J|A|C|S
ARTICLES
Published on Web 12/06/2003

Insight into G−T Mismatch Recognition Using Molecular Dynamics with Time-Averaged Restraints Derived from NMR Spectroscopy

Richard J. Isaacs and H. Peter Spielmann*

Contribution from the Department of Molecular and Cellular Biochemistry, Department of Chemistry, and Kentucky Center for Structural Biology, University of Kentucky, Lexington, Kentucky 40536-0084

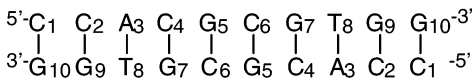
Received July 16, 2003; E-mail: hps@pop.uky.edu

Abstract: Molecular dynamics (MD) simulations were conducted for a G−T mismatch-containing DNA decamer, d(CCA*T*GC*G*TGG)$_2$, and its Watson−Crick parent sequence, d(CCACGCGTGG)$_2$. Dynamics in unrestrained MD trajectories were in poor agreement with prior ^{13}C NMR studies. However, the accuracy of the trajectories was improved by the use of time-averaged interatomic distance restraints derived from ^{1}H NMR. Postprocess smoothing of the trajectories further improved accuracy. Comparison of restrained and smoothed trajectories of the two DNA molecules revealed distinct differences in dynamics. The major groove width of the mismatched oligomer was more variable over the course of the simulation compared to its parent sequence. Greater variability in helical parameters stretch and opening for the mismatches indicated less kinetically stable base pairing. Interbase helical parameters rise, roll, and tilt were also more variable in certain base steps involving mismatched bases. These dynamic differences between normal and G−T mismatched DNA reflect differences in local flexibility that may play a role in mismatch recognition by the MutS. A potential alternate G−T mismatch binding mode for MutS is also proposed.

Introduction

Mismatches in DNA, such as guanidine opposite thymidine (G−T), are generally recognized by the MutS homodimer in prokaryotes and the MutSα heterodimer in eukaryotes to initiate their respective mismatch repair (MMR) pathways.[1] Since mismatches consist of normal DNA nucleotides, recognition of a mismatched base pair by MMR cannot rely on the detection of covalent chemical alteration of the bases themselves. This stands in contrast to base excision repair (BER) and nucleotide excision repair (NER) pathways that recognize covalently modified bases.[2] Mismatch recognition must therefore be achieved by detecting a more subtle difference between normal and mismatched DNA in the context of the double helix. In the X-ray cocrystal structure of MutS bound to a G−T mismatch-containing DNA oligomer, the DNA substrate is significantly deformed, including an approximately 60° bend into the major groove.[3] The bend is accompanied by insertion of a MutS side chain, Phe36, into the helix. Based on this observation, Lamers et al. suggest that MutS probes for DNA deformability and conformational flexibility at the mismatch site in order to accomplish mismatch recognition against the background of normal DNA flexibility.[3]

GC:



GT:

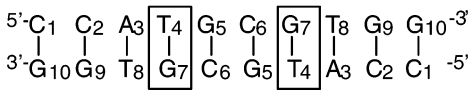


Figure 1. Sequences of the self-complementary decamers **GC** and **GT**. Boxes indicate mismatches in **GT**. Although the strands are degenerate in NMR spectra, they are distinct in MD simulations. Subsequently, one strand of each would be renumbered 11−20 when referring to MD results.

The differences in flexibility between normal and mismatched DNA are reflected in the internal dynamics of DNA oligomers that differ by a C → T transition. We have previously solved the solution structures (PDB accession numbers 1KKV and 1KKW) and described some of the internal dynamics of a normal and G−T mismatched DNA system[4] using ^{1}H homonuclear and ^{1}H−^{13}C heteronuclear NMR spectroscopy and the model free formalism of Lipari and Szabo.[5,6] The internal dynamics of molecules **GC** and **GT** (Figure 1), expressed as model free order parameters (S^2), were similar overall. However, we observed distinct order parameter differences in the vicinity

(1) Rasmussen, L. J.; Samson, L.; Marinus, M. G. In *DNA Damage and Repair, Vol. 1: DNA Repair in Prokaryotes and Lower Eukaryotes*; Nickoloff, J. A., Hoekstra, M. F., Eds.; Humana Press: Totowa, NJ, 1998; pp 205−228.
(2) Friedberg, E. C. *DNA repair and mutagenesis*; ASM Press: Washington, D.C., 1995.
(3) Lamers, M. H.; Perrakis, A.; Enzlin, J. H.; Winterwerp, H. H.; de Wind, N.; Sixma, T. K. *Nature* **2000**, *407*, 711−717.
(4) Isaacs, R. J.; Rayens, W. S.; Spielmann, H. P. *J. Mol. Biol*. **2002**, *319*, 191−207.
(5) Lipari, G.; Szabo, A. *J. Am. Chem. Soc*. **1982**, *104*, 4546−4559.
(6) Lipari, G.; Szabo, A. *J. Am. Chem. Soc*. **1982**, *104*, 4559−4570.

of the mismatch that may be important to lesion recognition by MutS. In our previous study, we were unable to arrive at a detailed description of specific motional modes within the DNA, how these modes are modulated by the introduction of G−T mismatches, and how differences in internal dynamics might affect recognition by repair proteins such as MutS, due to the model free nature of the NMR-derived dynamics information. In this report, we have analyzed fully solvated 5 ns duration molecular dynamics (MD) trajectories of **GC** and **GT** with respect to ^{13}C order parameters derived from NMR. This approach provided a more detailed understanding of the differences in dynamics between normal and G−T mismatch-containing DNA.

Computational simulations of DNA molecular dynamics are a powerful complement to experimental studies. However, the accuracy of MD trajectories is difficult to assess in the absence of relevant experimental data. Contemporary MD force field equations are Newtonian approximations of quantized molecular systems[7] and thus may not accurately reproduce fine physical details of atomic interactions and motions. DNA typically adopts only the double helix structure under physiological conditions.[8] Hence, differences in structure and internal dynamics due to sequence context, base mismatches, DNA damage, drug binding, etc. can be quite subtle. A high level of accuracy is therefore required for interpretations of DNA MD simulations to be relevant to the real DNA systems they represent. We have therefore assessed the accuracy of MD simulations of the **GC** and **GT** DNA systems by comparing the internal dynamics in MD trajectories to those derived from NMR.

In this study, we conducted unrestrained MD simulations of **GC** and **GT** using the AMBER 7.0[9] suite of programs and found many discrepancies between dynamics in the MD trajectories and experimentally derived dynamics. The agreement between simulation and experiment was substantially improved by adding ^{1}H NMR-derived interatomic distance restraints to the simulations in a time-averaged manner that interferes less with dynamic motion than conventional approaches to structural refinement using MD.[10−15] The application of a postprocessing smoothing operation on the restrained trajectories further improved their agreement with dynamics derived from ^{13}C NMR relaxation experiments. To the best of our knowledge, this is the first report of MD simulations of DNA with ^{1}H NMR-derived restraints that have been experimentally validated using independently derived ^{13}C NMR order parameters.

Armed with these accurate simulations, we were able to analyze the internal dynamics of our DNA molecules with confidence and come to important conclusions regarding the functional effects of G−T mismatches in DNA. We found that the groove widths of **GT** are more variable than those of **GC**, the mismatched base pairs have much greater conformational variability over the course of the trajectories than their parent Watson−Crick base pairs, and **GT** is more prone to bending than **GC** in ways that may relate to MutS recognition of mismatched DNA.

Results

Unrestrained MD Is Inaccurate. Two approaches to fully solvated unrestrained molecular dynamics simulations of **GC** and **GT** were investigated. The first was under constant volume, constant total energy (NVE) conditions, and the second was under constant pressure, constant temperature (NPT) conditions. The accuracy of the unrestrained trajectories was evaluated by comparing the internal dynamics in the trajectories to those calculated from ^{13}C NMR data.

The model free formalism of Lipari and Szabo[5,6] simplifies the description of complex atomic motion on the picosecond−nanosecond time scale. The dimensionless model free order parameter (S^2) describes the amplitude of spatial displacement for a particular interatomic vector with an associated correlation time (τ_e). Higher S^2 implies a more ordered vector. In the extended version of this formalism,[16] two time regimes of motion are considered, each with its own order parameter and associated correlation time. In this model, S_f^2 and τ_f relate motion on a faster time scale than S_s^2 and τ_s. For the four-parameter model, a composite S^2 is calculated by multiplying S_f^2 and S_s^2. No analysis of the internal correlation times was performed in this study, since it is unclear how the correlation times from these fits are best interpreted in the context of the complex internal dynamics of biomolecules.

Experimentally, S^2 values were calculated for methine C−H vectors in **GC** and **GT** based on ^{13}C relaxation measurements made via ^{1}H−^{13}C heteronuclear NMR.[4] Since the positions of all atoms in an MD trajectory are known, order parameters can be calculated by computing the autocorrelation function for C−H vectors at increasing time intervals (eq 2) to produce an autocorrelation map to which exponential decay functions can be fitted by nonlinear least-squares methods.[17] The NMR-derived order parameters for deoxyribose C1′−H1′ vectors and the aromatic base C6−H6 (for pyrimidines)/C8−H8 (for purines) vectors were previously reported for both **GC** and **GT** and were compared to the MD results. We found that S^2 calculation from the unrestrained trajectories was very sensitive to the extent of the autocorrelation maps used for the fitting (Figure 2), suggesting that these trajectories are not stable on the nanosecond time scale and their suitability for order parameter calculation and analysis of dynamics is very limited. Also, the order parameters for C−H vectors in the terminal nucleotides were consistently less accurate compared to experimental order parameters than those of interior nucleotides, whether restraints were included. Due to these persistent inaccuracies, the dynamics of the terminal nucleotides were excluded from the analysis.

(7) Cornell, W. D.; Cieplak, 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.
(8) Saenger, W. *Principles of nucleic acid structure*; Springer-Verlag: New York, 1984.
(9) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; P. K., W.; Kollman, P. A. University of California, San Francisco, 2002.
(10) Yao, L. J.; James, T. L.; Kealey, J. T.; Santi, D. V.; Schmitz, U. *J. Biomol. NMR* **1997**, *9*, 229−244.
(11) Schmitz, U.; Ulyanov, N. B.; Kumar, A.; James, T. L. *J. Mol. Biol.* **1993**, *234*, 373−389.
(12) Aramini, J. M.; Mujeeb, A.; Ulyanov, N. B.; Germann, M. W. *J. Biomol. NMR* **2000**, *18*, 287−302.
(13) Mujeeb, A.; Kerwin, S. M.; Kenyon, G. L.; James, T. L. *Biochemistry* **1993**, *32*, 13419−13431.
(14) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457−479.
(15) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235.
(16) Clore, G. M.; Szabo, A.; Bax, A.; Kay, L. E.; Driscoll, P. C.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1990**, *112*, 4989−4991.
(17) Chandrasekhar, I.; Clore, G. M.; Szabo, A.; Gronenborn, A. M.; Brooks, B. R. *J. Mol. Biol.* **1992**, *226*, 239−250.

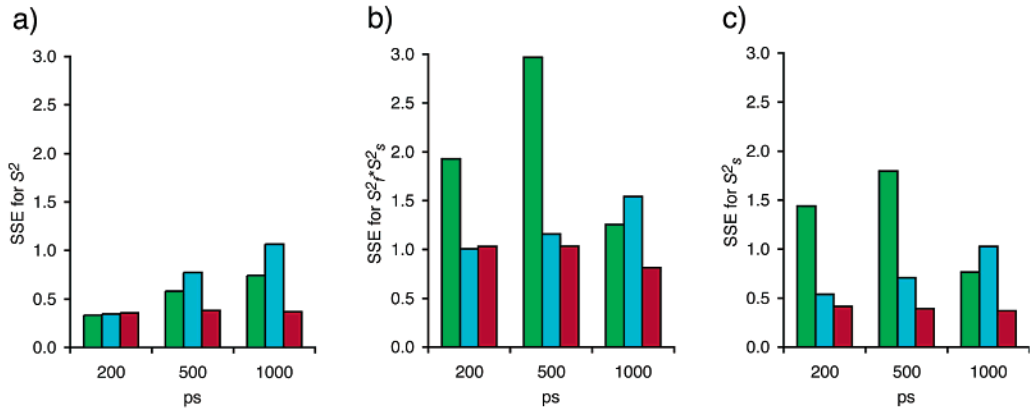


Figure 2. Sum of squared errors (SSE) compared to NMR results for computed S^2 from C1′−H1′ vectors in nonterminal nucleotides of **GT** using the first 200, 500, and 1000 ps of the autocorrelation maps for nonlinear least-squares fitting. Results from the unrestrained NVE trajectory are represented by green bars, unrestrained NPT, by cyan bars, and NPT with time-averaged distance restraints, by red bars. Frames represent (a) SSE for S^2 values from the two-parameter fit, (b) $S^2_f * S^2_s$ from the four-parameter fit, and (c) S^2_s alone from the four-parameter fit, which is equivalent to fixing S^2_f at 1.

Table 1. Sum of Squared Errors (SSE) of S^2 Values for C1′−H1′ Vectors and C6−H6 (for Pyrimidines) or C8−H8 (for Purines) Vectors in Nonterminal Residues of GT MD Trajectories[a] Compared to Those S^2 Values Derived from NMR

	NVE[b]	NPT[c]	TAR[d]	NPT-S[e]	TAR-S[f]
C1′-H1′	0.74	1.07	0.37	1.18	0.09
C6−H6/C8−H8	0.09	0.28	0.08	0.35	0.00

[a] Using the two-parameter model and fitting to 1000 ps of the computed autocorrelations. [b] Unrestrained constant volume, constant total energy. [c] Unrestrained constant pressure, constant temperature. [d] Time-averaged restrained constant pressure, constant temperature. [e] Smoothed NPT. [f] Smoothed TAR.

Compared to experimental order parameters, unrestrained simulations resulted in high sum-of-squared-errors (SSE) values (Table 1, Figure 2a−c), especially for the C1′−H1′ groups. In general, order parameters were underestimated for the majority of C−H vectors, indicating that there is more dynamic motion (disorder) present in the unrestrained MD simulations than is suggested by the ^{13}C NMR-derived order parameters in the real samples. MD-derived order parameters for aromatic C−H vectors were typically in better agreement with experiment than those of C1′−H1′ groups (Table 1). Overall, both approaches to fully solvated unrestrained MD gave average structures and internal dynamics that were in poor agreement with features determined from NMR experiments.[40]

Improvement of MD Accuracy Using Time-Averaged Restraints. We found that including ^{1}H−^{1}H NOE-based interatomic distance restraints to the MD simulations significantly improved the accuracy of the trajectories. This improvement was evident in the better agreement of the ^{13}C relaxation derived S^2 values and the corresponding S^2 calculated from the MD trajectories (Table 1). We have previously observed quantitative relationships between conformational features of solution structures and corresponding order parameters,[4,18] suggesting that internal dynamics are dependent on local structural and conformational features.

Complete relaxation matrix analysis of interproton NOE volumes was previously used to calculate upper and lower bounds for interatomic distance restraints.[4] Hydrogen bonding restraints between base pairs were also included, with upper and lower bounds of 2.10 and 1.74 Å, respectively. A total of 472 restraints were used for **GC** and 466 for **GT**. In a typical NMR-based structural refinement, the restraints are implemented such that any deviation of an interatomic distance outside its dictated upper or lower bound invokes an energy penalty that increases exponentially with the degree of violation. The implementation of the restraints in this form is expected to significantly dampen the normal dynamic motion that, in real DNA, results in a particular average distance that is reflected in the NOE buildup. In this study, we have applied our restraints in a time-averaged manner, including the hydrogen bonding restraints between strands. This approach allows dynamics to occur largely unimpeded. The implementation of the time-averaged interproton restraint applies no energy penalty as long as the average interatomic distance remains within specified upper and lower bounds over a particular time interval[10−15] (eq 1).

We found that inclusion of time-averaged restraints improved the agreement of the MD trajectories with NMR-derived ^{13}C S^2 results, as indicated by overall lower SSE values (Table 1). The self-complementary nature of the **GC** and **GT** sequences was originally chosen for simplification of the NMR spectra. However, for the MD trajectories, the consistency of results for symmetry related nucleotides can be used as a check on the convergence of the simulation to a stable equilibrium. The MD S^2 results for symmetry related nucleotides in the complementary strands are much more similar to each other in the restrained trajectories than in the unrestrained trajectories. For the NVE and NPT unrestrained trajectories of **GT**, the average ranges in computed S^2 for symmetry related C−H vectors under conditions described in Table 1 were 0.038 and 0.031, respectively. The average range for the restrained trajectory of **GT** was much lower at 0.007. Also, the accuracy of the order parameters calculated from the restrained trajectories is much less dependent on the extent of the correlation maps used for computing S^2 (Figure 2a−c). These observations suggest that the restrained trajectories have achieved equilibrium, unlike the unrestrained trajectories. In both the unrestrained and restrained approaches, the calculated C1′−H1′ S^2 was less accurate overall than S^2 of the aromatic C6−H6/C8−H8 groups, as compared to experimental values (Table 1). This suggests that DNA backbone motion and sugar repuckering are less well reproduced in the MD simulations than are motions of the bases.

(18) Isaacs, R. J.; Spielmann, H. P. *J. Mol. Biol.* **2001**, *307*, 525−540.

In most cases, calculation of S^2 from the trajectories using the simpler two-parameter fit (eq 3) produced overall more accurate results compared to experimental order parameters (Figure 2a) than did the four-parameter fit (eq 4) (Figure 2b). Inclusion of the higher-frequency S_f^2 component in the composite S^2 reported in the four-parameter results appeared to be the main cause of their lower accuracy, since S_s^2 alone (equivalent to fixing S_f^2 at 1) (Figure 2c) was more accurate than the composite S^2 (Figure 2b). Thus, the S_f^2 component is too low for the majority of C−H vectors, which is due in large part to nonphysical high-frequency components in the trajectories. Examples of such nonphysical motion apparent in the trajectories include out-of-plane motions of individual heavy atoms and twisting in the aromatic bases.

Smoothing of the Trajectories Further Improves MD Accuracy. The restrained **GC** and **GT** trajectories were smoothed by averaging atomic coordinates over a sliding 5 ps interval, which removed some of the nonphysical high-frequency motion and significantly improving their accuracy (Table 1). Aromatic rings were observed to be much more planar throughout the trajectories after smoothing. For each C−H vector examined in the restrained trajectories, the computed S^2 was higher after smoothing than before, generally bringing them into better agreement with ^{13}C NMR-derived S^2. This stands in contrast to the results from the unrestrained trajectories, where the S^2 of some C−H vectors was higher after smoothing, while others had lowered S^2. Smoothing only improved the overall accuracy of the restrained trajectories, while the unrestrained trajectories became even less accurate after smoothing (Table 1). The poor agreement between NMR and MD results in the unrestrained trajectories was probably due to a lack of stable equilibrium in the unrestrained trajectories. The use of time-averaged restraints coupled with the smoothing operation resulted in trajectories that were more accurate than unrestrained trajectories when compared to ^{13}C dynamics data derived from NMR. To the best of our knowledge, this is the first report of a comparison between experimental and simulated order in DNA and methods to improve agreement between simulation and experiment. We have shown that the use of experimentally derived time-averaged restraints improves agreement between NMR- and MD-derived order parameters. Experimentally verified trajectories lend more confidence to the analysis of the internal dynamics of **GC** and **GT** and how differences between the two molecules play a role in G−T mismatch recognition.

Discussion

Implications of Dynamics for Biological Mismatch Recognition. The utility of accurate MD simulations of DNA lies in the ability to characterize complex dynamics that cannot be discerned from static solution or crystal structures and are not well described by NMR-derived dynamics measurements such as model free order parameters. In the MutS/G−T mismatched DNA cocrystal structure, the bound DNA has a compressed major groove due to bending of the helix and a widened minor groove due to insertion of MutS amino acid side chains[3]. In our previous study,[4] it was evident that the major groove is wider in **GT** than in **GC**. This observation is inconsistent with the idea that mismatch-containing DNA is preorganized for MutS binding by having a more compressed major groove than the nonmismatched parent DNA. However, the restrained and

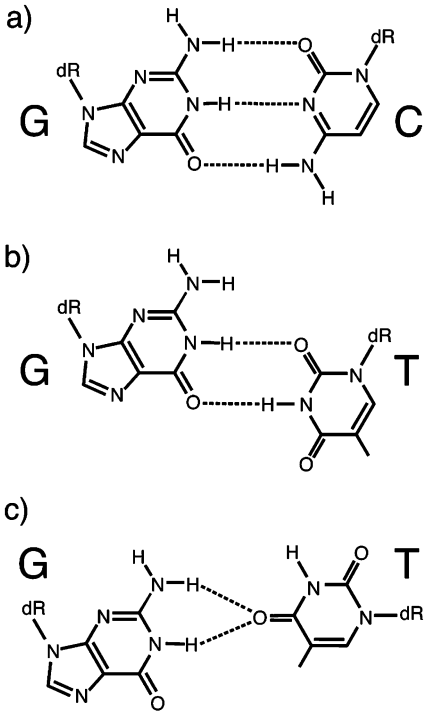


Figure 3. Hydrogen bonding schemes for G−C (a) and G−T (b) base pairs showing the highly sheared wobble base pairing of G−T mismatches. The G−T mismatch of the DNA substrate bound by MutS in the cocrystal structure[3] (c) has an unusual hydrogen bonding scheme.

smoothed MD trajectories show that **GT** has a much more variable major groove width than **GC** (Figure 4a). This suggests a greater flexibility in the **GT** major groove that would more easily allow expansion or compression in response to MutS binding. The root-mean-squared deviations (RMSDs) of the minor groove widths of **GC** and **GT** are much more similar to each other than those of the major grooves (Figure 4b). The minor groove width variability of **GC** is also very similar to that of its major groove, whereas the minor groove width of **GT** is much less variable than its major groove. This implies a coupling between major and minor groove widths in **GC** that is absent or significantly altered in **GT** due to the presence of the G−T mismatches. The differences in groove width dynamics between **GC** and **GT** are demonstrated in a movie available as Supporting Information. Considering the unusual conformational properties of the major and minor grooves in the DNA bound by MutS in the cocrystal structure, the differences in groove width dynamics we observe between normal and G−T mismatched DNA may play a role in mismatch recognition by MutS.

Another striking feature of the MutS cocrystal structure is disruption of the G−T base pair such that the mismatched bases are destacked and adopt an unusual hydrogen bonding scheme (Figure 3c) distinct from the wobble pairing typically seen in G−T mismatch structures[4,19−21] (Figure 3b). Such disruption implies that MutS recognition of and binding to the mismatch takes advantage of less favorable stacking interactions and lower thermodynamic and/or kinetic stability of base pairing for a G−T mismatch compared to a G−C pair. Previous NMR studies

(19) Allawi, H. T.; SantaLucia, J., Jr. *Nucleic Acids Res.* **1998**, *26*, 4925−4934.
(20) Hare, D.; Shapiro, L.; Patel, D. J. *Biochemistry* **1986**, *25*, 7445−7456.
(21) Hunter, W. N.; Brown, T.; Kneale, G.; Anand, N. N.; Rabinovich, D.; Kennard, O. *J. Biol. Chem.* **1987**, *262*, 9962−9970.

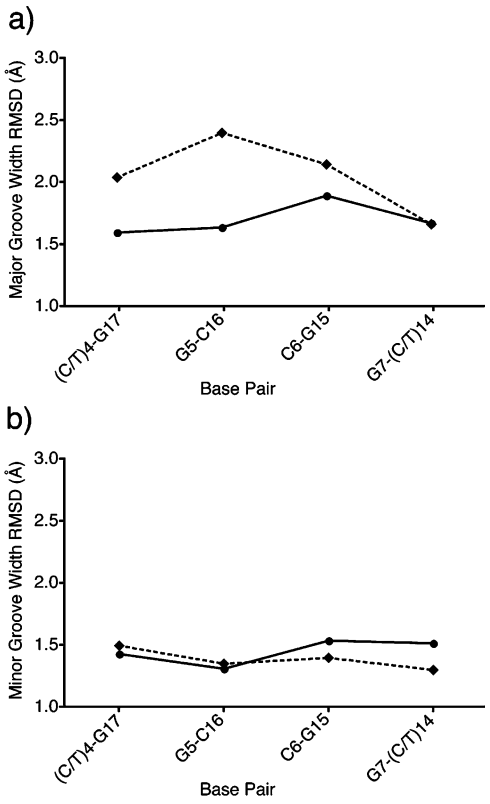


Figure 4. RMSD of the major groove (a) and minor groove (b) widths for the middle four base pairs of **GC** (solid line) and **GT** (dashed line) over the last 4 ns of the restrained and smoothed MD trajectories.

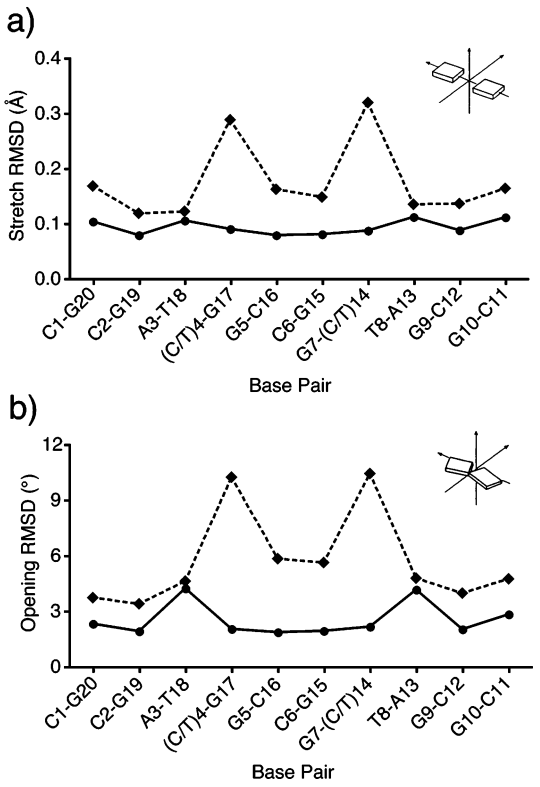


Figure 5. RMSD of the helical parameters (a) stretch and (b) opening over the last 4 ns of restrained and smoothed **GC** (solid line) and **GT** (dashed line) MD trajectories.

of G−T mismatch-containing DNA molecules found increased solvent exchange rates for the imino protons of the mismatch, and UV melting studies show that mismatches are thermodynamically less stable than Watson−Crick base pairs in the same sequence context.[4,22−24] However, it is unlikely that thermodynamics alone direct binding of MutS to mismatches, since G−G mispairs in certain sequence contexts[25] are more thermodynamically stable than many T−A rich sequences in normal DNA.[26] Thus, increased kinetic instability or flexibility in the vicinity of mismatches may also play an important role in recognition. Previously published structures of G−T mismatched DNA[4,19,21] showed a high shear value for the mismatch. Although this in and of itself does not imply decreased kinetic stability, since the wobble hydrogen bonding scheme of the mismatch requires this high shear value.[20] Comparison of the RMSD of helical parameters over the last 4 ns of the **GC** and **GT** MD trajectories showed that this high shear value is relatively stable over time. The shear RMSDs are 0.23 and 0.24 Å for the T4-G17 and T14-G7 mismatches, respectively. These values are actually the lowest of any shear RMSD in **GT**, demonstrating that the wobble configuration is kinetically stable on the picosecond−nanosecond time scale. The shear RMSDs of the corresponding G−C pairs in **GC** are only slightly lower at 0.19 and 0.18 Å, respectively. However, the mismatches show much higher variability in other conformational descriptors than their parent G−C pairs and other base pairs. The RSMDs of helical parameters stretch (Figure 5a) and opening (Figure 5b) are much higher for the mismatched base pairs than any other base pairs in **GC** and **GT**. This clearly is a result of more factors than simply losing a hydrogen bond when going from a G−C to G−T base pair. Although the doubly hydrogen bonded T−A base pairs of **GC** and **GT** have generally higher variability in stretch and opening than trebly hydrogen bonded G−C pairs, the variability is not nearly as high as is found in the mismatched base pairs. Higher variability in stretch and opening reflects lower kinetic stability and greater flexibility for the mismatches in directions described by these parameters. The stretch and opening parameters of the mismatch in the MutS cocrystal structure (−1.38 Å and −61.5°, respectively) are well outside the norm observed in B and A form DNA.[27] More flexibility in these parameters for a mismatched base pair, such as what we observe when comparing **GC** and **GT** trajectories, suggests that the mismatch can be more easily disrupted and forced into this unusual conformation than normal base pairs as an element of MutS mismatch recognition.

Interestingly, all 10 base pairs of **GT** have lower kinetic stability over the MD trajectories than the analogous base pairs of **GC**. This is evidenced by the larger RMSD in stretch and opening for each base pair of **GT** compared to its corresponding base pair in **GC** (Figure 5). This observation suggests that overall base pairing in **GT** is not as strong as in **GC**, which is in agreement with previous NMR results showing that **GT** has

(22) Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1997**, *36*, 10581−10594.
(23) Patel, D. J.; Kozlowski, S. A.; Ikuta, S.; Itakura, K. *Fed. Proc.* **1984**, *43*, 2663−2670.
(24) Sugimoto, N.; Honda, K.-i.; Sasaki, M. *Nucleosides Nucleotides* **1994**, *13*, 1311−1317.
(25) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J., Jr. *Biochemistry* **1999**, *38*, 3468−3477.
(26) SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A. *Biochemistry* **1996**, *35*, 3555−3562.
(27) Arnott, S.; Hukins, D. W. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504−1509.

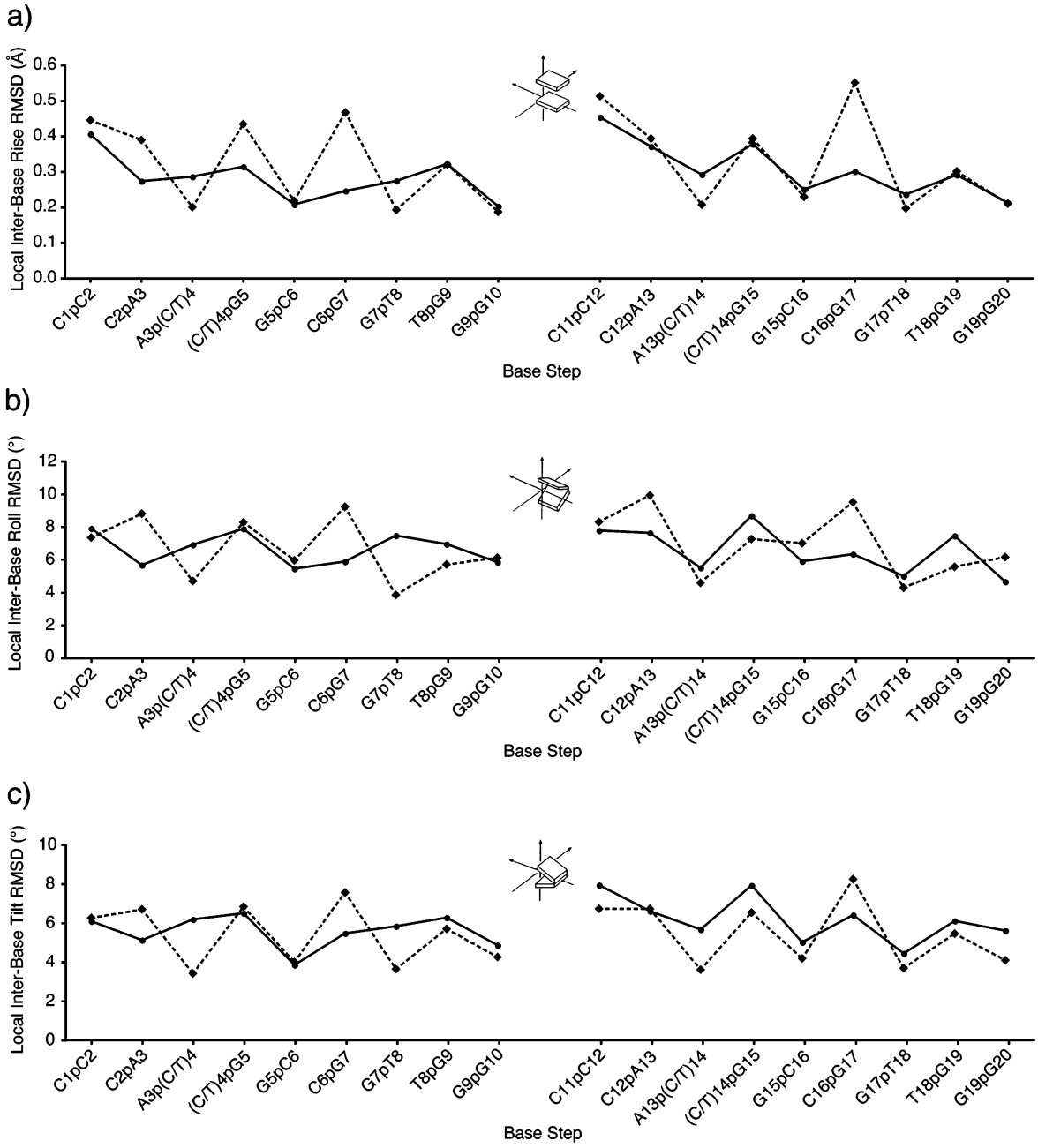


Figure 6. RMSD of the local interbase helical parameters (a) rise, (b) roll, and (c) tilt over the last 4 ns of restrained and smoothed **GC** (solid line) and **GT** (dashed line) MD trajectories.

a lower melting temperature than **GC**[4] and is also in accord with published thermodynamic studies of G−T mismatch-containing DNA.[22,24]

DNA bending, such as that resulting from MutS binding, is characterized by changes in interbase helical parameters such as rise, roll, and tilt in each strand.[28] We would expect to see greater flexibility (variability) in these parameters in base steps containing mismatched bases over the course of the trajectories if **GT** has a lower barrier to bending upon MutS binding than does **GC**. Indeed, certain base steps containing mismatches have higher than average RMSD values for local interbase rise, roll, and tilt (Figure 6). This is especially true for C6pG7 and C16pG17 steps of **GT**, which contain a mismatched G as the 3′ base of the step. The DNA bound by MutS in the cocrystal structure also has values well outside the norm for these parameters in steps containing a mismatched base. For the sequence surrounding the mismatch, $^{5'}$A**G**G$^{3'/3'}$T**T**C$^{5'}$ (where mismatched bases are in bold), the Ap**G** step has rise, roll, and tilt values of 6.76 Å, 49.1°, and −22.7°, respectively. Phe36 of MutS is inserted into the **T**pT step, resulting in an even higher rise of 9.58 Å. Roll and tilt values at this step are 44.6° and 45.7°, respectively. The observation that steps containing mismatched bases have greater variability in rise, roll, and tilt suggests that these steps in **GT** could adopt unusual conformations that are preorganized for binding by MutS.

A Potential Alternate G−T Mismatch Binding Mode for MutS. Although the cocrystal structure of MutS bound to G−T

(28) Dickerson, R. E. *Nucleic Acids Res.* **1998**, *26*, 1906−1926.

mismatched DNA[3] has provided much insight into mismatch recognition, it is unclear whether G−T mismatches in all sequence contexts would be bound in the same manner. One feature of the MutS−DNA complex revealed by the cocrystal structure is the insertion of Phe36 into the helix, which stacks with the mismatched and destacked T. Presumably, Phe36 would also insert into the helix and stack with a purine when MutS binds G−A and G−G mismatch-containing DNA.[29,30] Therefore, MutS could potentially bind G−T mismatched DNA in such a way that Phe36 inserts into either the G or T strand. This may depend on which strand is more susceptible to destacking or better preorganized for insertion of Phe36 into the helix. These conformational characteristics would in turn depend on the sequence dependent flexibility of the two strands.

G−T mismatches in different sequence contexts have distinct thermodynamic[22] and structural[4,19,21] outcomes in the context of the DNA double helix. The G−T mismatch in the DNA bound by MutS in the cocrystal structure is found in the local sequence context $^{5'}$A**G**G$^{3'/3'}$T**T**C$^{5'}$, while the sequence context of the mismatch in **GT** is $^{5'}$C**G**T$^{3'/3'}$G**T**A$^{5'}$. Using the nearest neighbor model, it is predicted that the MutS-bound trimer is destabilizing to the DNA double helix, with a ΔG at 37 °C of 0.39 kcal/mol. In contrast, the **GT** trimer stabilizes the DNA double helix, with a ΔG at 37 °C of −0.40 kcal/mol. For comparison, the ΔG values at 37 °C of the G−C containing parent sequences of these trimers are −3.12 and −3.61 kcal/mol, respectively. We have shown that the thermodynamic destabilization of G−T mismatches relative to G−C base pairs is accompanied by kinetic destabilization; therefore it is expected that the thermodynamically different $^{5'}$A**G**G$^{3'/3'}$T**T**C$^{5'}$ and $^{5'}$C**G**T$^{3'/3'}$G**T**A$^{5'}$ sequences will have different kinetic behaviors. Since it is also clear that the affinity of MutS for mismatches is sequence dependent,[31] MutS may interact differently with the **GT** sequence than the sequence found in the cocrystal structure.

In our previous work, we noted that the mismatched G7 of **GT** was structurally perturbed and more dynamic compared to its parent G7 in **GC**.[4] However, the mismatched T4 of **GT** was relatively unperturbed structurally and dynamically compared to its parent C4 in **GC**. The increased dynamics of G7 are also evident in the large RMSD of certain helical parameters involving the G7 (and symmetry related G17) in the MD trajectories. The observation that C6pG7 and C16pG17 steps of **GT** have, in general, greater variability in rise, roll, and tilt than steps involving the mismatched T's (Figures 6a−c) implies that these steps may be better preorganized for insertion of an aromatic ring between the bases. It is therefore possible that Phe36 can be more easily inserted into the helix at these mismatched G-containing steps than a step involving the mismatched T for our particular **GT** sequence. This is an alternative mode for MutS binding to G−T mismatches in other sequence contexts.

Experimental Section

MD Simulations. MD simulations of **GC** and **GT** were performed using the "sander" module of the AMBER 7.0 software package with an improved Cornell et al. force field.[7,32] Canonical DNA structures were generated with the AMBER "nucgen" module and solvated in a periodic truncated octahedron of TIP3P water[33] such that the minimum distance from the DNA to the periodic boundary was 10 Å. Sodium counterions were added iteratively via an automated routine in a shell around the DNA using a Coulombic potential on a grid, replacing solvent molecules where steric conflicts arose, until electroneutrality for the system was achieved. The systems were equilibrated in stages of 500 step steepest descent minimizations followed by 10 ps molecular dynamics with a 1 fs time step, electrostatics evaluated via the particle mesh Ewald method[34] and a 9 Å nonbonded cutoff, constant temperature (300 K) and pressure (1 atm) maintained via weak coupling[35] with time constants of 0.2 ps, and SHAKE (tolerance = 0.0001) applied to all bonds involving hydrogen.[36] Positional restraints of 1000 kcal/mol were applied to the DNA in the first stage, those of 25 kcal/mol were applied in the second stage, and no restraints were applied in the final stage. After a final 500 step steepest descent minimization, a 100 ps MD run was used to bring the systems slowly up to 300 K, followed by a 5 ns production run using the end state of the 100 ps run as its initial configuration. For constant volume, constant total energy simulations, the volume in the 5 ns production runs was fixed at the value of the last 100 ps run and no coupling to an external temperature bath was used. For the constant pressure, constant temperature simulations, the pressure was regulated at 1 atm, and the temperature, at 300 K as in the equilibration steps, but using coupling constants of 10 ps, so as to not interfere with dynamics more than necessary. The coordinates were saved every 500 steps (1 ps), and only the final 4 ns of each trajectory were used for analysis to minimize potential artifacts from the starting conformations.

Time-Averaged Restraints. Restraints were formulated as a standard flat-bottomed well with the edges defined by upper and lower bounds derived from complete matrix relaxation analysis of NMR-derived ^{1}H NOE volumes.[4] Hydrogen bonding restraints between complementary bases were also included. Penalty forces were dependent on adjustable force constants that were assigned an initial value of 32 kcal/mol and were ramped from 0.1 to 1000 times this amount over the first 100 ps of the production runs. The restraints were applied in a time-averaged fashion, instead of being instantaneously employed for transient violations. The sander module has functionality for this type of restraint built in and computes time-averaged distances as

$$\bar{r} = (1/C)\{\int_0^t \exp[(t' - t)/\tau]\, r(t')^{-i}\, dt'\}^{-1/i} \quad (1)$$

where $\bar{r}$ is the time-averaged value of the internal coordinate (distance or angle), t is the current time, τ is the exponential decay constant, $r(t')$ is the value of the internal coordinate at time t', the average is over internals to the inverse of i (we found a value of 1 to give the best results), and C is a normalization integral.

Smoothing of MD Trajectories. To remove some of the nonphysical high-frequency motion from the MD trajectories, the coordinate averages from each contiguous 5 frame (5 ps) group over the final 4 ns of the production runs was output as a new frame. These averaged frames were reassembled into new trajectories of 3996 frames.

Calculation of Order Parameters. The autocorrelation function $C(t)$ was evaluated for each C1′ and C6/C8 C−H vector in **GC** and **GT** as

$$C(t) = \left\langle \frac{P_2(\mu(\tau)\, \mu(\tau + t))}{r^3(\tau) r^3(\tau + t)} \right\rangle \quad (2)$$

where $P_2(x) = (3x^2 - 1)/2$ and is the second-order Legendre polynomial,

(29) Su, S. S.; Modrich, P. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5057−5061.
(30) Kramer, B.; Kramer, W.; Fritz, H. J. *Cell* **1984**, *38*, 879−887.
(31) Joshi, A.; Rao, B. J. *J. Biosci.* **2001**, *26*, 595−606.
(32) Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(34) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(35) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(36) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

μ is the unit vector along a C−H bond, τ corresponds to the time index of a MD frame, and t is the time between frames being compared.[17] The internuclear distance r between methine C and H atoms was fixed in the simulations via the SHAKE algorithm.[36] The "ptraj" module of AMBER was used to evaluate $C(t)$ over the course of the simulations.

Correlation curves were fit to

$$C(t) = S^2 + (1 - S^2)\exp(-t/\tau_e) \quad (3)$$

or

$$C(t) = S_s^2 + (1 - S_f^2)\exp(-t/\tau_f) + (S_f^2 - S_s^2)\exp(-t/\tau_s) \quad (4)$$

The two-parameter model (eq 3) corresponds to the original Lipari and Szabo model free formalism,[5] while the four-parameter model (eq 4) corresponds to the extended model of Clore et al.[16] Equation 4 becomes equivalent to eq 3 when S_f^2 is fixed at 1. Nonlinear least-squares fitting of $C(t)$ from the simulations to the model equations above was done using the "Solver" module of Microsoft Excel 2001.[37] Both two-parameter and four-parameter models were evaluated over t ranges up to 200, 500, and 1000 ps. For the four-parameter model, composite S^2 values were computed by multiplying S_f^2 and S_s^2.

Helical Parameter and Groove Width Analysis. Helical parameters for base pairs and base steps and major and minor groove widths for the middle four base pairs of each MD frame of **GC** and **GT** over the final 4 ns of the time-averaged restrained and smoothed production runs and for the DNA molecule of the MutS cocrystal structure were calculated using CURVES 5.3.[38,39] In frames where the groove width could not be calculated, the frame was ignored, but these cases were only a very small percentage of the total. Variability in helical parameters and groove widths over the course of the MD trajectories was defined as the root-mean-squared deviation (RMSD).

Acknowledgment. We gratefully acknowledge Dr. Thomas E. Cheatham III for help in compiling and running AMBER software on the University of Kentucky's HP N-4000 computer cluster. We also thank the Kentucky Center for Computational Sciences and the National Center for Supercomputing Applications (NCSA) for allowing us the use of the HP N-4000 and HP Superdome supercomputers. This work was supported by NSF Grant MCB-9808633, awarded to H.P.S.

Supporting Information Available: A Quicktime movie of a portion of the restrained and smoothed **GC** and **GT** trajectories is available for download. This material is available free of charge via the Internet at http://pubs.acs.org.

JA037333R

(37) Harris, D. C. *J. Chem. Educ.* **1998**, *75*, 119.
(38) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.
(39) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655−667.
(40) Konerding, D. E.; Cheatham, T. E., III; Kollman, P. A.; James, T. L. *J. Biomol. NMR* **1999**, *13*, 119−131.