6b).

4.3.5 Implications for Function

It is known that changes in the relative stability of the HP and PK conformations of the hTR pseudoknot domain compromise the enzyme activity. The estimate of the crowding effect in a typical cellular environment, $\Delta\Delta G = -1.4 \text{ kcal mol}^{-1}$ (see above), allows us to assess the extent to which macromolecules could regulate telomerase activity. We showed^[46] that the enzyme activity of the hTR mutants^[72,73] decreases exponentially as a function of $|\Delta\Delta G^*| = |\Delta G_{\text{PK}}^*(0) - \Delta G_{\text{PK}}(0)|$, where $\Delta G_{\text{PK}}^*(0)$ and $\Delta G_{\text{PK}}(0)$ are the stabilities of mutant and wild-type pseudoknots at $\phi = 0$ (Figure 7). The majority of mutations destabilize the PK, i.e., $\Delta\Delta G^* > 0$ (black squares in Figure 7), and only two mutants have stabilizing effects, i.e., $\Delta\Delta G^* < 0$ (red stars in Figure 7). For destabilizing mutants, the reduction in activity, α , was shown to follow the exponential dependence,^[46] $\alpha = \exp(-0.6\Delta\Delta G^*)$ (thick curve in Figure 7). The naturally occurring destabilizing mutations DKC and C116U have been linked to the diseases dyskeratosis congenita and aplastic anemia, respectively.^[77,78] The DKC and the stabilizing ΔU177 mutations have been studied *in vivo* (green symbols in Figure 7), as well as *in vitro*. In both cases, mutant telomerase *in vivo*

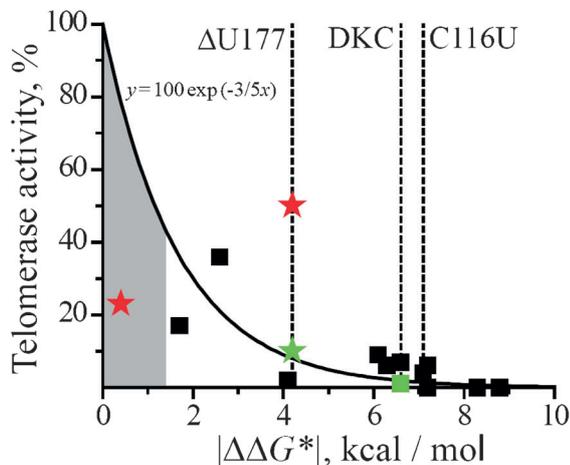


Figure 7. Activity of mutant telomerase normalized to wild-type activity (100%) as a function of the magnitude of the stability difference between mutant and wild-type pseudoknots. Three mutants, ΔU177 , DKC, and C116U, are explicitly marked. The *in vitro* data for destabilizing (black squares) and stabilizing (red stars) mutations are from Table 2 in Theimer *et al.*^[73] Green symbols: *in vivo* data for ΔU177 and DKC from Comolli *et al.*^[72] The experimental data for destabilizing mutations is fit to the exponential function (solid curve). Gray area: range of stability differences that can be accommodated by crowding. The figure adapted from Denesyuk and Thirumalai^[46] (reproduced with copyright permission from the American Chemical Society).

was found to be significantly less active than the corresponding construct *in vitro*, suggesting that a number of factors determine the activity of telomerase *in vivo*.

Although macromolecular crowding enhances the stability of the PK state, the crowding effect ($\Delta\Delta G = -1.4 \text{ kcal mol}^{-1}$) is less than the stability changes caused by mutations. In Figure 7, the grey area marks the domain of potential mutants with $\Delta\Delta G^* > 0$, whose activity may be completely restored by macromolecular crowding. All experimentally studied mutants fall outside the marked domain, including the two disease related mutants DKC and C116U. Nevertheless, due to the strong dependence of enzyme activity on $\Delta\Delta G^*$, the effect of crowding on telomerase function may be significant. We estimate that the activity of telomerase can be up- or down-regulated by a more than two-fold in response to density fluctuations in its immediate environment. Furthermore, due to the expected dynamical heterogeneities in cells, there will be variations in enzyme activity in different cell regions. It should be emphasized that the enhancement in the stability of PK relative to HP is a maximum for a given r_c and ϕ in the excluded volume dominated regime. All specific interactions between crowders and nucleotides will invariably decrease the extent of stability changes. Thus, the predicted enhancement is an upper bound.

4.3.6 Crowding Effects on RNA at Different Ionic Strengths

The entropic stabilization mechanism^[64] implies that crowding increases the stability of the folded state by reducing the population of expanded conformations in the unfolded state. Therefore, we expect that the magnitude of the crowding effect will be sensitive to the ionic strength of the RNA buffer, since the latter determines the size of the unfolded RNA. The quantitative discussion above assumed the limit of high ionic strength. As the buffer ionic concentration, c , is lowered, the screening of the negative charge on the RNA sugar-phosphate backbone becomes less efficient, which in turn, increases the mean radius of gyration of conformations in the unfolded state. The function $R_G(c)$ for the unfolded PK is shown in Figure 8a in the absence (black diamonds) and presence (green circles) of crowding. The same general trend is observed in both cases, with the R_G values being consistently smaller when crowders are present for the entire range of c . In accordance with our predictions, the crowder-induced stabilization of the folded PK becomes more significant at low ionic strengths (red squares in Figure 8a).

Interestingly, the stabilization effect increases rapidly upon lowering c from 1 M to 0.1 M, but shows little change when c is lowered further to below 0.1 M. The underlying reason for such behavior can be traced to the probability distributions $p(R_G)$ in the unfolded state (Figure 8b). For a given crowder solution, we can identify a typical size of the cavity which will be free of any crow-

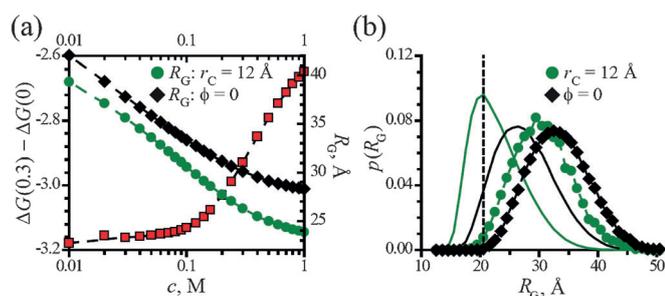


Figure 8. (a) The radius of gyration R_G of the unfolded PK (right axis) as a function of the monovalent ion concentration c for $\phi=0$ (black diamonds) and $\phi=0.3$, $r_c=12$ Å (green circles). Red squares show the excess stability of the PK due to crowding, $\Delta G_{PK}(0.3) - \Delta G_{PK}(0)$, for $r_c=12$ Å (left axis). (b) Probability distributions $p(R_G)$ of the radius of gyration of the unfolded PK for $\phi=0$ and $c=1$ M (black solid line), for $\phi=0$ and $c=0.1$ M (black diamonds), for $\phi=0.3$, $r_c=12$ Å and $c=1$ M (green solid line), and for $\phi=0.3$, $r_c=12$ Å and $c=0.1$ M (green circles). Vertical dashed line indicates the smallest size of RNA conformations that will be perturbed by crowders with $\phi=0.3$ and $r_c=12$ Å.

ders. If the radius of gyration of RNA conformations is such that they fit into the cavity, these conformations will not be perturbed by crowding. On the other hand, the population of conformations with R_G larger than the typical cavity size will be significantly depleted by crowding. For $\phi=0.3$ and $r_c=12$ Å, we can infer the size of a standard empty cavity from the distributions $p(R_G)$ at high ionic strength (Figure 8b). Note that $p(R_G)$ decreases for $R_G > 20$ Å when crowders are present (green solid line in Figure 8b), but $p(R_G)$ increases with R_G around 20 Å in the absence of crowders (black solid line in Figure 8b). This indicates that the crowders significantly perturb the unfolded conformations with R_G larger than 20 Å, which can serve as an upper estimate of the smallest RNA size affected by crowders. When c decreases, the distribution $p(R_G)$ shifts to larger R_G , increasing the fraction of the unfolded conformations affected by crowders. At $c=0.1$ M, all statistically significant values of R_G in the unfolded state fall within the range $R_G > 20$ Å, so that the entire distribution $p(R_G)$ is depleted due to crowding (symbols in Figure 8b). This explains why the crowding-induced stabilization is almost constant below 0.1 M, even if the average R_G continues to increase rapidly all the way to 0.01 M (symbols in Figure 8a). This explanation also shows that the relative ratio between the size of the unfolded RNA and the crowder size dictates the extent of stability changes.

4.3.7 RNA Becomes Compact as ϕ Increases

The reduction in conformational space accessible to RNA should increase with ϕ , for a fixed r_c . Thus, we expect that R_G should decrease as ϕ increases. This is precisely what is observed in experiments, which show that at all concentrations of Mg^{2+} the *Azoarcus* ribozyme becomes

more compact as the volume fraction of the crowding agent (PEG) increases^[68] (Figure 9). Interestingly, the midpoint of the folding transition c_m – the concentration of Mg^{2+} at which the folded and unfolded states of the *Azoarcus* ribozyme have equal populations – also decreases as ϕ increases (Figure 9). This finding can be

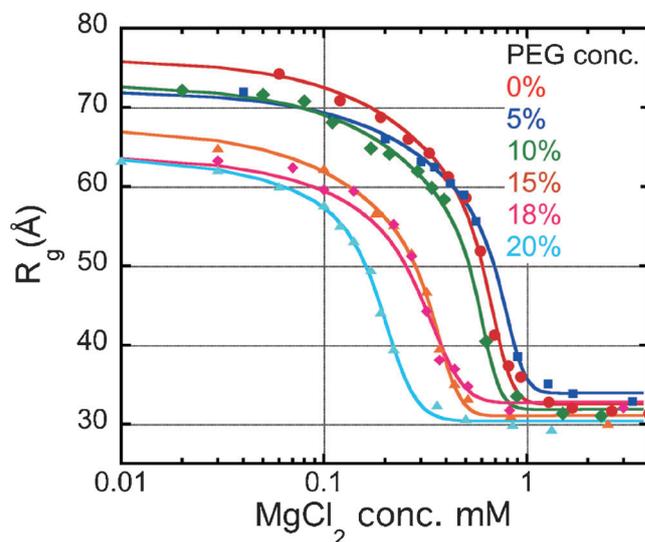


Figure 9. Small-angle X-ray scattering measurements of the radius of gyration of the *Azoarcus* ribozyme for different concentrations of Mg^{2+} ions and polyethylene glycol (PEG). Graphic adopted from Kilburn *et al.*^[68]

readily explained in terms of the entropic stabilization mechanism^[64] and suggests that, to a first approximation, PEG behaves as an inert hard sphere crowding agent. Based on our considerations from the previous section, we predict that the shift in c_m due to crowding will also depend on the concentration of monovalent counterions in the RNA buffer, a prediction that is amenable to experimental test. In addition, it would be of interest to perform experiments at a fixed ϕ but varying r_c , which can be changed by decreasing or increasing the molecular weight of PEG.

5. Outlook

There are compelling reasons to develop CG models further to study many complex phenomena associated with RNA. The two most obvious ones are: (i) Current atomistic RNA force fields do not appear to be accurate enough to predict folding thermodynamics of even a tetra loop.^[3] It is unclear if the deficiency is in the parameterization of RNA or water models or both. (ii) The time scales associated with the processes that we have touched upon here are far too great for even the most specialized computers. By essentially following a methodology of calibrating energy functions by direct comparison with experiments,

we have shown that CG models can not only reproduce experimental measurements, but also can make testable predictions. Although not discussed here, the folding landscape of add-riboswitch (which activates translation in *E. Coli.* by binding to adenine) under force was predicted accurately^[79] three years before experimental validation. More recently, we have predicted that a single point mutation should alter the folding landscape of add-riboswitch to resemble that of pbuE-riboswitch, which activates transcription.^[80] These results show that discovery of design principles of riboswitches, which control gene expression, can be made using CG models but remain, at present, outside the scope of atomically detailed simulations. The concept of modeling at the appropriate level, prevalent in the study of synthetic materials, which also applies to biological problems, will surely spur us on to develop suitable CG models and theories that capture the essence of the problem at hand without being encumbered by unnecessary details, and will continue to grow because there is an appetite to understand the workings of a cell.

Appendix 1: Manning-Oosawa Mechanism

Counterion condensation onto a polyelectrolyte is analogous to vapor-to-liquid phase transition.^[81] Consider a system consisting of N_{PE} polyelectrolyte molecules in a solution of volume V . Each PE, whose contour length is l , has n charged groups on the backbone onto which n monovalent counterions can condense. Counterions are divided into n_f free (vapor), and n_b , bound (liquid), counterions. Figure 10 shows that the volume occupied by the n_b bound counterions around each PE is approximated as a cylinder, with thickness, R_b , and volume v , ($\pi R_b^2 l \approx v$). The average volume for n counterions released from each

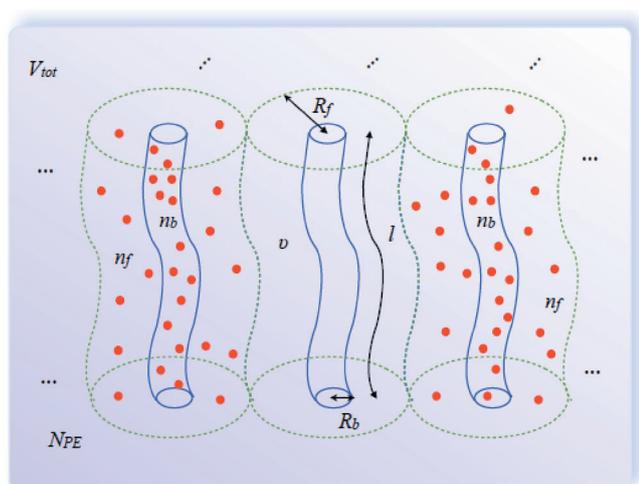


Figure 10. The volume occupied by the bound counterions around each polyelectrolyte.

PE is $\pi R_b^2 l \approx V/N_{PE}$. Therefore, the volume fraction of PE ($\phi \equiv N_{PE} v/V$) is related to R_b and R_f as $(R_f/R_b)^2 = \phi$. At equilibrium, chemical potentials of free (μ_f) and bound ions (μ_b) are equal, $\Delta\mu = \mu_b - \mu_f = 0$. The energetic contribution, $\Delta\varepsilon$, results from the difference in the Coulomb potentials ($-e\psi$) between the two phases. For a PE modeled as a cylinder $\Delta\varepsilon = -e(\psi_b - \psi_f) = -e(\psi(R_b) - \psi(R_f)) = 2k_B T \beta \xi \log(R_b/R_f) \approx k_B T \beta \xi \log \phi$, where $\beta \equiv n_f/n$ and $\xi = n l_B/l$, $l_B = e^2/k_B T$ is the Bjerrum length. The entropic contribution $T\Delta s$ to $\Delta\mu$ is determined by the ratio of accessible volumes for the counterions in each phase, V_b for bound ions and V_f for free ions. Since $V_b \approx N_{PE} v$ and $V_f = V - N_{PE} v$, $T\Delta s = k_B T \log(V_b/V_f) = k_B T \log(\phi/(1-\phi))$. For $\phi \rightarrow 0$ $\Delta\mu \approx (\beta\xi - 1)k_B T \log \phi$. This suggests that the fraction of free ions at equilibrium ($\Delta\mu = 0$) is $\beta = 1/\xi$ and the fraction of ions bound to PE (or the fraction of backbone charge neutralized by the counterions) is $1 - \xi^{-1}$. One can use similar arguments to quantify the effect of counterion condensation onto RNA, which is roughly spherical in the folded state.^[82] Surprisingly, this concept, familiar in PE theory, has been very successful in estimating renormalized charges on the phosphate groups.

Appendix 2: Genre of CG Models

In response to the challenge of describing biological processes that span several orders of magnitude in time and length scales, a variety of CG models for DNA, RNA, and proteins have been proposed. Although CG models have been prevalent in the polymer literature for over fifty years, their use in proteins began in earnest with the pioneering work of Levitt and Warshel.^[11] The efficacy of off-lattice models for protein folding kinetics was first demonstrated by Honeycutt and Thirumalai.^[83] More recently, in the same spirit, we introduced CG models for RNA. In all the CG models, polypeptide chains and nucleic acids are represented using a reduced description. Figure 11 shows: (a) three interaction site model for an RNA hairpin, obtained by representing each nucleotide by three sites, corresponding to phosphate, ribose, and base (the center panel). The rightmost panel shows further reduction to the SOP model. The associated energy functions are described in the main text. (b) Definition of the distance between the bases (r) and of the two dihedral angles (ϕ_1, ϕ_2) for stacking interactions in the TIS model (Eq. 5). Phosphates, sugars and bases are shown in black, green and red, respectively. The indices refer to different nucleotides. (c) Definition of the distance (r), angles (θ_1, θ_2) and dihedral angles (ψ, ψ_1, ψ_2) for hydrogen bond interactions between the bases in the TIS model (Eq. 6). Definitions of the structural parameters for hydrogen bonds involving phosphates or sugars and further details on the energy functions can be found in Ref. [47].

A major advantage of CG models is that their conformational space can be exhaustively sampled. However,

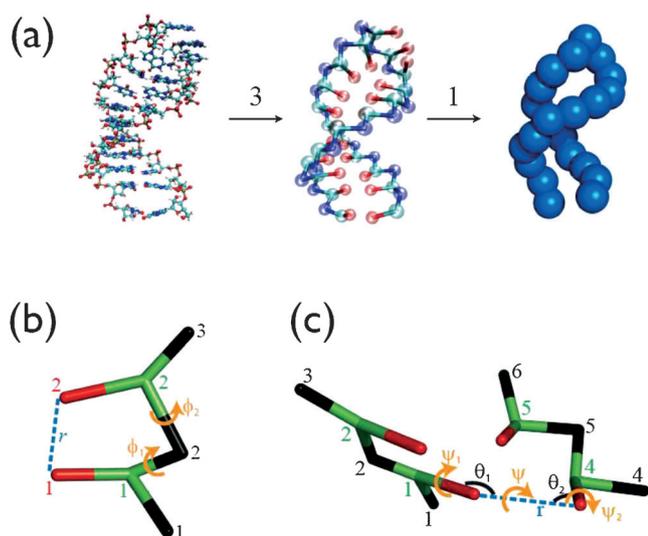


Figure 11. (a) RNA hairpin, with TIS model (center) and SOP model (right) representations. (b) Definition of the distance between the bases (r) and of the two dihedral angles (ϕ_1 , ϕ_2) for stacking interactions in the TIS model. (c) Definition of the distance (r), angles (θ_1 , θ_2) and dihedral angles (ψ , ψ_1 , ψ_2) for hydrogen bond interactions between the bases in the TIS model.

even with simplification, accurate results for thermodynamics might require enhanced sampling methods. Towards this end, simulation of CG models have used replica exchange methods and multicanonical methods. In addition, low friction Langevin dynamics has also been used to efficiently sample conformational space. These methods are especially necessary in simulating proteins with complex topology. To obtain kinetic information for folding, typically Brownian dynamics (BD) simulations are performed. In BD simulations, the Brownian time is $\tau_H \approx \zeta_H a^2 / k_B T_s$, where ζ_H is the friction constant, a is roughly the size of a coarse-grained bead, and T_s is the simulation temperature. Estimates of these quantities^[84] have been used to map simulation times to real times in a number of applications.

Appendix 3: Depletion Force

If the spacing between two colloidal particles in a suspension containing inert polymers is less than the polymer size, osmotic pressure of the solution produces an attractive force, referred to as the depletion force, between the colloidal particles.^[74] The depletion force arises so as to maximize the entropy of macromolecules. Consider two colloidal particles (red spheres with diameter, D) suspended in a solution of volume, V , with macromolecules (orange spheres with diameter, d) ($D > d$). If both colloids and macromolecules are hard spheres, the partition function for the system when two colloidal particles are separated by a distance, a , is the accessible volume for macro-

molecules, i.e., $Q(a) = \int_V dx \exp(-\omega(x, a)) = V_{\text{eff}}(a)$, where $\omega(x, a)$ is the potential energy between the two colloidal particles at separation a , and x denotes the orientation (θ , ϕ) between them. Depending on the distance between the two colloidal particles, a , the accessible volume for the macromolecules is

$$V_{\text{eff}}(a) = \begin{cases} V - \pi \left[\frac{4}{3} \left(\frac{D+d}{2} \right)^3 + \left(\frac{D+d}{2} \right)^2 a - \frac{1}{12} a^3 \right] & (D \leq a \leq D+d) \\ V - \frac{8\pi}{3} \left(\frac{D+d}{2} \right)^3 & (a \geq D+d) \end{cases} \quad (11)$$

For $D \leq a \leq D+d$, the two spheres with radius $(D+d)/2$ overlap, reducing the accessible volume for macromolecules. Given the partition function, one can calculate the depletion force between the spheres using $f = Nk_B T \partial \log Q(a) / \partial a$. For $V \rightarrow \infty$,

$$V_{\text{eff}}(a) = \begin{cases} -p \times \frac{\pi}{3} [(D+d)^2 - a^2] & (D \leq a \leq D+d) \\ 0 & (a \geq D+d) \end{cases} \quad (12)$$

Here $p = Nk_B T / V$ is the osmotic pressure due to the macromolecules. Note that $\frac{\pi}{3} [(D+d)^2 - a^2]$ is the area of circular intersection formed by the overlap of two spheres, with $\sqrt{\left(\frac{D+d}{2}\right)^2 - \left(\frac{a}{2}\right)^2}$ being the radius of intersection. The attractive interaction between two colloids (green arrows in Figure 12) due to the depletion force manifests itself only if the separation of the colloidal particles ($a-D$) is less than the size of macromolecules (d). From Eq. 12, it is expected that for a given osmotic pressure (p) the depletion force will be larger for bigger-size

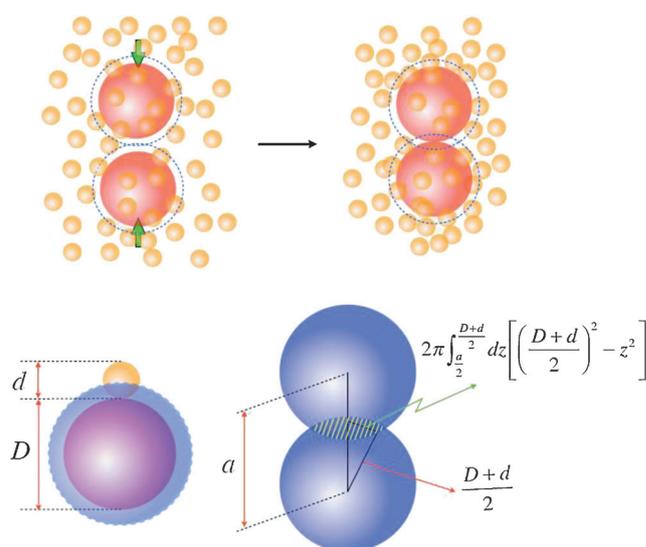


Figure 12. The attractive interaction between two colloids (green arrows) due to the depletion force manifests itself only if the separation of the colloidal particles ($a-D$) is less than the size of macromolecules (d).

macromolecular crowders. For a given volume fraction of crowders $\varphi_c = \frac{4\pi}{3}(d/2)^3 N/V$, the osmotic pressure is inversely proportional to d^3 ($p = 6\varphi_c k_B T / \pi d^3$); thus depletion force becomes stronger for crowders with a smaller size, d , although the range of non-zero depletion force [$a \in (D, D+d)$] is smaller.

Acknowledgements

This work was supported by a grant from the National Science Foundation (CHE09-10433) (D. T.). The list of references is not exhaustive and should be viewed as a guide to the literature.

References

- [1] M. Levitt, *Nature* **1969**, 224, 759–763.
- [2] I. Tinoco Jr., C. Bustamante, *J. Mol. Biol.* **1999**, 293, 271–281.
- [3] A. A. Chen, A. E. García, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, 110, 16820–16825.
- [4] M. Mézard, G. Parisi, M. Virasoro, *Spin Glass Theory and Beyond*, World Scientific, Singapore, **1988**.
- [5] T. R. Kirkpatrick, D. Thirumalai, *J. Phys. A.: Math. Gen.* **1989**, 22, L149–L155.
- [6] G. Parisi, F. Zamponi, *Rev. Mod. Phys.* **2010**, 82, 789.
- [7] G. Biroli, J. P. Garrahan, *J. Chem. Phys.* **2013**, 138, 12A301.
- [8] D. Thirumalai, C. Hyeon, *Biochemistry* **2005**, 44, 4957–4970.
- [9] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, 103, 227–249.
- [10] M. Levitt, *J. Mol. Biol.* **1976**, 104, 59–107.
- [11] M. Levitt, A. Warshel, *Nature* **1975**, 253, 694–698.
- [12] J. McCammon, B. Gelin, M. Karplus, *Nature* **1977**, 267, 585–590.
- [13] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, *Science* **2010**, 330, 341–346.
- [14] K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, *Science* **2011**, 334, 517–520.
- [15] T. J. Lane, D. Shukla, K. A. Beauchamp, V. S. Pande, *Curr. Opin. Struct. Biol.* **2012**, 23, 58–65.
- [16] M. E. Fisher, *Rev. Mod. Phys.* **1974**, 46, 597–616.
- [17] P. G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, **1979**.
- [18] C. Bustamante, J. F. Marko, E. D. Siggia, S. Smith, *Science* **1994**, 265, 1599–1600.
- [19] F. Valle, M. Favre, P. de Los Rios, A. Rosa, G. Dietler, *Phys. Rev. Lett.* **2005**, 95, 158105.
- [20] D. Thirumalai, N. Lee, S. A. Woodson, D. K. Klimov, *Annu. Rev. Phys. Chem.* **2001**, 52, 751–762.
- [21] R. I. Dima, C. Hyeon, D. Thirumalai, *J. Mol. Biol.* **2005**, 347, 53–69.
- [22] S. L. Heilman-Miller, J. Pan, D. Thirumalai, S. A. Woodson, *J. Mol. Biol.* **2001**, 306, 1157–1166.
- [23] S. Moghaddam, G. Caliskan, S. Chauhan, C. Hyeon, R. Briber, D. Thirumalai, S. Woodson, *J. Mol. Biol.* **2009**, 393, 753–764.
- [24] E. Koculi, C. Hyeon, D. Thirumalai, S. A. Woodson, *J. Am. Chem. Soc.* **2007**, 129, 2676–2682.
- [25] D. Lambert, D. Draper, *J. Mol. Biol.* **2007**, 370, 993–1005.
- [26] I. Tinoco Jr., K. Sauer, J. C. Wang, J. D. Puglisi, *Physical Chemistry Principles and Applications in Biological Sciences*, Prentice-Hall, Englewood Cliffs, **2002**.
- [27] V. A. Bloomfield, D. M. Crothers, I. Tinoco Jr., *Nucleic Acids, Structures, Properties and Functions*, University Science Books, Sausalito, **2000**.
- [28] D. Porschke, *Biophys. Chem.* **1974**, 1, 381–386.
- [29] D. H. Turner, N. Sugimoto, S. M. Freier, *Annu. Rev. Biophys. Biophys. Chem.* **1988**, 17, 167–192.
- [30] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, M. Zuker, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 9218–9222.
- [31] T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, D. H. Turner, *Biochemistry* **1998**, 37, 14719–14735.
- [32] J. Stombaugh, C. L. Zirbel, E. Westhof, N. B. Leontis, *Nucleic Acids Res.* **2009**, 37, 2294–2312.
- [33] L. T. Fang, A. M. Yoffe, W. M. Gelbart, A. Ben-Shaul, *J. Phys. Chem. B.* **2011**, 115, 3193–3199.
- [34] C. Hyeon, D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 6789–6794.
- [35] P. C. Whitford, A. Schug, J. Saunders, S. P. Hennelly, J. N. Onuchic, K. Y. Sanbonmatsu, *Biophys. J.* **2009**, 96, L7–L9.
- [36] J. Feng, N. G. Walter, C. L. Brooks III, *J. Am. Chem. Soc.* **2011**, 133, 4196–4199.
- [37] P. C. Whitford, P. Geggler, R. B. Altman, S. C. Blanchard, J. N. Onuchic, K. Y. Sanbonmatsu, *RNA* **2010**, 16, 1196–1204.
- [38] R. L. Hayes, J. K. Noel, U. Mohanty, P. C. Whitford, S. P. Hennelly, J. N. Onuchic, K. Y. Sanbonmatsu, *J. Am. Chem. Soc.* **2012**, 134, 12043–12053.
- [39] C. Hyeon, D. Thirumalai, *Biophys. J.* **2006**, 90, 3410–3427.
- [40] C. Hyeon, D. Thirumalai, *Biophys. J.* **2007**, 92, 731–743.
- [41] C. Hyeon, G. Morrison, G. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, 105, 9604–9606.
- [42] C. Hyeon, G. Morrison, D. L. Pincus, D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, 106, 20288–20293.
- [43] S. Cho, D. Pincus, D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, 106, 17349.
- [44] S. Biyun, S. S. Cho, D. Thirumalai, *J. Am. Chem. Soc.* **2011**, 133, 20634–20643.
- [45] D. Pincus, C. Hyeon, D. Thirumalai, *J. Am. Chem. Soc.* **2008**, 130, 7364–7372.
- [46] N. A. Denesyuk, D. Thirumalai, *J. Am. Chem. Soc.* **2011**, 133, 11858–11861.
- [47] N. A. Denesyuk, D. Thirumalai, *J. Phys. Chem. B* **2013**, 117, 4901–4911.
- [48] D. Chandler, J. D. Weeks, H. C. Andersen, *Science* **1983**, 220, 787–794.
- [49] R. Li, H. W. Ge, S. S. Cho, *J. Phys. Chem. B.* **2013**, 117, 12943–12952.
- [50] G. S. Manning, *J. Chem. Phys.* **1969**, 51, 924–933.
- [51] H. Ma, D. J. Proctor, E. Kierzek, R. Kierzek, P. C. Bevilacqua, M. Gruebele, *J. Am. Chem. Soc.* **2006**, 128, 1523–1530.
- [52] C. Hyeon, D. Thirumalai, *J. Am. Chem. Soc.* **2008**, 130, 1538–1539.
- [53] J. Pan, D. Thirumalai, S. A. Woodson, *J. Mol. Biol.* **1997**, 273, 7–13.
- [54] S. J. Chen, K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, 97, 646–651.
- [55] R. Das, L. W. Kwok, I. S. Millett, Y. Bai, T. T. Mills, J. Jacob, G. S. Maskel, S. Seifert, S. J. G. Mochrie, P. Thiagarajan, S. Doniach, L. Pollack, D. Herschlag, *J. Mol. Biol.* **2003**, 332, 311–319.

- [56] D. K. Treiber, J. R. Williamson, *Curr. Opin. Struct. Biol.* **2001**, *11*, 309–314.
- [57] B. Onoa, S. Dumont, J. Liphardt, S. B. Smith, I. Tinoco Jr., C. Bustamante, *Science* **2003**, *299*, 1892–1895.
- [58] a) V. Lehnert, L. Jaeger, F. Michel, E. Westhof, *Chem. Biol.* **1996**, *3*, 993–1009; b) TtLSU.pdb, available from the website <http://www-ibmc.u-strasbg.fr/upr9002/westhof>.
- [59] C. Hyeon, J. Lee, J. Yoon, S. Hohng, D. Thirumalai, *Nat. Chem.* **2012**, *4*, 907–914.
- [60] X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, S. Chu, *Science* **2002**, *296*, 1473–1476.
- [61] S. V. Solomatin, M. Greenfeld, S. Chu, D. Herschlag, *Nature* **2010**, *463*, 681–684.
- [62] C. Hyeon, R. I. Dima, D. Thirumalai, *Structure* **2006**, *14*, 1633–1645.
- [63] A. Elcock, *Curr. Opin. Struct. Biol.* **2010**, *20*, 196–206.
- [64] M. S. Cheung, D. Klimov, D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4753–4758.
- [65] A. Dhar, A. Samiotakis, S. Ebbinghaus, L. Nienhaus, D. Homouz, M. Gruebele, M. S. Cheung, *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17586–17591.
- [66] D. Homouz, M. Perham, A. Samiotakis, M. S. Cheung, P. Wittung-Stafshede, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11759–11781.
- [67] A. P. Minton, *Biophys. J.* **2005**, *88*, 971–985.
- [68] D. Kilburn, J. H. Roh, L. Guo, R. M. Briber, S. A. Woodson, *J. Am. Chem. Soc.* **2010**, *132*, 8690–8696.
- [69] C. A. Theimer, L. D. Finger, L. Trantirek, J. Feigon, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 449–454.
- [70] M. A. Blasco, *Curr. Opin. Genet. Dev.* **2003**, *13*, 70–76.
- [71] J.-L. Chen, C. W. Greider, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14683–14684.
- [72] L. R. Comolli, I. Smirnov, L. Xu, E. H. Blackburn, T. L. James, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16998–17003.
- [73] C. A. Theimer, C. A. Blois, J. Feigon, *Mol. Cell* **2005**, *17*, 671–682.
- [74] S. Asakura, F. Oosawa, *J. Polym. Sci.* **1958**, *33*, 183–192.
- [75] M. R. Shaw, D. Thirumalai, *Phys. Rev. A: At. Mol. Opt. Phys.* **1991**, *44*, R4797.
- [76] N. A. Denesyuk, D. Thirumalai, *Biophys. Rev.* **2013**, *5*, 225–232.
- [77] T. Vulliamy, A. Marrone, I. Dokal, P. J. Mason, *Lancet* **2002**, *359*, 2168–2170.
- [78] P. F. Fogarty, H. Yamaguchi, A. Wiestner, G. M. Baerlocher, E. Sloand, W. S. Zeng, E. J. Read, P. M. Lansdorp, N. S. Young, *Lancet* **2003**, *362*, 1628–1630.
- [79] J. Lin, D. Thirumalai, *J. Am. Chem. Soc.* **2008**, *130*, 14080–14081.
- [80] J.-C. Lin, C. Hyeon, D. Thirumalai, *Phys. Chem. Chem. Phys.* **2014**.
- [81] F. Oosawa, *Polyelectrolytes*, Marcel Dekker, Inc., New York, **1971**.
- [82] C. Hyeon, R. I. Dima, D. Thirumalai, *J. Chem. Phys.* **2006**, *125*, 194905.
- [83] J. D. Honeycutt, D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526–3529.
- [84] T. Veitshans, D. Klimov, D. Thirumalai, *Folding Des.* **1997**, *2*, 1–22.

Received: January 26, 2014
Accepted: February 26, 2014
Published online: August 21, 2014