J|A|C|S
ARTICLES
Published on Web 09/21/2002

Conformational Heterogeneity Observed in Simulations of a Pyrene-Substituted DNA

Guanglei Cui and Carlos Simmerling*

Contribution from the Department of Chemistry and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794-3400

Received May 7, 2002

Abstract: NMR studies previously carried out for a DNA system with a pyrene-substituted base did not observe NOEs involving the adenine located 5′ to the pyrene, and thus the conformation of the adenine was poorly defined in the resulting family of refined structures. However, chemical shift data suggested that an AT base pair may be present. We have carried out fully unrestrained molecular dynamics simulations starting from several members of the family of structures, and these simulations support the existence of an AT base pair for this region. Simulations in both explicit and implicit solvent were carried out, with each converging to either anti or syn conformation for adenine and base pairing in all cases. During these simulations, large and dramatic conformational changes are observed that suggest pathways for complex conformational changes in the highly packed DNA interior. Our analysis reveals little difference in the energies of these syn and anti conformations, in contrast to control calculations carried out for standard DNA (in the absence of a neighboring pyrene). While no interconversion between the conformations was observed in standard simulations, reversible anti/syn exchange was directly simulated using the locally enhanced sampling approach. No exchange was seen in the non-pyrene control sequence. Together, these results suggest that an increased flexibility is introduced as a consequence of the pyrene substitution, offering an explanation that is consistent with the available experimental data. These results increase our optimism that simulations in atomic detail may provide accurate models for experimental observations in complex systems.

Introduction

Pyrene DNA is a nonconventional DNA motif, which contains a bulky "base"— deoxypyrene nucleotide—paired with an abasic site (Figure 1). The pyrene is roughly equivalent in size to a traditional base pair. Observations that it is selectively assembled into a duplex DNA opposite abasic sites demonstrate that the Watson−Crick hydrogen bond pattern may not be required to entail the fidelity of DNA replication[1] and suggest a strategy for characterization of abasic lesions in damaged DNA. The structure of this unconventional DNA duplex has been studied using NMR methods.[2]

This particular system attracted our interest not only due to the importance of the system in contributing to the understanding of nucleic acid structure but also because of a lack of potentially important structural information. The structural ensemble derived from the NMR data failed to reach agreement on the conformation of adenine 8 (ADE8), located at the 5′-end of the pyrene "base", due to a lack of NOE information involving that base (Figure 2). In fact, a wide variety of conformations for ADE8 are present in the family of structures, including both anti and syn ADE8 conformations, with ADE8 often observed in the major or minor groove without hydrogen bonding to the partner THY19. However, observed chemical shift data suggest that the THY19 imino proton is involved in hydrogen-bonding interactions.[2] Thus, while the NMR data provide a detailed view of the conformation of the pyrene, the influence of this motif on the local structure and dynamics of the DNA remains unclear.

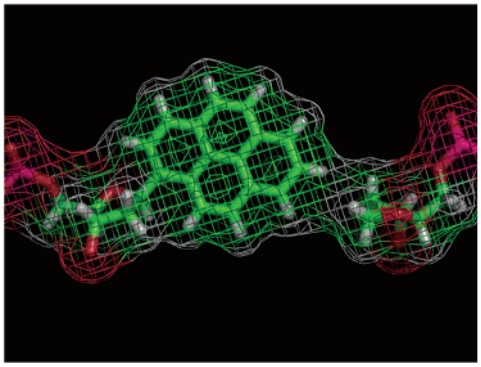

Figure 1. Pyrene−tetrahydrofuran "base pair", shown with a solvent-accessible surface. This pair is nearly as large as a standard base pair.

Lack of specific detail for portions of the structure is not a unique feature of this system. Structure determination for biomolecular systems is usually accomplished through X-ray diffraction or NMR spectroscopy. Both methods regenerate the structure from the measurables of the experiments, typically through the use of computational approaches that generate structures consistent with these data. However, many factors

* To whom correspondence should be addressed. E-mail: carlos.simmerling@stonybrook.edu.

(1) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399*, 704−708.
(2) Smirnov, S.; Matray, T. J.; Kool, E. T.; De Los Santos, C. To be published.

10.1021/ja026825l CCC: $22.00

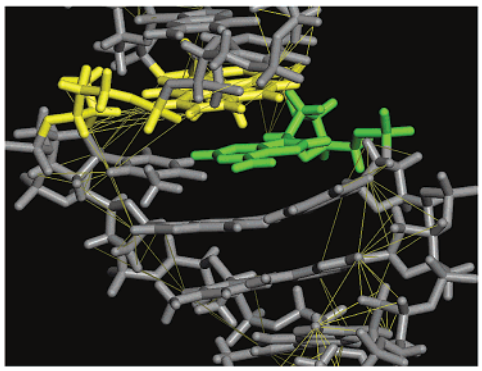

Figure 2. Member of the family of NMR-derived structures, with the NOE restraints shown as yellow lines. Adenine 8 (green), 5′ to the pyrene (paired with abasic furan, both in yellow), has no NOEs involving the base.

can influence the accuracy of the final structure. In particular, conformational heterogeneity in the sample studied can dramatically reduce the quality of the final structure. Such disorder typically results in averaged or incomplete data, which complicates the reconstruction process and can result in partially solved or low-resolution structures, which can be of little value.

More advanced biomolecular simulation approaches can be complementary to experimental data, providing an atomic-detail picture of molecular behavior in a manner that need not be time or ensemble averaged. For flexible molecules in the condensed phase, such as biomolecules in aqueous solution, molecular mechanics potential functions have made calculations tractable.[3−5] Our goal is to reduce the amount of experimentally obtained data needed for successful refinement and to show that state-of-the-art biomolecular simulation methods can give reliable insight into the atomic structures of complex systems.

The major obstacle is conformational sampling; few algorithms exist for navigating through the resulting complex energy landscape. Molecular dynamics simulation (MD) is the most widely accepted technique for efficient simulation of solvated molecules in atomic detail. The main drawback to use of MD simulations is that they are typically too complex to extend beyond nanoseconds in time scale, and traversing large energy barriers and reliably reaching an equilibrium state for systems with many degrees of freedom has not been achieved. In addition, a statistical problem exists when attempting to extend computational "observations" made on a single molecule to the ensemble-averaged data involved in experiments. Computational methods have thus provided very limited success, and new algorithms as well as faster computers are required.

Several important successes have been reported and have been recently summarized.[6,7] Particularly relevant to this study are cases where unrestrained simulations have been able to model structural changes, such as the conversion between A-DNA and B-DNA[8−12] and the description of sequence-dependent DNA structural properties such as bending.[13,14] The influence of abasic sites on the twisting and bending of DNA has also been studied[15−17] Individual base-pair "breathing" events have been reported in simulations.[18,19] Here we present an example to further demonstrate the power of MD simulation, especially when applied to flexible systems.

One particularly attractive approach to improving simulation efficiency is through the use of continuum solvent models such as the generalized Born (GB) approach.[20] GB increases simulation efficiency through several contributions: It removes the computational expense of calculating the motion of explicit solvent molecules, it directly provides a solvation free energy without the need for averaging over solvent configurations, and the lack of solvent friction can accelerate escape from local energy minima during dynamics simulations. This makes it attractive for the study of conformational changes and structural refinement when electrostatic interactions are dominant, such as in the case of nucleic acids. The GB solvation model has been very successful in reproducing the electrostatic contribution to solvation free energy of small organic compounds.[21] Although limited success with proteins has been reported,[22] the utility of GB solvation has been demonstrated on nucleic acid systems.[23−27]

To contribute to the understanding of the influence of pyrene on DNA structure, and also to investigate the possibility of using GB to study nonstandard DNA systems, we carried out calculations for the pyrene−DNA system. Fully unrestrained simulations were performed for several structure models spanning the range of structures satisfying the NMR data, with both explicit and implicit solvent molecular dynamics. The behavior with the two solvent models is comparable, providing additional evidence that the GB model can be a useful approach to the study of DNA dynamics. Our results decrease the possible conformations of ADE8 and strongly support the existence of base pairing regardless of the solvent model employed. However, both anti and syn ADE8 conformations are observed, in contrast to canonical DNA motifs.

To gain further insight into the presence of the syn conformation, local interactions of this particular A:T pair with the surroundings were calculated for each case and no significant difference was found. Analysis of local interactions for A:T pairs

(3) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.
(4) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(5) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487−509.
(6) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182−196.
(7) Cheatham, T. E.; Young, M. A. *Biopolymers* **2000**, *56*, 232−256.
(8) Cheatham, T. E.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434−444.
(9) Cheatham, T. E.; Kollman, P. A. *Structure* **1997**, *5*, 1297−1311.
(10) Cheatham, T. E.; Crowley, M. F.; Fox, T.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9626−9630.
(11) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Am. Chem. Soc.* **1998**, *120*, 10629−10633.
(12) Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 4658−4667.
(13) Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1998**, *281*, 675−687.
(14) Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1999**, *285*, 1623−1632.
(15) Ayadi, L.; Jourdan, M.; Coulombeau, C.; Garcia, J.; Lavery, R. *J. Biomol. Struct. Dyn.* **1999**, *17*, 245−257.
(16) Ayadi, L.; Coulombeau, C.; Lavery, R. *J. Biomol. Struct. Dyn.* **2000**, *17*, 645−653.
(17) Barsky, D.; Foloppe, N.; Ahmadia, S.; Wilson, D. M.; MacKerell, A. D. *Nucleic Acids Res.* **2000**, *28*, 2613−2626.
(18) Cieplak, P.; Cheatham, T. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6722−6730.
(19) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653−8654.
(20) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(21) Jeancharles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. *J. Am. Chem. Soc.* **1991**, *113*, 1454−1455.
(22) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243−252.
(23) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571−9576.
(24) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 251−265.
(25) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426−434.
(26) Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275−291.
(27) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.

in both standard and pyrene modified DNA suggests that van der Waals interactions with the bulky pyrene may contribute to the increased stability of syn ADE8 in this environment. The distribution of the two patterns was further studied with the locally enhanced sampling (LES) technique,[28] and direct interconversion of syn/anti ADE is observed during these simulations. This provides additional support to the possibility of syn ADE8. In contrast to the reported syn/anti equilibrium in simulations of enzo[*a*]pyrene diol epoxide (BPDE) DNA adducts,[29] syn and anti in this case are clearly separate minima and do not represent a simple shift in the position of the anti minimum. This is reasonable since in this case the change is seen for a standard base that can maintain canonical hydrogen bond patterns in normal syn/anti glycosidic torsion values. Interconversion between these conformers may explain the lack of NMR data for this region and should be investigated in more detail.

Methods

The system studied (PDB code 1FZL) consists of 13 base pairs (5′-D(Cp-Ap-Cp-Ap-Ap-Ap-Cp-Ap-(PYP)p-Gp-Cp-Ap-C)-3′ and 5′-D(Gp-Tp-Gp-Cp-(FUR)p-Tp-Gp-Tp-Tp-Tp-Gp-Tp-G)-3′), with 4 and 8 base pairs on each side of the pyrene. In the remainder of this article we refer to these as the short and long ends, respectively, and refer to the system as "pyrene DNA". Root-mean-square deviations (rmsd) were calculated on the basis of all heavy atoms of the 9 central base pairs, since the terminal 2 base pairs on each end showed significant fluctuation among the family of NMR-based structures and during simulations. Groove widths were calculated as the distance between the closest phosphorus pairs across the groove. Another sequence was simulated as a non-pyrene control with 10 base pairs: (5′-D(Cp-Cp-Ap-Ap-Cp-Gp-Tp-Tp-Gp-G)-3′ and 5′-D(Cp-Cp-Ap-Ap-Cp-Gp-Tp-Tp-Gp-G)-3′). This will be referred to as "standard DNA", since it is composed entirely of the 4 standard bases (A, T, G, and C). Simulations for this control sequence have been reported using the same force field as employed in this study.[10,27] All molecular dynamics simulations presented in this study were performed using the *sander* module in AMBER version 6.[30]

Force Field Parameters for the Nonstandard Residues. Partial charges for the pyrene and tetrahydrofuran analogue nucleotides were obtained using the restrained electrostatic potential fitting method (RESP[31,32]), in a manner similar to that employed for the standard DNA residues in the ff94 force field.[4,32] Each nucleotide was excised from the NMR duplex structure and optimized in Gaussian98[33] (HF/6-31G*). The electrostatic potential was then calculated and used in a two-stage RESP fit. The partial charges of all atoms, except the bases, C1′ and H1′ on the sugar ring, were held fixed at their values in the ff94 force field. Multiple conformations were not used in the charge derivation for both nucleotides, considering the limited flexibility and highly conjugated structure of pyrene and the simplicity of tetrahydrofuran. Assignment of atom types and missing bond angle and torsion angle parameters were made by analogy to existing atom types in the ff94 parameter set. The resulting partial charges, atom types, and additional parameters are provided as Supporting Information.

Generalized Born Solvent Simulations. An implementation of the generalized Born (GB) solvent model[21,34,35] has been incorporated into AMBER version 6, and testing of this model has been carried out for DNA systems.[27] All of the GB simulations described in this article use a similar approach as that study, employing a modified set of Bondi radii[36] (only hydrogen atoms are modified) for all bases, screening parameters taken from the Tinker program,[37] and 0.13 Å offset for the effective Born radii. Explicit counterions were not used; an ionic strength of 0.2 M was employed in the continuum solvent calculations. SHAKE was used in all dynamics calculations to fix the length of covalent bonds in which hydrogen atoms were involved. Nonbonded interactions were fully evaluated every time step with no cutoff distance used.

Each GB simulation consisted of two stages, equilibration and production, neither of which used the experimentally determined NOE distance restraints. In the equilibration stage, the starting structures were first minimized for 1000 cycles, with atoms restrained to the starting positions with a harmonic force constant 5.0 kcal/(mol·Å^2). Incrementally reduced force constants were used in four subsequent 1000-cycle minimizations, which gradually brought the system to the closest energy minimum. The temperature of the resulting system was raised to 300 K over 60 ps while the minimized structure was positionally restrained. The restraints were released incrementally in a subsequent 750-ps molecular dynamics (MD) with a 1 fs time step. This is longer than the equilibration procedure we typically use with explicit solvent, but simulations with GB appeared to be more sensitive since large oscillating changes in the structure were observed if less careful equilibration was carried out (data not shown). In the production stage, temperature coupling was not employed and the simulations were carried out in the microcanonical ensemble. Structure snapshots were saved every 10 ps.

Explicit Solvent Simulations. The particle mesh Ewald (PME) method[38] was used in all explicit solvent simulations to evaluate electrostatic interactions. The default parameter values in AMBER were used in PME calculations (8 Å cutoff for the real-space nonbonded interactions and a reciprocal space grid spacing of approximately 1 Å). The NMR structures were solvated in a roughly $55 \times 70 \times 50$ Å^3 box of TIP3P[39] water molecules with a clearance of at least 9 Å between the DNA atoms and each side of the box. The number of water molecules required to solvate each of the systems varied, with total system sizes of 12 000−14 000 atoms. Sodium ions were added to neutralize the system. All bonds with hydrogen atoms involved were constrained with SHAKE. Rigid body motion of the system as a whole (not just the solute) was removed.[40] First, 60 ps of dynamics was carried out at 300 K and 1 atm pressure, in which only water molecules and sodium ions were allowed to move. The whole system was then minimized for 5000 cycles with incrementally reduced positional restraints, similar to the GB simulations described above. Four 10 ps restrained MD simulations allowed both solvent and solute to reach a

(28) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(29) Yan, S. X.; Shapiro, R.; Geacintov, N. E.; Broyde, S. *J. Am. Chem. Soc.* **2001**, *123*, 7054−7066.
(30) Case, D. A.; D. A. P., J. W.; Caldwell, T. E.; Cheatham, W. S., III; Ross, C. L. Simmerling, T. A.; Darden, K. M.; Merz, R. V.; Stanton, A. L.; Cheng, J. J.; Vincent, M.; Crowley, V.; Tsui, R. J.; Radmer, Y.; Duan, J.; Pitera, I.; Massova, G. L.; Seibel, U. C.; Singh, P. K.; Weiner; Kollman, P. A. *AMBER 6*; University of California: San Francisco, CA, 1999.
(31) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(32) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision 7.x; Gaussian, Inc.: Pittsburgh, PA, 1998.
(34) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1−11.
(35) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(36) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(37) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016−1024.
(38) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(40) Harvey, S. C.; Tan, R. K. Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726−740.

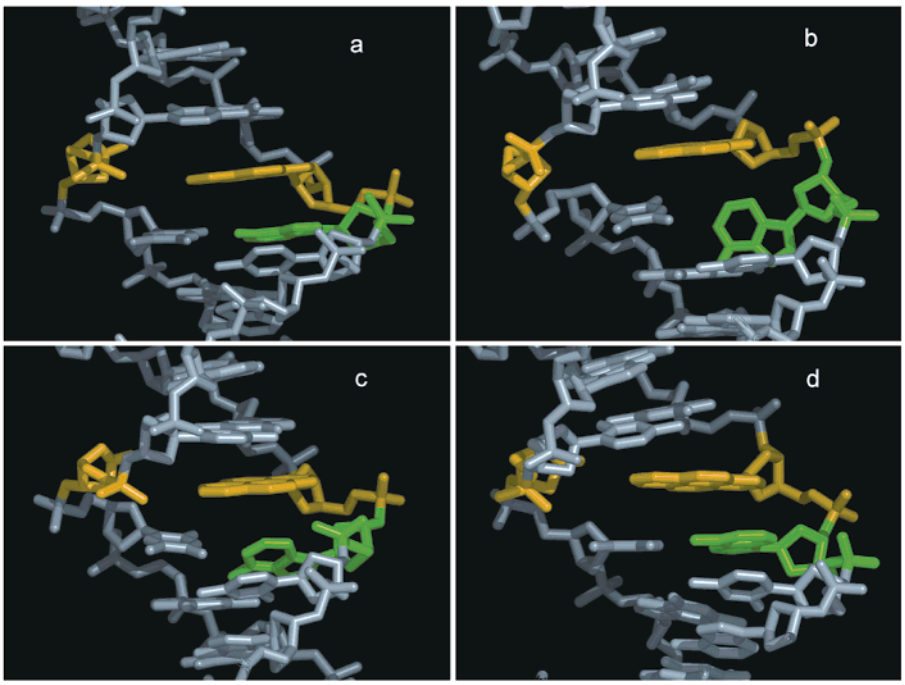

Figure 3. Closeup of the region near the pyrene in several structures from the refined NMR family that were used in the simulations: (a) antiHB; (b) antiminor; (c) syn1; (d) syn2.

local equilibrium. In the subsequent production dynamics, a 2 fs time step was used.

Explicit Solvent Simulations with Locally Enhanced Sampling. We replaced the ADE8−THY19 base pair (excluding P, O1P, O2P, O3′, O5′, C5′, H5′1, and H5′2) with 10 copies by using the *addles* module in AMBER6. The template structure was taken from equilibration dynamics. All copies belonged to the same region and were given the identical initial conformation as the template structure but altered initial velocity information to permit divergence. The AMBER code was modified to permit coupling of LES and non-LES regions to separate thermal baths so that their temperatures could be controlled independently.

MM−PB Calculation. MM−PB/SA (molecular mechanics−Poisson−Boltzmann with surface area)[41,42] is a relatively new method to estimate free energy differences by adding the contribution from intrasolute interactions, solvation energy, and hydrophobic effect. It was used in this study to compute the free energy difference of two alternate conformations of the system. For each conformation, 100 equally spaced snapshots were collected from a 1 ns explicit solvent simulation. The MM energy was calculated as the average of the sum of all bonded and nonbonded interactions using ff94 force field parameters. The electrostatic contribution to the solvation free energy of each conformation was calculated by solving the Poisson−Boltzmann equation using the Delphi program.[43] The cubic grid was constructed so that each side was twice as long as the longest dimension of the structure. A grid spacing of 0.25 Å was used. The cavitation, van der Waals, and hydrophobic contributions to the solvation free energy can be estimated using solvent-accessible surface area, but this contribution is typically small compared to polarization and screening effects for highly charged systems such as nucleic acids. We found that the surface areas for average anti and syn conformations differed by less than 0.1 Å^2 (out of 5200 Å^2); therefore this contribution was not included in subsequent MM−PB/SA calculations.

(41) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(42) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(43) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.

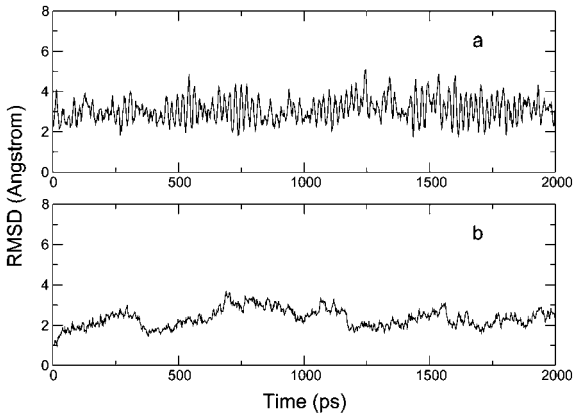

Figure 4. Graphs showing results of simulations starting from the Watson−Crick structure: (a) antiHB with GB; (b) antiHB with explicit solvent. Average rmsd values are ∼3 Å (compared to the initial structure) in both cases, and the average structures from the 2 solvent models differ by 1.7 Å.

Results and Discussion

GB and Explicit Solvent Simulations. The family of structures resulting from refinement with NMR-derived restraints[2] shows significant diversity in the conformation of ADE8 (Figure 3). The structure with conventional Watson−Crick base pairing for ADE8 (Figure 3a, hereafter denoted "antiHB") was selected and examined with unrestrained GB and explicit solvent simulations. Despite the expected slightly larger positional fluctuations and rmsd, the 2 ns GB simulation generally reproduced the results of the explicit solvent simulation of the same length. The rmsd values in each case average ∼3 Å from the initial structure (Figure 4). Helicoidal parameters in the two solvent models are similar (Figure 5), with increased fluctuation in pucker of furan17. The average structures differ by 1.7 Å with a slight widening of the major groove in GB (Figure 6). As one might expect for a continuum solvent model

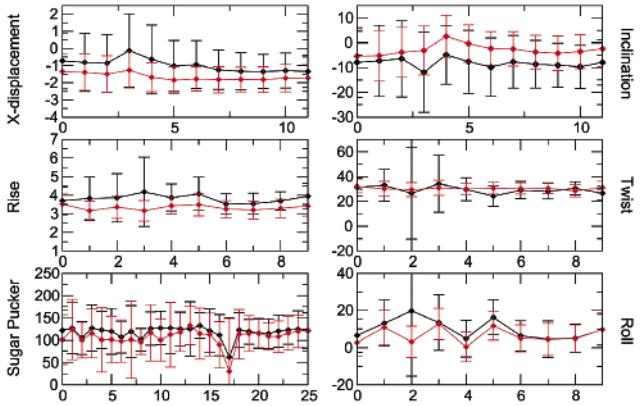

Figure 5. Helicoidal parameter comparison of antiHB GB (black) and explicit solvent (red) calculations. Similar results are obtained for the two solvent models.

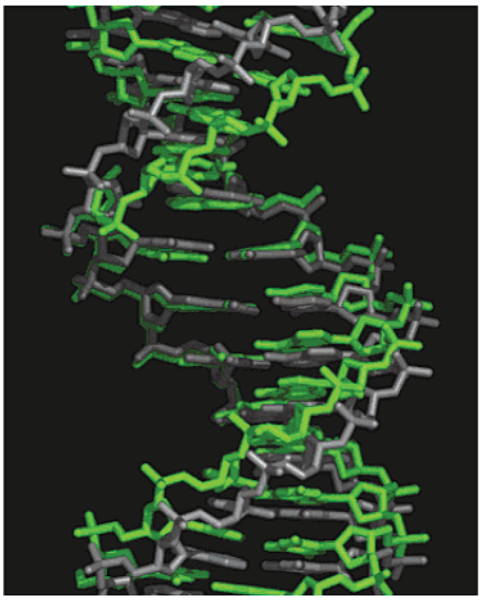

Figure 6. Average structure comparison of antiHB GB (gray) and explicit solvent simulation (green), with an rmsd of 1.7 Å. A slight widening of the major groove is observed with GB.

without frictional terms, the fluctuations in rmsd and helicoidal parameters are significantly larger with GB. These observations are consistent with the results reported previously for GB simulation of a standard DNA sequence.[27] The modified intrinsic radii for hydrogen atoms of different types appeared to work well with this modified DNA even though they were previously tested using only standard nucleic acid sequences. With both solvent models, the A:T pair was stable in the initial Watson−Crick conformation during the entire simulation. Even if alternate conformations should be sampled, the time scale for breaking this interaction was inaccessible in these simulations.

We next repeated this process for a member of the NMR family in which no A:T pair was present. We selected a structure with ADE8 in the minor groove of the DNA (Figure 1b, hereafter denoted "antiminor"). This structure was chosen because it has the lowest rmsd (1.5 Å) among all structures in the family as compared to the Watson−Crick (antiHB) structure described above. The small deviation suggested that the transition to a base-paired structure might be confined to a local region of conformational space and therefore may be accessible during molecular dynamics.

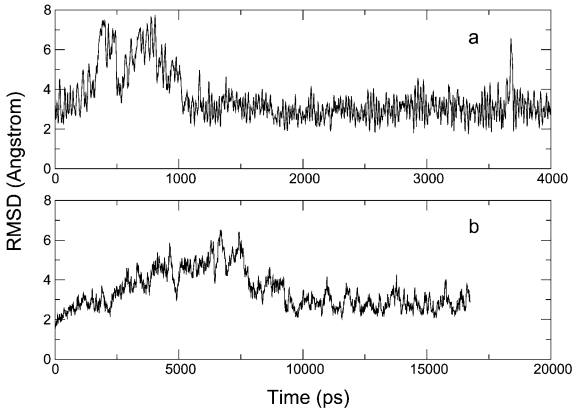

Figure 7. Graphs showing results of simulations starting with anti ADE8 in the minor groove (antiminor) with (a) GB solvation and (b) explicit solvent. Similar results are obtained: a large increase in rmsd values, followed by convergence to lower values. The time scale of the GB transition is much shorter.

Similar to the antiHB simulations, the general behavior was comparable during 4 ns GB and 16.8 ns explicit solvent simulations. While the antiHB model was stable regardless of solvent model, the antiminor model is likewise *unstable* in both solvent models. The deviations from the NMR structure during both simulations (Figure 7) clearly show that the unpaired minor-groove ADE8 in the NMR structure is nonoptimal. The DNA underwent significant conformational changes in both cases, with a maximum rmsd compared with the initial structure of 7.8 Å in GB and 6.4 Å in explicit solvent, with both simulations converging to rmsd values near 3 Å. This implies that our initial assumption in choosing antiminor (structures with very localized differences in structure may have a simple transition pathway) was not necessarily valid. The rmsd of average structures (after the transition) to average structures starting from antiHB are only 0.4 Å (GB) and 1.4 Å (explicit solvent).

Several important differences were noted when the results from GB and explicit solvent simulation were compared in detail. The most obvious was that when the durations of the transitions were compared, changes with GB occurred nearly 1 order of magnitude more rapidly than in explicit solvent (1 ns vs 10 ns, Figure 7). This is not surprising due to the lack of solvent-based friction in the GB model and has been reported in the past for nucleic acid simulations.[27,44] In this case such comparison would not be very meaningful if the stable structures after the transitions were dissimilar. In this case, the resulting average structures over the last 1 ns were very similar (Figure 8) and differed by only 1.2 Å.

The more rapid convergence to a base-paired structure in GB simulations is certainly impressive. However, a closer look at the final structures revealed that ADE8 was in the anti conformation forming a Watson−Crick-type base pair with GB, but in explicit solvent the syn conformation was located, forming a Hoogsteen-type base pair. This difference is subtle but important, because either hydrogen bond pattern could result in the imino proton chemical shift observed in the NMR experiment (but neither alone explains the lack of NOE data for ADE8).

Distinctive structural features from the transition process in GB are displayed via 8 snapshots (Figure 9) and the time

(44) Williams, D. J.; Hall, K. B. *Biophys. J.* **1999**, *76*, 3192−3205.

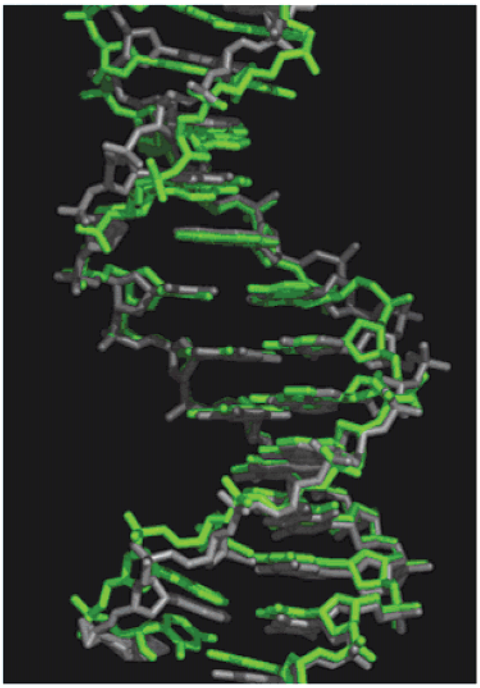

Figure 8. Average structure comparison of antiminor GB (gray) and explicit solvent simulations (green), after the formation of the A:T base pair. The rmsd between the two structures is 1.2 Å.

Table 1. Rmsd (in Å) of All Pairs of Average Structures from the AntiHB and Antiminor Simulations in GB and Explicit Solvent[a]

	antiHB (GB)	antiHB (explicit)	antiminor (GB)	antiminor (explicit)
antiHB (GB)		1.7	0.4	1.3
antiHB (explicit)			1.6	1.4
antiminor (GB)				1.2
antiminor (explicit)				

[a] The average structures are similar in all cases.

dependence of base pair hydrogen bonds (Figure 10). The two initial hydrogen bonds present when ADE8 occupied the minor groove, ADE8:N7−GUA20:N2 and ADE8:N6−THY21:O4′, were broken quickly during equilibration. ADE8 then moved to be coplanar with THY19 and formed reverse-Hoogsteen type hydrogen bonds to THY19 (ADE8:N6−THY19:O2 and ADE8:N7−THY19:N3).

As the simulation continued, the widths of the minor and major grooves changed continuously and simultaneously in multiple places, with the major groove widened by ∼8 Å and the minor groove narrowed by ∼4 Å, except at C11−T19 and A8−T22, where ADE8 entered the double helix (Figure 11). Severe stretching and unwinding was accompanied by the complete loss of 13 out of 17 Watson−Crick hydrogen bonds, involving 6 of the 7 base pairs in the longer end of the DNA (Figure 10). Perhaps surprisingly, the short end remained intact, even though the whole structure deviated from the NMR model by as much as 8 Å.

Next, the reverse-Hoogsteen hydrogen bonds between ADE8 and THY19 that had formed earlier were broken, and the bases shifted in-plane to a Watson−Crick configuration. Shortly after this local structure was formed, all of the base pairs previously lost were dramatically reestablished, and the rmsd value compared to the average structure obtained from the antiHB control simulation was reduced to only 0.4 Å. During the final 3 ns of this simulation, the overall structure remained stable.

It is remarkable that such a striking series of conformational changes could be observed in unrestrained molecular dynamics simulation of DNA with full atomic detail. As far as we know, it is the first time that such a transition has been reported. What is unusual in this case is not the relatively minor difference between initial and final states but rather the very large changes seen during the transition pathway. The system moved from the initial NMR model to a region quite distant in phase space and then returned to the neighborhood of the initial structure with a change in base pair geometry. In this sense the process is more comparable to an unfolding/refolding event than simple base pair formation. It is very exciting to see that simulations are starting to be able to model, in atomic detail, transient but key events that involve significant structural changes.

The transition to base-paired conformation was also observed in the explicit solvent simulation, but in this case it occurred over an almost 10 ns period. Snapshots from the simulated transition are shown in Figures 12 and 13. The rearrangement started with syn ADE8 forming reverse Watson−Crick hydrogen bonds using ADE8:N6−THY19:O2 and ADE8:N1−T19:N3. The major groove was again widened (by ∼10 Å, a greater extent than the GB simulations), but in contrast to the GB results the minor groove was also widened (Figure 14). At ∼3 ns, THY19 was flipped out into the solvent in the major groove. This motion opened sufficient space to permit a shift in the position of ADE8, allowing formation of Hoogsteen hydrogen bonds with the returning THY19. In the GB simulations, the space for a similar ADE8 shift had been provided by loss of the proximal C7:G20 base pair rather than motion in THY19. Here the base pair also re-formed at ∼10 ns and remained stable for the remaining 6.8 ns of the simulation. The other portions of the DNA did not show the instability observed during the GB transition; this could be a result of the spatial restrictions imposed by the explicit solvent cavity, deficiencies in the GB model or associated parameters, or possibly due to the different final structures. The rmsd of all pairs of average structures from the 4 simulations are given in Table 1; all are similar, with deviations of only ∼0.5−1.5 Å.

Our attention was immediately drawn to the inconsistent occurrence of the anti and syn ADE8 in the GB and explicit solvent simulations, respectively, because bases in syn conformation are not usually observed in standard base pairs and anti is generally considered to be thermodynamically more stable than syn. In this case it is possible that the observed syn conformation was simply kinetically trapped and does not reflect a conformation that is significantly populated in the equilibrium ensemble. To investigate this possibility, 100 structures for each conformation were collected from explicit solvent simulations, stripped of solvent and counterions, and used in MM−PB free energy evaluation. MM−GB energies were also calculated for comparison, and even though the absolute energy values differ from those of MM−PB by almost 300 kcal/mol, the correlation is good with a correlation coefficient of 0.956 and slope of 0.980. This provides additional reassurance that the results of the GB model are satisfactory in this case.

The energy distributions, shown in Figure 15, indicate no significant preference for either anti or syn, regardless of whether GB or PB was employed. This suggests that both anti and syn

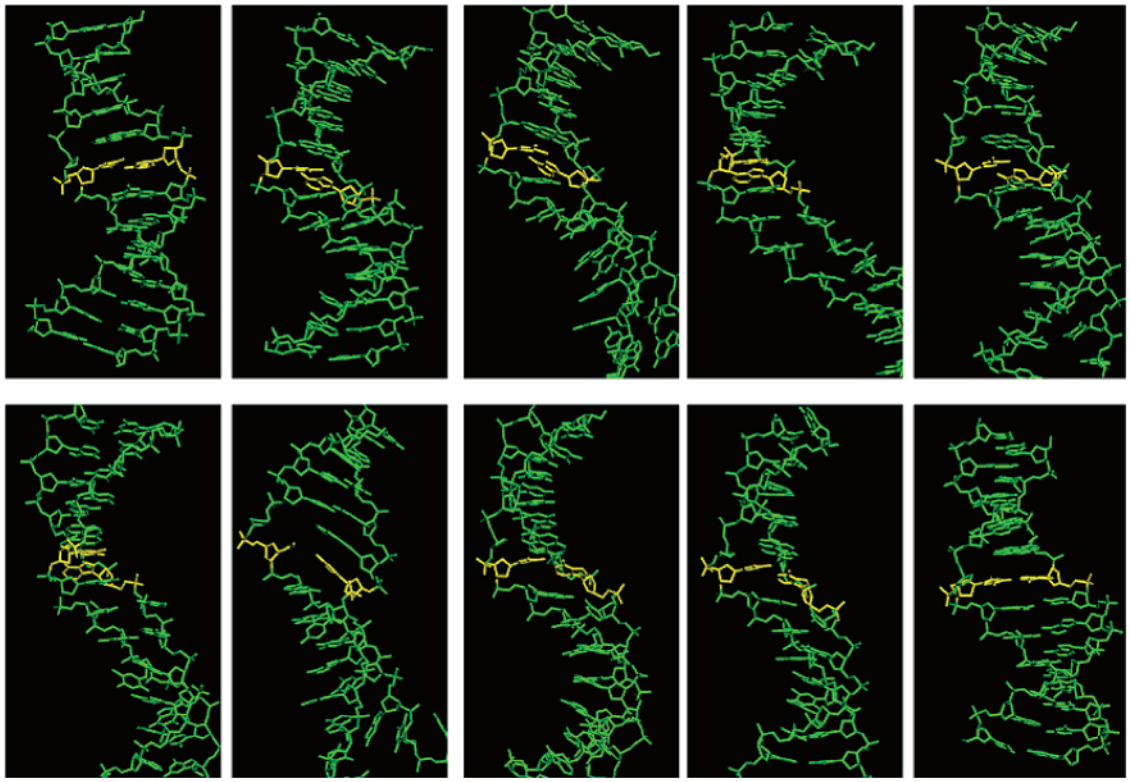

Figure 9. Transition snapshots from antiMinor GB simulations, spaced equally 80 ps apart. Large distortions extending several base pairs beyond the ADE8 region are evident during formation of the A:T base pair, but the final structure is similar to the initial structure with the exception of the A:T base pair geometry.

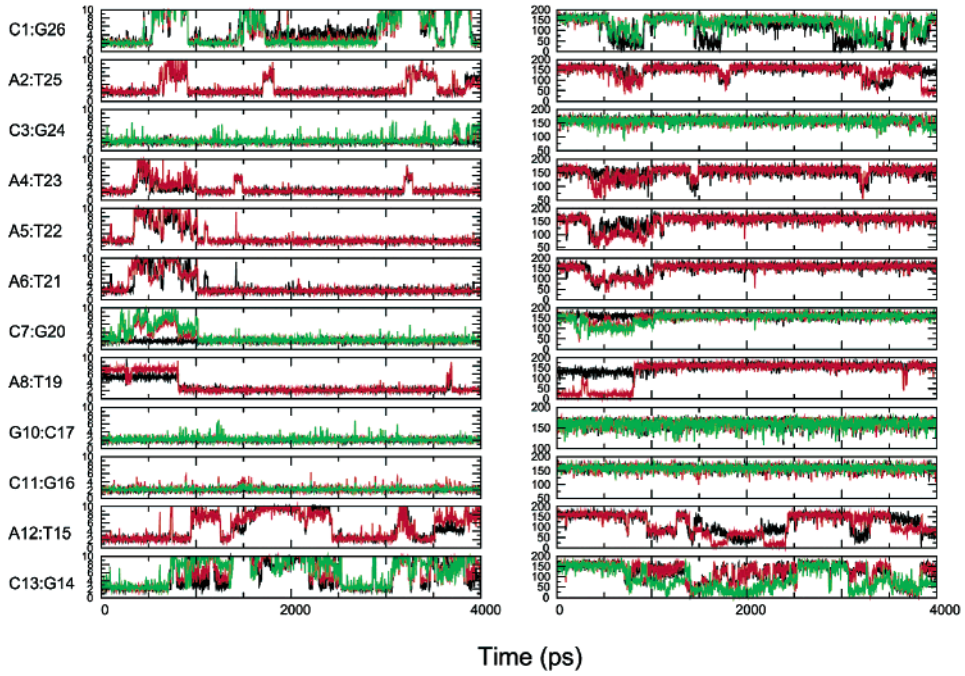

Figure 10. Hydrogen bond breaking and re-forming events illustrated by the acceptor−hydrogen distances and acceptor−hydrogen−donor angles in the antiMinor GB simulation. Six of seven base pairs in the long end are lost and then reestablished after formation of the A:T base pair.

conformations should be similarly populated at equilibrium in this specific environment. Among the individual components of MM−PB energies for this particular system, electrostatic and solvation terms are the largest in magnitude but favored different conformations. The effects nearly cancel; this is a common trend in such calculations and likely reflects the enhanced solvation of structures with unfavorable intramolecular electrostatics (such as parallel dipole alignment).

We speculated that the presence of the neighboring pyrene contributes to the unusually high stability of syn ADE8 and investigated this interaction in greater detail. The interactions of ADE8 with the rest of the system (Table 2) were calculated

Table 2. Group Contributions to the Local Interaction of the AT Base Pair[a]

struct	VDW (kcal/mol)	bonding (kcal/mol)	electrostatic (kcal/mol)	tot. (kcal/mol)
pyrene anti ADE8	−51.4	43.1	−17.5	−25.9
pyrene syn ADE8	−52.2	43.1	−15.9	−25.1
std anti ADE	−54.3	42.7	−16.6	−28.2
std syn ADE	−51.6	41.0	−15.3	−25.9

[a] Electrostatics favor anti in both pyrene and standard DNA, but the van der Waals interactions with pyrene preferentially stabilize the syn conformation, nearly canceling the electrostatic effects in the pyrene System.

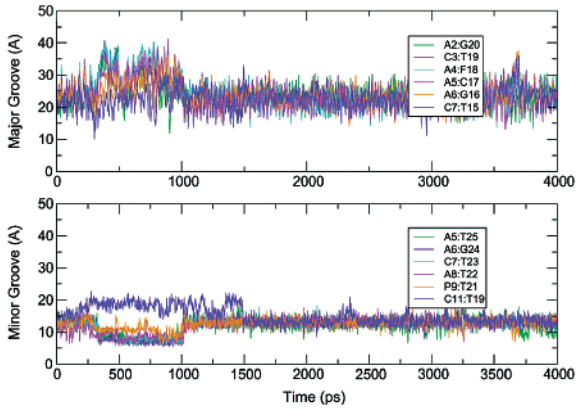

Figure 11. Groove width changes in antiMinor GB simulation.

for 4 average structures: anti and syn adenine in an A:T base pair, neighboring either the pyrene−abasic pair or surrounded by standard bases. The anti conformation is seen to be significantly more stable than syn in non-pyrene DNA (2.3 kcal/mol), consistent with typical experimental observations. However, the anti/syn energy difference is much smaller in the pyrene DNA (0.8 kcal/mol). The syn conformation in the pyrene DNA is stabilized by local interactions, predominantly from van der Waals contacts with the bulky pyrene residue.

To further explore this issue, GB simulations were initiated using 2 NMR structures with syn ADE8 (Figure 1c,d), with the expectation that a more stable anti conformation in GB would lead to a syn/anti transition. However, the syn conformation was preserved in both simulations despite the relatively long lengths of 75.8 and 37.4 ns long, respectively. GB simulations at 325 K were also performed for both anti and syn conformations to determine if the higher temperature would facilitate syn/anti interconversion during the time scale accessible by these simulations. However, the process was not observed until after the DNA structure became unstable; the anti conformation unfolded after ∼11 ns, and the syn, after ∼7 ns. This suggests that the anti/syn transition may have barriers comparable to the stability of the double helix, and thus increased temperature was abandoned as an approach to simulate this interconversion.

LES Simulations. The calculations described above lead us to a possible interpretation that both anti and syn conformers may exist in the solution and the flexibility in this portion of the structure was responsible for the difficulty in obtaining NMR-based restraints for ADE8. However, this interpretation may be invalid if our statistics are unconverged. Although both structures could be formed in simulations, and energy analysis suggests that they may have similar population, direct simulation of an anti/syn transition was not observed. The energy barrier to such transitions in a tightly packed duplex is likely to result in a time scale beyond that currently accessible by standard simulation techniques.

LES (locally enhanced sampling)[45] has been demonstrated in several applications to improve sampling when multiple energy minima are present.[46−48] Particularly relevant to this study was the observation by Simmerling and Kollman that LES simulations of an RNA tetraloop spontaneously converted from incorrect to correct structure[49] despite the inability of otherwise identical non-LES simulations to do so.[50] We employed LES here not to aid in locating a single optimal structure but rather to explore the possibility of directly simulating interconversion between anti and syn ADE8. A total of 10 copies of ADE8 and THY19 were placed in the pyrene DNA duplex in explicit solvent using one snapshot from the equilibration where the base pair had not yet formed.

Differences in the behavior of the LES system could be the result of the pyrene or an artifact of LES. Two control simulations were therefore used: In the first, an A:T base pair, ADE4:THY17, was duplicated in the non-pyrene DNA 10 base pair system in an identical manner as was done for the pyrene-containing 13 base pair sequence. This provides a test for the use of LES by comparing to our non-LES results for these same sequences. It is also possible that the change in anti/syn equilibrium is due to sequence-specific effects rather than the pyrene; an additional control was therefore carried out for the 13 base pair system in which ADE8 and THY19 were again replaced with 10 LES copies, but additionally the pyrene:furan pair was replaced with a non-LES A:T pair. Differences in these two simulations should be due to the pyrene substitution.

In the 4.8 ns control for the 10 base pair system, all copies remained in the anti conformation during the entire simulation and syn was never sampled. Very similar results were obtained during 4.5 ns simulation of the 13 base pair system without pyrene: syn conformations represented <1% of the total. The glycosidic torsion angle (χ) of each copy sampled during dynamics of the non-pyrene 13 base pair system and distribution histograms are shown in Figure 16.

LES results for the pyrene-containing system differed dramatically from the control simulations. All 10 LES copies in the pyrene DNA spontaneously formed A:T base pairs, consistent with the non-LES simulations described above. However, as we anticipated, more than one LES copy was able to sample the syn conformation in the 14 ns pyrene DNA simulation (Figure 17). A snapshot showing simultaneous LES population of anti and syn is shown in Figure 18. The base pair hydrogen bond patterns for each χ rotamer are consistent with those located at the end of the transition period in anti/syn non-LES simulations (in contrast to the less stable patterns observed shortly after base pair formation in the single copy MD). A single THY19 conformation is compatible with both ADE8 possibilities, consistent with the observation of NOE data for THY19. Each conformation was often stable for multiple nanoseconds before the next transition occurred. The total syn population was roughly 30% of the total structures sampled by

(45) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(46) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277−9287.
(47) Simmerling, C.; Elber, R. *J. Am. Chem. Soc.* **1994**, *116*, 2534−2547.
(48) Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 8392−8402.
(49) Simmerling, C.; Miller, J. L.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 7149−7155.
(50) Miller, J. L.; Kollman, P. A. *J. Mol. Biol.* **1997**, *270*, 436−450.

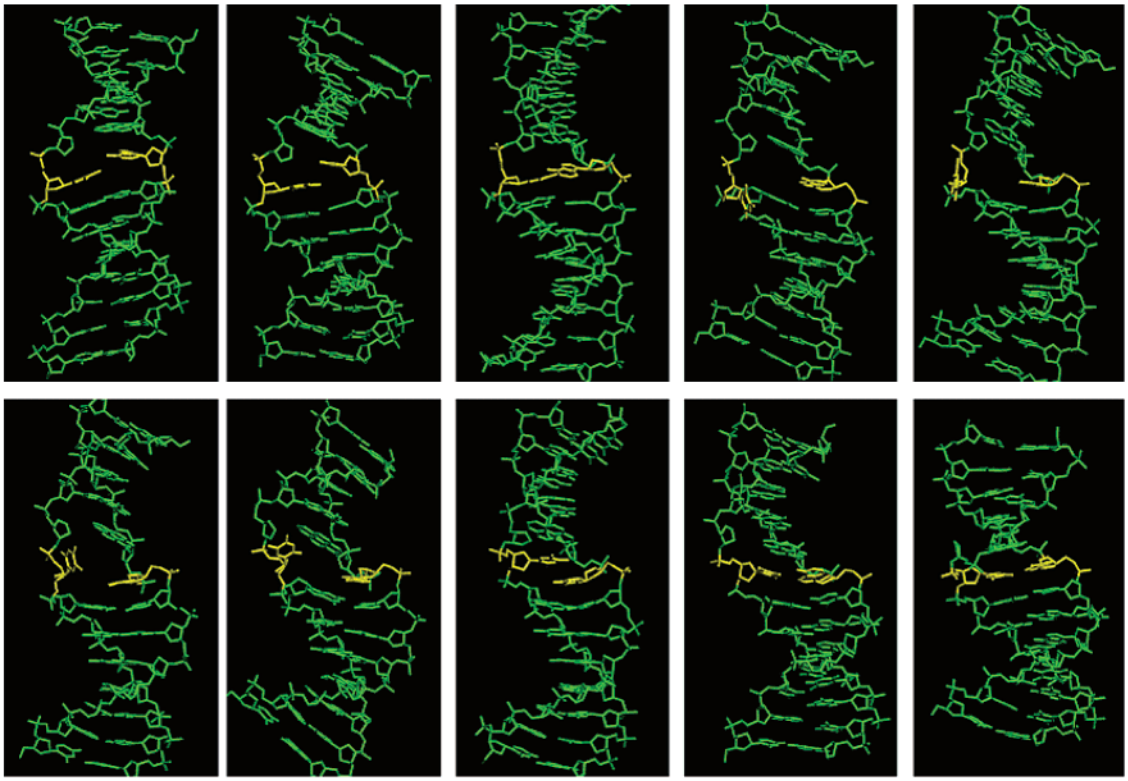

Figure 12. Transition snapshots (750 ps apart) from antiminor simulations in explicit solvent. Distortions are evident during the A:T transition but are less dramatic than observed in the GB simulation.

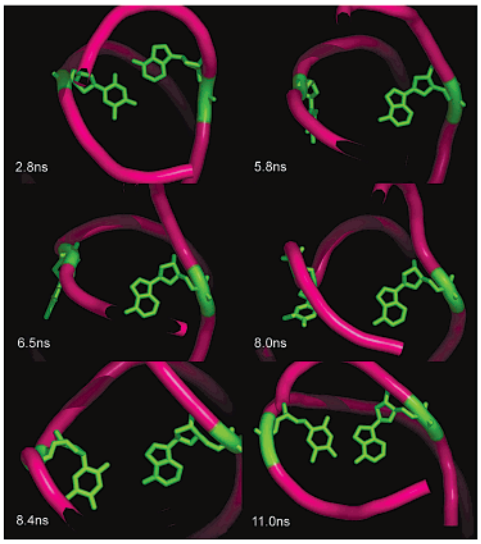

Figure 13. Snapshots showing the relative positions of ADE8 and THY19 during the A:T rearrangement in explicit solvent. Only the backbone and bases for this pair are shown.

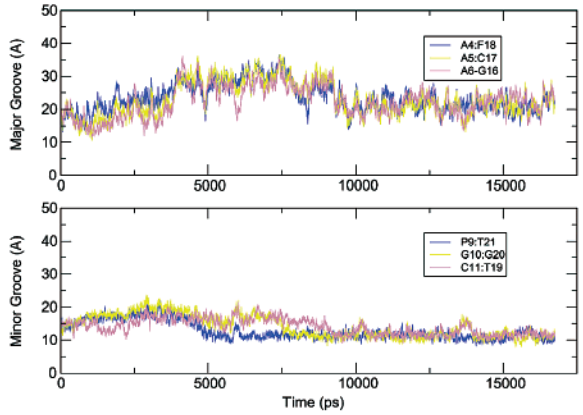

Figure 14. Groove width changes in antiminor simulation in explicit solvent. In contrast to GB simulations, no narrowing of the minor groove is seen during the transition.

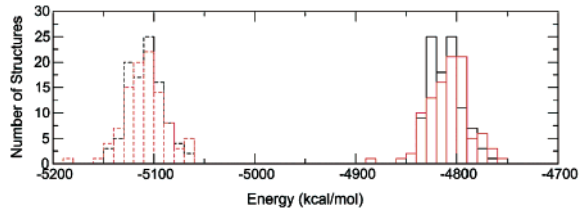

Figure 15. MM−PB (dashed lines on left) and MM−GB (solid lines on right) energy calculations for anti (black) and syn (red) conformations. The PB and GB energies are shifted, but both show no significant difference in energies for the anti and syn conformations.

all the copies, providing an estimated free energy difference of 0.5 kcal/mol favoring the anti conformation. This is consistent with the 0.8 kcal/mol determined by the energy analysis described above; in combination with the control simulation for the non-pyrene 13 base pair system with LES, this strongly suggests that the shift in anti/syn equilibrium is due to the presence of the pyrene.

With increased LES temperature (250 K), the distribution of angles was broader and transitions became more rapid (data not shown), but the conclusions obtained from the lower temperature remained valid. Exclusively anti conformation was still present in all copies for standard DNA at this temperature. Since LES is an approximate method, the transition frequency and anti/syn ratio are also approximate yet the influence of the pyrene is clearly perceivable. This provides further evidence

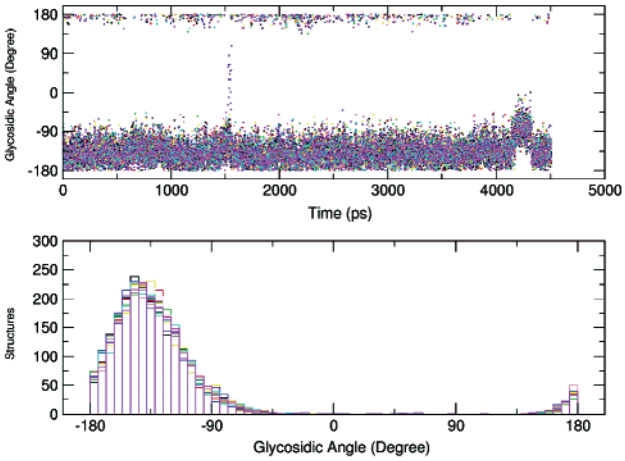

Figure 16. Time dependence (upper) and histogram of the adenine glycosidic torsion in LES simulation of non-pyrene DNA. LES copies are colored differently. Only anti is significantly sampled.

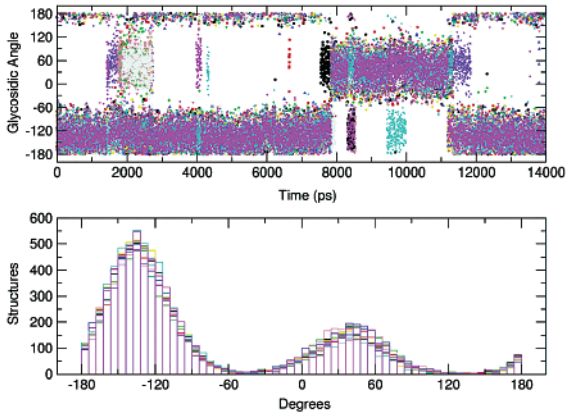

Figure 17. Adenine glycosidic torsion distribution, similar to Figure 16 but taken from a 14 ns LES simulation of pyrene DNA. In contrast to the results of LES with standard DNA, the LES copies (different colors) of ADE8 neighboring pyrene sample significant syn population. Syn and anti are clearly separate minima.

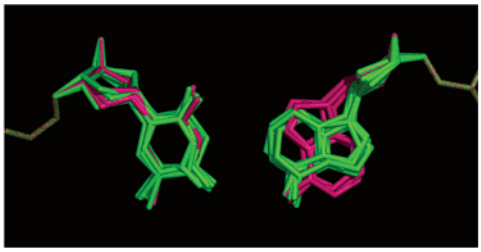

Figure 18. MD snapshot showing heavy atoms in the LES A:T pair neighboring pyrene. Copies are colored by conformation: anti with Watson−Crick hydrogen bonds (green); syn with Hoogsteen pattern (magenta).

that a distribution of anti and syn may indeed be present in this region and provides a potential explanation for the difficulties encountered in the NMR structure determination.

Conclusions

Structure determination using NMR techniques relies heavily on the accurate determination of relative geometry information, such as interatomic distances and dihedral angles. When such information is abundant, reliable structure models can be built using restrained energy minimization or molecular dynamics techniques, and sophisticated treatment of solvation and electrostatics may not be necessary. However, difficulties in data assignment or interpretation or motion on the time scale of the data collection can result in a lack of data for these regions and models of reduced quality. In such cases, simulations with more accurate treatment of inter- and intramolecular interactions may provide insight into the missing details.

In the present case, the conventional restraint-based model construction procedure failed to define the conformation of the adenine base proximal to the moiety of particular interest (pyrene). Our simulations strengthen the hypothesized presence of the A:T base pair and additionally provide a possible explanation for missing experimental data due to increased conformational heterogeneity in this region.

Our experience with this pyrene DNA suggests that investigation of sequence-dependent structure flexibility can be a very challenging task; transitions as simple as base flipping occur on a time scale that is currently inaccessible by standard molecular dynamics simulations. Transient base-pair opening events have been previously observed in simulations; in one case, a standard base pair at the end of the DNA strand opened,[18] and in another, an adenine−difluorotoluene interaction in the DNA interior was temporarily lost.[19] Our simulations in explicit solvent demonstrate not only loss and re-formation of a standard A:T base pair but transition between hydrogen bond patterns during the process. A continuum treatment of solvation appears to be very promising in this regard, with the potential to dramatically reduce the computational effort required to simulate a given time period, and also provide more rapid equilibration due to the lack of solvent friction. While the structural conclusions we draw from continuum and explicit solvent are quite similar in this case, the agreement may not be expected to generally hold. One must therefore be cautious in relying solely on continuum solvation until the simulation community gains more experience with the relative merits of the many variants of these solvent models.

Simulations with the GB model show transient coupled loss of 6 consecutive base pairs, resulting in a dramatic unfolding−refolding event for the majority of the DNA double helix. However, direct simulation of anti/syn transitions was not possible even with the help of a continuum solvent model, and this process was not observed in any of our standard MD simulations despite evidence of thermodynamic accessibility.

The use of approximations (such as LES) that reduce the energetic barriers to such transitions and permit direct observation may be a critical component of complex structural studies. In this study, the LES approximation with explicit solvation provided multiple observations of the base pair formation and reopening events, as well as anti/syn transitions and ensemble populations that are consistent with energy analysis of single copy simulation data. LES also gave results that were independent of starting model, unlike single copy simulations that can be poorly converged.

The effects of incorporation of unusual “bases”, such as pyrene, into DNA systems continues to provide new insight into the determinants of nucleic acid structure and flexibility and warrants further investigation. The introduction of the bulky pyrene residue not only preserves the fidelity of DNA replication, supporting the hypothesis that van der Waals interactions can be as important as specific hydrogen bonding in DNA replication, but as revealed in this study can also influence the

behavior of nearby residues. In this case the pyrene appears to stabilize a syn conformation in the adenine 5′ to the pyrene, providing a potential explanation for difficulty in determining a single structure for this region with a restraint-based refinement procedure. Similar effects on local structure and flexibility may have other biological implications in the recognition and repair of DNA lesions. Minor conformers have been observed experimentally for polycyclic aromatic hydrocarbon (PAH) substituted nucleic acid systems, such as benzo[*a*]pyrene diol epoxide (BPDE) DNA adducts, and these may be responsible for tumorigenic activity.[51] Characterization of these ensembles at the atomic level may be critical to understand both the effects of the damage and recognition by cellular repair machinery. We hope that approaches such as those we have presented will extend the range of conformational variability that is accessible in computer simulation and provide valuable models that complement experimental data for these systems.

Acknowledgment. We thank Carlos de los Santos for generously providing us unpublished structures and NMR data for the pyrene DNA system. G.C. thanks Vickie Tsui, Viktor Hornak, and Thomas E. Cheatham III for help and advice. Partial funding for this research was provided by the AMDeC Foundation of New York City, through its "Tartikoff/Perelman/EIF fund for Young Investigators in Women's Cancers", and by NIH Grant GM6167803.

Supporting Information Available: Force field parameters for pyrene and tetrahydrofuran moieties (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA026825L

(51) Schwartz, J. L.; Rice, J. S.; Luxon, B. A.; Sayer, J. M.; Xie, G.; Yeh, H. J. C.; Liu, X.; Jerina, D. M.; Gorenstein, D. G. *Biochemistry* **1997**, *36*, 11069−11076.