

# Experimental and Computational Analysis of the Unfolded State Ensemble in Nucleic Acid Folding (Supplementary Materials)

Yu Bai, Vincent B. Chu, Jan Lipfert, Vijay S. Pande,  
Daniel Herschlag, and Sebastian Doniach

June 30, 2008

## Hybrid Monte Carlo Sampling of the Tethered Duplex System

To sample the conformational ensemble of the unfolded tethered duplex system we employed a “hybrid” Monte Carlo (MC) algorithm to speed convergence. In this scheme, two Hamiltonians are sampled:  $H$ , an accurate but computationally-expensive Hamiltonian, and  $H'$ , its inexpensive approximation. For the tethered duplex system,  $H$  is the energy of a particular conformer and is composed of two terms:

$$H = H_E + H_{PEG} \quad (1)$$

where  $H_E$  is the inter-duplex repulsive energy, computed from Poisson-Boltzmann (PB) theory, and  $H_{PEG}$  is the contribution from the chain entropy of the PEG tether.  $H'$  is also similarly decomposed, with an approximation for  $H_E$ :

$$H' = H'_E + H_{PEG} \quad (2)$$

The hybrid MC algorithm is designed to speed convergence by substituting the computationally inexpensive  $H'$  for  $H$  for a few MC iterations during the simulation. The hybrid MC algorithm nests an inner loop of MC that samples  $H'$  within an outer loop that samples  $H$ . New candidate structures are generated for the inner loop by fixing one duplex of the tethered duplex and moving the other duplex with random rigid-body displacements and rotations [1]. After a fixed number of inner loop MC iterations, the structure is passed to the outer loop, with the inner loop acting as the “random move” for the outer loop. The standard Metropolis criterion was used to accept or reject structures in each loop [2]. At the end of each outer loop iteration, the atomic coordinates of the tethered duplex conformation are recorded for further analysis. The hybrid MC algorithm is written in the Nucleic Acid Builder (NAB) language to

take advantage of its molecular manipulation routines and is available from the authors by request [3].

In total, approximately 3000 outer loop iterations were run at each ion condition, with the first 1000 serving to equilibrate the system. The atomic coordinates of the last 2000 steps were recorded for analysis. Between each iteration of the outer loop, 200 inner iterations employing  $H'$  were executed. Calculations were performed on the Jacquard cluster at the National Energy Research Scientific Computing Center (NERSC). Visualization of the structural ensemble was performed using the PyMOL molecular graphics system (DeLano Scientific, USA).

### Calculation of $H_E$

In the outer loop, the electrostatic contribution was computed with the Adaptive Poisson-Boltzmann Solver (APBS v0.4.0), an open source software package, to solve the PB equation around all-atom representations of the DNA duplexes [4]. Atomistic models of the duplexes were generated using NAB [3] and atomic charges and radii were assigned for the atoms in the duplexes using PDB2PQR, a utility included in the standard APBS package [5].

Solutions to the PB equation were computed using the nonlinear form of the PB equation with boundary conditions set by the Debye-Hückel solution. The interior of the molecule was defined as the union of spheres centered on each atomic coordinate with radii equal to the sum of the van Der Waals radii and a solvent probe radius of 1.4 Å. The interior of the molecule was assigned a dielectric value of  $\epsilon = 2.0$  while the exterior was assigned  $\epsilon = 78.4$ . Charges were assigned to grid points using a cubic b-spline discretization. Grid resolution for all simulations was 0.5 Å.

### Calculation of $H'_E$

Due to the high computational cost of repeated solutions of the PB equation, a simple Yukawa-like interaction was used to compute the electrostatic contribution  $H'_E$  in the inner loop and speed convergence:

$$H'_E = \sum_{i,j} E_0 \frac{a}{r_{ij}} e^{-(r_{ij}-a)/\lambda} \quad (3)$$

where  $r_{ij}$  is the distance between phosphates,  $a$  sets the minimum distance between two phosphates (4 Å),  $E_0$  is the interaction energy between phosphates for  $r_{ij} = a$ , and  $i, j$  are indices over phosphate positions on separate duplexes.  $\lambda$  is a parameter, in Å, that sets the range of the interaction and can be adjusted to mimic the range of  $H_E$ .

### Calculation of $H_{PEG}$

$H_{PEG}$  is determined as the logarithm of the probability distribution of lengths observed in a separate MC simulation of the PEG tether in isolation. The PEG

tether in isolation was modeled as a freely-jointed, self-avoiding worm-like chain with geometric parameters (bond lengths, angles, and dihedral angles) for the C-C and C-O bonds taken from literature values [6]. 10,000 total MC steps were performed to obtain an end-to-end length distribution of the PEG tether in isolation; this distribution was then used to compute an empirical Hamiltonian for the tether as a function of the end-to-end distance of the chain.

## Supplementary Tables

Subensemble	Fitted Midpoints (mM)	
	Monovalent	Divalent
0	250	19.7
1	260	19.4
2	230	21.7
3	230	24.9
4	250	21.8
5	260	24.7
6	240	18.8
7	220	20.0
8	240	21.4
9	260	26.1
Mean	240	21.9
Full Ensemble	250	21.0
Variation	4%	4%

Table 2: Midpoints derived from 10-fold cross validation. The full ensemble was divided into 10 equal subensembles by random selection without repeats. From these subensembles, 10 scattering profiles were computed. These 10 scattering profiles were then substituted for the scattering profile computed from the full ensemble in the empirical Hill analysis (see “Critical assessment of PB theory”). The mean of the fitted midpoints from the 10 subensembles differed only slightly ( $\sim 4\%$ ) demonstrating that the simulation had converged.

## Supplemental Figures

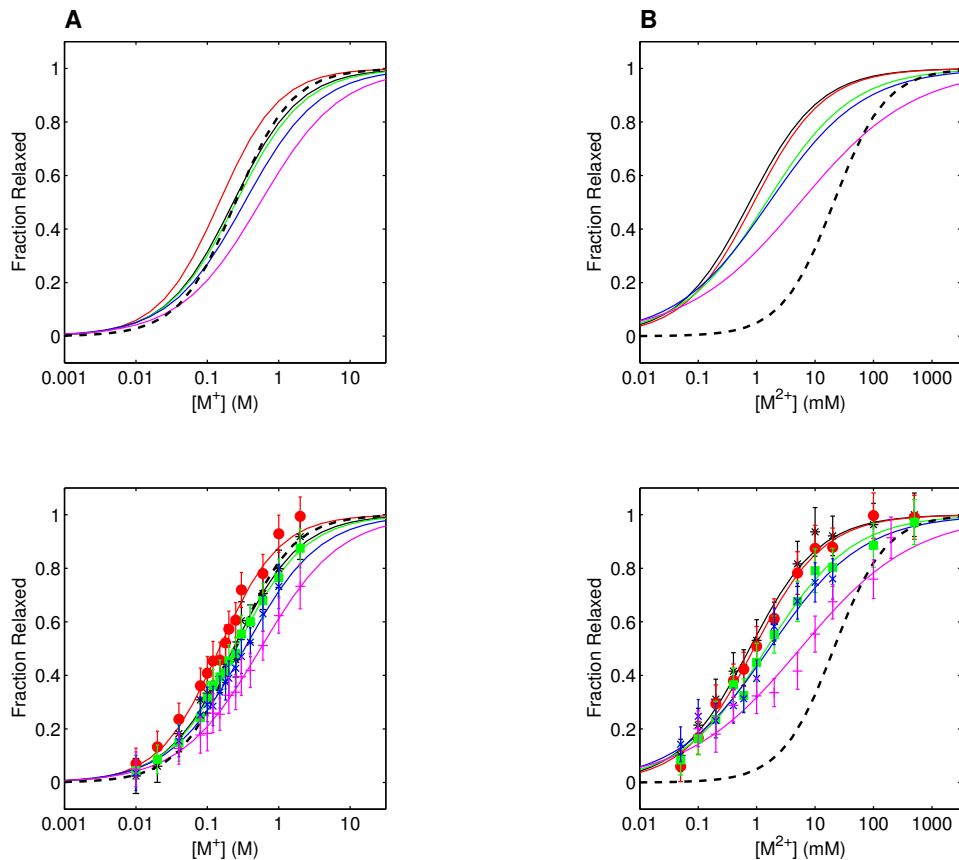


Figure 5: Salt-mediated structural transition fit to the empirical Hill model for different monovalent (A) and divalent (B) cations. For monovalent ions:  $\text{Li}^+$  (red),  $\text{Na}^+$  (black),  $\text{K}^+$  (green),  $\text{Rb}^+$  (blue),  $\text{TMA}^+$  (magenta). For divalent ions:  $\text{Mg}^{2+}$  (black),  $\text{Ca}^{2+}$  (red),  $\text{Sr}^{2+}$  (green),  $\text{Ba}^{2+}$  (blue), putresine $^{2+}$  (magenta). For both plots, transition derived from PB is shown by the black dashed lines. Plots on the top row are identical to plots on the bottom row except for the suppression of symbols in the top row for clarity.

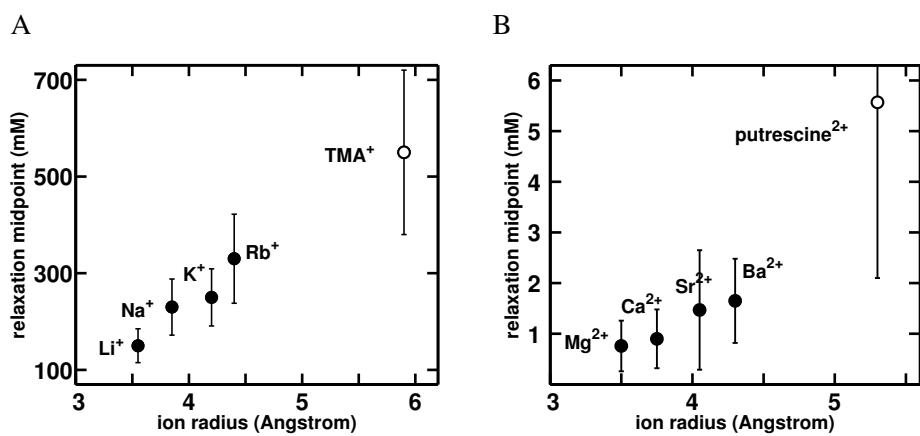


Figure 6: Observed correlation between relaxation midpoint and ion radius for monovalent ions (A) and divalent ions (B).

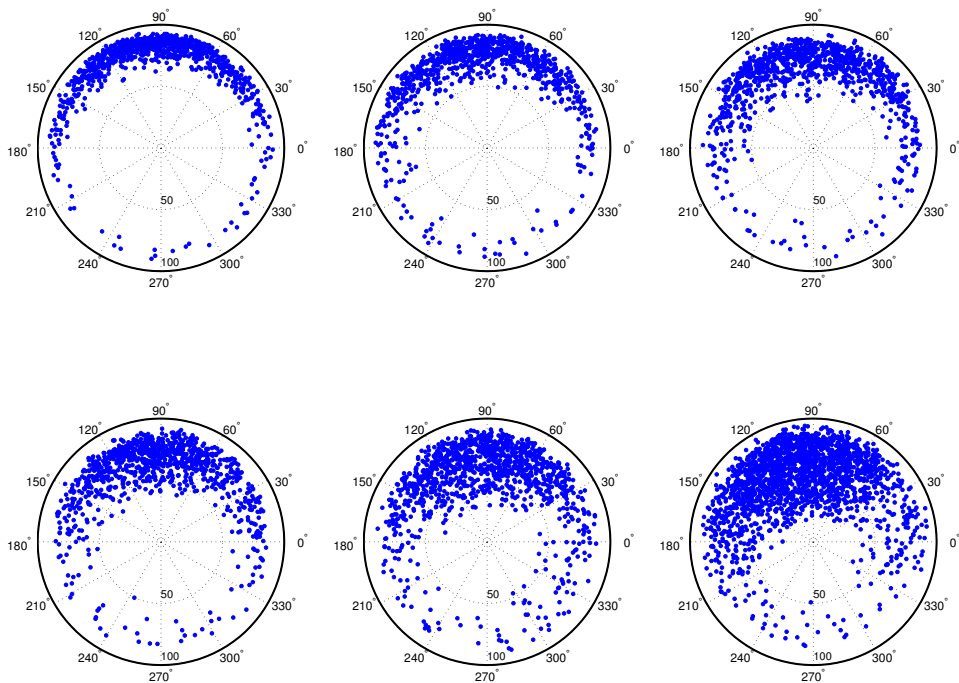


Figure 7: Polar projection of the conformational ensemble illustrated in Figure 3. The angle  $\theta$  ( $0 - 360^\circ$ ) represents the azimuthal angular coordinate of the proximal end of the mobile helix in a coordinate frame where the fixed duplex is aligned along  $+z$  axis and the tether attachment  $3'$  oxygen is on the  $+y$  axis. The radial coordinate represents the end-to-end distance between the two duplexes. From left to right, top to bottom: 0, 0.04, 0.15, 0.3, 2 M monovalent ion (in 16 mM monovalent background). The last figure shows the ensemble in the absence of electrostatics (i.e., steric effects only).

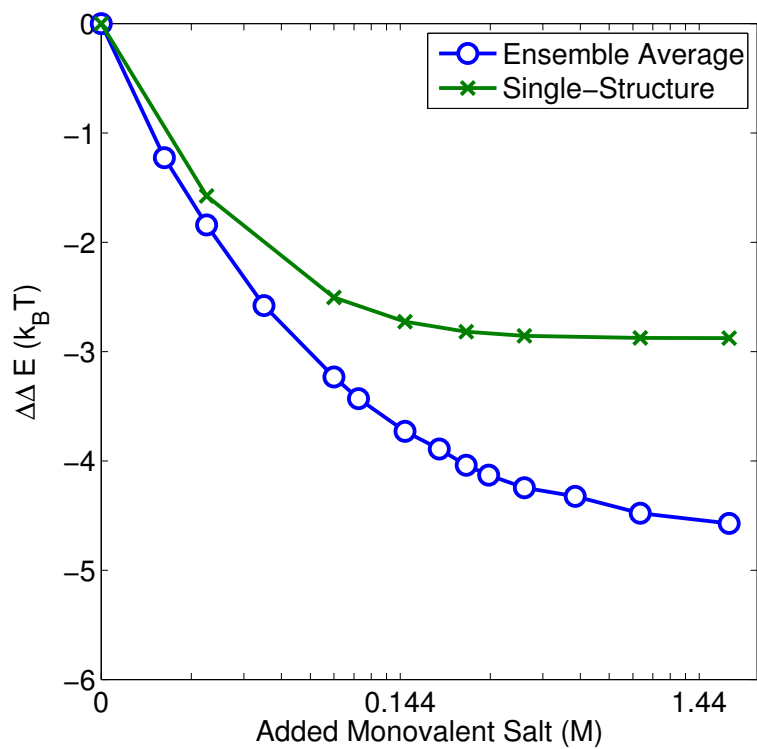


Figure 8: Change in mean electrostatic energy (relative to 16 mM monovalent background) as a function of ionic concentration.



## References

- [1] Leach, A. R. (2001) *Molecular Modelling*. (Prentice Hall), Second edition.
- [2] Metropolis, N & Ulam, S. (1949) *J. Am. Stat. Assoc.* **44**, 335.
- [3] Macke, T & Case, D. A. (1998) in *Molecular Modeling of Nucleic Acids*, eds. Leontes, N. B & Santa Lucia Jr., J. (American Chemical Society), pp. 379–393.
- [4] Baker, N. A, Sept, D, Joseph, S, Holst, M. J, & McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041.
- [5] Dolinsky, T. J, Nielsen, J. E, McCammon, J. A, & Baker, N. A. (2004) *Nucleic Acids Res.* **32**, 665–667.
- [6] Kienberger, F, Pastushenko, V. P, Kada, G, Gruber, H. J, Riener, C, Schindler, H, & Hinterdorfer, P. (2000) *Single Molecules* **1**, 123–128.