

# Ensemble Simulations: Folding, Unfolding, and Misfolding of a High-Efficiency Frameshifting RNA Pseudoknot

Khai K. Q. Nguyen<sup>†,1,2</sup>, Yessica K. Gomez<sup>†,1,3</sup>, Mona Bakhom<sup>1</sup>, Amethyst Radcliffe<sup>3</sup>,  
Phuc La<sup>1</sup>, Dakota Rochelle<sup>1</sup>, Ji Won Lee<sup>1</sup>, and Eric J. Sorin<sup>1,\*</sup>

Departments of <sup>1</sup>Chemistry & Biochemistry,  
<sup>2</sup>Computer Engineering & Computer Science, and  
<sup>3</sup>Physics & Astronomy  
California State University Long Beach,  
Long Beach, California 90840, United States

---

## SUPPLEMENTAL INFORMATION: Computational Approximations and Sources of Uncertainty

**I. Potential Energy Function:** This article reports the results of all-atom point-charge molecular dynamics simulations using the AMBER potential energy function (1) ported to the GROMACS 3.3 molecular dynamics suite (2). Our porting of the AMBER force fields and relevant water models have been shown to numerically reproduce the results of these parameters when compared to the AMBER software (3), and are now distributed within the GROMACS software suite. The AMBER energy function employs, from top to bottom in Equation [1] below, van der Waals contact between non-bonded atoms; Coulomb interactions between all pairs of non-bonded atoms; a Hooke's Law treatment for all bonds, and a similar treatment of all bond angles; and sinusoidal treatments of proper and improper dihedral angles, where the latter is defined across a planar group involving a central atom bonded to three other atoms, as in the case of  $-NH_2$  amine groups.

$$\begin{aligned} E = & \sum_{\text{nb pairs}} 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \\ & + \sum_{\text{nb pairs}} k_{qq} \frac{q_i q_j}{\epsilon r_{ij}} \\ & + \sum_{\text{bonds}} \frac{1}{2} k_r (r_{ij} - r_0)^2 \\ & + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2 \\ & + \sum_{\text{dihedrals \& improper}} k_d [1 + \cos(n_{ijkl}\phi - \phi_0)] \end{aligned} \quad [1]$$

Notably, this approach allows for the application of multiple sinusoidal functions to each specific quartet of linearly bonded atoms participating in a given dihedral torsion angle, allowing for highly accurate parameterizations of torsion angle energy profiles, which are known to greatly affect the overall conformational preferences of proteins and nucleic acids (4,5). To allow larger simulation timesteps, and thereby enhance the speed of our simulations, it is assumed that all bonds are in their ground state configurations, *sans* molecular vibration, and all bonds are therefore constrained using the Linear Constraint Solver (LINCS) algorithm of Hess and co-workers (6).

This treatment employs single point charges on each atom, as derived via the Restrained Electrostatic Potential (RESP) method, which fits the electrostatic field for a given chemical group or molecule to that predicted by *ab initio* calculations (1). This method thus sacrifices the effects of atomic polarizability, which greatly decreases real-time simulation speed, to make simulations of large biochemical systems tractable on reasonable research timescales. In this case, the collective data set reported herein required approximately eight months to collect using tens of thousands of processors, which would not have been possible using a polarizable model.

**II. Solvation:** To account for both the bulk and molecular properties of the ionic solvent, the RNA was centered in a 75 Å cubic box (with Na<sup>+</sup> counterions) and solvated with approximately 13,500 TIP3P water molecules (7). The TIP3P, or Transferrable Intermolecular Potential 3-Point, model is a rigid 3-point model of water that employs energy function [1] and thus assigns partial point-charges to each atom. Although this model excludes the polarizability of molecular water and cannot reproduce some physical features of liquid water across a broad temperature range (8), this water model is well-understood; outperforms more sophisticated water models in several ways; provides a physically relevant solvation environment at standard and biological temperatures; and therefore remains widely used within the biomolecular simulation community (8,9).

To avoid physical artifacts that would be expected from a hard-walled or van der Waals container (i.e. edge or fringe effects), periodic boundary conditions are applied in all simulations. This is a standard simulation tool that effectively treats the system as an infinite number of identical neighboring cells and allows particles to traverse the boundaries of the aforementioned 75 Å cubic box without physical artifacts while maintaining constant pressure, temperature, and particle concentration, and which is also commonly used in rigorous *ab initio* calculations (10).

**III. Data Analysis:** The authors have attempted to highlight approximations and sources of uncertainty inherent to analysis of the reported simulation data set within the associated article. As noted in the text, the following should be considered with respect to the integrity of the data set reported and the authors' analysis and interpretation of that data.

- **K-means clustering** (3) is a heuristic method and is used herein only to determine the approximate time at which the simulated ensemble reaches steady state behavior.
- **Counterion treatment:** As the pseudoknot of interest has been shown to fold in the absence of divalent ions (11), the reported simulations were performed using sodium counterions, and it is unclear what influence species such as Mg<sup>2+</sup> might have on the conformational behavior of this species. While beyond the scope of the current report, this is an important target for future studies.
- **Native and non-native contacts:** Quantitative and algorithmic definitions of “native contacts” may vary across studies based on the system(s) studied and the models employed. Such definitions are largely arbitrary and must, therefore, be physically logical and justifiable while avoiding the introduction of artifacts within the structural analysis. The reported model and parameters used to define native contacts have been assessed for artifacts and are consistent with previously reported models (12,13). It should be noted, however, that minor changes to the parameters used to define native contacts would impact the overall appearance of data that employs such metrics (Table 1 and Figures 4 and 7), which are meant largely to provide a qualitative view of the conformational states and behavior of the system studied. As noted in the associated article, normalization of non-native contacts is similarly arbitrary and should also be interpreted from a primarily qualitative perspective.
- **Sampling thoroughness:** Finally, as noted in the associated article, it is not possible to verify that sampling of the configurational space of the studied species is complete, and there remains the possibility that regions outside of the reported configurational space remain unobserved within our simulated ensemble. The methods employed herein, however, were designed to optimize the sampling of the entirety of the configurational space and the thoroughness of that sampling. It is thus expected that regions that remain unsampled would include only high-energy states and those not participating to any significant degree in the conformational equilibrium of the system studied.

## REFERENCES

1. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179-5197.
2. Lindahl, E., Hess, B. and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual*, **7**, 306-317.
3. Sorin, E.J. and Pande, V.S. (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.*, **88**, 2472-2493.
4. Sorin, E.J. and Pande, V.S. (2005) Empirical Force Field Assessment: the Interplay Between Backbone Torsions and Non-covalent Term Scaling. *J. Comput. Chem.*, **26**, 682-690.
5. Pande, V.S., Sorin, E.J., Snow, C.D. and Rhee, Y.M. (2008) In Munoz, V. (ed.), *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*. Royal Society of Chemistry - Biomolecular Sciences, pp. 161-187.
6. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463-1472.
7. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926.
8. Horn, H.W., Swope, W.C., Pitara, J.W., Madura, J.D., Dick, T.J., Hura, G.L. and Head-Gordon, T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, **120**, 9665-9678.
9. Mao, Y. and Zhang, Y. (2012) Thermal conductivity, shear viscosity and specific heat of rigid water models. *Chem. Phys. Lett.*, **542**, 37-41.
10. Nishizawa, H. and Okumura, H. (2016) Rapid QM/MM approach for biomolecular systems under periodic boundary conditions: Combination of the density-functional tight-binding theory and particle mesh Ewald method. *J. Comput. Chem.*, **37**, 2701-2711.
11. Nixon, P.L., Cornish, P.V., Suram, S.V. and Giedroc, D.P. (2002) Thermodynamic Analysis of Conserved Loop–Stem Interactions in P1–P2 Frameshifting RNA Pseudoknots from Plant Luteoviridae. *Biochemistry*, **41**, 10665-10674.
12. Sorin, E.J., Rhee, Y.M. and Pande, V.S. (2005) Does water play a structural role in the folding of small nucleic acids? *Biophys. J.*, **88**, 2516-2524.
13. Mouro, Paulo R., de Godoi Contessoto, V., Chahine, J., Junio de Oliveira, R. and Pereira Leite, Vitor B. (2016) Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape. *Biophys. J.*, **111**, 287-293.