

# Atomistic Simulation of the DNA Helix–Coil Transition<sup>†</sup>

Stefano Piana<sup>‡</sup>

Nanochemistry Research Institute, Curtin University of Technology, GPO Box U1987, 6845 Perth, WA, Australia

Received: July 19, 2007; In Final Form: August 25, 2007

The helix to coil transition, or DNA melting, is one of the fundamental processes of life. Nevertheless, it is difficult to achieve an atomistic description of this reaction in solution with current spectroscopic techniques. On the other hand, the computer simulation of DNA melting poses a formidable challenge for theoretical chemists as, even for short sequences, the process occurs in nature on the millisecond or longer time scale. For this reason, this type of simulation has not been attempted yet and the accuracy of force fields in reproducing the free energy of DNA hybridization is not known. Here it is shown how, by combining replica exchange and metadynamics, it is possible to simulate the helix to coil transition of DNA hexamers at room temperature and characterize the reaction intermediates and the relative free energy of hybridization. Three sequences were investigated with both the standard Amber99 force field and a version with modified phosphate torsion parameters [Peres, A.; et al. *Biophys. J.* 2007, 92, 3817–3829]. It is shown that the Amber99 force field overestimates the stability of single stranded and noncanonical DNA that are predicted to be thermodynamically more stable than canonical B-DNA. Therefore this force field is not suitable for DNA studies where large conformational changes occur. The changes introduced by Peres et al. significantly improve the agreement with the experiment. However, the stability of one of the sequences investigated is still underestimated by the modified force field. It is concluded that, although current force fields can provide a reasonable picture of the hybridization reaction, a polarizable force field may be required to obtain a more quantitative agreement with the experimental free energies of hybridization.

## Introduction

DNA melting is the dissociation of a DNA double helix to form two single strands of DNA. This is one of the fundamental processes of life as it provides the key for both the regulation of genome expression and the replication of the genetic information. The melting of short DNA tracts in solution has been characterized with spectroscopic techniques among which NMR has a prominent role.<sup>1,2</sup> Studies on short DNA sequences have shown that the sequence of events that leads to the formation of two single stranded DNA (ssDNA) filaments from a double stranded DNA (dsDNA) helix is sequence dependent<sup>3–5</sup> and can be characterized by the formation of partially melted intermediates.<sup>5,6</sup> In particular, AT-rich tracts appear to constitute “weak points” that are more prone to be opened with respect to the stronger CG-rich tracts of the sequence that act as anchor points.<sup>4</sup>

Computer simulations could in principle be used to provide an atomistic description of this reaction and characterize the sequence of events and rate-limiting steps. However, even for short sequences, DNA melting typically occurs on the millisecond to second time scale; that is, about 4–7 orders of magnitude slower than the events that could be observed in a typical MD simulation. For this reason it is necessary to resort to enhanced sampling techniques to be able to accelerate the reaction and observe it in a shorter time scale. Several enhanced sampling methods are available in the literature. With the notable exception of parallel tempering,<sup>7</sup> most of these methods require

the identification of a suitable reaction coordinate that describes the reaction. Unfortunately, this reaction is so complex that it is difficult to define a priori a small set of reaction coordinates to describe it.

A method to overcome this difficulty, *bias-exchange metadynamics* (BE-META), has been recently introduced by us to study protein folding.<sup>8</sup> BE-META is based on the combination of replica exchange and metadynamics and allows the exploration of free energy surfaces of high dimensionality. Here this method is applied to the simulation of the melting process of three short DNA sequences, namely d(CCATGG)<sub>2</sub>, d(ACATCG)<sub>2</sub>, and d(ATGCAT)<sub>2</sub>. It is found experimentally that the melting transition of these three sequences is characterized by the formation of one intermediate for sequences d(CCATGG)<sub>2</sub> and d(ACATCG)<sub>2</sub> and two for sequence d(ATGCAT)<sub>2</sub>. Simulations were performed with the standard Amber99 force field and with a version with modified phosphate torsion parameters.<sup>9</sup> Anticipating our results, it is found that the standard Amber99 force field severely overestimates the stability of single stranded and non canonical DNA. The modified force field provides results in better qualitative agreement with the available experimental data but still overestimates the stability of ssDNA. It is concluded that, although simulations of DNA hybridization are now indeed possible, great care should be taken in the choice of the force field parameters for these types of studies. It is possible that a polarizable force field may be required to correctly describe the energetic of the hybridization reaction.

## Methodology

**Molecular Dynamics Simulations of DNA.** Molecular dynamics simulations of the three DNA sequences d(C-

<sup>†</sup> Part of the “Giacinto Scoles Festschrift”.

<sup>‡</sup> Present address: DE Shaw Research, 120W 45<sup>th</sup> Street, New York, NY 10036. E-mail: pianaago@deshaw.com.

CATGG)<sub>2</sub>, d(ATGCAT)<sub>2</sub>, and d(ACATCG)<sub>2</sub> were performed with the program GROMACS,<sup>10</sup> the Amber99<sup>11,12</sup> force field, and the Amber99 force field as modified by M. Orozco and co-workers.<sup>9</sup> In the revised force field the torsion parameters of the phosphate backbone were modified on the basis of high-level *ab initio* calculations. Starting structures of canonical B-DNA double helices were generated with the program *nucgen*.<sup>13</sup> The three DNA sequences were solvated by 3130, 3126, and 3125 TIP3P water molecules<sup>14</sup> in a 4.5 × 4.4 × 4.9 nm cell. The corresponding duplex concentration is 16 mM. The negative charge of the phosphate backbone was neutralized with a stoichiometric amount of Na<sup>+</sup> counterions. At difference with the original parametrization of the Amber99 force field,<sup>11,12</sup> all bonds lengths were constrained to their equilibrium values with the LINCS algorithm.<sup>15</sup> This is necessary to reduce instabilities in the simulations driven by the time-dependent potential. Nonbonded interactions were evaluated out to a cutoff of 0.9 nm and the particle mesh Ewald method<sup>16</sup> was used to treat the long-range electrostatic interactions. The time step for the simulations was 2 fs, and the nonbonded pair list was updated every 5 steps. Constant temperature simulations at 293.15 K were performed by coupling the system to a Nose–Hoover thermostat<sup>17</sup> with a relaxation time of 2.0 ps. Constant pressure was obtained by coupling to an anisotropic Berendsen barostat<sup>18</sup> with relaxation time of 4.0 ps. The starting structures were relaxed with 1500 steps of conjugated gradient geometry optimization. Then, all systems were equilibrated by performing 40 ps of NPT simulation at 193.15 K followed by 1.0 ns of NPT simulation at 293.15 K. The final structures obtained from the equilibration were used as starting points for 16.0 ns of bias-exchange metadynamics simulations with the Amber99 force field and 50.0 ns with the modified Amber force field.

**Bias-Exchange Metadynamics Simulations.** The bias-exchange metadynamics (BE-META) approach is described elsewhere<sup>8,19</sup> and briefly outlined here. In BE-META several replica of the same system are simulated in parallel. Each replica is biased by a time-dependent potential<sup>20</sup> generated by a metadynamics<sup>21</sup> as the sum of small Gaussian functions of height  $h$  and width  $\omega$  deposited at regular time intervals  $\delta t$ . At fixed intervals, exchanges of conformations  $x^a$  and  $x^b$  are allowed in a replica exchange scheme.<sup>22</sup> The acceptance probability  $p_{ab}$  is determined by the time-dependent potentials  $V_G(x,t)$  of any two replicas:

$$p_{ab} = \min(1, \exp(\beta(V_G^a(x^a,t) + V_G^b(x^b,t) - V_G^a(x^b,t) - V_G^b(x^a,t))))$$

In this way, each replica explores a multidimensional free energy surface sequentially biased by low-dimensional forces and the diffusion properties of the system in the space of the collective variables are greatly increased, thus allowing a highly efficient exploration of the free energy surface.<sup>23</sup> One replica, the *neutral replica*, is not biased by any time-dependent force, i.e.,  $V_G(x,t) = 0$ . In the limit that the variation of the time-dependent force on the biased replicas is small, the statistic accumulated in the neutral replica approaches the canonical distribution<sup>8</sup> and can be used to calculate unbiased properties. In this work, bias exchange metadynamics simulations were performed using sixteen replicas and six collective variables. Six replicas were evolved in a one-dimensional space biased by one of the collective variable and nine replicas were evolved in a two-dimensional space. A neutral replica was used to accumulate unbiased statistics. The Gaussian deposition time  $\delta t$  was 1.0 ps; exchanges between all the replicas were allowed every 0.4 ps of metadynamics simulation.

**Collective Variables Used in the Bias Exchange Simulation.** To perform an enhanced sampling simulation, a set of variables that are function of the system coordinates must be selected. To be useful for BE-META simulations, the variables must be differentiable and computationally inexpensive. The helix to coil transition is a complex process that involves changes in the base–base pairing and changes in the DNA strand structure. Variables that describe both processes are required. Base pairing can be described by a step function where a pair is formed when the distance between two atoms is less than a given cutoff. This simple step function is not differentiable and in the BE-META simulations it is approximated by a sigmoidal function of the form

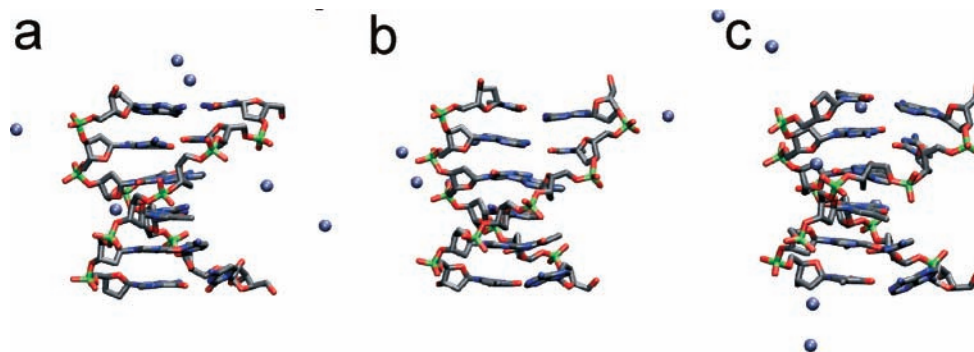
$$N_{AB} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^a}{1 - \left(\frac{r_{ij}}{r_0}\right)^b}$$

where  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ,  $r_0$  is a cutoff distance and  $a < b$ . Note that  $N_{AB}$  is continuous beside at  $r_{ij} = r_0$ , which is a removable singularity.  $N_{AB}$  is computationally inexpensive and its steepness can be tuned by changing the  $a$  and  $b$  exponents at the numerator and denominator. Large exponents with  $a \ll b$  can provide a good approximation of a step function but will also generate very large forces on the atoms in the neighborhood of  $r_0$  that deteriorate the quality of the simulations. In this work  $a = 8$  and  $b = 10$ . These values were chosen as a compromise between fitting the original step function and conserving the energy in the simulations. Two variables of this type were used: one to describe the formation of A-T pairs and one to describe the formation of C-G pairs. The distance used to define the formation of a pair is the distance between the central nitrogens of the A-T and of the C-G base pairs and the cutoff distance was  $r_0 = 0.4$  nm. The width  $\omega$  of the Gaussians for this collective variable was 0.5. The height  $h$  of the Gaussians was 0.1 kJ mol<sup>-1</sup> for all the collective variables.

A collective variable that describes the conformation of the phosphate backbone is also required. To this aim, a variable that detects the formation of regular structures was defined as follows:

$$\Phi = \sum_{i=2}^N \sqrt{1 + \cos^2\left(\frac{\varphi_i - \varphi_{i-1}}{2}\right)}$$

where the sum runs on all the nucleobases of a strand and  $\varphi_i$  is a phosphate backbone dihedral angle of residue  $i$ . Each member of this sum ranges between  $\sqrt{2}$ , when  $\varphi_i$  has the same value of the dihedral angle of the previous nucleobase  $\varphi_{i-1}$  and 1, when  $\varphi_i - \varphi_{i-1} = 180$ . As a consequence,  $\Phi$  is maximum in a regular structure, where all  $\varphi_i$  have the same value. In the Amber99 force field simulations 4 variables of this type were used: two for each DNA strand. In the first variable  $\varphi$  is the  $\angle O3'-P-O5'-C5'$  phosphate backbone dihedral angles (also called  $\alpha$  in the literature) and in the second it is the  $\angle C3'-O3'-P-O5'$  phosphate backbone dihedral angles (also called  $\zeta$ ). The width  $\omega$  of the Gaussians for these collective variables was 0.05. Calculations performed with the Amber99 force field indicated that the collective variables acting on  $\alpha$  and  $\zeta$  dihedrals are highly correlated. For this reason, in the subsequent set of calculations performed with the modified Amber99 force field



**Figure 1.** Structure of (a) d(CCATGG)<sub>2</sub>, (b) d(ATGCAT)<sub>2</sub>, and (c) d(ACATGT)<sub>2</sub> as obtained after 1.0 ns of MD simulation. Hydrogen atoms and water molecules are not shown for clarity.

the  $\zeta$  variable was not used. The reduction in the total number of collective variables improves the convergence properties.

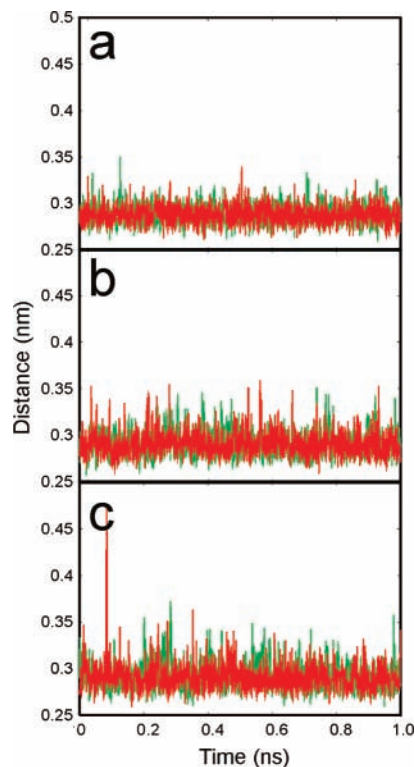
**Parallel Tempering Simulations.** Parallel tempering simulations<sup>24</sup> in the NPT ensemble<sup>25</sup> were carried out on 32 replicas spanning a temperature interval ranging from 298 to 420 K. A parabolic distribution was chosen for the temperatures of the replicas:  $T_i = 298 + i(2 + i/16)$ . The total time of the simulation was 50 ns. Exchanges between the replicas were attempted every 0.2 ps. With this choice the acceptance probability ranges between a maximum of  $\sim 40\%$  for the lower temperatures and a minimum of  $\sim 10\%$  for the higher temperatures.

**Structural Analysis.** The structural analysis of the DNA conformation was carried out following the scheme proposed by Dixit et al.<sup>26</sup> based on the values of the dihedral angles  $\alpha$  ( $\angle O3'-P-O5'-C5'$ ),  $\gamma$  ( $\angle O3'-P-O5'-C5'$ ),  $\epsilon$  ( $\angle O3'-P-O5'-C5'$ ), and  $\zeta$  ( $\angle O3'-P-O5'-C5'$ ).<sup>27</sup> According to this scheme, the DNA backbone can assume 7 different conformations (states 1–7). The canonical B-DNA form is state 1.

## Results and Discussion

**MD Simulation of dsDNA.** One nanosecond of preliminary molecular dynamics simulations was performed for each system using the Amber99 force field. During these simulations all base pairings are maintained, as indicated by the N1–N3 distances that oscillate around 0.29 nm (Figure 1a–c). This result shows that the three sequence are locally stable under the simulation conditions. The most flexible base pairs are the terminal A-T pairs of sequences d(ACATGT)<sub>2</sub> and d(ATGCAT)<sub>2</sub> where N1–N3 distances larger than 0.35 nm are observed during the simulation (Figure 2b,c). This result is consistent with the experimental observation that fraying of the terminal base pairs of these two sequences is observed at relatively low temperatures (288 and 307 K for sequence d(ACATGT)<sub>2</sub><sup>4</sup> and d(ATGCAT)<sub>2</sub><sup>5</sup>, respectively).

**BE-META Simulations with Amber99 Force Field.** The equilibrated structures of the MD simulations were used as starting points for BE-META simulations where the three systems are evolved with a time-dependent potential in a one and two-dimensional space defined by a combination of the following reaction coordinates: number of A-T base pairs ( $N_{AT}$ ), number of C-G base pairs ( $N_{CG}$ ), correlation between the phosphate dihedrals of strand 1 ( $\varphi_{\alpha 1}$  and  $\varphi_{\zeta 1}$ ), and correlation between the phosphate dihedrals of strand 2 ( $\varphi_{\alpha 2}$  and  $\varphi_{\zeta 2}$ ). BE-META simulations were performed on 16 processors. Six processors were allocated to one of the six collective variables each, nine processors were allocated to combinations of the base pair and backbone variables, and one was used to accumulate unbiased statistics. In the beginning of the BE-META simulations the time-dependent potentials are small and the acceptance



**Figure 2.** Distance between the terminal base pairs (nm) plotted as a function of time during 1.0 ns of MD simulation of (a) d(CCATGG)<sub>2</sub>, (b) d(ATGCAT)<sub>2</sub>, and (c) d(ACATGT)<sub>2</sub>. The central N1–N3 distance is taken as a measure of the base-pair distance.

probability for the replica exchange is close to 100%. Afterward the acceptance probability drops, but it remains larger than 10% in all simulations, giving an average exchange time  $< 4$  ps. This high exchange rate ensures that in the several nanoseconds of simulation performed, each trajectory diffuses through the entire replica space and is therefore sequentially biased by all the time-dependent potentials introduced. It is worth noting that the same Gaussian sizes are used to build both the one- and two-dimensional time-dependent potentials. As a consequence, the growth rates and also the forces generated by the one-dimensional potentials are  $\sim 8$  times larger than the two-dimensional ones. Therefore in the BE-META simulations the one-dimensional potentials are often responsible for the larger conformational changes but provide only a rough estimate of the free energy changes. On the other hand, the two-dimensional potentials produce smaller displacements on the system but at convergence allow obtaining more accurate estimates of the system free energy. During the simulation, the error on the relative population of dsDNA and ssDNA was monitored as the difference between the values calculated on the third and



fourth quarter of the trajectory. After  $\sim 12$  ns the error in the ssDNA population in the three simulations was less than 10% and the simulations were considered converged.

The dissociation constant  $K_d$  for dsDNA at 293.15 K can be calculated from the fraction of ssDNA ( $\alpha$ ):

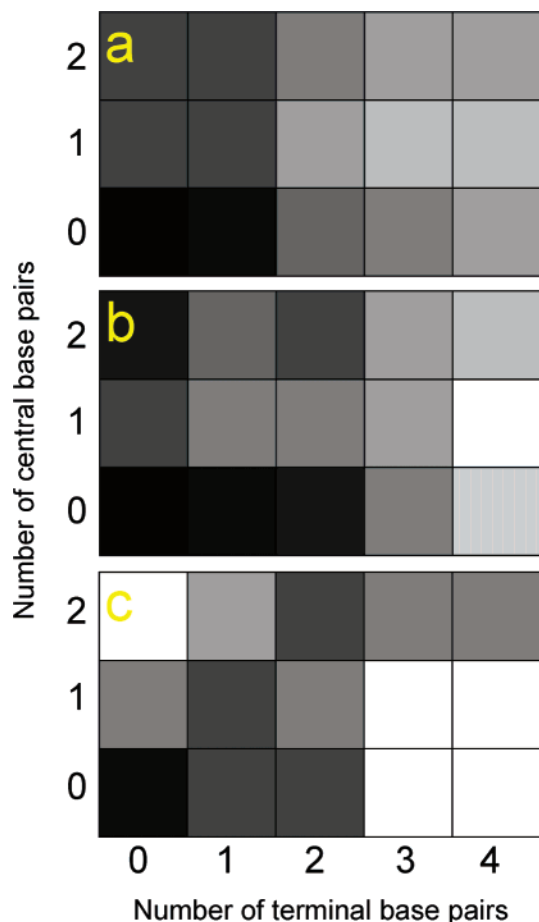
$$K_d = 2c \frac{\alpha^2}{1 - \alpha}$$

where  $c$  is the ssDNA concentration and is equal to 33 mM. The definition of ssDNA and dsDNA is somewhat arbitrary. Here ssDNA is defined as DNA where  $N_{AT} + N_{CG} < 2$ . This definition selects structures that are either unbound or weakly bound and are expected to produce signals typical of ssDNA in an NMR experiment.  $K_d$  values calculated with this definition are therefore expected to be comparable with the  $K_d$  values determined from NMR experiments. However, this definition is quite conservative and the calculated values are expected to be upper bounds to the dissociation constant. The calculated  $K_d$  values are 0.32, 0.25, and 0.18 M, corresponding to a hybridization free energy of 1.0, 0.9, and 0.8 kcal mol $^{-1}$  for d(CCATGG) $_2$ , d(ATGCAT) $_2$ , and d(ACATGT) $_2$ , respectively. These calculated  $K_d$  values are larger than the experimental values derived from NMR data of  $\sim 10^{-4}$  M. $^{3-5}$  A lower bound to the dissociation constant  $K_d$  was also calculated by defining ssDNA as DNA where  $N_{AT} + N_{CG} < 1$ . The calculated lower bounds to  $K_d$  values were  $3.6 \times 10^{-3}$ ,  $6.8 \times 10^{-2}$ , and  $5.6 \times 10^{-2}$  M for d(CCATGG) $_2$ , d(ATGCAT) $_2$ , and d(ACATGT) $_2$ , respectively. These correspond to an hybridization free energy of  $-0.8$ ,  $+0.3$ , and  $+0.2$  kcal mol $^{-1}$  for d(CCATGG) $_2$ , d(ATGCAT) $_2$ , and d(ACATGT) $_2$ , respectively. Even these lower bounds to the dissociation constants are larger than the experimental values, indicating that the force field underestimates the duplex stability. The experimental dissociation constant of  $\sim 10^{-4}$  M corresponds to a hybridization free energy of  $\sim -2$  kcal mol $^{-1}$ . It is concluded that the force field underestimates the duplex stability by 1.2–2.3 kcal mol $^{-1}$ , corresponding to 0.2–0.4 kcal mol $^{-1}$  per base pair.

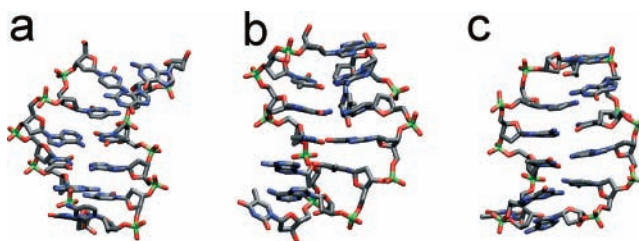
A plot of the free energy surface as a function of the number of central and terminal contacts is reported in (Figure 3a–c). A base pair is considered to be formed when the distance between the central nitrogen of an A-T or C-G base pair is less than 0.4 nm. Besides ds and ssDNA, intermediate states are present. In all the three cases most of the structures display fraying of the terminal base pairs that in d(ATGCAT) $_2$  and d(CCATGG) $_2$  can be further extended to the second terminal base pairs. The calculated stability of the partially melted intermediates in d(ATGCAT) $_2$  ( $-1.5$  and  $-0.9$  kcal mol $^{-1}$ , for melting of the first and second base pair, respectively) is  $\sim 2$  kcal mol $^{-1}$  larger than the experimental value ( $+0.2$  and  $+0.6$  kcal mol $^{-1}$ ). $^4$  This observation is consistent with the previous finding that the force field underestimates the stability of the double helix.

In all three cases a region with two terminal and no central contacts is largely populated (Figure 3a–c). This region correspond to sequences where melting has occurred sequentially starting from one side and has progressed to the core of the helix. Such a structure has not been observed experimentally and its large population is likely to be a force field artifact.

A structural analysis indicates that, despite the starting structure for all the calculation is the canonical B-DNA, a large amount of noncanonical B-DNA is present in all the simulations (Figure 4). Indeed, in the neutral replica structure ensemble 25–40% of the phosphate backbone angles  $\alpha$  ( $\angle O3'-P-O5'-C5'$ ) and  $\zeta$  ( $\angle C3'-O3'-P-O5'$ ) have values typical of Z-DNA

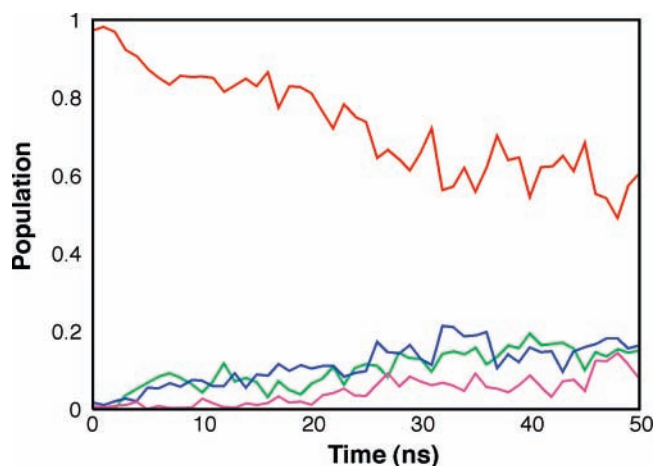


**Figure 3.** BE-META simulations with the Amber99 force field. Plot of the free energy profile as a function of the number of central (\*) and terminal (#) base pairs formed: (a) d(C#C#A\*T#G#G#) $_2$ ; (b) d(A#T#G#C#A#T#) $_2$ ; (c) d(A#C#A\*T#G#T#) $_2$ . Coloring goes from black (more negative energy) to white (more positive energy) in steps of 0.5 kcal mol $^{-1}$ .



**Figure 4.** BE-META simulations of (a) d(CCATGG) $_2$ , (b) d(ATGCAT) $_2$ , and (c) d(ACATGT) $_2$ . dsDNA structure after 16 ns of simulations. Note the irregular conformations of the DNA backbone produced by the rotation of the phosphate groups. Hydrogen atoms, water molecules and Na $^+$  counterions not shown for clarity.

( $+80(3)^\circ$  and  $+79(4)^\circ$ , respectively), instead of the  $-93^\circ$  and  $-68^\circ$  values typical of B-DNA. The backbone conformation was further analyzed with the method of Dixit et al. $^{26}$  This analysis reveals that only  $\sim 50$ – $60\%$  of the backbone is in state 1 (B-DNA) and large fractions of states 2 ( $\sim 8$ – $12\%$ ) and 3 ( $\sim 15$ – $20\%$ ) are present. A control 50 ns parallel tempering MD simulation was performed with 32 replicas spanning a temperature range between 298 and 420 K. The starting structure was the equilibrated structure of d(ACATGT) $_2$ . During this simulation the relative population of the 7 possible states of DNA was monitored. It turns out that the overall fraction of noncanonical states 2–4 is  $>40\%$ , and the fraction of state 1 (B-DNA) is only  $\sim 50\%$  (Figure 5), in good agreement with the BE-META results. These results indicate that the formation of

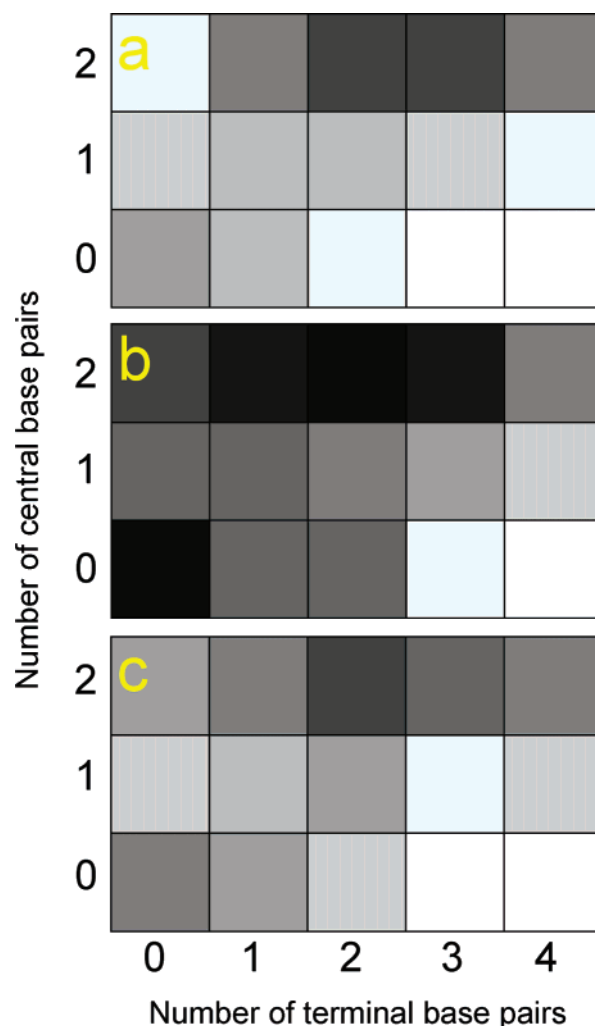


**Figure 5.** Parallel tempering MD simulation of  $d(\text{ACATGT})_2$  with the standard Amber99 force field. The phosphate backbone conformation were analyzed, and the fraction of states 1 (red), 2 (green), 3 (blue), and 4 (purple) in the 298 K replica averaged over 1 ns intervals is reported as a function of simulation time.

noncanonical DNA is not an artifact of the BE-META simulation protocol and that noncanonical DNA is a free energy minimum for this system described by the Amber99 force field. Z-DNA is typically observed in sequences of alternating purine and pyrimidine bases at high salt concentration.<sup>28</sup> However, all simulations were carried out at a relatively low salt concentration ( $\sim 0.16$  M  $\text{Na}^+$ ), and no indication for the presence of Z-DNA in these sequences can be found in the literature. Therefore it is concluded that this result, rather than reflecting a natural tendency of these sequences to assume non canonical DNA conformations, is a force-field artifact arising from an overestimation of the stability of noncanonical DNA states.

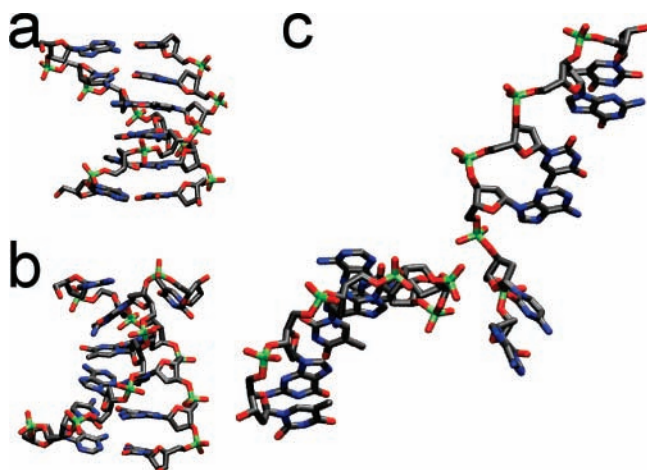
**BE-META Simulations with a Modified Amber99 Force Field.** The occasional population of nonstandard states has already been observed in MD simulations of longer DNA sequences.<sup>26</sup> Previous calculations of the possible  $\alpha/\gamma$  transitions suggest that the canonical B-DNA form is a global free energy minimum for a DNA dodecamer.<sup>27</sup> However, the most stable state of the three dsDNA hexamers investigated here is very different from B-DNA (Figure 4). To solve this deficiency, a version of Amber99 force field with modified backbone torsion parameters has been recently proposed by M. Orozco and co-workers.<sup>9</sup> BE-META simulations were repeated with this modified Amber99 force field. The space of conformations explored in these simulations is reported in Figure 6.

Comparison with the Amber99 simulations (Figure 2) shows that in the modified force field the population of intermediate states is much smaller. In particular, the asymmetric intermediates with a low number of base-pair contacts are now scarcely populated in all simulations. This result suggests that in the modified force field the melting transition is more cooperative. In all cases, fraying of the terminal base pairs is observed. In agreement with the experiment the largest amount of fraying is observed in  $d(\text{ATGCAT})_2$ , where it can be extended to the second base pair without breaking the duplex (Figure 6b). The relative free energies of the partially melted intermediates in  $d(\text{ATGCAT})_2$  were calculated from their population in the simulation. The free energy changes associated with fraying of the first and second base pairs are  $-0.6$  and  $+1.0$  kcal mol<sup>-1</sup>. Comparison with the values of 0.2 and 0.6 kcal mol<sup>-1</sup> determined experimentally<sup>5</sup> indicates that, although in the modified force field the overall agreement with experiment is



**Figure 6.** BE-META simulations with a modified Amber99 force field. Plot of the free energy profile as a function of the number of central (\*) and terminal (#) base pairs formed: (a)  $d(\text{C}^\# \text{C}^\# \text{A}^* \text{T}^* \text{G}^\# \text{G}^\#)_2$ ; (b)  $d(\text{A}^\# \text{T}^\# \text{G}^\# \text{C}^\# \text{A}^\# \text{T}^\#)_2$ ; (c)  $d(\text{A}^\# \text{C}^\# \text{A}^* \text{T}^* \text{G}^\# \text{T}^\#)_2$ . Coloring goes from black (more negative values) to white (more positive values) in steps of 0.5 kcal mol<sup>-1</sup>.

improved, the stability of the terminal base pairs in this sequence is still slightly underestimated. Figure 6 shows that in all cases the dissociation of the internal base pairs has a high-energy cost if the terminal base pairs are still formed. This indicates that even in these short sequences melting of the helix starts from the termini and proceeds toward the core. Dissociation constants for the three duplexes were calculated from the relative fraction of ssDNA present in the simulation. In the simulations performed with the modified Amber99 force field, the population of intermediates with a low number of base pairs is small (Figure 6). Therefore the calculated fraction of ssDNA present in the simulation is not dependent on the definition of ssDNA as in the previous set of simulations. The calculated dissociation constants are  $2.0 \times 10^{-5}$ ,  $3.0 \times 10^{-3}$  and  $1.1 \times 10^{-4}$  M, and the corresponding free energies of hybridization at 293.15 K are  $-1.1$ ,  $-0.5$  and  $-0.8$  kcal mol<sup>-1</sup> for  $d(\text{CCATGG})_2$ ,  $d(\text{ATGCAT})_2$ , and  $d(\text{ACATGT})_2$ , respectively. The improvement with respect to the standard Amber99 force field is considerable, and the calculated dissociation constants for  $d(\text{CCATGG})_2$  and  $d(\text{ACATGT})_2$  are now in good agreement with the experimental data,<sup>3–5</sup> whereas the stability of  $d(\text{ATGCAT})_2$  appears to be still slightly underestimated.



**Figure 7.** dsDNA and ssDNA conformations taken from the simulation of d(ACATGT)<sub>2</sub> with the modified amber99 force field: (a) the most populated DNA structure, corresponding to canonical B-DNA; (b) the second most populated structure, which shows fraying of one terminal base pair; (c) ssDNA.

The dsDNA and ssDNA conformations were analyzed. It turns out that 85–90% of the base pairs in dsDNA are in the canonical B-DNA conformation. The only bases that assume different conformations are the terminal base pairs, which are fairly flexible. Interestingly, the analysis of the phosphate backbone angles performed with the method of Dixit et al.<sup>26</sup> shows that the ssDNA backbone is mostly in state I, which is the conformation found in B-DNA. Indeed, the structure of the isolated ssDNA stretches closely resembles those of dsDNA (Figure 7). This is consistent with the experimental observation that for these small DNA sequences, the conformation of the sugar backbone is not strongly perturbed by hybridization.<sup>4</sup>

## Conclusions

BE-META simulations of melting of the DNA double helix have been carried out for three dsDNA hexamers, d(CCATGG)<sub>2</sub>, d(ATGCAT)<sub>2</sub>, and d(ACATGT)<sub>2</sub>. Simulations performed with the Amber99 force field are not in qualitative or quantitative agreement with the experiment as the stability of single strand and non canonical DNA is severely overestimated. The modifications to this force field proposed by Orozco and co-workers<sup>9</sup> were tested. Simulations performed with this modified force field yield structures and energies in good agreement with the experimental data. Its usage is therefore recommended if large conformational changes in DNA are simulated. However, for one of the sequences investigated here, the stability of ssDNA is still overestimated by a few kcal mol<sup>-1</sup>. A large effort has been recently put in the development of polarizable force fields for nucleic acids with very promising results.<sup>29,30</sup> Such force fields are expected to further improve the description of the backbone electrostatics and therefore may be required for obtaining more quantitative agreement between the calculated and experimentally measured stability of ssDNA and dsDNA.

**Acknowledgment.** The author thanks A. Laio, P. Carloni, and F. Gervasio for several helpful suggestions and M. Orozco for releasing the modified version of the Amber99 force field. The support of the ARC for a Research Fellowship for SP and iVEC and APAC for the provision of computing resources are gratefully acknowledged.

**Supporting Information Available:** Definition of the collective variables used in the bias exchange simulation. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Vandeven, F. J. M.; Hilbers, C. W. *Eur. J. Biochem.* **1988**, *178* (1), 1–38.
- (2) Gueron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *261*, 383–413.
- (3) Tran-Dinh, S.; Neumann, J. M.; Huynh-Dihn, T.; Lallemand, J. Y.; Igolen, J. *Nucleic Acids Res.* **1982**, *10*, 5319–5332.
- (4) Tran-Dinh, S.; Neumann, J. M.; Huynh-Dihn, T.; Geniessel, B.; Igolen, J.; Simonnot, G. *Eur. J. Biochem.* **1982**, *124*, 415–425.
- (5) Patel, D. J.; Hilbers, C. W. *Biochemistry* **1975**, *14* (12), 2651–2656.
- (6) Patel, D. J.; Kozlowski, S. A.; Marky, L. A.; Broka, C.; Rice, J. A.; Itakura, K.; Breslaner, K. J. *Biochemistry* **1982**, *21*, 428–436.
- (7) Hansmann, U. H. E. *Chem. Phys. Lett* **1997**, *281*, 140–150.
- (8) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
- (9) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (10) Lindhal, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306–317.
- (11) Wang, J.; Cieplack, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (12) Cornell, W. D.; Cieplack, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (13) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, T.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER*, 8; University of California: San Francisco, 2004.
- (14) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. P. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (15) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, G. E. M. J. *Comput. Chem.* **1997**, *18*, 1463–1472.
- (16) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T. A.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (17) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (18) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (19) Piana, S.; Laio, A.; Marinelli, F.; Van Troys, M.; Bourry, D.; Ampe, C.; Martins, J. C. *J. Mol. Biol.* **2007**, doi:10.1016/j.jmb.2007.10.020.
- (20) Hüber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput. Aided. Mol. Des.* **1994**, *8*, 695–708.
- (21) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (22) Mitsuoka, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96–123.
- (23) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.
- (24) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
- (25) Seibert, M. M.; Patriksson, A.; Hess, B.; van der Spoel, D. *J. Mol. Biol.* **2005**, *354* (1), 173–183.
- (26) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., III; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. *Biophys. J.* **2005**, *89*, 3721–3740.
- (27) Varnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. *Nucleic Acid Res.* **2002**, *30* (24), 5398–5406.
- (28) Rich, A.; Nordheim, A.; Wang, A. H. *J. Annu. Rev. Biochem.* **1984**, *53*, 791–846.
- (29) Baucom, J.; Transue, T.; Fuentes-Cabrera, M.; Krahn, J. M.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2004**, *121* (14), 6998–7008.
- (30) Babin, V.; Baucom, J.; Darden, T. A.; Sagui, C. *J. Phys. Chem. B. Condens. Matter. Mater. Surf. Interfaces Biophys.* **2006**, *110* (23), 11571–11581.