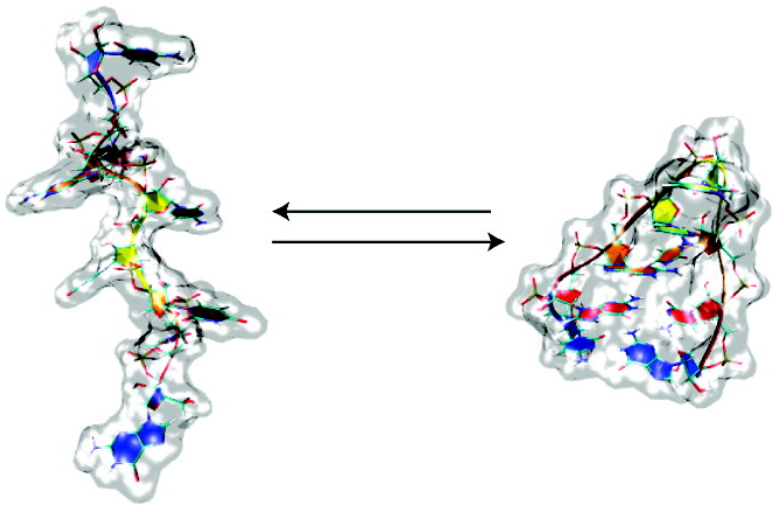
Communication

Simulation of the Pressure and Temperature Folding/Unfolding Equilibrium of a Small RNA Hairpin

Angel E. Garcia, and Dietmar Paschek

J. Am. Chem. Soc., **2008**, 130 (3), 815-817 • DOI: 10.1021/ja074191i



More About This Article

J|A|C|S
COMMUNICATIONS

Published on Web 12/23/2007

Simulation of the Pressure and Temperature Folding/Unfolding Equilibrium of a Small RNA Hairpin

Angel E. Garcia*,† and Dietmar Paschek‡

Department of Physics, Applied Physics and Astronomy, and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, and Physikalische Chemie, Otto-Hahn-Strasse 6, Universitaet Dortmund, D-44221 Dortmund, Germany

Received June 8, 2007; E-mail: angel@rpi.edu

We report the simulation of the folding/unfolding equilibrium thermodynamics of the r(GCUUCGGC) RNA tetraloop and calculate the free energy as a function of temperature and pressure. We predict the pressure and temperature elliptical stability diagram, $\Delta G(P,T)$, and show that this tetraloop will unfold at high hydrostatic pressures, similar to the behavior of globular proteins. The pressure-driven unfolding transition results in a small volume change of −4.1 mL/mol, favoring the unfolded state. The equilibrium thermodynamics is obtained with replica exchange molecular dynamics simulations, and the hairpin structures are formed without biasing the sampling due to initial conditions.

The simulation of the folding/unfolding of biomolecules has been one of the major computational challenges of molecular biology. The replica exchange molecular dynamics (REMD) simulations have been used to study the folding stability of peptides and small proteins in explicit water and using implicit solvent models.[1,2] The ability of performing unbiased folding of peptides and small proteins has enabled the further development and validation of force fields and computational methods.[3] The same success has been limited for nucleic acids.[4,5] The kinetics of folding of the r(GCUUCGGC) RNA studied by Ma et al. showed that the oligomer folded in the microsecond time scale and that the energy landscape was rugged, with the system adopting multiple conformations.[6] This sequence forms a stable UUCG tetraloop and a two CG base pair stem.[7] This eight nucleotide sequence is the smallest RNA that folds into a stable hairpin.[8,9] Motivated by Ma's work, we study the unbiased equilibrium folding of this RNA tetraloop, using the Amber force field (ff99)[10] in explicit TIP3P water.[11] We predict a pressure−temperature ($P-T$) free energy diagram for the oligomer and find that the RNA hairpin fold is destabilized by increases of hydrostatic pressure. The $P-T$ diagram has been extensively studied in proteins, but similar studies have not been conducted for RNA oligomers.[12,13]

The REMD is an enhanced sampling technique based on the parallel tempering Monte Carlo method[2,14] where multiple copies (or replicas) of identical systems are simulated in parallel at different temperatures. Periodically, state-exchange moves are attempted, where two neighboring replicas (in T) exchange their thermodynamic states (their temperature). The acceptance rule for each state-exchange move between two states i and j is chosen to be

$$P_{\mathrm{acc}} = \min\{1, \exp((\beta_i - \beta_j)(U(\vec{r}_i) - U(\vec{r}_j))\}$$

where $\beta_i = 1/k_B T_i$, T_i is the temperature of replica i, k_B is the Boltzmann constant, and $U(\vec{r}_i)$ represents the configurational energy of the system in state $\vec{r}_i$. The state-exchange acceptance probability, P_{acc}, has been shown to obey the detailed balance condition for an extended ensemble of canonical states.[15] Due to the temperature swaps, the extended ensemble does not describe the kinetics of the system. However, the kinetics of the system can be modeled by using a coarse-grained Langevin dynamics model.[16]

We perform REMD simulations of the r(GCUUCGGC) oligomer using 52 replicas and simulating for 226 ns per replica, for a total of over 11.75 μs of sampling. We calculate averages of the folding energy, folding specific volume, and pressure as a function of T. We simulated the RNA system starting from an extended, nonhelical, conformation with unstacked bases, solvated in 2557 waters, 14 Na^+ ions, and 7 Cl^- ions, representing 150 mM excess salt. The initial configurations of the systems do not have any bias toward the crystal structure configuration. The REMD simulations were conducted at constant volume, and changes in pressure result from changes in T.

The system was equilibrated for 2 ns at 330 K and 1 atm (0.1 MPa), resulting in a cubic box size of 4.3487 nm. We then performed 52, 5 ns simulations of the systems at various values of T to determine the average and standard deviation of the energy as a function of T needed to determine the optimal T spacing for the replicas.[17] We chose T in the range of 270−601 K, spread over 52 replicas. The temperatures of the replicas are chosen to maintain an exchange rate of 15−20%. State exchanges among replicas are attempted at random intervals with a probability of 5%, giving exchange rates of ∼2 ps. The entire simulation of 226 ns per replica adds up to a total simulation length of 11.75 ms, which took over 100 days when distributed over 52 Opteron processors.

To analyze the trajectories and classify the RNA states as folded or unfolded, we must use an order parameter. We explored many possible order parameters, including the rmsd distance from the crystal structure, the end-to-end distance, the radius of gyration (data not shown), and the number of native CG base pairs in the stem. We first use the rmsd distance from the crystal structure of the fragment (residues 31−37 of the structure with PDB code: 1F7Y) as an order parameter.[7] Figure 1 shows time series of the number of replicas that have reached the folded state for the first time (at time t) during the simulation and the total number of folded replicas (rmsd < 0.4 nm) in the ensemble of replicas. The rmsd < 0.4 nm criterion is very restrictive since it includes the low rmsd tail of the distribution of sampled distances, shown in Figure 2A. A single exponential fit of the curve in Figure 1a gives a folding time of 20 ns in the extended replica ensemble, and all replicas have folded at least once after 200 ns. The average number of folded replicas at any time reaches a steady state of 20 folded replicas after 50 ns. On the basis of these results, we chose the last 120 ns of the simulation to calculate ensemble averages and free energies.

Folding kinetics experiments on this tetraloop found that multiple states are sampled.[6] Indeed, the REMD calculation describes a diverse ensemble of conformations. Figure 2 shows representative structures populating the ensemble at low T (270 K). Structures

† Rensselaer Polytechnic Institute.
‡ Universitaet Dortmund.

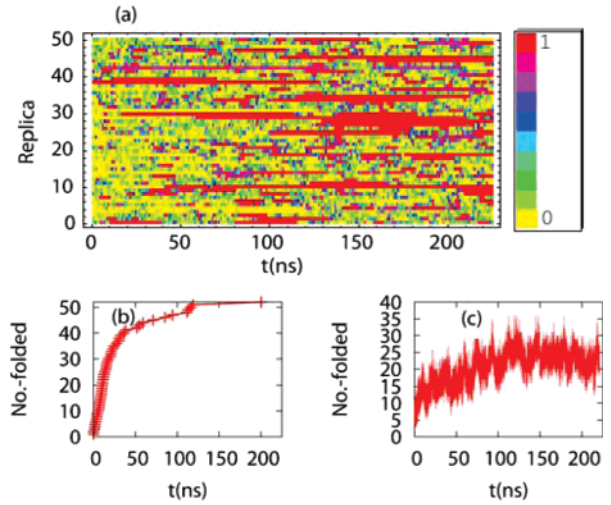


Figure 1. (a) Fraction of the time that each replica samples the folded state over 0.5 ns time blocks (red indicates a folded fraction close to 1, yellow indicates a folded fraction close to 0). (b) Time history of the number of replicas that have folded (rmsd < 0.4 nm) for the first time during the simulation. (c) Number of replicas sampling the folded state as a function of time (in ns).

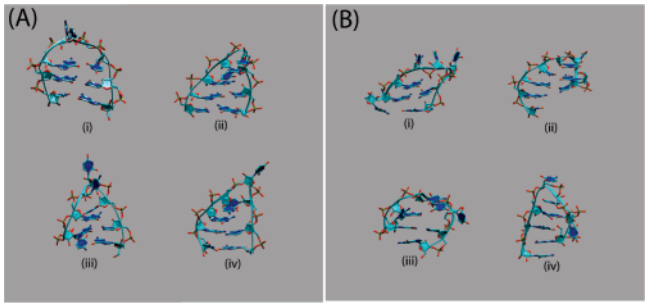


Figure 2. The RNA hairpin samples a diverse ensemble of conformations. (A) Representative configurations of cluster centroids with rmsd within 0.4 nm to the crystal structure.[7] The structures have different stacking arrangements of the loop bases. (B) Representative configurations of cluster centroids with rmsd within 0.4−0.6 nm to the crystal structure. The structures have different stacking and base pairs arrangements, including (i and iii) one native loop base pair (C2−G7), (ii) two non-native base pairs (C1−G6 and G2−C5), and (iv) a tightly stacked arrangement without base pairs. Clustering is done following the method by Daura.[18]

within 0.4 nm to the crystal conformation show the stem base pairing and different loop packing. Structures at higher distances show a variety of misfolded states.

Different order parameters highlight different structural features. Figure 3a shows histograms of the rmsd sampled by the ensembles at various T. Configurations sample rmsd values from 0.22 to 1.0 nm. These histograms show sharp distributions centered at 0.5 nm at low T and broad distributions centered at 0.6 nm and covering configurations from 0.4 to 1.0 nm at high T. Figure 3b shows similar histograms as a function of the end-to-end distance (measured from the C4′ atoms of the first and last nucleotide) at various T. Peaks at 0.6, 1.0, and 1.5 nm are present at low T and disappear at high T. The peak at 1.5 nm corresponds to the end-to-end distance for hydrogen-bonded bases. Shorter distances correspond to stacked end bases. At high T, the distributions are broad and peak at 2.3 nm, which corresponds to a stacked single-stranded helix. Figure 3c shows the rmsd distribution for configurations that contain at least one of the two native C−G base pair hydrogen bonds and shows that configuration with at least one hydrogen bond sample lowers rmsd.

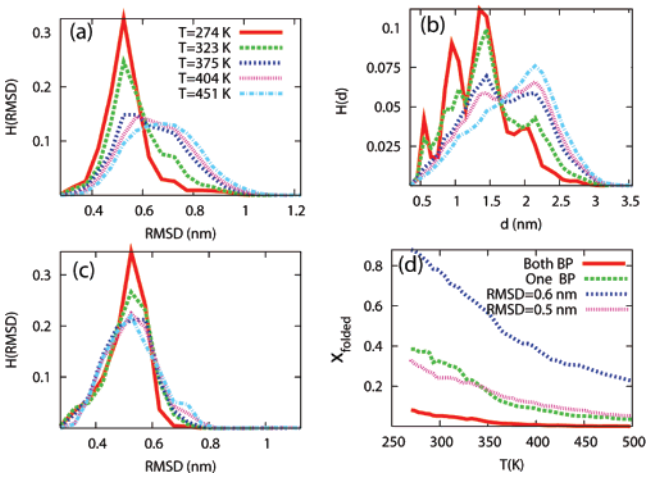


Figure 3. Distributions of the (a) rmsd, (b) end-to-end distance, and (c) rmsd of conformations that form at least one stem CG base pair. The distributions are shown for five temperatures. (d) Fraction of folded configurations as a function of temperature. Four different criteria are used to define the folded state: formation of both CG base pairs in the stem; formation of either base pair; rmsd < 0.5 nm; and rmsd < 0.6 nm.

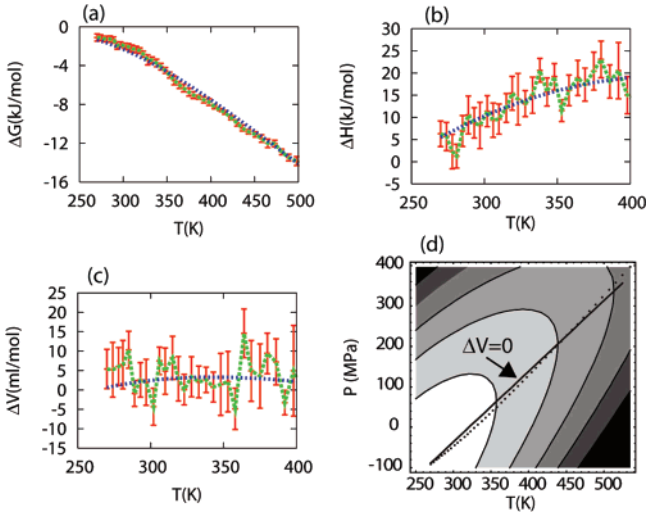


Figure 4. Ensemble averages of the difference upon unfolding of the (a) free energy, (b) enthalpy ($\Delta H = \Delta E + P\Delta V$, where ΔE is the difference in average potential energy), and (c) specific volume. (d) $P-T$ free energy diagram obtained by fitting the free energy and its derivatives (in b and c) to a polynomial in P and T. The green lines connect the calculated averages, and the blue lines describe the corresponding fits. Error bars are calculated using block averages.

Figure 3d shows the fraction of folded structures, x_{folded}, as a function of T obtained by using different order parameter criteria to define the folded state. Using the rmsd < 0.6 nm criterion for the folded state describes the transition from a hairpin to a stacked single helix. Using the rmsd < 0.5 nm criterion to characterize the folded state is compatible with the criterion of using the presence of either one of the two base pairs. At low T, <10% of the conformations have both stem CG base pairs formed. The thermodynamics of the folding/unfolding transition will depend on the order parameter used for defining the folded state. In what follows, we use the presence of one or more stem (native) base pairs as an order parameter that defines the folded state.

Figure 4 shows the ensemble average of the difference upon unfolding of the free energy, specific volume, and enthalpy as a function of T. The free energy of unfolding is defined as $\Delta G_u = G_{\text{unf.}} - G_{\text{folded}} = -RT \ln [(1 - x_{\text{folded}})/x_{\text{folded}}]$, where x_{folded} is the

fraction of folded states described in Figure 3d. The partial molar volume change of the RNA hairpin upon unfolding is obtained from calculations of the volume and number of water molecules at two different distances from the RNA atoms for ensemble of configurations in the folded and unfolded state. Figure 4c shows that the specific volume of RNA increases upon unfolding. The $P-T$ free energy diagram for the RNA hairpin unfolding is shown in Figure 4d. This $P-T$ diagram is obtained by simultaneously fitting the obtained changes in free energy, enthalpy, and volume along the isochore ($<P>$,T) points simulated by the REMD.

We used an approximation described by Smeller,

$$\Delta G_u(P,T) = \Delta G_u^{\,0} - \Delta S_u(T - T_0) + \Delta C_p(T - T_0) - \Delta C_p[T(\log(T/T_0) - 1) + T_0] + \Delta V_u(P - P_0) + \frac{1}{2}\Delta\beta(P - P_0)^2 + \Delta\alpha(P - P_0)(T - T_0)$$

where ΔC_p is the heat capacity, ΔS_u is the entropy, $\Delta G_u{}^0$ is the Gibbs free energy, $\Delta\beta$ is the compressibility change, ΔV_u is the volume change upon unfolding, and $\Delta\alpha$ is the temperature expansivity change upon unfolding.[19] All parameters are evaluated at a reference temperature T_0 (310 K) and pressure P_0 (0 MPa). The fitted values are $\Delta\alpha = 3.7 \times 10^{-4}$ kJ/mol·K·MPa, $\Delta\beta = -2.117 \times 10^{-4}$ kJ/mol·MPa,[2] $\Delta C_p = 0.31$ kJ/mol·K, $\Delta S_u = 0.034$ kJ/mol·K, $\Delta V_u = -4.1$ mL/mol, and $\Delta G_u{}^0 = -2.7$ kJ/mol. The calculated free energy of unfolding is small and similar to the measured value for this sequence (measured $\Delta G_u = -2.4$ kJ/mol).[9] However, our calculations are done at 150 mM NaCl salt, and the experiments are done at 10 mM sodium phosphate, 0.1 mM Na_2-EDTH, and pH 7.1.

In proteins and β-hairpins, the specific volume decreases upon unfolding, and proteins will unfold at high hydrostatic pressures, while for α-helical peptides, the specific volume increases upon unfolding.[13,20,21] We find that the RNA hairpin will unfold at high P and low T, with a similar degree of unfolding at high hydrostatic (~200 MPa) and 300 K pressures and at 400 K and 0 MPa, as indicated by the ellipsoidal $P-T$ diagram. The volume changes upon unfolding are positive for the (P, T) states sampled in our replica simulations (Figure 4c and dotted line curve in Figure 3d). This may appear to violate LeChatlier's principle, which requires lower volumes at higher pressures, but it does not. The corresponding $\Delta V = 0$ curve is shown as a solid line in Figure 4d. All (P, T) states above the $\Delta V = 0$ line will have a lower volume upon unfolding, thus favoring the unfolded state a high P and low T. Below the $\Delta V = 0$ line, states at high T and low P will show volume increases upon unfolding, which is typical of temperature unfolding.

Our calculations are the first atomic simulations of the folding/unfolding thermodynamics of an RNA hairpin where the hairpin is formed without biasing the sampling due to initial conditions. All our calculations were started from an extended unstacked conformation. Replica dynamics simulations aimed at studying the kinetics of folding of an RNA hairpin by Sorin et al. also reached the folded state.[4,5] These calculations and ours show that the Amber force field (ff99) is well suited to model the conformational transitions of the system.[10] The suitability of the ff99 to describe the dynamics of nuclei acids has been questioned by calculations on a DNA G-quadruplex, which found that the ff99 inadequately described this system.[22] It was found that when using the ff99 force field the backbone conformation of DNA irreversibly flipped to a minor (α,γ) = (g+,t) backbone conformation. However, these calculations spanned a much shorter time scale than ours, and for a larger system. Long (200 ns) MD simulations with the ff99 force field show that the sarcin−ricin loop of rRNA is stable, and that transitions in (α,γ) were reversible.[23] The ff99 has been modified recently to resolve the incorrect sampling of the α,γ backbone dihedral angles in DNA.[24] The effect of these modifications in the folding/unfolding thermodynamic stability of RNA hairpins needs to be explored further.

The studies presented here open the possibility of studying RNA conformational changes upon small molecule binding. Conformational transitions of RNA molecules upon binding of small molecules play an important role in regulating gene expression in bacteria.[25] Simulation studies can help understand the energetics of RNA conformational changes.

Acknowledgment. This work has been supported by the National Science Foundation (MCB-0543769).

Supporting Information Available: Details of the simulation, free energy fitting. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168−174.
(2) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(3) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782−2787.
(4) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Biophys. J* **2005**, *88*, 2516−2524.
(5) Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. *Biophys. J.* **2003**, *85*, 790−803.
(6) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530.
(7) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35−42.
(8) Molinaro, M.; Tinoco, I. *Nucleic Acids Res.* **1995**, *23*, 3056−3063.
(9) Proctor, D. J.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 14004−14014.
(10) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(12) Zipp, A.; Kauzmann, W. *Biochemistry* **1973**, *12*, 4217−4228.
(13) Panick, G.; Malessa, R.; Winter, R.; Rapp, G.; Frye, K. J.; Royer, C. A. *J. Mol. Biol.* **1998**, *275*, 389−402.
(14) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(15) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002.
(16) Yang, S.; Onuchic, J.; Garcia, A.; Levine, H. *J. Mol. Biol.* **2007**, *372*, 756−763.
(17) Garcia, A. E.; Herce, H.; Paschek, D. *Annu. Rep. Comput. Chem.* **2006**, *2*, 83−96.
(18) Daura, X.; Suter, R.; van Gunsteren, W. F. *J. Chem. Phys.* **1999**, *110*, 3049−3055.
(19) Smeller, L. *Biochim. Biophys. Acta* **2002**, *1595*, 11−29.
(20) Paschek, D.; Garcia, A. E. *Phys. Rev. Lett.* **2004**, *93*, 238105.
(21) Paschek, D.; Gnanakaran, S.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6765−6770.
(22) Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E., III; Sponer, J. *Biophys. J.* **2004**, *87*, 227−242.
(23) Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2006**, *34*, 697−708.
(24) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.
(25) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952−956.

JA074191I