Supporting Methods

Preparation of the System. The system that we study is a 3-bp oligomer with sequence CGC and its complement. The initial coordinates were taken from those determined experimentally for the first three base pairs in the Dickerson Dodecamer (PDB code 1BNA; ref. 1). After deleting the final nine base pairs, we constructed the proper 3' and 5' termini where the strands were cut with ideal internal coordinate values for the CHARMM c27 parameter set (2). There were no phosphate groups on the 5' terminus of either strand, so there were four phosphate groups in all.

We separately equilibrated 1,345 TIP3 water molecules (3) in a 34 x 34 x 34-Å box. We placed the center of mass of the DNA molecule in the center of this box and deleted any waters with oxygen atoms within 2.8 Å of DNA heavy atoms. We then inserted sodium ions 6.0 Å from each phosphorus atom, along the bisector of the oxygens and in the plane created by the oxygens and the phosphorus. This procedure followed the prescription in ref. 4. Waters with oxygens within 2.4 Å of the sodiums were deleted. The final system consisted of 1,227 water molecules, 4 sodium ions, and 187 DNA atoms.

Cubic periodic boundary conditions were implemented by using the CRYSTAL module in CHARMM. The particle mesh Ewald method (5) was used to treat the nonbonded interactions; the width of the Gaussian was 0.34 $Å^{-1}$, the fast Fourier transform used 32 points in each dimension, and a sixth-order B-spline was used for interpolation. The energy was minimized by using the steepest descent method for 3,000 steps followed by the Adopted Basis Newton Rapson (ABNR) method until the gradient was reduced to <0.001. Using the Leapfrog Verlet integrator with a time step of 1 fs, the system was heated to 300 K in increments of 30 K over 20 ps, and was then run for an additional 20 ps, during which the velocities were rescaled if the temperatures fell out of the range $295 < T < 305$ K. Dynamics were then run for 800 ps without velocity rescaling. The energy and temperature were found to be reasonably constant over this time (the ratio of the average fluctuation in the total energy to the average total energy was $<10^{-4}$). In addition, all dihedral angles fluctuated within the appropriate ranges (6), and the average rms deviation of the heavy atoms from the crystal structure was ≈0.9 Å.

It is interesting to note that, although, on average, three of the four ions could be found near the DNA, the ions were extremely mobile during the dynamics. The diffusion constants for sodium ions were averaged for 95 trajectories taken from TPS simulations (set B, discussed below) which resulted in a average value of 0.21 $Å^2$/ps. This is higher than the value for the diffusion coefficient of sodium in bulk water found in experiment (0.12 $Å^2$/ps; ref. 7), but reasonably close to the value of 0.17 $Å^2$/ps measured for a simulation of DNA, ions, and water (8) that used the AMBER forcefield (9).

Because the simulation was carried out in a periodic box that was only 34 Å in each dimension, we were concerned that interactions between the DNA molecule and its images might become significant as the chains were separated. To check this possibility, we carried out a simulation in which we applied a harmonic bias potential that pulled the two strands of DNA apart and measured the interactions between the primary and image DNA molecules. The bias was applied based on the average

hydrogen bond heavy atom distance of the middle base pair (G2 and C5) and the minimum was incremented from 3.0 to 9.2 Å in steps of 0.4 Å. The hydrogen bond distances of the other two bases were generally similar to, or larger than, those of the middle base pair during the simulation. We found that the (primary DNA)–(DNA image) electrostatic energies were on average 0.5% of the overall DNA–DNA electrostatic interactions, although a fluctuation as large as 10% was observed.

High-Temperature Melting Simulations. To obtain an initial set of reactive trajectories and determine the behaviors of various order parameters during base pairing and unpairing, we simulated the dynamics of the system at temperatures ranging from 350 to 1,100 K in increments of 50 K. At each temperature, we performed one simulation, which began with the final positions and velocities from the equilibrium native state simulation described above. In each case, the system was heated from 300 K to the desired temperature over 10 ps, and then velocities were scaled to maintain this temperature for another 10 ps. The system was then simulated for 500 ps without velocity rescaling. In these simulations and all subsequent ones, a time step of 2 fs was used. The lowest temperature at which a base pair unbinding event occurred was 400 K.

Based on these results, we carried out 43 molecular dynamics simulations at 400 K, as described in the main text. The frequency with which each base pair unfolded is shown in Table 1. Every unfolding event was initiated at one of the end base pairs. Structures with only one base pair unfolded were observed for long periods of time (120 ps on average) before another base pair unfolded, which suggests that such states are metastable at 400 K. The remaining two base pairs unfolded in either order and sometimes simultaneously. These observations suggest that unfolding of the 3-bp molecule at 400 K could be considered a two-step process: unfolding of an end base pair, followed by unfolding of the remaining two base pairs. Although some structures with two base pairs unfolded also persisted for long periods of time, with an average lifetime of about 75 ps, we chose to focus on the unfolding of and end base pair (C1 and G6).

During the simulations, we monitored several order parameters to evaluate their ability to distinguish between the native and unfolded states. These coordinates are described in *Coordinates* and their variations during unfolding are shown in Table 3. We initially decided to define an unbound base pair by an average hydrogen bond heavy atom distance (l_{16}) >7.0 Å, because such structures visually appeared unfolded. However, it was recognized that this criterion needed to be refined, because ≈20% of trajectories that reached this state refolded. A histogram of values of l_{16} shows a shallow minimum ≈7.0 Å (see Fig. 6).

Coordinates. The following coordinates were considered during the high-temperature simulations and/or the TPS simulations.

1. The heavy atom distance averaged over the three hydrogen bonds between the unbinding bases C1 and G6 (l_{16}).

2. The average of the three out-of-line hydrogen bond angles between C1 and G6 (a_{16}). In the case of the bonds involving NH_2 groups, which contain two possible donor hydrogens, the hydrogen that had the smaller angle was

considered to be the participating species.

3. The number of waters around bases (n_w). Specifically, we used the continuous function:

$$n_w = \frac{1}{3}\Sigma_i\Sigma_j \frac{1}{1+\exp[\alpha(r_{ij}-3.5)]}. \quad \textbf{[1]}$$

The first sum (i) is over DNA heavy atoms, the second sum (j) is over water oxygen atoms, and r_{ij} is the distance between atoms i and j. The paramater, α, is a smoothing parameter (we used $\alpha = 2.0$). Because distances are calculated between water atoms and each DNA heavy atom, some water atoms contribute to the sum in Eq. **1** more than once. We found empirically that the factor of 1/3 made n_w approximately equal to a discrete count of the number of waters within 3.5 Å of any heavy atom in C1 or G6 (with no overcounting).

4. A pseudo dihedral angle based on the centers of mass of four groups (a_d), as described in ref. 10. For example, in the case of base 1 unbinding, the groups are: the nonbackbone atoms of bases 5 and 2, the sugar in base 2, the sugar in base 1, and the nonbackbone atoms in base 1. Thus, the axis for the dihedral is the line between the sugars of bases C1 and G2.

5. The Lee and Richards surface area (11) of the DNA molecule (S_{LR}).

6. The rms deviations of the DNA heavy atoms from the crystal structure (d_{rms}).

7. The total hydrogen bond energy of the unbinding bases (including both electrostatic and van der Waals energies), given by a sum over each hydrogen bond of the interaction energy of the acceptor atom, the donor hydrogen, and the attached heavy atom (h_{16}).

8. The angle between the normals to the base planes (a_n).

9. Properties of the helix as calculated with the program CURVES (12): shear (Sx), stretch (Sy), stagger (Sz), buckle (κ), twist (Ω), opening (σ), shift (Dx), slide (Dy), rise (Dz), tilt (τ), roll (ρ), and propeller twist (ω).

10. The dihedral angles along the DNA backbone ($\alpha,\beta,\gamma,\delta,\varepsilon,\zeta$) and for the glycosidic bond (χ) (13).

11. The density of waters between (instead of around as in n_w) the unbinding bases (ρ_b). Here, the center of mass of bases 1 and 6 was determined and a cylindrical coordinate system was defined with the line connecting the two centers of mass as the z axis. All waters with a z coordinate between the two bases were then selected and counted according to a function similar to that used for n_w:

$$n_b = \frac{1}{1 + \exp[\alpha(r - r_0)]}, \quad [2]$$

where α was chosen to be 5 and r denotes the radial distance from the z axis. The value of r_0 was set to 3.0. The density was then determined by $\rho_b = n_b / V$, where V is the volume of the cylinder between the centers of mass of C1 and G6.

12. The number of waters in the region above bases 2 or 5 (n_a) was counted in a similar manner. Here, a cylindrical coordinate system was defined for each ring structure in bases C5 (one ring) and G2 (two rings) with the z axis defined as the normal to the plane defined by three atoms in the ring. Waters were counted by using Eq. **2**, with r_0 given by:

$$r_0 = r_{ring} + a, \quad [3]$$

where r_{ring} denotes the average distance from the center of mass to the position of the atoms in the ring and a was empirically determined. Values of α from 2 to 5 were tried and the height of the cylinder ranged from 5 to 6 Å. In the native state, there is almost no water in this region for reasonable values of the parameters,(eg, α = 5, the cylinder height = 5.5 Å, and a = 0.5 Å).

13. The degree of tetrahedral ordering of waters in the vicinity of the unpairing bases, as measured by the order parameter considered in refs. 14 and 15:

$$\psi = 1 - \frac{3}{8} \Sigma_{j=1}^{3} \Sigma_{k=j}^{4} (\cos\theta_{jk} + \frac{1}{3})^2, \quad [4]$$

where θ_{jk} is the angle formed by the lines joining the oxygen atom of a given water molecule and those of its nearest neighbors j and k. For perfect ice, this gives $\psi = 1.0$, whereas it is zero for complete isotropy (15). In the case of liquid water at room temperature, $\psi \approx 0.6$.

14. We have already considered the breaking of DNA–DNA hydrogen bonds with h_{16}, but we felt that the formation of hydrogen bonds between water and DNA might also be important. Therefore, we considered the overall interaction energy between all waters and the DNA hydrogen bonding groups (h_w). A hydrogen bonding group consists of either a donor hydrogen and its heavy atom or an acceptor heavy atom. The values given in Table 3 and Fig. 11 sum over all the hydrogen bonding groups in bases C1 and G6. We additionally considered the sum of the largest N interactions between individual water molecules and DNA hydrogen bonding groups, where N ranged from 1 to 4. This is not shown in Table 3 because it did not provide any further information than h_w.

15. The number of hydrogen bonds formed between DNA groups and water molecules (n_h). For the DNA, the hydrogen bond donor and acceptor definitions in the CHARMM c27 residue topology file were used. We count a

hydrogen bond when the heavy atom distance in a donor acceptor pair is <3.4 Å and the out-of-line angle is $< 70^{\circ}$.

16. The density of dangling or unformed hydrogen bonds (including both DNA and water groups) in the vicinity of the DNA (ρ_d). Hydrogen bonds are calculated as described above.

17. The radius of gyration of all atoms in bases C1 and G6 (R_g).

18. Two angles describing rotation of C1 around the phosphorous atom (θ and ϕ). This is the analog of a_d for and end base; because there is more freedom of motion two angles are required. The angle θ is given by the angle between the nucleotide center of mass for C1, the phosphorous on the 5' side of G2 and the position of the C1 nucleotide center of mass in the crystal structure (see Fig. 7). The latter vector is determined at any point along a trajectory by predetermining its value in a coordinate system based on the plane formed by the phosphorous, the center of mass of the sugar, and the center of mass of the nucleotide of base G2. By arbitrarily defining this vector as the z axis and then choosing an orthoganal x axis, we also determine an azimuthal angle, ϕ. It was found that the relative positions of the three groups in G2 that form the reference plane essentially do not change during the unfolding of C1, thus it was possible to calculate $\theta$and ϕ at any point along a trajectory.

19. The overall electrostatic and van der Waals interaction energy between bases C1 and G6 (E_{16}). All atoms in C1 and G6 are included except backbone atoms. This coordinate detects attractive interactions in conformations in which C1 stacks above G6 as well as the hydrogen bonding energy that stabilizes the native state.

20. The overall electrostatic and van der Waals interaction energy between bases C1 and G2 (E_{12}). This coordinate monitors the stacking interactions that provide important energetic stabilization in the native state.

21. The distance between the centers of mass of bases C1 and G6 (d_{16}).

22. The distance between the centers of mass of bases C1 and G2 (d_{12}).

TPS of the End Base Pair Unfolding. The basin designations and acceptance statistics for the TPS simulations we carried out are shown in Table 2.

TPS set A. We focus on reactions in which the hydrogen bonds between C1 and G6 dissolve and the base C1 flips out of the DNA base stack. The unfolding trajectories described above were used to initialize a set of transition path sampling simulations 400 K (set A). Trajectories were harvested in which one endpoint was in the native state and the other end point was in the unfolded state. It was found that the coordinates n_w and h_{16} serve as better order parameters than l_{16} and a_{16}. The frequency of observing values of these coordinates in this set of trajectories is shown in Fig. 8.

TPS sets B and C. These trajectories resulting from set A were annealed to 300 K by rescaling the velocities during shooting moves. The acceptance criteria was further refined over several sets of TPS simulations. In these and subsequent simulations, we harvested trajectories in which one end point was in the folded basin and at least one point anywhere along the trajectory visited the unfolded basin. Two types of moves were employed to generate trial paths: shifting and shooting. Because the acceptance rates for full shooting moves (set B) were smaller than 5% at 300 K, half shooting moves were carried out in which no velocity perturbation was applied (set C); rather, the motions of the waters far away (>15.0 Å) from the center of the DNA molecule were simulated with a stochastic (Langevin) algorithm and the path was only updated in one direction. Shooting in only one direction at a time significantly increased the acceptance rate to ≈0.4, which is consistent with ref. 16. Paths were recorded every five trial moves.

We first verified that the half path shooting moves yielded results consistent with those from the basic shooting method. Points taken from simulations with full shoots and from ones with half shoots are shown in Fig. 9, which were overlayed on a projection of the free energy onto these coordinates. We see that the two sets of dynamics are similar, as are properties calculated from the two methods, such as $\langle h(x_t)\rangle$ (see next paragraph), although the statistics are poor for the full shooting method because the acceptance rate was so low.

To determine whether our paths were long enough, we calculated (using the notation of ref. 17) $\langle h(x_t)\rangle_{AB}^{*}$, where $\langle ? \rangle_{AB}^{*}$ denotes an average over paths in which one end point resides in the A basin (the native state) and there is some point in the B basin (the unfolded state) and x_t denotes the coordinates at time, t. The indicator function ($h(x_t)$) is defined such that:

$$h(x_t) = \begin{cases} 1 & \text{if } x_t \in \text{unfolded state} \\ 0 & \text{otherwise} \end{cases} . \qquad \textbf{[5]}$$

It is shown in ref. 17 that the path length should be long enough so that $\frac{d\langle h(x_t)\rangle_{AB}^{*}}{dt} = \text{constant}$. In addition, because we are not restricting the end of the paths, this mode of sampling allows us to check "on the fly" whether we have effectively described the B basin. If some structures that are designated as being in basin B have not really committed, we should see these trajectories quickly return to state A.

The calculated $\langle h(x_t)\rangle_{AB}^{*}$ is shown in Fig. 10; note that it saturates at a value of ≈0.27. This means that ≈73% of the trajectories that reached the unfolded state left within the simulation time. In fact, we found that most of these trajectories returned to the native state. This means that many of these structures that have been designated as being in the unfolded state must still be connected to the native state. In other words, there is another important coordinate that must be considered, so we do not have a discriminating set of order parameters.

Determination of descriminating order parameters. Unfortunately, there is no systematic way to obtain a discriminating set of order parameters; rather, the process requires a combination of intuition and trial and error.

We began our analysis by running short (20 ps) trajectories from structures taken from the path sampling runs and calculated a probability for the end of the trajectory to be in the unfolded state, P_u. [This calculation was similar to that described in *Results*, in the main text, except those committor estimations used longer (200 ps) trajectories]. Although the value of P_u was correlated to h_{16}, there was essentially no dependence on n_w. Thus, n_w does not appear to be an important factor in designating the unfolded state. Although they were correlated to the value of h_{16}, committor estimations showed that this coordinate alone is not sufficient to determine whether a structure is committed to the unfolded basin, as the TPS simulations already indicated.

Because the committor values appear to be correlated to h_{16}, but not n_w, we decided to try alternative order parameter sets that retained h_{16} but replaced n_w. To evaluate potential coordinates in an efficient fashion, we selected two setsof structures from trajectories that had already been sampled. The first set consisted of structures from TPS trajectories for which $h_{16} > -0.5$ kcal/mol, but in which the trajectory returned to the native state within 30 ps. For the second set, we selected structures that appeared to be truly committed to the unfolded state. First, we selected from the trajectories already sampled, two in which it was clear that the unfolded state had been reached (as judged from the fact that either C1 or G6 remained out of the DNA base stack for >60 ps). We additionally selected several trajectories run during umbrella sampling calculations in which the system was biased toward values of $h_{16} = 0$. As these trajectories were equilibrated for 800 ps at values of $h_{16} \approx 0$, we assumed that all coordinates had committed to the unfolded state in these trajectories. (This assumption was later checked by running committor estimations from some structures from these trajectories. We found that committor values were almost always $P_u = 1.0$ and always $P_u \geq 0.9$.) We then examined whether potential order parameters were helpful in differentiating between the two sets of structures. The potential order parameters that we investigated aregiven in *Coordinates* and the average values of these parameters in different states, as calculated during the final set of TPS simulations (set D), are shown in Table 3.

We found that most of the candidate coordinates exhibited a change in value as h_{16} increased above –0.5 kcal/mol. However, if we restrict our investigation to regions where $h_{16} > -0.5$ kcal/mol, there is no difference in the distributions of most of these coordinates (including all coordinates that explicitly involve solvent properties) for structures committed to the unfolded state and those that were not. An example of this is shown in Fig. 11, where distributions of h_w are shown for uncommitted TPS trajectories and committed umbrella sampling trajectories. The distributions are statistically identical. Thus, these coordinates are not useful for designating the unfolded state; rather, they are slaved to the coordinate h_{16}.

This was not the case, however, for one coordinate, E_{12} (see Fig 12). This coordinate monitors the stacking of C1 and G2. Based on analysis of the trajectories at 400 K, it had appeared that stacking interactions were broken concomitantly with the hydrogen bonds. However, as Fig. 12 shows, there was often significant stacking energy remaining after the hydrogen bonds were mostly broken at 300 K. In addition, we found that it was necessary to consider stacking interactions between C1 and G6. Therefore, the coordinate E_{16} was used in place of h_{16}. As described in the main text,

we found that E_{12} and E_{16} were no longer hydrodynamic modes when C1 was almost completely unfolded, so the coordinates d_{12} and d_{16} were also used.

Set D TPS simulations. Trajectories were first annealed to the new order parameters (E_{12}, E_{16}, d_{12}, and d_{16}) by carrying out several sets of TPS simulations in which trajectories were lengthened and forced deeper into the unfolded state. The trajectories resulting from this process were then used to initialize the final set of simulations, set D, which were 400 ps in length. This length was chosen to be longer than the relaxation time for crossing the barrier, which was found to be ≈200 ps. This can be seen in Fig. 13, which shows $\langle h(x_t)\rangle^*_{AB}$ for the set D simulations. As mentioned in *TPS of the End Base Pair Unfolding*, the barrier crossing times can be associated with the point at which $\frac{d\langle h(x_t)\rangle^*_{AB}}{dt} = \text{constant}$ (17). As mentioned in the main text, these long barrier crossing times result because there is only a weak driving force for motion from a region of $d_{12} \approx d_{16} \approx 10$ to the completely unfolded region ($d_{12} \approx 12.5, d_{16} \approx 17$). The results of the TPS simulations with the final set of order parameters are described in the main text.

Free Energy. As described above and in the main text, we have projected the coordinates onto several sets of coordinates at 300 K. Harmonic (except as described below) bias potentials were used to constrain the system near particular values of these coordinates. In the case of the energy coordinates and n_w, it was necessary to modify the CHARMM user subroutine to incorporate the bias potential. The bias potential for the distance variables was applied using the RXNCOR module. In all cases, umbrella sampling was carried out with CPT dynamics. Pressure was controlled using the Langevin piston method (18), with a reference pressure of 1 ATM, a piston mass of 400 atomic mass units (amu), and a friction coefficient of 20.0 ps^{-1}. Each simulation consisted of a 200 ps equilibration run in which the temperature was controlled with a Nose–Hoover thermostat (the mass of the fictitious degree of freedom was 1,000 amu), followed by a 600-ps production run without the Nose–Hoover thermostat. In all cases, the temperature was close to 300 K. Values of the coordinates were saved every 20 fs during the production runs.

The free-energy was obtained from these data by using the weighted histogram analysis method (WHAM) (19). Window-free energies were iterated until the maximum change in free energies was <0.0015 kcal/mol.

Projection onto h_{16} and n_w. Sampling was carried out in windows whose centers were separated by 1 kcal/mol in h_{16} and 2/3 of a water in n_w. Harmonic constraints were applied for each coordinate with force constants of 1.0 $(\text{kcal/mol})^{-1}$ for h_{16} and 0.6 kcal/mol for n_w. The starting conformations for windows were obtained as follows. The final coordinates and velocities from the long equilibration at 300 K described above were used as the initial structure for the first window. A 4-ps run was carried out with force constants of 2.0 $(\text{kcal/mol})^{-1}$ for h_{16} and 8.0 kcal/mol for n_w. The final structure from this simulation was used as the initial structure for the next window and the procedure was repeated for each window. Additional simulations were carried out in which the bias was centered around values of h_{16} = 0.5 kcal/mol. Windows were not carried out in regions where preceding windows indicated that the free energy was large. Results are shown in Fig. 9.

Projection onto E_{12} ***and*** E_{16}. Simulations were carried out as described above, except the initial structures were taken from TPS simulations. Simulations were carried out with windows centered in increments of 2.0 kcal/mol in E_{16} ranging from -16.0 to 2.0 kcal/mol, with a force constant of 0.25 (kcal/mol)$^{-1}$. Then, three additional sets of windows were centered at values of -1.0, 0.0, and 1.0 kcal/mol with force constants of 0.75 (kcal/mol)$^{-1}$. For each of these values of E_{16}, windows were centered in the dimension E_{12} from -7.0 to 1.0 kcal/mol in increments of 1.0 kcal/mol and with a force constant of 0.75 (kcal/mol)$^{-1}$. To improve sampling in regions of $|E_{1i}| < 1.0$, we also carried out simulations in which the bias force was linear in the logarithm of the coordinates:

$$U_{bias} = K_{bias}(|E_{1i}| \log |E_{1i}| - |E_{1i}| - \log E_{1i}^{0} + E_{1i}^{0}), \quad [6]$$

where i refers to 2 or 6, K_{bias} is the force constant, and the last term (E_{1i}^{0}) ensures that the potential is zero at the point of zero force. Using this bias potential, we carried out simulations in which $E_{1i,0} = 1.0, 10^{-1}, \ldots 10^{-4}$. Force constants were given by $K_{bias} = 0.75 E_{1i,0}^{-1}$. Results are shown in the main text.

Projection onto d_{12} ***and*** d_{16}. Simulations were carried out as described above. Windows were separated by 2 Å in both coordinates, in a grid encompassing $d_{12} = 2$ Å to $d_{12} = 15$ and $d_{16} = 2$ to $d_{16} = 24$ Å with force constants of 0.25 kcal/mol/Å. A number of additional windows were carried with force constants of 1.0 kcal/mol/Å in regions with large free energy gradients. Windows were not carried out in regions where preceding windows indicated that the free energy was large. Results are shown in the main text.

1. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2179–2183.

2. MacKerrel, A. D. J. & Banavali, N. K. (2000) *J. Comput. Chem.* **21,** 105–120.

3. Jorgensen, W. L, Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983) *J. Chem. Phys.* **79,** 926–935.

4. Martinez, J., Elmroth, S. K. C. & Kloo, L. (2001) *J. Am. Chem. Soc.* **123,** 12279–12289.

5. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995) *J. Chem. Phys.* **103,** 8577–8593.

6. Foloppe, N. & MacKerrel, A. D. J. (2000) *J. Comput. Chem.* **21,** 86–104.

7. Tyrell, H. J. & Harris, K. R. (1984) in *Diffusion in Liquids* (Butterworths, London), pp. 387–437.

8. Makarov, V. A., Feig, M., Andrews, B. K. & Pettitt, B. M. (1998) *Biophys. J.* **75,** 150–158.

9. 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. & Kollman, P. A. (1995) *J. Am. Chem. Soc.* **117,** 5179–5197.

10. Banavali, N. K. & MacKerrel, A. D. J. (2002) *J. Mol. Biol.* **319,** 141–160.

11. Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.

12. Lavery, R. & Sklenar, H. (1989) *J. Biomol. Struct. Dyn.* **6,** 655–667.

13. Bloomfield, V. A., Crothers, D. M. & Tinoco, I. J. (2000) *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausilito, CA).

14. Chau, P.-L. & Hardwick, A. J. (1998) *Mol. Phys.* **93,** 511–518.

15. Errington, J. R., Debenedetti, P. G. & Torquato, S. (2001) *Nature* **409,** 318–321.

16. Bolhuis, P. G. (2003) *J. Phys. Condens. Mat.* **15,** S113–S120.

17. Dellago, C., Bolhuis, P. G. & Geissler, P. L. (2001) *Adv. Chem. Phys.* **123,** 1–78.

18. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. (1995) *J. Chem. Phys.* **103,** 4613–4612.

19. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. (1992) *J. Comput. Chem.* **13,** 1011–1021.