Published in final edited form as:

J Phys Chem B. 2013 May 23; 117(20): 6194–6205. doi:10.1021/jp403127a.

DNA Bending Propensity in the Presence of Base Mismatches: Implications for DNA Repair

Monika Sharma¹, Alexander V. Predeus¹, Shayantani Mukherjee¹, and Michael Feig^{1,2,*}

¹Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI, 48824, United States

²Department of Chemistry, Michigan State University, East Lansing, MI, 48824, United States

Abstract

DNA bending is believed to facilitate the initial recognition of the mismatched base for repair. The repair efficiencies are dependent on both the mismatch type and neighboring nucleotide sequence. We have studied bending of several DNA duplexes containing canonical matches: A:T, G:C, various mismatches: A:A, A:C, G:A, G:G, G:T, C:C, C:T, T:T, and a bis-abasic site: X:X. Free energy profiles were generated for DNA bending using umbrella sampling. The highest energetic cost associated with DNA bending is observed for canonical matches while bending free energies are lower in the presence of mismatches, with the lowest value for the abasic site. In all of the sequences, DNA duplexes bend towards the major groove with widening of the minor groove. For homoduplexes, DNA bending is observed to occur via smooth deformations, whereas for heteroduplexes, kinks are observed at the mismatch site during strong bending. In general, pyrimidine:pyrimidine mismatches are the most destabilizing, while purine:purine mismatches lead to intermediate destabilization and purine:pyrimidine mismatches are the least destabilizing. The ease of bending is partially correlated with the binding affinity of MutS to the mismatch pairs and subsequent repair efficiencies, indicating that intrinsic DNA bending propensities are a key factor of mismatch recognition.

Keywords

Umbrella sampling; free energy; mismatch recognition; DNA repair

INTRODUCTION

DNA structure is commonly characterized in terms of its helical form. However, deviations from the regular helix structure are observed during many biological processes, such as gene expression,^{1,2} DNA repair,^{3–6} or even under mechanical stress.⁷ These deformations can range from the wrapping of DNA around histones^{1,2} to the local bending of DNA during protein-DNA recognition.^{4–6} One prominent example is the highly bent DNA structure found in complexes with MutS and MutS α , the enzymes responsible for the first echelon of post-replication mismatch repair.^{3–6} DNA bending is believed to be a key feature by which base mismatches or base insertions/deletions are recognized. One hypothesis suggests that the mismatch repair proteins initially bind to DNA non-specifically and then probe for

*Corresponding Author: Department of Biochemistry & Molecular Biology, Michigan State University, 603 Wilson Road, BCH 218, East Lansing, MI, 48824, United States. Tel: 1-517-432-7439 feig@msu.edu.

SUPPLEMENTARY DATA

Supplementary Tables S11–S14 and Supplementary Figures S11–S15. This material is available free of charge via the Internet at <http://pubs.acs.org>.

increased local flexibility in the DNA due to the presence of a mismatch.^{8–12} The crystal structures of mismatch DNA bound to the mismatch recognition proteins for both prokaryotes^{13–15} and eukaryotes^{16,17} show highly specific contacts with the mismatch, involving a conserved motif that is inserted into the minor groove of the DNA. This raises the question of whether DNA mismatch recognition is achieved primarily because of intrinsic DNA properties or through specific protein-DNA interactions at the mismatch site. The broader question is how the mismatches alter the flexibility of DNA duplexes.

A DNA base mismatch occurs when other base pairs are present in place of the correct canonical matched pairs adenine:thymine (A:T) or guanine:cytosine (G:C). This may result most frequently from nucleotide misincorporation during replication. Interestingly, the mismatch repair system does not repair all types of mismatches with equal efficiency,^{3–5,18–23} suggesting that the differences in the intrinsic properties of DNA owing to different mismatches may play a role. Several biophysical,^{24–26} biochemical,²⁷ and molecular biology experiments^{18–21} have been carried out to characterize the mismatch efficiencies for different mismatch types. It was found that the probability of repair is correlated with both the mismatch and the neighboring nucleotide sequence.^{18,19,28} A general conclusion from these studies is that, for the methyl-directed mismatch repair system in *E. coli*, G:T, A:C, G:G, and A:A mismatches are repaired with the highest efficiency, followed by T:T, C:T, and G:A that are repaired with intermediate efficiencies, while C:C mismatches are the least efficiently repaired. The order of affinity for the binding of synthetic DNA fragments containing possible mismatches to *E. coli* MutS was found to be as: G:T > G:G > A:A ≈ T:T ≈ T:C > C:A > G:A > C:C > G:C from band shift analysis.²⁹ The order of affinity largely matches the repair efficiencies suggesting that the initial binding of DNA to MutS is a major factor. Similar trends are observed for the eukaryotic mismatch repair efficiencies,^{4,21,30} with the general conclusion that purine:pyrimidine and purine:purine mismatches are repaired more effectively than pyrimidine:pyrimidine mismatches.^{5,18,31} Biophysical characterizations of DNA duplexes containing mismatches, using gel electrophoresis, NMR, and calorimetric techniques^{24,25,32,33} furthermore suggest an influence of the neighboring sequence context on the stability of the mismatch pairs. The stability of mismatch-containing DNA in turn is roughly correlated with their observed repair efficiencies. However, these experiments fail to establish a clear mechanistic picture for how exactly the mismatches are identified by the DNA repair enzymes. In particular, the intrinsic bending propensity of mismatch-containing DNA remains unclear, which is the focus of the present work.

Sequence-induced DNA deformations have been studied previously.^{34–37} For bent DNA structures,^{38,39} different types of deformations have been observed including kinked structures,^{40,41} flipped-out nucleotides,⁴² and the formation of local bubbles.^{39,41} However, the inherent flexibility and transient stability of bent DNA structures limit the role of experiments when aiming to characterize such structures in atomistic detail. Instead, molecular dynamics (MD) studies have been used extensively to provide structural details for bent DNA, involving DNA minicircles^{40,41} and DNA oligomers, such as A-tract DNA^{43–46} and B-DNA oligomers.^{44,45,47–49}

A number of studies have previously examined bending in DNA duplexes. Using the helical parameters roll, tilt, and propeller twist to evaluate the magnitude and direction of bending in B-DNA^{47,48} and A-tracts,^{43,47} it was observed that the precise magnitude of the bend is sequence dependent and bending is facilitated by a smooth deformation that is induced via rolling of adjacent base pairs. For small DNA minicircles, Lankas *et al.*⁴⁰ observed that in addition to smooth deformations, both, local base unpairing and kinking, enhance the DNA flexibility; and kinking remains a feasible route for relaxing the elastic energy in strongly bent DNA giving rise to anharmonic behavior⁵⁰. Two types of kinks were observed at the

positions of high curvature: type I and type II. The type I kink is characterized by the significant bending at a single base pair in the DNA structure⁵¹. The type II kink is characterized by the localized melting of the DNA,^{39,52,53} where the central base pair of three consecutive base pairs is broken and its bases are stacked onto the 5' neighboring bases of the corresponding strand.

Curuksu *et al.*^{44,45} used a screw-axis orientation approach to induce continuous bending deformations of B-DNA, which is based on the calculation of a set of screw axes of the adjacent base pairs, and restraining the angle between two axes using umbrella sampling. Such controlled bending studies on oligomers with alternating GC and AT sequences and A-tracts⁴⁵ indicate that while moderate bending occurs mainly through the coupled rolling of adjacent base pairs, strong bending leads to local type II kinks. Adiabatic mapping along the screw-axis orientation for oligomers containing G:A, G:T, C:C mismatches, or an abasic site⁴⁴ suggested that a greater variety of bent conformations with different directions and magnitudes of global bending are possible for the G:A and C:C mismatches compared to regular B-DNA. In another approach, Lankas *et al.*⁴⁶ used averaging of two or more base-fixed coordinate frames to quantitate the global bending of A₄T₄ or T₄A₄ tracts. They observed that A₄T₄ tracts are bent significantly more than T₄A₄ tracts. In a recent study,⁴⁹ Spiriti *et al.* used an adaptive umbrella sampling approach to determine the free energy surface of DNA dodecamers using a pseudoroll angle definition to describe bending. In these simulations, DNA bending was studied in both directions at a central base pair step, exploring roll angles between -70° to +70°, with positive and negative roll angles corresponding to bending of the DNA towards the major and minor grooves, respectively. They observed that most of the DNA bending flexibility is due to changes in the roll angles. The key finding from this study was that bending at high roll angles appeared to be facilitated by ionic screening. While the focus of the previous studies has been on bending in regular B-DNA structures,^{44,45,49} very little is known about bending propensities of mismatch-containing DNA.

Here, we focus specifically on the intrinsic ability of mismatch-containing DNA to assume bent conformations. Because spontaneous bending of DNA in the absence of proteins is a rare event, we employ umbrella sampling simulations using the bending angle as a restraining degree of freedom, and report relative free energy profiles as a function of the mismatch type. The results are then further interpreted in the context of DNA repair and recognition of mismatch-containing DNA by repair enzymes.

MATERIALS AND METHODS

Molecular Dynamics Protocol

Molecular dynamics simulations were performed with NAMD 2.7⁵⁴ using the CHARMM27 force field for nucleic acids.^{55,56} Parameters for protonated cytosine (at N3) were taken from supplementary parameters provided in the CYTH residue patch developed by the MacKerell group and distributed with the recent CHARMM c36 and c37 packages. Parameters for protonated adenine (at N1) were developed following the same protocol as the original nucleic acid parameters. Further details are provided in the supplementary material. The CHARMM program (version c36a6)⁵⁷ was used for the initial setup and to prepare the protein structure files (PSF). Each structure was solvated in a truncated octahedral box of TIP3P water molecules in such a way that the minimum distance between solute images would be 20 Å. This resulted in a box size of approximately (65 Å)³. The negative charge of the overall system was then neutralized by adding sodium ions (28 for the regular DNA duplex and 27 for protonated nucleic bases, corresponding to 0.16 M). All simulations were carried out using periodic boundary conditions. Electrostatic interactions were calculated using the particle-mesh Ewald (PME) summation scheme with the PME grid spaced evenly

at 1 Å. The simulations were performed in the NPT ensemble at a constant temperature of 298 K using a Langevin thermostat applied to non-hydrogen atoms only, with a damping coefficient of 5 ps⁻¹. A constant pressure of 1 atm was maintained using a Langevin pressure piston with a piston decay time of 100 fs, a piston oscillation time of 200 fs, and with a piston temperature of 298 K. All simulations were performed under holonomic constraints placed on all bonds that included a hydrogen atom, using the SETTLE algorithm⁵⁸ with a tolerance of 10⁻⁸ Å.

Preparation of DNA duplexes

A total of 15 pentadecamer DNA duplexes with the sequence 5'-GAACCGCXCGCTAGG-3'/5'-CCTAGCGYGC GGTTTC-3' were considered where the central base pair X8:Y23 was varied to include canonical and mismatched pairings as well as a bis-abasic variant without a base at either X or Y sites. Throughout this paper, each of these DNA oligomers is referred to by the name of its central base pair. For instance, the G:T notation will refer to the DNA heteroduplex consisting of guanine and thymine as the central mismatch base pair. Initial structures were prepared from the crystal structure of the G:T-mismatch DNA in complex with MutSα (PDB ID: 2O8B¹⁶). The bent mismatched DNA fragment was extracted, minimized, solvated in a rectangular (48 × 55 × 70 Å³) box of explicit (TIP3P) water molecules, and then simulated (unrestrained MD, 298 K, Langevin thermostat with damping coefficient of 5 ps⁻¹, 20 ns) in the absence of the protein. This resulted in a slightly curved B-DNA structure with a bending angle of around ~170°. This equilibrated structure was then used as a starting point to model the structures of the other mismatches.

The mismatches A:A, A:C, G:A, G:G, C:T, C:C, and T:T, as well as canonically paired structures (G:C, A:T) were prepared by mutating the bases using *mutateNA.pl* from the MMTSB Tool Set.⁵⁹ For the three purine-purine mismatches A:A, G:A, and G:G, both, *anti*, *anti* and *anti*, *syn* forms, are possible. NMR studies suggest that only the *anti*, *anti* form with transient hydrogen bonds is seen for A:A.²⁴ For G:A⁶⁰ and G:G base pairs,⁶¹ there appears to be a dynamic equilibrium. The *anti*, *anti* form is believed to be dominant for G:A,^{60,62} but the G(*anti*):A(*syn*) form is also observed by X-ray analysis.¹⁵ In the case of G:G, the G(*anti*):G(*syn*) form is believed to be dominant.^{15,26,62,63} For mismatches C:C and A:C, the protonation states of cytosine and adenine require special attention since the pK_a values of protonated A:C and C:C were estimated at 7.20³³ and 6.95,⁶⁴ respectively, which is close to physiological pH.

To consider alternate mismatch conformations, four additional simulations were set up. One simulation each was run with G(*anti*):X(*syn*) (X = G or A), where the equilibrated and solvated structures of DNA duplex containing G(*anti*):X(*anti*) were taken as the starting point but with the atoms of the base X rotated around the C1'-N9 axis by 180°. Two simulations were run with protonated pairs C(+):C and A(+):C, which were built from the equilibrated and solvated structures of DNA duplexes containing C:C and A:C but with the extra hydrogen atom added using CHARMM. The four additional systems were minimized in vacuum for 500 steps. Restraints were initially placed on the following bonds: in G(*anti*):G(*syn*), N1(GUA8)-O6(GUA23), N2(GUA8)-N7(GUA23); in C(+):C, N4(CYTH8)-N3(CYT23), N3(CYTH8)-O2(CYT23), H3(CYTH8)-O2(CYT23), H41(CYTH8)-N3(CYT23); in A(+):C, N1(ADEH8)-O2(CYT23), N6(ADEH8)-N3(CYT23), H1(ADEH8)-O2(CYT23), H62(ADEH8)-N3(CYT23). Then, the structures were simulated using the setup described above (Langevin thermostat, 5 ps⁻¹ at 298 K, Langevin barostat at 1 atm) with force constants that were varied as follows: k = 2 kcal/mol/Å² for 20 ps, k = 10 kcal/mol/Å² for 20 ps, k = 20 kcal/mol/Å² for 50 ps, k = 5 kcal/mol/Å² for 1.0 ns, k = 1 kcal/mol/Å² for 0.5 ns, followed by an unrestrained simulation for 0.5 ns. The G(*anti*):A(*syn*) simulation created a stable wobble base pair without any restraints and

was simply equilibrated with conventional MD for 1.1 ns. All of the hydrogen bonds remained stable in the unrestrained simulations. The relative position of the bases (*anti/syn*) and the hydrogen bonding patterns of the resulting mismatches adequately represents known experimental structures of matched G:C, A:T, as well as the A(*anti*):A(*anti*),^{24,65,66} G(*anti*):A(*anti*),^{60,62} G(*anti*):G(*anti*),^{24,61,62,67} G(*anti*):A(*syn*),¹⁵ G(*anti*):G(*syn*),^{15,61–63} A:C,^{33,68} A(+):C,^{33,68} C:T,⁶⁴ C:C,^{24,64} C(+):C,^{24,64} T:T,²⁴ and G:T^{13–16,69} mismatches (Figure 1).

Finally, a system with an bis-abasic oligomer, where both nucleobases 8 and 23 were completely eliminated and replaced with a hydrogen atom, was set up for comparison of the bending energies. The resulting system was minimized for 150 steps and then equilibrated over 40 ps.

In order to determine which protonation state is most relevant during bending we compared bending free energy profiles for protonated and unprotonated cases. In the case of A:C, we find that the bending is energetically more favorable for A(+):C compared to A:C (Figure SII). Given a pK_a value of 7.2 for adenine, this suggests that at pH = 7 there may be an equilibrium between A and A(+) for unbent DNA, but upon bending, A(+):C is likely to be the preferred protonation state for the sequence we are considering here. For C:C base pairs, the bending free energy profiles are similar for both C:C and C(+):C (Figure SII) suggesting that the protonation is secondary to the focus of this paper. Since the pK_a is 6.95, we will focus the subsequent analysis on the C:C base pair.

Umbrella sampling using DNA bending angle restraints

To perform umbrella sampling simulations, a bending angle (ξ) was used as the reaction coordinate. The angle was calculated between the centers of mass of heavy atoms of three nucleotide blocks of the pentadecamer: bases 2–5 / 26–29, 6–10 / 21–25, and 11–14 / 17–20 (Figure 2). The bending angle restraint was applied using the *angle* component of the NAMD *colvar* module. The reference bending angle, ξ_{ref} was changed from 90° to 180° with increments of 5° for a total of 19 windows. All simulations used a simple harmonic biasing potential of the form $V(\xi) = \frac{1}{2}k(\xi - \xi_{ref})^2$. To prepare the initial structures for each window, the equilibrated starting structures (see preparation above) with a bending angle of 170° were simulated with restraints gradually changing in both directions, towards 180° and towards 90°. Each window was sampled initially for 0.6 ns with a force constant of 0.1 kcal mol⁻¹ degree⁻². After the initial umbrella run, an additional 20 ns of production sampling was carried out with a force constant of 0.4 kcal mol⁻¹ degree⁻².

The free energy associated with DNA bending was extracted from the umbrella simulations using the weighted histogram analysis method (WHAM).^{70,71} WHAM expresses the optimal unbiased probability distribution $P(\xi)$, from a set of simulations performed along the reaction coordinate ξ with biased potentials, $V(\xi)$. The free energy or potential of mean force (PMF) at a given ξ is then obtained as $W(\xi) = -k_B T \ln P(\xi)$.

For the calculation of free energy landscapes along additional degrees of freedom that are not part of the biasing function, such as helicoidal parameters, multi-dimensional WHAM analysis was carried out where the first dimension is the bending angle and the additional coordinates were included as additional dimensions. 1D PMFs were generated using in-house generalized WHAM code, based on a formalism described earlier.^{72,73}

Structural and energetic analysis

The resulting DNA conformations were analyzed in terms of their helicoidal and DNA backbone parameters using 3DNA.⁷⁴ Averaging of the parameters was done after bias elimination using Boltzmann averaging. A hydrogen bonding analysis was performed using

VMD.⁷⁵ Cluster analysis of conformations, sampled at 4 ps intervals, was performed using the *kclust* program in the MMTSB Tool Set.⁵⁹ Structures were grouped based on mutual root mean square deviations (RMSD) of heavy atoms of the central nucleotides 5–11 and 20–26 using a 3.0 Å radius cutoff.

The van der Waals components of the stacking energies of the nucleobases of the central three base pair segments consisting of nucleotides 7–9 and 22–24 were calculated using CHARMM's interaction energy module (version c36a6).

Much of the discussion in this paper focuses on the effect of bending on DNA energetics and structure. We have covered here bending angles ranging from 90° to 180° which ranges from highly bent DNA to unbent DNA. In order to separate the effects of different bending regimes we separate the discussion into three ranges that will be referred to throughout the paper: (a) strong bending, corresponding to $\xi < 110^\circ$; (b) moderate bending, corresponding to $110^\circ \leq \xi < 130^\circ$, and (c) weak bending, corresponding to $130^\circ \leq \xi \leq 180^\circ$. As can be seen from Table SII, the moderate bending regime corresponds to the bending angles observed for the DNA when bound to the mismatch recognition proteins.

RESULTS AND DISCUSSION

The primary goals of this study are to examine the energetics as a function of DNA helix bending and to describe the structural perturbations in the DNA that result from bending in the presence of different mismatches. To address these questions, we have performed umbrella sampling simulations of mismatch-containing DNA duplexes along a helix bending reaction coordinate. Eight different mismatches at the central base pair are compared with the two canonical base pairs and bis-abasic DNA. The resulting energetic and structural analysis is presented in detail in the following sections.

Free energy profile associated with DNA bending

The free energy of bending for the DNA duplexes is shown in Figure 3. The profiles adopt an overall similar shape with a global minimum near 160° that corresponds to a slightly bent structure and rising energetic cost of bending towards smaller or larger angles. The shape of the profile is relatively independent of the presence and type of mismatch pairing for bending angles between 130° and 180°. For angles below 130°, differences in the bending free energy become more pronounced. The highest energetic cost for bending DNA is observed for the canonical matches reaching more than 15 kcal/mol at 90°, and the lowest cost is found for the abasic DNA with about 8 kcal/mol at 90°. The energetic cost of bending mismatch-containing DNA is distributed between these two extremes.

The overall shape of the free energy curves is similar to those obtained by Curuksu *et al.* using screw axis orientations^{44,45} and Spiriti *et al.* using pseudoroll angles,⁴⁹ but with differences in the magnitude of the bending angle due to different definitions. Large values in our bending reaction coordinate correspond to low positive bending angles in the studies by Curuksu *et al.*^{44,45} and Spiriti *et al.*⁴⁹ Both of those studies^{44,45,49} determined the minimum of free energy minima surface at $\sim 10^\circ$ – 20° , which corresponds to $\sim 160^\circ$ – 170° in our reaction coordinate. The free energy costs associated with stronger bending of DNA are higher in our study than those found by Curuksu *et al.* (they obtained a value of about 5.5 kcal/mol at 90°),^{44,45} but similar to those obtained from the simulations using pseudoroll angle restraints.⁴⁹ The difference may be because of sequence effects that play an increasing role as DNA is strongly bent at angles less than 130° or because of differences in force fields and computational methodology.

A particular interesting observation is that differences between canonical and mismatched base pairs do not become pronounced until bending angles of 120° are reached. Such angles correspond to the bending angles observed for the bent DNA conformations when bound to mismatch recognition proteins (Table SI3). A comparison of the bending free energy values at the bending angles seen in the DNA-mismatch recognition protein crystal structures (110° < ξ < 130°) shows increasing values in the order: X:X < A(+):C < C:C, A:A < C:T < G(*anti*):G(*anti*) < G:T < G(*anti*):A(*anti*) < G(*anti*):A(*syn*) < T:T < G(*anti*):G(*syn*), A:T, G:C (Table 1). It is clear that less energetic cost is associated with bending of mismatch-containing DNA in comparison with canonical pairs. For the glycosidic rotamers, the free energies associated with the bending of DNA duplexes with G:A and G:G in the *anti*, *syn* conformations are much less favorable than the *anti*, *anti* conformers. This suggests that the *anti*, *anti* conformations of G:A and G:G may be the relevant conformations in bent DNA. Furthermore, the overall trend in the relative cost of bending DNA for different mismatches is in line with the thermodynamic studies which indicate that (a) purine:pyrimidine and purine:purine mismatches are stable mismatches;^{3,4,32} (b) pyrimidine:pyrimidine mismatches are destabilizing in all sequence contexts.^{24,32} However, there is less of a correlation with experimental MutS binding affinities (Table 1) suggesting that bending propensity alone may not be sufficient to explain MutS binding preferences.

Because our definition of the bending angle is somewhat arbitrary, we have also calculated bending angles using the commonly used DNA analysis software 3DNA and generated the corresponding free energy profiles (see supplementary material, Figure SI2). There is generally a good correlation between our definition and the 3DNA analysis with the overall trends being similar except for a systematic shift in the angles by ~10° due to differences in how DNA bending is defined. Therefore, we believe that our findings are robust irrespective of the exact bending angle definition.

Directionality of bending

The DNA duplex is free to bend in any direction during our simulations since we only restrain the bending angle but not its direction. In our simulations, all DNA duplexes bend towards the major groove. This is reflected in the DNA groove widths (Figure 4). For weak bending, we observed that the major groove widths of all DNA heteroduplexes are wider than the minor groove widths as expected for canonical DNA structures. As the DNA structure bends more, we observe narrowing of the major grooves and widening of the minor grooves for all of the duplexes (Figure 4). The effect is most pronounced for A(+):C, A:A, G:A, and C:C mismatches where the minor groove is wider than the major groove by 5 to 10 Å. For an ideal B-DNA structure, the minor and major groove widths are about 12 Å and 17 Å, respectively. Structural studies of mismatch DNA-protein complexes^{13–16} indicate that the mismatch recognition proteins MutS and MSH6 insert a conserved “Phe-X-Glu” motif into the minor groove of DNA. This is accompanied by bending of the DNA towards the major groove side and widening of the minor groove around the mismatch site (Table SI3). In the complex structures, the minor grooves are about 5 Å wider than the major grooves similar to what we observe for strongly bent DNA.

We have also calculated the curvature of the central seven base pairs from roll and tilt angles (Figure 4). As would be expected, higher curvature values are observed for more highly bent DNA duplexes. The curvature angle is a measure of local DNA bending away from straight DNA (0° degrees curvature). The average values of 50–70° for highly bent DNA therefore suggest that most of the overall DNA bending angle is accounted for by changes in the base pair roll and tilt angles in the three base pairs – the mismatch itself, as well as the pairs before and after the mismatch site.

Variations in helicoidal parameters

To analyze the influence of bending on the helical conformations of the DNA duplexes, we have calculated the intra base pair (shear, stretch, stagger, buckle, propeller, and opening) and inter base pair step (roll, tilt, twist, rise, slide, and shift) parameters (Figure 5) using 3DNA. The intra base pair parameters account for the deformations of base pairs, and the inter base pair step parameters account for the alterations in stacking. Local variations are observed for the central 5 base pairs at and around the mismatch site (−2 to +2), with the rest of the duplex conformation showing negligible perturbations. Comparing average values between weakly, moderately, and strongly bent DNA duplexes, we find that for canonical base pairs there are only small changes upon bending except for roll and tilt, which change to accommodate for the overall bending (Figure 4). In the presence of the mismatches, the DNA structure differs significantly from the structure with canonical base pairs for most mismatches, even for weakly bent DNA. Furthermore, DNA bending induces significant changes from the unbent structures for many mismatches. The specific changes in helicoidal parameters upon bending depend strongly on the type of mismatch and it is difficult to discern common trends. For example, the A:A, G:A, and G:G mismatches lead to a greatly increased propeller twist upon bending while C:T and T:T mismatches exhibit significantly altered stagger and buckle upon bending. The only common observations are significant variations in opening and shear parameters of the mismatch pairs upon bending, and a nearly uniform increase in roll angles that is reflective of DNA bending towards the major groove and widening of the minor groove.^{43,47,48}

Comparing our simulated helicoidal parameters with those observed crystallographically at the mismatch site when heteroduplex is bound to MutS or MutSa proteins, there is close agreement for shear, stretch, and stagger, but buckle and opening values differ significantly (Figure 5A). In the crystal structures of DNA heteroduplexes, the mismatch base orients to form hydrogen bonds with a conserved glutamate residue and the mismatched base pairing opens up, which leads to a significant increase in buckle and opening angle values. For the inter-base pair step parameters (Figure 5B), increased rise, slide, and curvature are observed for the crystal structures at the base pair step (C7X8/Y23G24). No significant variations are observed for the consecutive base pair step (X8C9/G22Y23) parameters. This may be attributed to the fact that in the protein-bound crystal structures, phenylalanine is inserted between Y23 and G24, which leads to a type I kink at the mismatch site and alters the slide, rise, and curvature.

Fluctuations in phosphodiester backbone torsions

We further analyzed the DNA backbone by calculating ribose pseudorotation and backbone torsion angles (Figure 6, Supplementary Table SI4). At low bending, the sugar pucker conformations are predominantly C2'-endo, as expected from regular B-DNA structures. As the DNA bends, frequent transitions of sugar pucker at the mismatch site are observed from C2'-endo to C3'-endo, primarily for Y23 (Figure 6A, Table SI4). These transitions are observed more frequently for the pyrimidine nucleobases, reflecting previous findings that pyrimidines have a greater tendency towards A-type sugar conformations than purines.⁷⁶ For the crystal structure bending angle range ($\xi = 120^\circ\text{--}125^\circ$), we observe that for X8, the preferred sugar pucker conformation is C2'-endo and for Y23, the sugar puckers are mostly simulated as O4'-endo. This is in agreement with the crystallographic data where in structures of DNA bound to MutS or MutSa proteins, mismatch pairs have primarily C2'-endo sugars on the plus strand (X8) and O4'-endo sugars on the minus strand (Y23).

We also analyzed BI/BII transitions that are characterized by the difference in the phosphodiester torsions, $\epsilon\text{-}\xi$, where ϵ is the torsion angle defined by C4'-C3'-O3'-P atoms and ξ is defined by C3'-O3'-P-O5' atoms. We find that in general, BII conformations are

more frequent during strong bending (Figure 6B, Table SI4), as suggested earlier.⁷⁶ For the mismatches G:T, A(+):C, C:C, G(*anti*):G(*anti*), we observed BII conformations for Y23 during low DNA bending. The mismatch pairs observed in the crystal structures of mismatch DNA bound with MutS or MutSa proteins show a preference for the BII conformation at X8 and for BI at Y23. Interestingly, Y23 has a uniformly high propensity for BI for all base pairs between 120° and 140°, close to the bending angle of the DNA when bound to MutS, while BII conformations are sampled increasingly at X8 for angles below 120° for many mismatches.

The simulations reported here were carried out with the CHARMM27 force field to be compatible with previous simulations of MutS⁷⁷ and MSH2-MSH6 bound DNA.^{78,79} Recently, modifications to the CHARMM nucleic acid force field were published to improve the sampling of BI/BII conformations of DNA.⁸⁰ The modifications increase the amount of BII sampling relative to BI but otherwise have only marginal effects on backbone torsion sampling and essentially no effect on the helicoidal parameters.⁸⁰ In this study, we focus on the energetics of DNA bending which is driven primarily by changes in roll and tilt angles as discussed above. Therefore, we expect that the updated parameters would have little difference on the bending energetics and overall structural features upon bending. We would expect an overall increase in the sampling of BII conformations, but it is not clear that the updated force field parameters would change the conclusions from our study that DNA bending increases the sampling of BII conformations.

Conformational space sampled at various bending angles

To further characterize the different conformations visited by bent DNA duplexes as a function of mismatches, we pooled the conformations obtained from the simulations at all bending angles, 90° ≤ ξ ≤ 180°, and focused the analysis on the base pairs −3 to +3 around the central base pair. Clustering was then carried out to identify the major conformations that are sampled at different bending angles. The centroids of the clusters are presented as representative structures in Figures SI3, SI4, and SI5 along with their relative occurrences as a function of bending angle. It should be pointed out that the relative populations of different clusters are only qualitatively meaningful since they result from sampling under biased conditions.

To analyze whether the DNA conformations similar to those observed in the crystal structure are also visited during moderate bending, we calculated RMSD values with respect to the DNA in the MutS crystal structure. The RMSD calculation was carried out for backbone atoms as well as heavy atoms of the central seven base pair fragment of the DNA in sampled conformations with bending angles between 110° ≤ ξ < 130°. The analysis further focused on the structures closest to the crystal structure and the fraction of structures below a certain cutoff (2 Å RMSD for backbone atoms and 4.2 Å for heavy atoms). The results are given in Table 2. We observe that for most of the mismatches, backbone conformations as close to ~1.5 Å are visited during the simulations as indicated by the lowest RMSD values (R_{BB}, Table 2). Interestingly, for certain mismatches such as A:A and G:A, significant populations of backbone conformations (P_{BB}) sample the crystal structure conformation. This suggests that crystal-structure-like conformations are sampled even in the absence of the protein and that once the DNA is bent binding to MutS may involve conformational selection rather than an induced fit mechanism. However, the closest heavy atom RMSD values are relatively large suggesting that the nucleobases do not sample conformations that are consistent with the complex-bound structures and that an induced fit mechanism is required to orient the bases appropriately.

For most of the mismatches, we observe that type II kinks are found at strong bending angles (ξ ≥ 120°). All type II kinked base-pair steps are associated with highly positive roll

angles and highly negative tilt angles (Figure 5B). In crystal structures of protein-DNA complexes of mismatch recognition proteins, a conserved phenylalanine residue stacks between the mismatch base and its 5' flanking base of minus strand of DNA. The mismatched base is then displaced towards the widened minor groove and type I kinks are observed as the mismatched pair unstacks from 5' flanking base pair. The type I kinks, similar to those observed in crystal structure and characterized by unstacking of two base pairs, are not observed in our simulations. Presumably, type I kinks would be introduced only in the presence of the protein once the phenylalanine is inserted to stack with the mismatch base.

In previous studies on DNA bending,⁴⁵ the presence of kinks was observed to be force field-dependent. For simulations with the parmbsc0 force field,⁸¹ only type II kinks were observed during bending simulations.⁴⁵ However, during simulations with the parm94 force field, which artificially stabilizes the γ *trans* angles, type I kinks were observed.⁴⁵ In a separate study, Spiriti *et al.* used capping potential to exclude the effect of type II kinks and study only the type I kinks with both CHARMM27 and AMBER parmbsc0 force fields.⁴⁹ Thus some of our conclusions presented here may also depend on the choice of force field.

Hydrogen bonding and stacking interactions between base pairs

To further assess the stability of the central base pairs in the presence of mismatches, we have analyzed intra base pair hydrogen bonding for X8:Y23 as a function of DNA bending (Figure 7). In general, the base pairs which form more stable hydrogen bonds have been observed to have higher resistance to bending (Figure 3). An interesting case is observed for G:G, where no consistent hydrogen bonding interaction is present between the guanines (Figure 7) in *anti*, *anti* conformer. The representative structures for G:G simulations indicate that in *anti*, *anti* conformation, larger disruptions in the DNA structure, also indicated by higher backbone fluctuations (Figure 6), are observed even at weaker bending. G:G pairing in *anti*, *syn* conformation on the other hand, stabilizes the DNA duplex with persistent hydrogen bonding interactions.

For weakly bent DNA, the experimentally known hydrogen bonds (Figure 1) are largely maintained. However, upon bending many of the hydrogen bonds are weakened significantly as bending increases to angles below 120–130° for the mismatches. In contrast, canonical base pairs retain hydrogen bonding until bending angles of about 100°. This suggests that MutS could identify mismatches through weakened hydrogen bonding at bending angles around 120°, allowing for favorable interactions with protein residues in the context of MutS. In fact, a conserved glutamate residue has been identified that interacts in the crystal structures with the mismatch base pair in the minor groove in a position to compete for hydrogen bonding with the mismatch bases.^{13–16}

Besides hydrogen bonding interactions between the bases, the stacking interactions of the nucleobases also stabilize the DNA duplex. Bending reduces the ability of bases to form optimal stacking interactions and an interesting question is to what extent the energetics of losing such interactions contributes to the cost of bending. We have analyzed the van der Waals contributions to the energy for the nucleobases of the central three base pair fragment (Figure 8) as a measure of stacking between the base pairs –1 to +1 around the mismatch site. For all of the duplexes, the van der Waals energies become more favorable as bending decreases, indicating enhanced stacking interactions among the central base pairs in unbent DNA duplexes. The increase in van der Waals energies upon bending ranges from 2 to 10 kcal/mol, on the same order as the increase in free energy. This suggests that the loss of stacking interactions is a major contributor to the overall increase in free energy upon bending. However, the ordering of different base pairs in terms of bending free energy and

in particular the much larger cost of bending DNA with canonical base pairs is not reproduced by the change in van der Waals energies alone.

CONCLUSIONS

The results from this study indicate that the free energies associated with the bending of DNA duplex are lower for mismatch containing heteroduplexes than for homoduplexes with canonical base pairs. For homoduplexes, DNA bending is observed to occur via smooth bending of base pair steps, whereas for heteroduplexes, kinks are observed at the mismatch site during strong bending. Our results are in line with the thermodynamic studies^{3,4,32} which indicate that purine:pyrimidine and purine:purine mismatches are more stable mismatches than pyrimidine:pyrimidine mismatches. Furthermore, the relative ease of bending mismatch containing DNA suggests at least a partial mechanism for discriminating heteroduplex from homoduplex DNA.

From the PMF profiles associated with DNA bending (Figure 3), it is evident that the bending of either homoduplex or heteroduplex DNA is not a spontaneous process. In the absence of proteins, the global minimum for bending is observed near $\sim 160^\circ$ and even for the most easily bendable mismatch containing heteroduplex the free energy cost of bending beyond 120° exceeds 3 kcal/mol. This indicates that in order to achieve bent DNA as observed in the crystal structures in complex with MutS or MutS α , the interactions with the proteins are necessary to induce and stabilize the bent conformations.

Our conformational analysis (Table 2) for structures in the moderate bending regime, corresponding to the structures in complex with MutS, indicates that a fraction of the backbone conformations assumed upon forced bending are quite similar to the crystal structure of protein-bound DNA. Therefore, after DNA is bent by the proteins, the crystallographically observed backbone conformation could be assumed simply through conformational selection. However, the orientations of the base pairs upon bending in the absence of the proteins do not closely match the crystal structure conformation. Hence, an induced fit mechanism is required to assume the protein-bound conformation. In the crystal structures,^{13–16} there are two key interactions: The mismatched base stacks with the aromatic ring of a conserved phenylalanine, which requires base unstacking and an increase of helical rise. Furthermore, a conserved glutamate forms a hydrogen bond with the N3 atom of a mismatched thymine or the N7 atom of a mismatched purine. We did not gain insight from our simulations whether phenylalanine insertion may be more favorable for heteroduplex DNA. However, a reduction in hydrogen bonding upon bending in our simulations for mismatch-containing DNA suggests that competition for base-base hydrogen bonding by the conserved glutamate residue would provide additional discrimination for mismatch-containing DNA by preferentially stabilizing a complex with heteroduplex over homoduplex DNA.

An interesting question arises from the *anti*, *anti* – *anti*, *syn* equilibrium of purine:purine mismatches. In order to form the latter interaction with the glutamate residue, the mismatched purines are observed in *syn* conformations. On the other hand, our results indicate that the *anti*, *anti* conformers of the purine:purine mismatches are relatively easier to bend than the corresponding *anti*, *syn* conformers. This suggests that upon initial bending of DNA the purine:purine mismatch may be in the *anti*, *anti* conformation but then would have to convert to the *anti*, *syn* conformation to assume the crystallographically observed structure. Close inspection of the crystal structures of DNA duplexes consisting of purine-purine mismatch pairs bound to MutS¹⁵ shows that there would be enough space for rotation of the purine rings from *anti* to *syn* at the mismatch site (Figure 9). Further studies of heteroduplex DNA bound to MutS would be required to examine this idea in more detail.

In vivo, the mismatch recognition proteins have variable affinities for different mismatches.^{3,4,6,18–20,22,23} The strongest binding to the MutS is observed for G:T, A:C, and G:G mismatches, whereas weaker binding is observed for C:C mismatch. Correlating our mismatches bending energetics (observed for the crystal structure bending angles) with the binding affinities of MutS to different mismatches (Table 2) and the order of repair efficiencies, we observe that the easily bent mismatches, such as G:T, A:A, G:G, and A:C are good substrates for the MutS recognition, and are repaired efficiently. T:T and G:A mismatches which are relatively resistant to bending are repaired with intermediate efficiencies. This indicates that the intrinsic ease of bending mismatch-containing DNA is likely a key contribution to mismatch recognition by MutS. However, the less efficiently repaired C:C and C:T mismatches do not agree with this trend since their bending free energies are relatively low. This indicates that there are likely additional factors beyond the intrinsic properties of heteroduplex that ultimately determine repair efficiencies.

Our simulation results combined with existing single molecule^{8–11} and biochemical data,^{4–6,12,28} allow us to propose the following mechanism for mismatch recognition by MutS and related homologs: First, MutS scans the DNA duplex via one-dimensional diffusion^{8–10}. While doing so, the protein attempts to induce bending in the DNA. In the presence of canonical base pairs, the energetic cost is high and a stable complex is not formed. However, when a mismatch is encountered, bending the DNA becomes easier and formation of a stable complex stops the diffusion process. Second, upon successful binding to bent heteroduplex DNA more specific interactions involving the “Phe-Xaa-Glu” motif are formed with the DNA. The mismatched base hydrogen bonds with the glutamate; and forms stacking interaction as the phenylalanine is inserted to introduce a type I kink at the mismatch site. These conformations would presumably further stabilize the MutS-heteroduplex DNA complex and allow repair to be initiated. Our previous works on the post-mismatch recognition by MutS or MutS α ^{77,79} proteins suggest that the kink introduced at the mismatch site will destabilize the DNA base pairing at the 5' adjacent base pair. The opening of the 5' adjacent base next to the mismatched base was observed to promote the ATP hydrolysis,⁷⁷ which is associated with the initiation of the sliding-clamp formation.^{9,10} In this sliding mode, the protein is capable of sliding along the DNA via ATP hydrolysis, and eventually, signals the repair machinery to repair.^{9,10,12}

Collectively, our results demonstrate that DNA bending is an essential aspect of the mismatch recognition. Here, we have simulated free DNA structures and the next step is to study DNA bending in the presence of MutS to further understand the details of the mismatch recognition process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Dr. Sean Law for helpful discussions.

FUNDING

Funding from NIH GM092949 is acknowledged. Computer resources were used at XSEDE facilities (TG-MCB090003).

ABBREVIATIONS

Amber	assisted model building with energy refinement
ATP	adenosine triphosphate
CHARMM	chemistry at Harvard molecular mechanics
DNA	deoxyribonucleic acid
MD	molecular dynamics
MMTSB	molecular modeling tools for structural biology
NMR	nuclear magnetic resonance
PME	particle-mesh Ewald
PMF	potential of mean force
RMSD	root mean square deviation
VMD	visual molecular dynamics
WHAM	weighted histogram analysis method

References

1. Vilar JM, Saiz L. *Curr Opin Genet Dev.* 2005; 15:136–144. [PubMed: 15797196]
2. Richmond TJ, Davey CA. *Nature.* 2003; 423:145–150. [PubMed: 12736678]
3. Modrich P. *Annu Rev Genet.* 1991; 25:229–253. [PubMed: 1812808]
4. Modrich P. *J Biol Chem.* 2006; 281:30305–30309. [PubMed: 16905530]
5. Kunkel TA, Erie DA. *Annu Rev Biochem.* 2005; 74:681–710. [PubMed: 15952900]
6. Li GM. *Cell Res.* 2008; 18:85–98. [PubMed: 18157157]
7. Lionberger TA, Demurtas D, Witz G, Dorier J, Lillian T, Meyhofer E, Stasiak A. *Nucleic Acids Res.* 2011; 39:9820–9832. [PubMed: 21917856]
8. Gorman J, Chowdhury A, Surtees JA, Shimada J, Reichman DR, Alani E, Greene EC. *Mol Cell.* 2007; 28:359–370. [PubMed: 17996701]
9. Jeong C, Cho WK, Song KM, Cook C, Yoon TY, Ban C, Fishel R, Lee JB. *Nat Struct Mol Biol.* 2011; 18:379–385. [PubMed: 21278758]
10. Cho WK, Jeong C, Kim D, Chang M, Song KM, Hanne J, Ban C, Fishel R, Lee JB. *Structure.* 2012; 20:1264–1274. [PubMed: 22682745]
11. Wang H, Yang Y, Schofield MJ, Du C, Fridman Y, Lee SD, Larson ED, Drummond JT, Alani E, Hsieh P, et al. *Proc Natl Acad Sci U S A.* 2003; 100:14822–14827. [PubMed: 14634210]
12. Tessmer I, Yang Y, Zhai J, Du C, Hsieh P, Hingorani MM, Erie DA. *J Biol Chem.* 2008; 283:36646–36654. [PubMed: 18854319]
13. Obmolova G, Ban C, Hsieh P, Yang W. *Nature.* 2000; 407:703–710. [PubMed: 11048710]
14. Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK. *Nature.* 2000; 407:711–717. [PubMed: 11048711]
15. Natrajan G, Lamers MH, Enzlin JH, Winterwerp HH, Perrakis A, Sixma TK. *Nucleic Acids Res.* 2003; 31:4814–4821. [PubMed: 12907723]
16. Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. *Mol Cell.* 2007; 26:579–592. [PubMed: 17531815]
17. Gupta S, Gellert M, Yang W. *Nat Struct Mol Biol.* 2012; 19:72–78. [PubMed: 22179786]
18. Dohet C, Wagner R, Radman M. *Proc Natl Acad Sci U S A.* 1985; 82:503–505. [PubMed: 16593539]
19. Schaaper RM. *J Biol Chem.* 1993; 268:23762–23765. [PubMed: 8226906]
20. Kramer B, Kramer W, Fritz HJ. *Cell.* 1984; 38:879–887. [PubMed: 6386179]

21. Varlet I, Radman M, Brooks P. *P Natl Acad Sci U S A*. 1990; 87:7883–7887.
22. Fazakerley GV, Quignard E, Woisard A, Guschlbauer W, Vandermarel GA, Vanboom JH, Jones M, Radman M. *Embo J*. 1986; 5:3697–3703. [PubMed: 2951250]
23. Radman M, Wagner R. *Annu Rev Genet*. 1986; 20:523–538. [PubMed: 2949693]
24. Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. *Biochemistry*. 1999; 38:3468–3477. [PubMed: 10090733]
25. Tikhomirova A, Beletskaya IV, Chalikian TV. *Biochemistry*. 2006; 45:10563–10571. [PubMed: 16939208]
26. Brown T, Hunter WN. *Biopolymers*. 1997; 44:91–103.
27. Brown J, Brown T, Fox KR. *Biochem J*. 2003; 371:697–708. [PubMed: 12558499]
28. Rajski SR, Jackson BA, Barton JK. *Mutat Res*. 2000; 447:49–72. [PubMed: 10686306]
29. Brown J, Brown T, Fox KR. *Biochem J*. 2001; 354:627–633. [PubMed: 11237867]
30. David P, Efrati E, Tocco G, Krauss SW, Goodman MF. *J Neurosci*. 1997; 17:8711–8720. [PubMed: 9348340]
31. Marra G, Schar P. *Biochem J*. 1999; 338(Pt 1):1–13. [PubMed: 9931291]
32. Ke SH, Wartell RM. *Nucleic Acids Res*. 1993; 21:5137–5143. [PubMed: 8255768]
33. Allawi HT, SantaLucia J Jr. *Biochemistry*. 1998; 37:9435–9444. [PubMed: 9649326]
34. Okonogi TM, Alley SC, Reese AW, Hopkins PB, Robinson BH. *Biophys J*. 2002; 83:3446–3459. [PubMed: 12496111]
35. Lankas F, Sponer J, Langowski J, Cheatham TE 3rd. *Biophys J*. 2003; 85:2872–2883. [PubMed: 14581192]
36. Lankas F. *Biopolymers*. 2004; 73:327–339. [PubMed: 14755569]
37. Geggier S, Vologodskii A. *Proc Natl Acad Sci U S A*. 2010; 107:15421–15426. [PubMed: 20702767]
38. Wiggins PA, van der Heijden T, Moreno-Herrero F, Spakowitz A, Phillips R, Widom J, Dekker C, Nelson PC. *Nat Nanotechnol*. 2006; 1:137–141. [PubMed: 18654166]
39. Yuan C, Rhoades E, Lou XW, Archer LA. *Nucleic Acids Res*. 2006; 34:4554–4560. [PubMed: 16954151]
40. Lankas F, Lavery R, Maddocks JH. *Structure*. 2006; 14:1527–1534. [PubMed: 17027501]
41. Mitchell JS, Laughton CA, Harris SA. *Nucleic Acids Res*. 2011; 39:3928–3938. [PubMed: 21247872]
42. Giudice E, Varnai P, Lavery R. *Nucleic Acids Res*. 2003; 31:1434–1443. [PubMed: 12595551]
43. Strahs D, Schlick T. *J Mol Biol*. 2000; 301:643–663. [PubMed: 10966775]
44. Curuksu J, Zakrzewska K, Zacharias M. *Nucleic Acids Res*. 2008; 36:2268–2283. [PubMed: 18287117]
45. Curuksu J, Zacharias M, Lavery R, Zakrzewska K. *Nucleic Acids Res*. 2009; 37:3766–3773. [PubMed: 19380377]
46. Lankas F, Spackova N, Moakher M, Enkhbayar P, Sponer J. *Nucleic Acids Res*. 2010; 38:3414–3422. [PubMed: 20123729]
47. Goodsell DS, Dickerson RE. *Nucleic Acids Res*. 1994; 22:5497–5503. [PubMed: 7816643]
48. Dickerson RE, Kopka ML, Pjura P. *Proc Natl Acad Sci U S A*. 1983; 80:7099–7103. [PubMed: 6580629]
49. Spiriti J, Kamberaj H, de Graff AMR, Thorpe MF, van der Vaart A. *J Chem Theory Comput*. 2012; 8:2145–2156.
50. Ruscio JZ, Onufriev A. *Biophys J*. 2006; 91:4121–4132. [PubMed: 16891359]
51. Crick FH, Klug A. *Nature*. 1975; 255:530–533. [PubMed: 1095931]
52. Travers A. *Curr Biol*. 2005; 15:R377–379. [PubMed: 15916938]
53. Yan J, Marko JF. *Phys Rev Lett*. 2004; 93:108108. [PubMed: 15447460]
54. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. *J Comput Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
55. Foloppe N, MacKerell AD. *J Comput Chem*. 2000; 21:86–104.

56. MacKerell AD, Banavali NK. *J Comput Chem*. 2000; 21:105–120.
57. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
58. Miyamoto S, Kollman PA. *J Comput Chem*. 1992; 13:952–962.
59. Feig M, Karanicolas J, Brooks CL. *J Mol Graph Model*. 2004; 22:377–395. [PubMed: 15099834]
60. Allawi HT, SantaLucia J Jr. *Biochemistry*. 1998; 37:2170–2179. [PubMed: 9485363]
61. Lane AN, Peck B. *Eur J Biochem*. 1995; 230:1073–1087. [PubMed: 7601138]
62. Johnson SJ, Beese LS. *Cell*. 2004; 116:803–816. [PubMed: 15035983]
63. Skelly JV, Edwards KJ, Jenkins TC, Neidle S. *Proc Natl Acad Sci U S A*. 1993; 90:804–808. [PubMed: 8430089]
64. Boulard Y, Cognet JA, Fazakerley GV. *J Mol Biol*. 1997; 268:331–347. [PubMed: 9159474]
65. Arnold FH, Wolk S, Cruz P, Tinoco I Jr. *Biochemistry*. 1987; 26:4068–4075. [PubMed: 3651437]
66. Gervais V, Cognet JA, Le Bret M, Sowers LC, Fazakerley GV. *Eur J Biochem*. 1995; 228:279–290. [PubMed: 7705340]
67. Borden KL, Jenkins TC, Skelly JV, Brown T, Lane AN. *Biochemistry*. 1992; 31:5411–5422. [PubMed: 1606167]
68. Hunter WN, Brown T, Kennard O. *Nucleic Acids Res*. 1987; 15:6589–6606. [PubMed: 3627999]
69. Hunter WN, Brown T, Kneale G, Anand NN, Rabinovich D, Kennard O. *J Biol Chem*. 1987; 262:9962–9970. [PubMed: 3611072]
70. Roux B. *Computer Physics Communications*. 1995; 91:275–282.
71. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. *J Comput Chem*. 1992; 13:1011–1021.
72. Banavali NK, Roux B. *Journal of the American Chemical Society*. 2005; 127:6866–6876. [PubMed: 15869310]
73. Banavali NK, MacKerell AD. *Plos One*. 2009; 4.
74. Lu XJ, Olson WK. *Nucleic Acids Res*. 2003; 31:5108–5121. [PubMed: 12930962]
75. Humphrey W, Dalke A, Schulten K. *J Mol Graph Model*. 1996; 14:33–38.
76. Djuranovic D, Hartmann B. *J Biomol Struct Dyn*. 2003; 20:771–788. [PubMed: 12744707]
77. Law SM, Feig M. *Biophys J*. 2011; 101:2223–2231. [PubMed: 22067162]
78. Mukherjee S, Feig M. *Biophys J*. 2009; 96:L63–65. [PubMed: 19486659]
79. Mukherjee S, Law SM, Feig M. *Biophys J*. 2009; 96:1707–1720. [PubMed: 19254532]
80. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, Mackerell AD Jr. *J Chem Theory Comput*. 2012; 8:348–362. [PubMed: 22368531]
81. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE 3rd, Loughton CA, Orozco M. *Biophys J*. 2007; 92:3817–3829. [PubMed: 17351000]

Figure 1.

Structures of the base pairs studied. Hydrogen bond pairings known experimentally and maintained during initial structure preparation are shown for A. canonical matches, B. purine:pyrimidine, C. purine:purine, D. pyrimidine:pyrimidine mismatches.

Figure 2. Definition of the bending angle. Bending angle, ξ , is calculated between the center of masses of three blocks of the nucleotides: Block I : 2–5/26–29 (pink), Block II: 6–10/21–25 (orange) and Block III: 11–14/17–20 (blue). The angle increases from 90° to 180° with increments of 5° during the umbrella sampling simulations

Figure 3.

Free energy profiles associated with bending angles ξ . The solid lines corresponds to DNA duplexes containing canonical pairs: A:T (red), G:C (blue) and abasic site, X:X (grey). The dotted lines correspond to purine:pyrimidine pairs: A(+):C (dark green) and G:T (purple). The dashed lines correspond to the purine:purine mismatches: A:A (dark brown), G(*anti*):A(*anti*) (light brown), G(*anti*):G(*anti*) (red), short dash/long dash lines to G(*anti*):A(*syn*) (black) and G(*anti*):G(*syn*) (bright green). The dot-dashed lines correspond to pyrimidine:pyrimidine mismatches: C:C (indigo), C:T (magenta), and T:T (green). (a) and (s) corresponds to *anti* and *syn* glycosidic conformations of nucleotides. A+:C, G:A and G:G denotes A(+):C, G(*anti*):A(*anti*) and G(*anti*):G(*anti*) mismatches, respectively. Bending free energy profiles for alternative conformations are shown in Figure S11.

Figure 4.

Average groove widths (in Å) and curvatures (in °) observed for mismatches in three regimes of strong, moderate and weak bending. Averages are taken from the DNA conformations with bending angles, $\xi < 110^\circ$ for stronger bending (denoted by red circles), $110^\circ < \xi < 130^\circ$ for moderate bending (denoted by blue circles), and in the range $130^\circ < \xi < 180^\circ$ for weaker bending (denoted by black circles). The curvature value for X:X is not shown since the nucleobases of the central base pair are missing. (a) and (s) corresponds to *anti* and *syn* glycosidic conformations of nucleotides.

Figure 5.

Variations in average helicoidal parameters in three regimes of strong, moderate, and weak bending. Panel A shows the averages for base pair parameters: shear, stagger, opening, and propeller of central mismatch pair X8:Y23. Panel B shows the averages for base step parameters: rise, shift, slide, and twist for base pair steps C7X8/Y23G24 (−1/0 step) and X8C9/G22Y23 (0/+1 step). Averages are calculated from DNA conformations with bending angles, $\xi < 110^\circ$ for stronger bending (denoted by filled red triangles), in the range of 110° $\xi < 130^\circ$ for moderate bending (denoted by filled violet circles), and in the range of 130° $\xi \geq 180^\circ$ for weaker bending (denoted by filled blue squares). The crystal structure helicoidal values for bent DNA with G:T mismatch bound with MutS and MutSa proteins are indicated by black diamonds and orange inverted triangles, respectively; (a) and (s) corresponds to *anti* and *syn* glycosidic conformations of nucleotides.

Figure 6.

A. Sugar pucker conformations observed for X8:Y23 mismatch pair at all bending angles, ξ . B. BI/BII transitions for mismatched pair, X8:Y23 as observed by the variations in ϵ - ξ values at all the bending angles. Normalized probabilities of different pucker conformations for sugars of X8 and Y23 are shown as stacked histograms; (a) and (s) denote the *anti* and *syn* glycosidic conformations of nucleotides.

Figure 7.

Heat maps depicting the occupancies of hydrogen bonds at different bending angles, ξ . The range of the occupancy values is depicted as color gradient white to red, with dark red corresponding to the highest value and white as zero occupancy. The atom naming is similar to that shown in Figure 1. (a) and (s) denote the *anti* and *syn* glycosidic conformations of nucleotides.

Figure 8.

Variations in the van der Waals contributions, E_{VDW} , of the base-base stacking interactions for the central three base pair segment. Averages are calculated from DNA conformations with bending angles, $\xi < 110^\circ$ for stronger bending (denoted by filled red triangles), in the range of $110^\circ < \xi < 130^\circ$ for moderate bending (denoted by filled violet circles), and in the range of $130^\circ < \xi < 180^\circ$ for weaker bending (denoted by filled blue squares). (a) and (s) correspond to *anti* and *syn* glycosidic conformations of nucleotides.

Figure 9.

Superposition of crystal structure of MutS (PDBID: 1OH7) bound with DNA duplex consisting of G(*anti*):G(*syn*) mutation (DNA duplex shown in green, protein shown in white) and representative structure of G(*anti*):G(*anti*) with a bending angle around 125° (G8:G23 shown in magenta). The crystal structure interaction of the N7 atom (as blue sphere) with glutamate (in yellow) is shown as a dashed line. The crystal structure stacking interaction of G22 with Phe (in yellow) is also shown.

Table 1

Average free energy values associated with the bending angles ($110^\circ \leq \xi < 130^\circ$) for different sequences of central base pair, X:Y. Binding affinity energies were calculated from K_d values reported for binding of MutS protein to DNA duplexes containing mismatches in (24). (a) and (s) denote the *anti* and *syn* glycosidic conformations of nucleotides.

X8:Y23	Bending free energy (kcal/mol)	Binding affinity $kT \ln K_d$ (kcal/mol)
G:T	4.74	-0.98
G(a):G(a)	4.32	-0.28
G(a):G(s)	6.15	
A:A	3.59	0.06
T:T	5.74	0.06
C:T	4.13	0.16
A(+):C	3.36	0.73
A:C	5.04	
G(a):A(a)	4.95	0.93
G(a):A(s)	5.70	
C:C	3.74	1.14
C(+):C	3.94	
A:T	5.85	1.61
G:C	6.00	
X:X	2.84	--

Table 2

Sampling of DNA conformations within the moderate bending regime that are similar to those observed in the crystal structure. The root mean square deviations (RMSD) for the backbone and heavy atoms of the central seven base pair fragment of DNA conformations with bending angles within the range of $110^\circ \leq \xi < 130^\circ$ are calculated with respect to the DNA coordinates taken from the MutS crystal structure (PDBID: 2O8B). The lowest RMSD values observed for backbone atoms (R_{BB}) and heavy atoms (R_{HA}) are shown for each mismatch along with the percentage occurrences of conformations that have RMSD values lower than a certain cutoff limit. For backbone atoms (P_{BB}), an RMSD cutoff of 2.0 Å is used; for heavy atoms (P_{HA}), a cutoff of 4.2 Å is used. (a) and (s) corresponds to *anti* and *syn* glycosidic conformations of the nucleotides.

X8:Y23	R_{BB} (Å)	P_{BB} % (R_{BB} 2.0Å)	R_{HA} (Å)	P_{HA} % (R_{HA} 4.2Å)
A:T	1.55	3.62	4.09	1.23
G:C	1.50	5.00	3.98	3.25
A(+):C	1.20	10.72	3.96	9.83
G:T	1.37	4.40	4.00	3.24
A:A	1.06	43.98	3.89	48.49
G(a):A(a)	1.12	30.75	3.91	28.62
G(a):A(s)	1.63	3.33	4.05	1.75
G(a):G(a)	1.42	7.68	3.98	6.27
G(a):G(s)	1.33	11.36	3.95	7.37
C:C	1.24	9.5	3.90	9.15
C:T	1.30	13.91	3.92	10.24
T:T	1.22	13.95	3.93	15.34