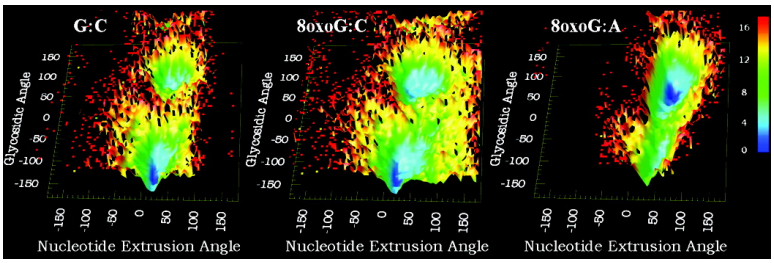
Article

Dynamic Behavior of DNA Base Pairs Containing 8-Oxoguanine

Xiaolin Cheng, Catherine Kelso, Viktor Hornak, Carlos de los Santos, Arthur P. Grollman, and Carlos Simmerling

J. Am. Chem. Soc., **2005**, 127 (40), 13906-13918• DOI: 10.1021/ja052542s • Publication Date (Web): 17 September 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 10 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 09/17/2005

Dynamic Behavior of DNA Base Pairs Containing 8-Oxoguanine

Xiaolin Cheng,[†] Catherine Kelso,[§,||] Viktor Hornak,[§] Carlos de los Santos,[‡] Arthur P. Grollman,[‡] and Carlos Simmerling*[,†,§]

Contribution from the Department of Chemistry, Department of Pharmacological Sciences, and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794, and Ward Melville High School, East Setauket, New York 11733

Received April 19, 2005; E-mail: carlos.simmerling@stonybrook.edu

Abstract: The process by which DNA repair enzymes recognize and selectively excise damaged bases in duplex DNA is fundamental to our mechanistic understanding of these critical biological reactions. 8-Oxoguanine (8-oxoG) is the most common form of oxidative DNA damage; unrepaired, this lesion generates a G:C→T:A mutation. Central to the recognition and repair of DNA damage is base extrusion, a process in which the damaged base lesion or, in some cases, its partner disengages from the helix and is bound to the enzyme's active site where base excision takes place. The conformation adopted by 8-oxoG in duplex DNA is affected by the base positioned opposite this lesion; conformational changes may also take place when the damaged base binds to its cognate repair enzyme. We performed unrestrained molecular dynamics simulations for several 13-mer DNA duplexes. Oligomers containing G:C and 8oxoG:C pairs adopted Watson−Crick geometries in stable B-form duplexes; 8oxoG showed increased local and global flexibility and a reduced barrier to base extrusion. Duplexes containing the G:A mismatch showed much larger structural fluctuations and failed to adopt a well-defined structure. For the 8oxoG:A mismatch that is recognized by the DNA glycosylase MutY, the damaged nucleoside underwent spontaneous and reproducible anti→syn transitions. The syn conformation is thermodynamically preferred. Steric hindrance and unfavorable electrostatics associated with the 8oxoG O8 atom in the anti conformation were the major driving forces for this transition. Transition events follow two qualitatively different pathways. The overall anti→syn transition rate and relative probability of the two transition paths were dependent on local sequence context. These simulations indicate that both the dynamic and equilibrium behavior of the duplex change as a result of oxidation; these differences may provide valuable new insight into the selective action of enzymes on damaged DNA.

Introduction

The structural biology of damaged DNA systems is of interest due to their altered stability and biological significance for DNA repair.[1,2] Of particular significance is the process by which DNA repair enzymes recognize and act preferentially on lesion-containing base pairs. 8-Oxoguanine (8-oxoG) is the most common form of oxidative DNA damage; in duplex DNA, 8-oxoG forms a stable base pair with dC. In mammalian cells, 8oxoG is recognized and excised by the DNA glycosylase, hOGG1,[3] while its counterpart, Fpg, is operative in bacteria.[4,5] While dG pairs predominantly with dC, employing a Watson−Crick hydrogen-bonding pattern, 8oxoG also can form a stable Hoogsteen pair with dA (Figure 1). dA may be incorporated opposite 8oxoG during DNA replication,[6,7] in which event hMYH (MutY in bacteria) initiates the repair process by excising dA. In the absence of these enzymes, replication of the adenosine-containing strand generates G:C→T:A transversions, potentially contributing to carcinogenesis and/or other age-related degenerative diseases[8,9]

A key step in the dynamic recognition of damaged DNA is nucleotide extrusion (nucleotide flipping), a general mechanism whereby a 2′-deoxynucleotide swings out of the DNA helix and

[†] Department of Chemistry, Stony Brook University.
[‡] Department of Pharmacological Sciences, Stony Brook University.
[§] Center for Structural Biology, Stony Brook University.
[||] Ward Melville High School.

(1) Burrows, C. J.; Muller, J. G. Oxidative Nucleobase Modifications Leading to Strand Scission. *Chem. Rev.* **1998**, *98* (3), 1109−1152.
(2) Stivers, J. T.; Jiang, Y. L. A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729−2760.
(3) Bruner, S. D.; Norman, D. P. G.; Verdine, G. L. Structural basis for recognition and repair of endogenous mutagen 8-oxoguanine in DNA. *Nature* **2000**, *403*, 859−866.
(4) Gilboa, R.; Zharkov, D. O.; Golan, G.; Fernandes, A. S.; Gerchman, S. E.; E. Matz, J. H. K.; Grollman, A. P.; Shoham, G. Structure of Formamidopyrimidine-DNA Glycosylase Covalently Complexed to DNA. *J. Biol. Chem.* **2002**, *277*, 19811−19816.
(5) Serre, L. P. d. J. K.; Boiteux, S.; Zelwer, C.; Castaing, B. Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. *EMBO J.* **2002**, *21* (12), 2854−65.
(6) Moriya, M. O. C.; Bodepudi, V.; Johnson, F.; Takeshita, M.; Grollman, A. P. Site-specific mutagenesis using a gapped duplex vector: a study of translesion synthesis past 8-oxodeoxyguanosine in *E. coli*. *Mutat Res.* **1991**, *254*, 281−8.
(7) Shibutani, S. B. V.; Johnson, F.; Grollman, A. P. Translesional synthesis on DNA templates containing 8-oxo-7,8-dihydrodeoxyadenosine. *Biochemistry* **1993**, *32*, 4615−21.
(8) Floyd, R. The role of 8-hydroxyguanine in carcinogenesis. *Carcinogenesis* **1990**, *11*, 1447−1450.
(9) Fraga, C.; Shienaga, M.; Park, J.; Degan, P.; Ames, B. Oxidative damage to DNA during aging: 8-hydroxy-2′-deoxyguanosine in rat organ DNA and urine. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4533−4537.

10.1021/ja052542s CCC: $30.25

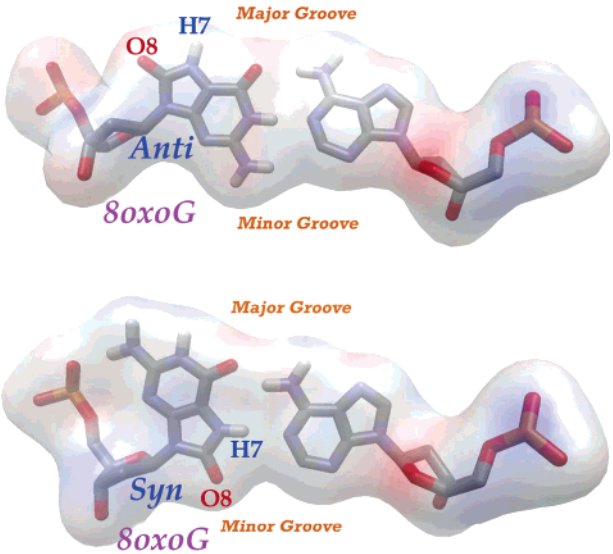


Figure 1. The upper image shows the 8oxoG:A pair with 8oxoG in the anti conformation, and the lower image shows 8oxoG in the syn conformation which allows formation of an 8oxoG(H7)−A(N1) hydrogen bond.

is accommodated in the catalytic pocket of the cognate repair enzyme. DNA N-glycosylases and DNA cytosine-5 methylases employ this mechanism for recognition of damaged residues or cytosine bases targeted for methylation, respectively.[10,11] The extrusion rate of uracil is accelerated in the presence of uracil DNA glycosylase, suggesting that the efficiency of this enzyme can be ascribed, in part, to extrusion.[12]

MutY is of particular interest in that the *undamaged* base is excised from the 8oxoG:A mismatch. Based on stopped-flow fluorescence measurements, it was suggested that MutY extrudes both the damaged 8oxoG and complementary adenosine residues with rates of 108 s^{-1} and 16 s^{-1}, respectively.[13] However, the subsequently published crystal structure of a DNA/MutY complex revealed that only the scissile dA residue extrudes into the catalytic pocket of the enzyme while 8oxoG remains intrahelical, adopting the anti conformation for its glycosidic torsion angle.[11] In contrast, 8-oxoG adopts the syn conformation in the 8oxoG:A pair in solution. The computational experiments that we report here explore the processes by which MutY results in a syn→anti transition and how rotation of the glycosidic bond of 8oxoG relates to extrusion of dA.

Nucleotide extrusion has not been observed in the absence of enzymes that participate in this process. In duplex DNA, hydrogen bonds transiently break, allowing base pairs to continuously open and close. Base pair lifetimes are reported to be in the 4−40 ms range; DNA structural features, such as A- or G-tracts, can profoundly affect base pair opening rates.[14,15] The degree of nucleotide extrusion during imino proton exchange is unknown; however, a small degree of base displacement toward the major groove of the helix may be sufficient to mediate this process. Base pair opening and base extrusion rates are of similar magnitudes, suggesting that both reactions share common initial structural intermediates.

Computational studies have provided insight into this process through calculations of the potential mean force (PMF) of base pair opening in free DNA duplexes and in duplexes bound to various proteins.[16−18] These studies suggest that pathways for extrusion through either the major or minor groove are feasible. Energy barriers for base extrusion from canonical DNA duplexes were calculated to be in the range 15−20 kcal/mol. When DNA binds to the closed form of the cytosine-5-methyltransferase, M.HhaI, the barrier was lowered by about 5 kcal/mol, resulting in preferential extrusion through the major groove.[19]

In this article, we report results obtained with multiple unrestrained MD simulations. Explicit and implicit solvent models were used to study the structures and dynamics of 13-mer DNA duplexes containing a single 8oxoG residue paired with dC or dA. All simulations were initiated from canonical B-DNA duplexes; well-converged simulations provided appropriate conformations for each system. Replica-exchange molecular dynamics[20,21] was used to obtain equilibrium distributions for comparison with ensemble-averaged experimental data, while multiple standard MD simulations were used to investigate the nonequilibrium behavior of selected systems. All simulations were unrestrained and, therefore, free of possible bias introduced through imposition of a particular reaction coordinate for conformational transitions.

The free energy landscape for the 13-mer containing 8oxoG:C showed the same global free energy minimum (Watson−Crick base pair geometry) as a control containing G:C at that position. However, the oxidation of dG results in a notable broadening of the free energy basins and a reduction in the free energy barrier for extrusion of 8oxoG. These differences may be critical to the specific recognition of the lesion for enzymatic action. Focusing on the mechanism by which MutY recognizes 8oxoG:A, we also studied the 8oxoG anti↔syn interconversion in this base pair. Results of our simulations clearly show the predominance of the Hoogsteen syn:anti 8oxoG:A, in contrast to the anti:anti form of the base pair populated in simulations for 8oxoG positioned opposite dC. In addition, two different anti→syn transition pathways are reproducibly observed in our MD simulations: a dominant path in which 8oxoG is extruded from the duplex prior to rotation around the glycosidic bond, and another in which rotation occurs inside the duplex after transient breaking of a flanking base pair. Both pathways are consistent with free energy landscapes derived from our equilibrium simulation data. The influence of base pairs flanking

(10) Roberts, R. J.; Cheng, X. Base flipping. *Annu. Rev. Biochem.* **1998**, *67*, 181−198.
(11) Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **2004**, *427*, 652−656.
(12) Stivers, J. T.; Pankiewicz, K. W.; Watanabe, K. A. Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry* **1999**, *38*, 952−963.
(13) Bernards, A. S.; Miller, J. K.; Bao, K. K.; I., W. Flipping duplex DNA inside out: a double base-flipping reaction mechanism by *Escherichia coli* MutY adenine glycosylase. *J. Biol. Chem.* **2002**, *277*, 20960−20964.
(14) Leroy, J.-L.; Charretier, E.; Kochoyan, M.; Gueron, M. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. *Biochemistry* **1988**, *27*, 8894−8898.
(15) Dornberger, U.; Leijon, M.; Fritzsche, H. High base pair opening rates in tracts of GC base pairs. *J. Biol. Chem.* **1999**, *274*, 6957−6962.
(16) Varnai, P.; Lavery, R. Base Flipping in DNA: Pathways and Energetics Studied with Molecular Dynamic Simulations *J. Am. Chem. Soc.* **2002**, *124* (25), 7272−7273.
(17) Giudice, E.; Varnai, P.; Lavery, R. Energetic and conformational aspects of A:T base-pair opening within the DNA double helix. *ChemPhysChem* **2001**, *2* (11), 673ff.
(18) Giudice, E.; Lavery, R. *J. Am. Chem. Soc.* **2003**, *125*, 4998−4999.
(19) Huang, N.; Banavali, N. K.; MacKerell, A. D. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (1), 68−73.
(20) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281* (1−3), 140−150.
(21) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.

8oxoG:A was studied by varying the flanking sequence and repeating the transition simulations.

Methods

DNA Sequences. The systems studied consist of 13-mer duplexes 5′-d(CCAGGAXGAAGCC)-3′ and 5′-d(GGCTTCYTCCTGG)-3′, with four specific X:Y base pairs. The first, with X = 8oxoG and Y = C, has the form shown by Grollman et al. to be cleaved efficiently by the DNA glycosylase Fpg.[4] In the remainder of this article, we refer to this system as 8oxoG:C. Replacement of the cytosine opposite to 8oxoG with adenosine gives the 8oxoG:A duplex, which is a substrate of MutY and will be the focus of most of the simulations presented here. Two other duplexes, abbreviated as G:C and G:A, respectively, are used as controls. The four sequences used in our simulations are listed as follows, with bold text indicating the bases varied and G* indicating 8oxoG:

G:C	(CCAGGAGGAAGCC)	·	(GGCTTC**C**TCCTGG)
G:A	(CCAGGAGGAAGCC)	·	(GGCTTC**A**TCCTGG)
8oxoG:C	(CCAGGAG*GAAGCC)	·	(GGCTTC**C**TCCTGG)
8oxoG:A	(CCAGGAG*GAAGCC)	·	(GGCTTC**A**TCCTGG)

To study the impact of the flanking sequence on the 8oxoG:A DNA structure and dynamics, three additional duplexes, referred to as AG*A, GG*A, and GG*G, were simulated. These sequences are shown below.

AG*A (CCAGG**A**G***A**AAGCC)·(GGCTT**T**A**T**CCTGG)

GG*A (CCAGG**G**G***A**AAGCC)·(GGCTT**T**A**C**CCTGG)

GG*G (CCAGG**G**G***G**AAGCC)·(GGCTT**C**A**C**CCTGG)

Molecular Dynamics. All molecular dynamics simulations were carried out with the SANDER module in Amber (version 8)[22] using the ff94 force field.[23] The time step was 2 fs, and all bonds involving hydrogen were constrained with the SHAKE[24] algorithm with a tolerance of 10^{-4} Å.[24] All possible nonbonded interactions were fully evaluated every time step. Parameters for 8oxoG were obtained from Miller et al.[25] Canonical B-form DNA structures were constructed with the NUCGEN module. The starting structure for DNA containing 8oxoG was obtained from canonical B-form DNA by replacing the hydrogen at C8 with oxygen and adding hydrogen to N7.

Generalized Born Solvent Simulations. The Generalized Born (GB) model is used extensively in this study because it provides improved conformational sampling as compared to simulations with explicit representation of solvent molecules. This arises both from the longer time scales accessible to simulation through reduced cost of calculating forces, as well as an acceleration of transitions due to the lack of solvent viscosity. The reduced computational cost enables us to obtain statistically significant, well-converged data for these systems. Likewise, the reduced cost of GB simulations has enabled comparison of absolute folding rates[26] for small proteins at a level that remains intractable with explicit representation of solvent molecules.

However, GB methods are approximate, with drawbacks[27] arising from both the lack of solvent molecules of finite size and the ad hoc nature of the GB equations as compared to more rigorous continuum descriptions such as those based on the Poisson−Boltzmann formalism. Experiments and simulation both have indicated[28,29] that DNA is associated with structured water molecules that have relatively long residence times.[30] However, previous nucleic acid studies have shown that GB provides reasonable results[27,31−34] and also converges to experimentally observed conformations of DNA and RNA more rapidly than similar simulations in explicit solvent.[31,32] Biomolecular simulations continue to require a compromise between the accuracy/expense of the energy function employed and the precision of the results that they provide, particularly when samplings of conformations distant from the initial structure are required.

The Born radii were adopted from Bondi with modification and an offset of 0.13 Å.[31] The scaling factors for Born radii were taken from the Tinker modeling package.[35,36] Each GB simulation consisted of two stages, equilibration and production. During equilibration, the starting structures built from canonical B-form DNA were minimized for 2000 steps to remove close contacts. The system was then gradually heated to 320 K at 40 K intervals during eight simulations of 300 ps. We note that the GB simulations seemed more sensitive to careful equilibration as compared to those performed with explicit solvent molecules. After the equilibration, coordinates were archived every 2 ps during the production phase. Each production run was 15 ns in length.

Explicit Solvent Simulations. For the explicit solvent simulations, the canonical B-form DNA structures were immersed into TIP3P[37] water in a truncated octahedron periodic box, with a minimum initial distance of 10 Å between the solute and the box boundary. This resulted in a total of ∼6400 water molecules, with the exact number varying slightly among different DNA systems. Twenty-four sodium ions were added to neutralize the net charge on the 13-mer. The solvated system was then equilibrated in three steps: (i) 5000 steps of MD simulations of water and ions (with DNA frozen), (ii) five cycles of 500-step minimizations of the system with decreasing positional restraints on the DNA, and (iii) four cycles of 5000-step MD simulations of the system, with decreasing positional restraints on the DNA and no restraints during the last step. Simulations were carried out in the NPT ensemble at 320 K. A short-range cutoff of 9 Å was used for nonbonded interactions, and long-range electrostatic interactions were treated with a particle-mesh Ewald PME method.[38,39] All other parameters were the same as those used with the GB model.

Replica Exchange and Local Replica Exchange Simulations. Replica Exchange Molecular Dynamics (REMD) has recently been widely used to study the structure and folding landscapes of peptides

(22) Case, D. A. et al. *AMBER 8*; University of California: San Francisco, 2004.
(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field For the Simulation Of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197.
(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints − Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(25) Miller, J. H.; Chiang, C. P.; Straatsma, T. P.; Kennedy, M. A. 8-Oxoguanine Enhances Bending of DNA that Favors Binding to Glycosylases. *J. Am. Chem. Soc.* **2003**, *125*, 6331−6336.
(26) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420* (6911), 102−106.
(27) Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 243−252.
(28) Drew, H. R.; Dickerson, R. E. Structure of a B-DNA Dodecamer. 3. Geometry of Hydration. *J. Mol. Biol.* **1981**, *151* (3), 535−556.
(29) Subramanian, P. S.; Ravishanker, G.; Beveridge, D. L. Theoretical Considerations on the Spine of Hydration in the Minor Groove of D(Cgcgaattcgcg)·D(Gcgcttaagcgc) − Monte Carlo Computer Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85* (6), 1836−1840.
(30) Liepinsh, E.; Otting, G.; Wuthrich, K. NMR Observation of Individual Molecules of Hydration Water Bound to DNA Duplexes − Direct Evidence for a Spine of Hydration Water Present in Aqueous Solution. *Nucleic Acids Res.* **1992**, *20* (24), 6549−6553.
(31) Tsui, V.; Case, D. A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **2000**, *122* (11), 2489−2498.
(32) Williams, D. J.; Hall, K. B. Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys. J.* **1999**, *76* (6), 3192−3205.
(33) Cui, G. L.; Simmerling, C. Conformational heterogeneity observed in simulations of a pyrene-substituted DNA. *J. Am. Chem. Soc.* **2002**, *124* (41), 12154−12164.
(34) Tsui, V.; Case, D. A. Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. *J. Phys. Chem. B* **2001**, *105* (45), 11314−11325.
(35) Ponder, J. W. *TINKER: Software Tools for Molecular Design*, 3.9; St. Louis, MO, 2001.
(36) Ponder, J. W.; Richards, F. M. An Efficient Newton-Like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem.* **1987**, *8* (7), 1016−1024.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

and small proteins. As we describe below, application of REMD to large systems becomes prohibitively expensive in terms of computational requirements. On the other hand, the enhanced sampling is frequently only needed for a subset of atoms. For these cases, we recently derived a local replica exchange method (LREMD).[40] While full details of the LREMD method have recently been provided elsewhere,[40] we summarize here the essential similarities and differences between REMD and LREMD.

Standard REMD employs replicas of the entire system, with each replica realized as a fully independent simulation coupled to a bath at a different temperature. Importantly, these temperatures span a range from low temperatures of interest (such as 280 K or 300 K) up to high temperatures (such as 600 K) at which the system can rapidly overcome potential energy barriers that would otherwise impede conformational transitions on the time scale simulated.

Periodically, the potential energies of structures being sampled at neighboring temperatures are compared. A Metropolis-type Monte Carlo procedure is used to determine the probability that each of the structures would be sampled at the other temperature. If successful, the temperatures of the replicas are exchanged. In this manner, simulations at low temperature have less tendency to become kinetically trapped through the ability to "jump" directly to alternate minima being sampled at higher temperatures. Likewise, the structures sampled at high temperatures can anneal by being transferred to successively lower temperatures. The net result is more rapid convergence to Boltzmann-weighted ensembles of structures at each of the simulated temperatures. For large systems, however, REMD becomes intractable since the number of replicas needed to span a given temperature range increases with the square root of the number of degrees of freedom in the system.[41,42]

This approach is desirable when enhanced sampling is needed for the entire system and the system is small enough that the required number of replicas can be accommodated with available computational resources. In some cases, however, enhanced sampling is only needed for a small subset of atoms, such as several base pairs out of many in a DNA duplex (as is the case in the present study). In addition, increasing sampling of regions outside the area of interest by using standard REMD may actually be counterproductive. In the present case, loss of the duplex would be detrimental to obtaining conformational data for the mismatch region, but this would likely occur for replicas at higher temperatures unless restraints were employed. Imposing restraints, however, may have other unintended consequences on motion in the region of interest.

Our LREMD method addresses this need by only replicating those atoms for which the enhanced sampling is desired. LREMD is an adaptation of REMD to Locally Enhanced Sampling (LES),[43−45] a mean-field simulation method in which a subset of the system is replicated and interacts in an average way with the rest of this system. LES provides multiple trajectories of the region of interest at a similar computational cost as compared to a single standard simulation. We and others have previously shown that LES provides excellent results when applied to nucleic acid systems, including those with unusual structures or modified bases.[33,45−47]

The combination of LES and LREMD is straightforward. Whereas in standard REMD the "replicas" consist of a set of fully independent simulations, in LREMD one replicates only a subset of atoms, and each copy is coupled to a separate temperature bath. The remaining atoms are maintained at a single temperature and feel the average force of the copies. Thus LREMD is appropriate only when the conformational changes in the region of interest are expected to be rather weakly coupled to the remainder of the system. This can be adjusted by increasing the size of the region of "interest"; for example, in the present case the base pairs neighboring the mismatch are also included in the LREMD copies since their stacking interactions and covalent link through the backbone likely couples them to the local conformation at the mismatch. Since we do not apply restraints to the nonreplicated atoms, they are free to move in response to the average behavior of the replicated region.

Since the number of degrees of freedom in the replicated sub-system in LREMD is smaller than the full system, acceptable exchange rates can be obtained with fewer replicas than with traditional REMD. It is also possible to obtain well-converged conformations at a range of temperatures using only a single simulation, rather than the multiple simulations needed for standard REMD. However, we note that the LES method is a mean-field approximation and results in potential drawbacks in addition to these advantages. In particular, the averaging of forces between the LES copies and the noncopied region results in a smoother energy landscape, with reduced barriers to conformational transitions. For this reason, the properties of the free energy landscape sampled by LES or LREMD do not exactly match those of the original system. This reduction in barriers and an accompanying scaling of the relative energies of local minima result in the LES copies having behavior that corresponds to a higher effective temperature.[48,49] We previously reported that three LES copies of an RNA hairpin needed to be maintained at a lower temperature (150 K−200 K) in order to adopt a well-defined structure in standard molecular dynamics using LES.[46] Choosing an appropriate temperature (often through trial and error) has thus presented an additional obstacle to the use of LES. The ability of LREMD to provide ensembles at a range of temperatures provides at least a partial solution to this challenge.

We showed that the REMD and LREMD free energy landscapes for the RNA hairpin showed high qualitative similarity and identical global free energy minima.[38] In the present study, we employ LREMD to locate the global free energy minimum for several mismatches and to map and compare the sequence dependence of the general features of the landscape. Free energy landscapes were calculated from the structure ensemble sampled during LREMD simulations, but detailed transition pathways are mapped using standard molecular dynamics.

The free energy was calculated as $F = -k_B T ln P(\theta, \chi)$, where $P(\theta, \chi)$ is the probability distribution function for each pair of θ (8oxoG nucleotide extrusion angle) and χ (8oxoG glycosidic angle) values. As described above, a limitation to this approach is that since the LES

(38) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald − an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(39) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular Dynamics Simulations On Solvated Biomolecular Systems − the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* **1995**, *117* (14), 4193−4194.
(40) Cheng, X. L.; Cui, G. L.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **2005**, *109* (16), 8220−8230.
(41) Fukunishi, H. W. O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058−9067.
(42) Kofke, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **2002**, *117* (15), 6911−6914.
(43) Elber, R.; Karplus, M. Enhanced Sampling in Molecular Dynamics − Use of the Time-Dependent Hartree Approximation For a Simulation of Carbon Monoxide Diffusion Through Myoglobin. *J. Am. Chem. Soc.* **1990**, *112* (25), 9161−9175.
(44) Roitberg, A.; Elber, R. Modeling Side Chains in Peptides and Proteins − Application of the Locally Enhanced Sampling and the Simulated Annealing Methods to Find Minimum Energy Conformations. *J. Chem. Phys.* **1991**, *95* (12), 9277−9287.
(45) Simmerling, C.; Miller, J. L.; Kollman, P. A. Combined Locally Enhanced Sampling and Particle Mesh Ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid. *J. Am. Chem. Soc.* **1998**, *120* (29), 7149−7155.
(46) Cheng, X. L.; Hornak, V.; Simmerling, C. Improved conformational sampling through an efficient combination of mean-field simulation approaches. *J. Phys. Chem. B* **2004**, *108* (1), 426−437.
(47) Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E., III; Sponer, J. Molecular dynamics simulations of Guanine quadruplex loops: advances and force field limitations. *Biophys. J.* **2004**, *87* (1), 227−42.
(48) Straub, J. E.; Karplus, M. Energy Equipartitioning in the Classical Time-Dependent Hartree Approximation. *J. Chem. Phys.* **1991**, *94* (10), 6737−6739.
(49) Ulitsky, A.; Elber, R. The Thermal Equilibrium Aspects of the Time Dependent Hartree and the Locally Enhanced Sampling Approximations − Formal Properties, a Correction, and Computational Examples For Rare Gas Clusters. *J. Chem. Phys.* **1993**, *98* (4), 3380−3388.

Hamiltonian is scaled as compared to the original one, the relative free energy values obtained will not directly match those of the original system. However, we estimate values based on the relationship that relative LES energies are approximately reduced[44] by a factor of $1/N$ (where N is the number of LREMD replicas used for simulation). Therefore, we scaled our free energy values by a factor of 5 for the five LREMD replicas, though we note that the features of the landscapes are not affected by this scaling factor. For these reasons, we used the free energy surfaces for comparison of LREMD data for different sequences, or for alternate paths on a landscape, but the free energy values themselves should be used for qualitative analysis only. This is consistent with the approximations already inherent in LES and GB.

The ADDLES module of Amber was employed to construct the LREMD system. The two bases at the mismatch site were replaced by five LES replicas coupled to thermostats at 80 K, 93 K, 108 K, 125 K, and 145 K, with an anticipated overall 15% acceptance ratio of exchange of temperatures between replicas. This temperature range was found to be successful[40] for RNA hairpin LREMD simulations with five LES copies of the loop region. Exchange was attempted every 1 ps, and simulations were run until 8000 exchange attempts were performed (8 ns) for each of the four sequences. Coordinates were archived after each exchange attempt, and further thermodynamic analysis was performed on the ensemble of structures at a given temperature that were sampled by any of the replicas.

All LREMD simulations were performed using GB solvation. We previously reported[46] the derivation of modified GB equations that provide solvation energies and forces for LES copies that are identical to those obtained using standard GB solvation of the isolated conformations of the individual copies. Since the calculation of the energies for the LREMD copies employs the LES routines in Amber, the combination of LREMD with GB involves no additional approximations beyond those already present in GB or LES.

Data Analysis. Root-mean-square deviations (RMSD) were calculated on the basis of all heavy atoms of the central 11 base pairs, since the terminal base pairs on each end showed fluctuations and fraying during the simulations. Groove widths were calculated as the distance between the closest phosphorus pairs across the minor groove (P_i and P_{j+3} for base pair i and j), with 5.8 Å subtracted for the radii of the P atoms. The nucleotide extrusion angle was defined following MacKerell et al.[19] using the pseudo dihedral angle of the mass centers of four bases. Local bending near the mismatch was calculated with the ANGLE option in the program CURVES[50] as the angle between the local helical axis segments of A6/G8. This provided information on local rather than global helical bending. Purine glycosidic angles were calculated using the O4′−C1′−N9−C4 atoms. The geometric criteria for hydrogen bonds were hydrogen−acceptor distances less than 2.5 Å and donor−hydrogen−acceptor angles of 120° to 180°. Base pairs were considered to be broken when the hydrogen bond distances were 3.5 Å (the larger cutoff reduced the effect of transient fluctuations).

Decomposition of the various contributions to stability of alternate conformations were estimated by separation of the total potential energy into three individual groups: (1) the base pair interaction energy, which includes all nonbonded interactions between 8oxoG and its adenosine partner; (2) the stacking energy, which consists of nonbonded interactions between 8oxoG and its two flanking bases, adenosine and cytosine; and (3) an 8oxoG steric/electrostatic term for the O8 oxygen atom, involving all interactions with the nearby backbone oxygen and phosphorus atoms.

Results and Discussion

Simulations of G:C and 8oxo:C. As controls, both G:C and 8oxoG:C are stable, showing no major conformational changes

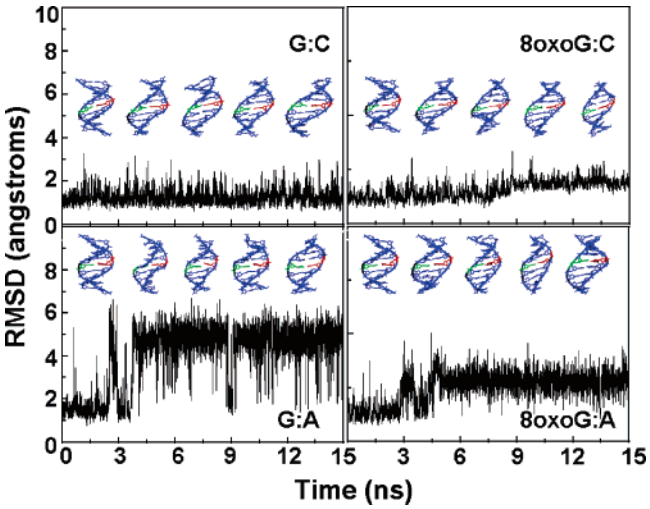


Figure 2. The heavy-atom RMSD (compared to the initial canonical B-form DNA structure) for four DNA sequences during the first 15 ns of MD simulation. Representative structures at the various points in time are shown above the RMSD values. The G or G* is shown in red, and its partner C or A is shown in green. The terminal base pairs show fraying in all sequences and are not shown to improve image clarity.

(Figure 2). As monitored by the base pair hydrogen bond distances (Figure 3), no evidence for opening of the G:C or 8oxoG:C base pair is found in either simulation. Neither do we find any significant change for minor groove width, helical bending, or G/8oxoG glycosidic angles (Figures 3 and 4). These observations are consistent with the results reported previously for MD simulation with GB for a standard DNA sequence[31,33] and for an 8oxoG:C containing sequence in explicit solvent.[25]

Simulations of G:A. Small differences in CD spectra suggest that the G:A mismatch does not alter the global duplex conformation, but local variation in minor groove width, propeller twist, and ϵ and ξ angles are evident from NMR and X-ray studies.[51,52] The G:A duplex also exhibits ∼12 °C reduction in T_m compared to G:C and 8−9 °C reduction compared to 8oxoG:C.[53] Consistent with these data, the simulation of the G:A sequence shows greater variability than G:C and 8oxoG:C (Figure 2) and has multiple periods of large change where the DNA bending increases to nearly 60° at the mismatch (Figure 4). Pairing of G with A also causes a remarkable disruption of the duplex, in agreement with previous reports of reduced thermodynamic stability for sequences with mismatched bases.[54,55] The register of normal base pairs near the mismatch is unstable during the simulation, sampling alternate purine−pyrimidine pairs. Two consecutive base pairs 3′ to G:A were

(50) Lavery, R.; Sklenar, H. Defining the Structure of Irregular Nucleic-Acids − Conventions and Principles. *J. Biomol. Struct. Dyn.* **1989**, *6* (4), 655−667.

(51) Gao, X.; Patel, D. J. G(syn).A(anti) Mismatch Formation in DNA Dodecamers at Acidic pH: pH-Dependent Conformational Transition of G:A Mispairs Detected by Proton NMR. *J. Am. Chem. Soc.* **1988**, *110*, 5178−5182.

(52) Prive, G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.; Kopka, M.; Dickerson, R. Helix Geometry, Hydration, and G:A Mismatch in a B-DNA Decamer. *Science* **1987**, *238*, 498−503.

(53) Plum, G. E.; Grollman, A. P.; Johnson, F.; Breslauer, K. J. Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex. *Biochemistry* **1995**, *34* (49), 16148−60.

(54) Barbara, L.; Gaffney, R. A. J. Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O6-methylguanine with reference both to Watson−Crick pairs and to mismatched pairs. *Biochemistry* **1989**, *28* (14), 5881−5889.

(55) Aboul-ela, F.; Koh, D.; Tinoco, I.; Martin, F. H. Base−base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A, C, G, T). *Nucleic Acids Res.* **1985**, *13*, 4811−4824.

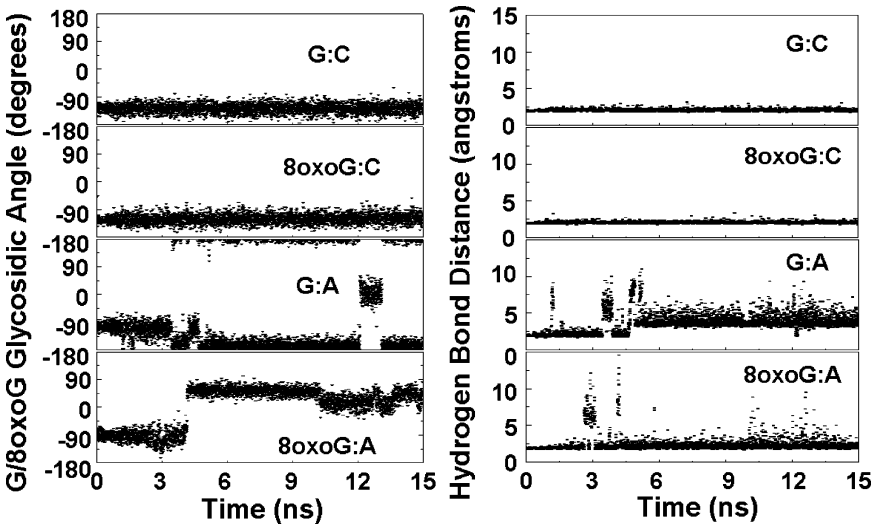


Figure 3. G7 or 8oxoG glycosidic torsions (left) and hydrogen bond distance (G7:H1−C20:N3 in G*:C and G:C; G*7:H1−A20:N1 in G*:A and G:A) (right) during the simulations. The G:A base pair was lost and never regained due to large buckle and twist. The 8oxoG:A base pair was lost and then re-established after a anti→syn transition for 8oxoG:A.

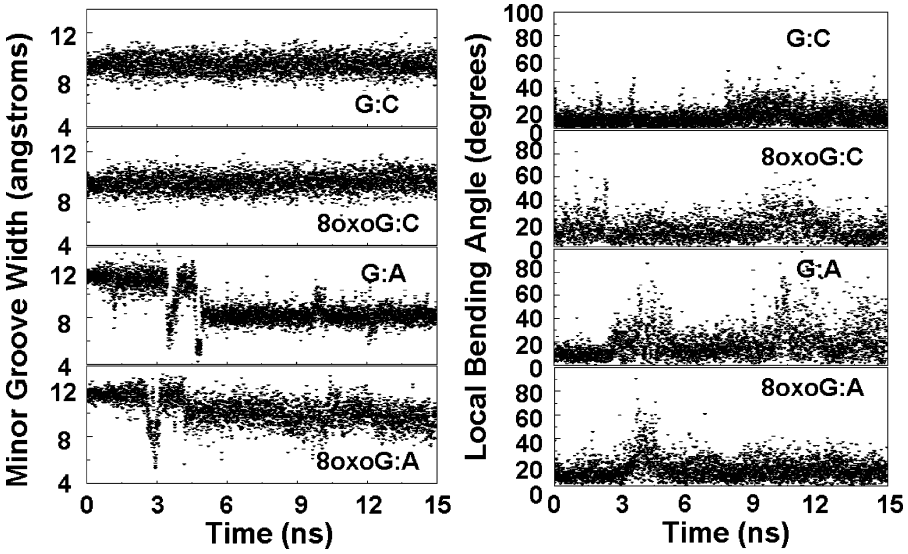


Figure 4. Minor groove width (left) and local bending at the mismatch (right) as function of time during the simulations. G:C and 8oxoG:C are stable. The initially wider groove in both purine:purine pairs narrows by 4−5 Å.

transiently lost at about 5.4 ns, resulting in three new base pairs (G7:T21, A6:C22, and G5:C23) before returning to the previous pattern after ∼4 ns.

The glycosidic angles of the G:A mismatch were previously shown by NMR and X-ray to adopt different conformations in different sequence context and solution conditions, such as G(anti):A(anti),[56] G(syn):A(anti),[51] and G(anti):A(syn),[57] with the G(syn):A(anti) conformation dominating at low pH.[51] Our simulation started with G(anti):A(anti) and converted to G(anti):A(syn) at ∼10 ns. Figure 4 also shows that this sequence increases local bending at the mismatch (27 ± 17° after 10 ns). The G:A pair also adopted an ∼30° propeller twist, while maintaining two hydrogen bonds. Average values for backbone torsions (−175° for ϵ and −100° for ξ) deviated significantly from values in canonical B DNA (−133° and −157°, respectively).

Simulations of 8oxoG:A. As with our other sequences, the 8oxoG:A simulation started with the anti:anti conformation (Figure 1a). During the first several nanoseconds of simulation, a stable 8oxoG(anti)•A(anti) alignment was maintained. At ∼4 ns, we observed a spontaneous opening of the 8oxoG:A base pair, extrusion of the 8oxoG base into the major groove, rotation of 180° around the glycosidic angle, and subsequent re-establishment of the base pair, resulting in an anti → syn transition for 8oxoG (Figure 1b). This opening into the major groove is in agreement with potential of mean force for base pair opening reported[58] by Giudice et al., in which guanine shows a 7−9 kcal/mol lower barrier for opening into the major rather than minor groove.

(56) Patel, D.; Kozlowski, S.; Ikuta, S.; Itakura, K. Deoxyguanosine-deoxyadenosine pairing in the d(C-G-A-G-A-A-T-T-C-G-C-G) duplex: conformation and dynamics at and adjacent to the dG X dA mismatch site. *Biochemistry* **1984**, *23* (14), 3207−17.

(57) Hunter, W.; Brown, T.; Kennard, O. Structural features and hydration of d(C-G-C-G-A-A-T-T-A-G-C-G); a double helix containing two G.A mispairs. *J. Biomol. Struct. Dyn.* **1986**, *4* (2), 173−91.

(58) Giudice, E.; Varnai, P.; Lavery, R. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* **2003**, *31* (5), 1434−43.

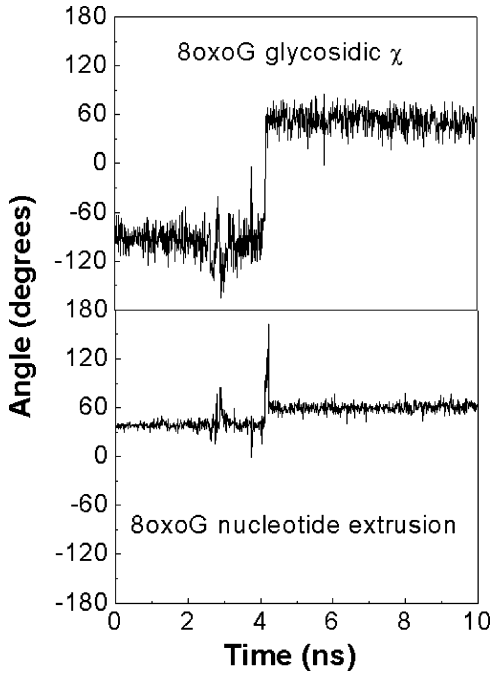


Figure 5. 8oxoG glycosidic angle (top) and nucleotide extrusion angle (bottom) as a function of time during 8oxoG anti→syn transition in 8oxoG:A.

The time dependences of the 8oxoG glycosidic angle and nucleotide extrusion angle during the transition are shown in Figure 5. Interestingly, rotations of the base about the glycosidic angle and base pair opening occur simultaneously. The implication of this coupling is that the rate of glycosidic angle rotation can thus become the same magnitude as that of nucleotide extrusion. Without nucleotide extrusion, the barrier for glycosidic rotation is expected to be much higher due to the closely packed environment of the stacked base pairs. However, we will revisit this issue below. Significant but transient changes were also observed during the transition for 8oxoG backbone torsions ϵ and ξ (data not shown). These motions were not a consequence of the transition; rather they caused compression of the helix at the anti orientation, perhaps facilitating the extrusion process. X-ray structures of damaged DNA bound to enzymes such as MutS,[59] MutY,[11] and Fpg[4] have ϵ and ξ backbone torsion and bending parameters well outside the range of values observed for standard DNA, consistent with our simulation data and suggesting that these experimental observations may not solely be due to formation of a complex with the enzyme.

Energy Analysis. As described above, 8oxoG:A is known to adopt a stable conformation, in contrast to G:A. In an anti:anti conformation of the base pair, both strands would have considerable backbone strain due to widening of the major groove and increased cross-strand distance needed to accommodate the purine:purine mismatch. Both sequences underwent larger initial fluctuations in groove width and bending than those seen for the purine:pyrimidine pairs. 8oxoG adopts a syn conformation, with two stabilizing hydrogen bonds and a narrower groove. The G in G:A also transiently sampled a syn conformation (Figure 3) but cannot form one of the Hoogsteen

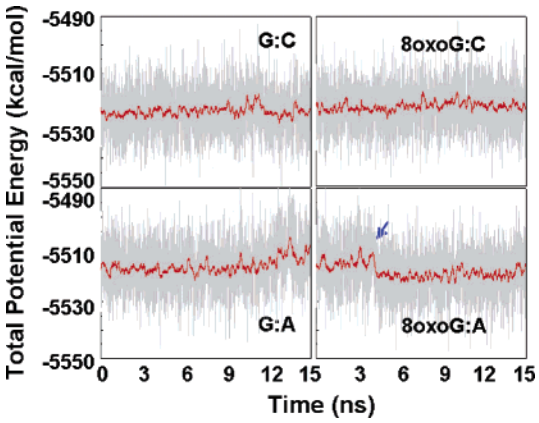


Figure 6. Potential energy (including solvation free energy) as a function of time during MD. The red line is a running average over 10ps. The blue arrow indicates the point of the anti→syn transition in 8oxoG:A. A drop in potential energy of ~5 kcal/mol is seen, compared to the relatively flat profile seen for other sequences.

hydrogen bonds due to lack of the H7 atom. Another factor in the preferred conformation may be due to the change in orientation of the dipole of the purine ring in guanosine vs 8oxoG. This will potentially alter the strength of the 8oxoG stacking interaction, affecting the relative stability of anti and syn in each case. Additionally, bulky substitutions at the C8 position in purines (Br, Cl, F) have been shown to increase the percentage of the syn rotamer at the mononucleotide level.[60] This observation has been interpreted as a result of steric clash between the C8 substituent and the phosphate backbone and would be expected to favor the syn conformation of 8oxoG.

To investigate the relative importance of these factors (base pairing, dipole stacking, and C8 substitution), we examined potential energies during the simulation of each sequence (Figure 6). For 8oxoG:A, there is a reduction in potential energy (which includes solvation free energy) of ~5−6 kcal/mol after the transition, while, for G:A, the average value remains relatively constant, suggesting an inability to locate a conformation that reduces the initial strain of the purine:purine pair. We next defined three energy components (Figure 7) that approximate the effects described above: (1) the base pair interaction energy, which includes the nonbonded interactions between 8oxoG and its adenosine partner; (2) the stacking energy, which consists of nonbonded interactions between 8oxoG and its two flanking bases, adenosine and cytosine; and (3) an 8oxoG steric/electrostatic term for the O8 oxygen atom, involving all interactions with the nearby backbone oxygen and phosphorus atoms. The time evolution of these components during the transition in 8oxoG:A is shown in Figure 7, and average values before and after transition are provided in Table 1. The peak near 4 ns shows the transition point where the 8oxoG is extruded from the duplex, with a loss of both stacking and base pairing energies.

All three components favor the syn conformation, but to different degrees. The component that makes the largest contribution (~2.7 of 5.1 kcal/mol overall) is the nonbonded clash of the 8oxoG O8 atom in the anti conformation, predominantly with the ribose O4′. Base pair interactions are ~1.3 kcal/mol more favorable for the syn conformation. This

(59) Lamers, M.; Perrakis, A.; Enzlin, J.; Winterwerp, H.; Wind, N. d.; Sixma, T. The crystal structure of DNA mismatch repair protein MutS binding to a GT mismatch. *Nature* **2000**, *407* (6805), 711−7.

(60) Uesugi, S.; Ikehara, M. Carbon-13 magnetic resonance spectra of 8-substituted purine nucleosides. Characteristic shifts for the syn conformation. *J. Am. Chem. Soc.* **1977**, *99* (10), 3250−3.

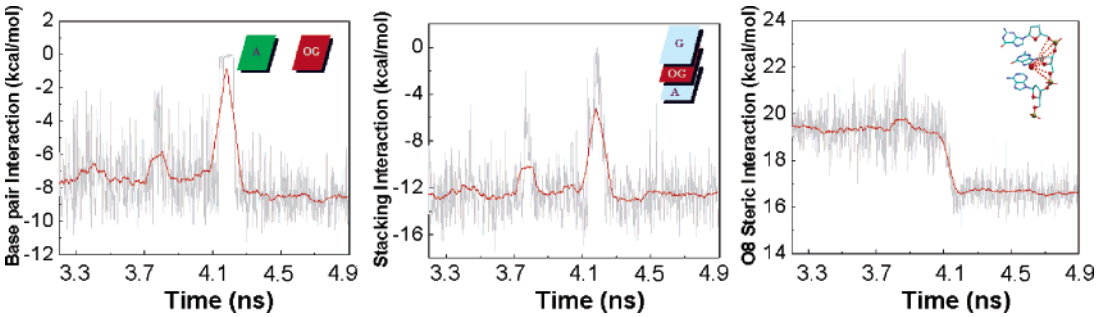


Figure 7. Energy decomposition analysis for the 8oxoG anti→syn transition, which shows potential energies (including solvation free energy) as a function of time near the transition (which occurred at ∼4.1 ns). (Left) Base pair interaction energy of 8oxoG with the partner A20; (middle) stacking energy of 8oxoG with the flanking A6 and G8; (right) 8oxoG O8 steric interactions with backbone oxygen and phosphorus atoms. Energy differences are discussed in the text.

Table 1. Energy Decomposition (in kcal/mol) during 8oxoG Anti → Syn Transition Simulation[a]

	stacking	hydrogen bond	o8 clash	total energy
anti	−11.9 ± 2.5	−7.2 ± 1.8	19.5 ± 0.9	80.6 ± 1.8
syn	−12.4 ± 1.7	−8.5 ± 1.2	16.8 ± 0.8	75.5 ± 1.5
difference	−0.5	−1.3	−2.7	−5.1

[a] Values shown are average energies and include solvation free energy; uncertainties reflect the standard deviation during simulation.

difference arises from a disparity in hydrogen bond strength in the anti and syn conformations. Although 8oxoG can form two hydrogen bonds in both conformations, a notable difference in the population of these hydrogen bonds is present (60% for anti and 90% for syn), due to nonideal geometry in the strained anti:anti conformation of this mismatch. Both anti and syn structures have similar stacking energies (∼0.5 kcal/mol difference).

Simulations of 8oxoG:A in Explicit Solvent. Even though the GB model has been shown in many cases to provide a reasonable model for probing the structure and dynamics of DNA, simulations have also shown that the presence of water and/or ions in both grooves can be strongly coupled to DNA bending and minor groove width.[61] On the other hand, motions in implicit solvent can occur more rapidly due to absence of solvent viscosity, providing the ability to model events that occur on time scales currently inaccessible in explicit solvent. To investigate the role of the solvent model in our observations, we carried out two additional ∼15ns simulations of 8oxoG:A in explicit water. One simulation was initiated from the same canonical B-form DNA as the GB simulation (anti:anti), while the other used a snapshot extracted from our GB simulation after the 8oxoG anti→syn transition. The syn simulation was stable, with RMSD values fluctuating in the range 1−3 Å. In contrast, the anti simulation showed significantly increased RMSD values (2−5 Å). No evidence for significant glycosidic bond rotation or base opening is seen in either simulation (Figure 8). Rather, the gradual increase of the 8oxoG glycosidic torsion seen at the tail of the anti plot results from increased 8oxoG:A buckle (κ), with the 8oxoG base bending about 15° toward the 5′ flanking AT pair. This type of structural change was also observed in several of our GB simulations (described below), but in explicit solvent the change was less dramatic and occurred on a slower time scale. To improve our statistics, we performed 11 additional 20-ns-long simulations in explicit solvent starting from the anti conformation for 8oxoG; all showed the same trend as shown in Figure 8, with gradual increase in 8oxoG glycosidic angle and poor convergence on this time scale. Therefore, to observe the anti→syn transition in explicit water, we expect that longer simulations and/or improved sampling methods will be required, such as those previously reported necessary by Simmerling et al. for observation of transitions in RNA hairpin loop conformations[45] and anti/syn transitions in pyrene-substituted DNA.[33] However, our limited observations in explicit solvent are consistent with our data obtained in implicit solvent.

Multiple Anti→Syn Transition Pathways in 8oxoG:A. Our MD simulation showed spontaneous conversion to the 8oxoG-(syn):A(anti) conformation without imposition of any restraints or other bias toward this conformation (other than underlying energetic factors in the force field). This preference for 8oxoG-(syn):A(anti) is consistent with available experimental data, but the transition process itself has not previously been characterized, either through computation or experiment. Our single simulation indicated that the 8oxoG base is extruded into the major groove while simultaneously rotating about the glycosidic bond to the syn conformation, followed by reformation of the base pair. However, the results of this single observation have little statistical significance; thus we further validated this observation by performing an additional 44 GB MD simulations, all starting from the anti:anti conformation. The simulation results are summarized in Table 2.

Overall, in about 30% of 45 trajectories, the anti→syn transition was observed within the 15 ns simulation length. These transitions tend to follow two pathways. One important difference between the pathways is apparent in Figure 9, which shows the base pair distances for the 8oxoG:A and two flanking base pairs as a function of time during two representative simulations. Figure 9a shows a pathway similar to what was described above, with no dramatic change in the flanking pairs; only the 8oxoG:A pair breaks and reforms. Figure 9b shows a substantially different pathway involving opening of the 5′ flanking A:T base pair first, followed by breaking of the 8oxoG:A pair. The A:T pair is not re-established until *after* rotation of 8oxoG about the glycosidic bond and formation of the 8oxoG(syn):A(anti) conformation. However, the 3′ flanking G:C base pair never opens in either pathway.

The first transition path is referred to as the 8oxoG extrusion path (Figure 9a). Detailed structural features for this process are displayed via eight snapshots in Figure 10a. The two initial hydrogen bonds present in the anti:anti conformation (G*7:O6−A20:H61 and G*7:H1−A20:N1) are broken quickly during

(61) McConnell, K. J.; Beveridge, D. L. DNA structure: What is in charge? *J. Mol. Biol.* **2000**, *304* (5), 803−820.

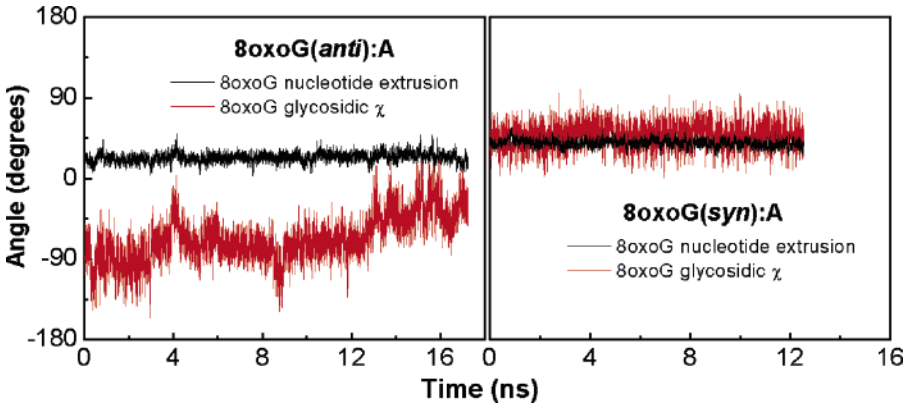


Figure 8. The 8oxoG glycosidic angle and nucleotide extrusion angle during 8oxoG:A simulations in explicit solvent. The left panel is for anti 8oxoG, and the right panel is for syn 8oxoG. Consistent with our results using GB solvation, the anti conformation is notably less stable than syn.

Table 2. Summary of Anti→Syn Transition Events Occurring in 45 Independent MD Simulations for 8oxoG:A Starting with Anti:Anti Orientation

	path	events	probability (%)
anti → syn	8oxoG extrusion	9	20
	internal rotation	4	9
total anti → syn		13	29

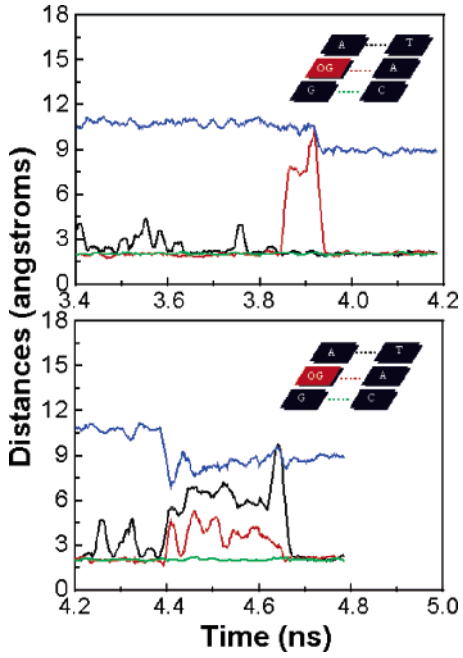


Figure 9. Distances between bases for the 8oxoG:A and the 5′ and 3′ flanking base pairs as a function of time for the two transition pathways. Minor groove width is shown in blue; 5′ A:T base pair distance, in black; 8oxoG:A base pair distance, in red; and 3′ G:C base pair distance, in green. Figure 9a (upper) shows a pathway where only the 8oxoG:A pair breaks, while Figure 9b (lower) shows a pathway where the 5′ A:T pair is coupled to the 8oxoG:A conformational transition. Only the time region at the point of transition in each simulation is shown.

equilibration. Next, accompanied by the significant changes described above in ϵ and ξ backbone torsions, the 8oxoG:A pair opens and 8oxoG is extruded into the major groove. The base rotates 180° around the glycosidic angle and inserts back into the duplex with a syn configuration, accompanied by a decrease in major groove width by ∼3 Å. The changes seen during our explicit solvent simulations are consistent with the early portions of this pathway.

The other path (Figure 9b), which is referred to as the internal rotation path, shows a fundamentally different transition process. Snapshots sampled from this path are shown in Figure 10b. The transition also starts from the partial loss of two hydrogen bonds (G*7:O6−A20:H61 and G*7:H1−A20:N1). As the simulation continues, severe stretching and unwinding are observed (data not shown). The major groove becomes significantly compressed (Figure 9b), followed by breaking of the 5′ flanking A:T base pair. This motion opened sufficient space in the duplex to permit rotation of 8oxoG about the glycosidic bond without requiring extrusion into the major groove. Shortly after this local structure rearrangement was completed, the AT base pair was re-established.

The statistics of occurrence of each pathway (summarized in Table 2) reveal that the base extrusion path occurs more than twice as frequently as the internal rotation path, suggesting that the latter pathway has a larger transition barrier than the former one. Although the groove compression occurs at different points of transition for two paths, in both cases the groove width is decreased by ∼3 Å after the transition, bringing it to a value comparable to a standard DNA sequence.

Free Energy Landscapes for the Transition. The sections above show two examples of how the anti→syn transition can occur in 8oxoG:A. However, normal MD simulations have extremely limited ability to provide information about the relative probability of the two conformers. Since none of the simulations showed reversible sampling of the base pairing schemes, the relative populations sampled are not reliable indicators of thermodynamic stability. To improve the sampling efficiency, we applied replica exchange molecular dynamics to each of the DNA sequences and obtained equilibrium probability distributions for alternate base pair conformations in each case.

Free energy surfaces were constructed using the 8oxoG glycosidic and nucleotide extrusion angles as two reaction coordinates. For each sequence, pairs of values of these coordinates were obtained for all of the structures sampled at a given temperature, and a two-dimensional histogram was constructed. These histograms were converted to free energy values, scaled to approximately correct for LREMD surface smoothing effects, and plotted as surfaces as shown in Figure 11 (see Methods for more detail).

It is important to note that all of the simulations were initiated from 8oxoG(anti), yet free energy minima are apparent on the landscape for both anti and syn in all sequences. The relative

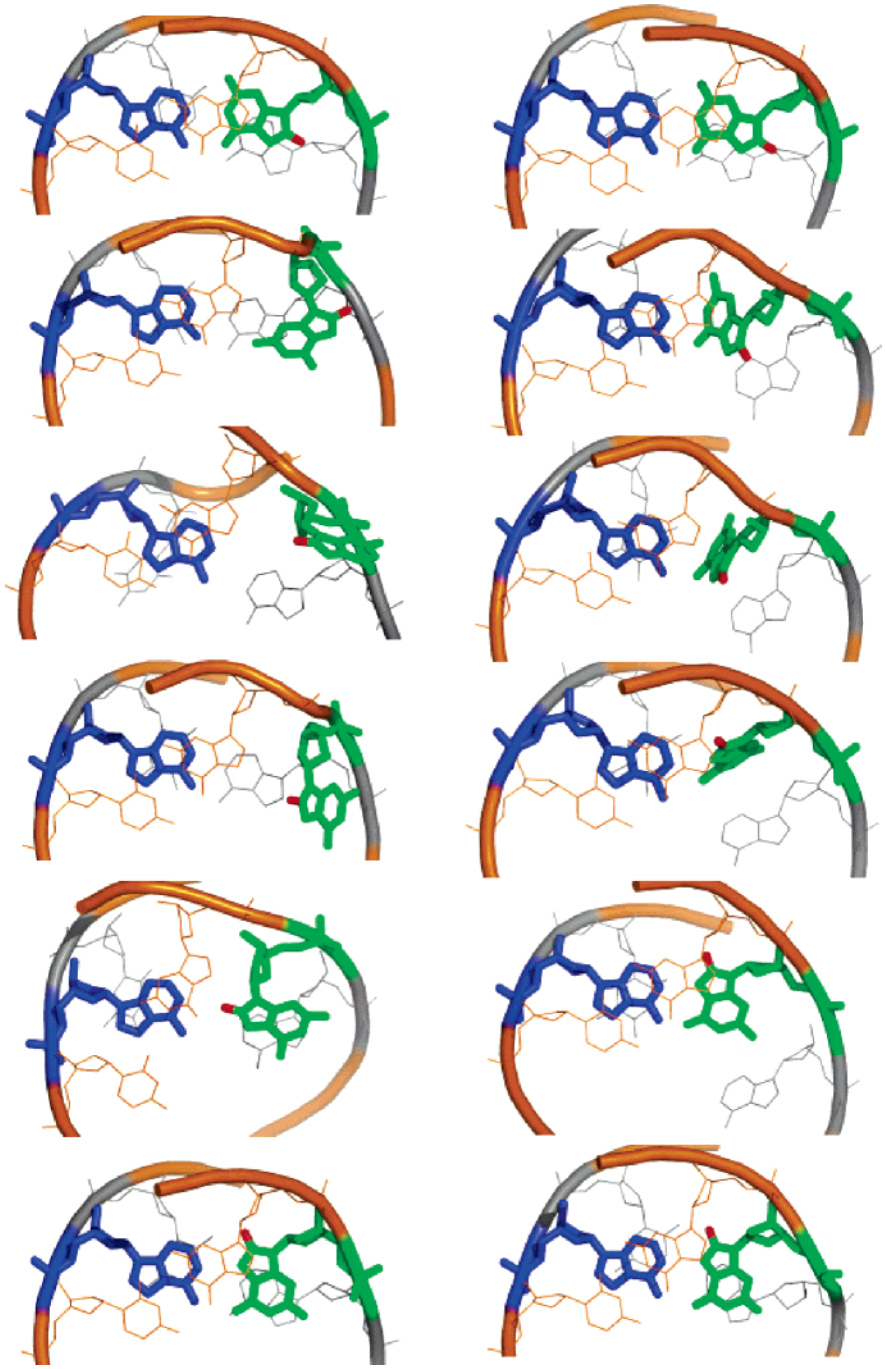


Figure 10. Snapshots during two transition paths, viewed along the helix axis. 8oxoG is shown in green; the O8 atom of the base, in red; and the partner A, in blue. G8 and C19 are shown in orange; A6 and T21 are shown in gray. (Left column) The 8oxoG extrusion path (shown in Figure 9a). (Right column): The internal rotation path (shown in Figure 9b).

free energies clearly show that syn 8oxoG is dominant for 8oxoG:A, in contrast to the dominant anti in G:C and 8oxoG:C. These results are all consistent with the available experimental data. We estimate that the free energy change for the anti to syn transition is 3.8 (G:C), 2.8 (8oxoG:C), and −2.8 (8oxoG:A) kcal/mol. However, we note that these values are approximate due to our use of LREMD.

A noticeable broadening of the basins is observed for 8oxoG:C as compared to G:C. This suggests that greater flexibility and variation occur within each conformation type as a result of the lesion despite adoption of the same base pair geometry in both cases. This broadening also results in a reduced free energy barrier to base extrusion for 8oxoG:C as compared to G:C. These differences may provide useful insights and an atomic-detail model for understanding the processes of damage recognition and binding in enzyme/DNA complexes. While directly comparable experimental data are not yet available, these observations are in good agreement with our nonequilibrium MD data described above. While a difference in flexibility was noted in previous MD simulations,[62] it is much more evident here, likely due to the improved sampling.

(62) Ishida, H. Molecular dynamics simulation of 7,8-dihydro-8-oxoguanine DNA. *J. Biomol. Struct. Dyn.* **2002**, *19* (5), 839−51.

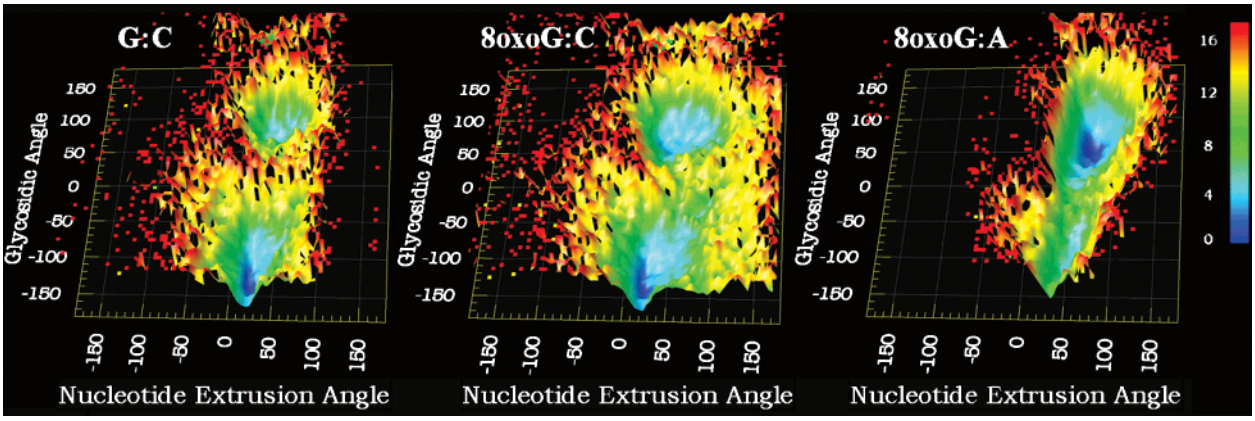


Figure 11. Surfaces with height and color corresponding to relative free energy, constructed from LREMD simulations. Reaction coordinates are the nucleotide extrusion and glycosidic rotation angles. Standard DNA with G:C (left); 8oxoG:C (middle); 8oxoG:A (right). The preferred conformation of the 8oxoG glycosidic torsion (the global free energy minimum) varies in the three sequences, with anti G/8oxoG corresponding to the lower basin and syn in the upper basin. The legend represents free energy values that have been scaled to approximately correct for the surface smoothing effect of LREMD (see Methods).

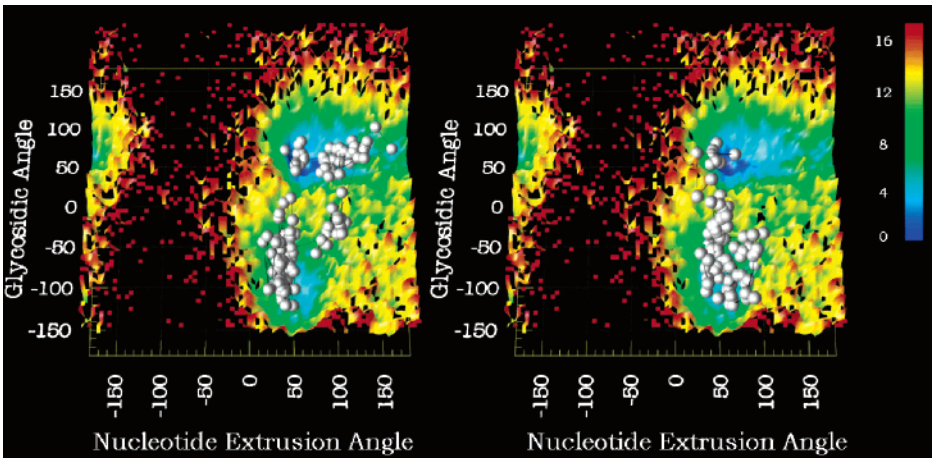


Figure 12. Free energy surfaces for 8oxoG:A as shown in Figure 11, with anti 8oxoG corresponding to the lower basin and syn in the upper basin. Each white sphere indicates the values of the reaction coordinates for a structure sampled during an MD simulation that started in the lower basin. Lines connect consecutive snapshots to indicate the path. The underlying free energy surface is the same for both images. On the left is the 8oxoG extrusion pathway, with an increase in nucleotide extrusion angle while crossing the barrier between the basins; on the right is the internal rotation pathway where little change is observed in nucleotide extrusion angle as the barrier is crossed.

Another remarkable variation between the landscapes for each sequence is the difference in barrier heights for rotation of the glycosidic bond. 8oxoG:A has the lowest barrier height, followed by 8oxoG:C, while G:C has the most significant barrier. The approximate barrier for G:C obtained from our LREMD data is consistent with the range 15−20 kcal/mol obtained from experiment[63,64] and potential of mean force calculations.[58,65] The latter required simulations with an imposed reaction coordinate, while our data were obtained without need for specification of a particular transition pathway.

We next link the free energy data obtained from equilibrium simulations shown above with the conformational transitions observed in our nonequilibrium MD simulations. In Figure 12, we show the free energy surface for 8oxoG:A with white spheres superimposed on the surface. Each sphere represents the values of these reaction coordinates for consecutive snapshots during an anti→syn transition, indicating the pathway taken on the landscape. We estimate that the free energy barrier for the base flipping pathway to be ∼9 kcal/mol, while that for the internal rotation pathway to be ∼12 kcal/mol. The higher free energy barrier for the internal pathway is consistent with the lower probability of sampling this path in the set of 45 nonequilibrium MD simulations. This correspondence, along with the correct location of the global free energy minimum for each sequence, provides us with additional confidence that our data are well converged and our model describing the effect of the lesion on DNA structure and dynamics is at least qualitatively accurate.

Effect of the Flanking Base Pair Sequence. In our multiple simulations of the 8oxoG:A sequence, we noted breaking of the 5′ flanking AT base pair in 19 out of 45 simulations, representing 42% of the trajectories (Table 3). The breaking of the 5′ AT pair was required for the transition to occur without flipping of the 8oxoG base (following the internal rotation pathway). In contrast, the 3′ flanking G:C base pair was much more stable with no G:C opening in any of the 45 trajectories, consistent with the increased stability of G:C vs A:T pairs as indicated by experimentally determined base pair opening

(63) Gueron, M.; Leroy, J. L. Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol.* **1995**, *261*, 383−413.
(64) Gueron, M. K. M.; Leroy, J. L. A single mode of DNA base-pair opening drives imino proton exchange *Nature* **1987**, *328*, 89.
(65) Ramstein, J.; Lavery, R. Energetic Coupling between DNA Bending and Base Pair Opening. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85* (19), 7231−7235.

Table 3. Summary of Base Pair Breaking and Anti→Syn Transition Events Occurred in Independent Simulations for Four Alternate Flanking Sequences

	no. simulations	probability breaking either BP	events anti→syn (total)	events anti→syn (extrusion)	events anti→syn (internal)
AG*G	45	42%	13	9	4
GG*A	45	53%	13	11	2
GG*G	45	31%	20	20	0
AG*A	12	92%	1	0	1

equilibrium constants (K_{open} $\sim 10^{-7}$ for G:C, $\sim 10^{-5}$ for A:T).[63,64] The observation led us to hypothesize that sequence context might have a measurable impact on the anti→syn glycosidic bond transition for 8oxoG:A mismatches. Our observations for the preferred rotation of the 8oxoG (toward the 5′ pair) could arise from two possible scenarios: either the 8oxoG base prefers the 5′ rotation due to intrinsic energetics or the weaker AT pair undergoes more frequent fluctuations that permit 8oxoG rotation. To test these hypotheses, additional simulations for three alternate sequences were initiated, with different combinations of AT and GC at the flanking positions as compared to the original 5′A:T−G*:A−G:C3′ (AG*G) sequence. These include two flanking GC pairs (GG*G), two flanking AT pairs (AG*A), and one of each pair with the previous positions switched (GG*A).

The switched GG*A sequence shows similar behavior as compared to AG*G. Opening of the AT flanking pair is again predominant, which occurs in 24 out of 45 simulations. Overall, this observation remains consistent with the relative weakness of AT vs GC pairs, and moving the position of the AT pair shifts the position of the opening events to the 3′ side of 8oxoG. Accordingly, the internal rotation pathway is still sampled, but the direction of 8oxoG rotation changes direction compared to that seen in AG*G.

As expected, fewer opening events were sampled in 45 GG*G simulations and all transitions followed the 8oxoG extrusion pathway. In contrast, the AG*A sequence is more structurally perturbed and dynamic, with the strain of the mismatch resulting in melting of the duplex in that region. The fluctuations in RMSD and local bending at the mismatch were also significantly increased (RMSD fluctuation and average bending were 3.5 Å and 49° compared to 1.6 Å and 30° for AG*G). This structural instability resulted in significant ambiguity in the interpretation of the transition pathway; thus only 12 simulations were run for this sequence. Out of 12, 9 show opening of both flanking AT pairs, and 11 out of 12 (92%) showed opening of at least one AT pair, significantly higher than the 42% we observed for the 5′AT-GC3′ parent sequence.

When all four sequences are considered (Table 3), an unexpected inverse correlation is observed between flanking base pair opening probabilities and spontaneous anti→syn transition rates. The GG*A and AG*G sequences have intermediate flanking pair opening and anti→syn transition rates, but the GG*G flanking pair opening rate is lower yet the anti→syn transition rate is higher. AG*A has the highest flanking pair opening rate but the lowest transition rate. This may simply reflect a longer time scale for completing the transition in sequences in which the duplex is less stable in the region near the mismatch, presenting alternate local minima in which the system can become kinetically trapped. As expected, however, the sequences with more frequent flanking pair opening do tend to increase the probability of sampling the internal rotation pathway. Internal rotation was a minor pathway for all but the AG*A sequence, for which our uncertainty is higher due to the smaller number of simulations and disorder in the mismatch region.

Conclusions

Simulations were performed on 13-mer DNA duplexes containing G:C, G:A, 8oxoG:C or 8oxoG:A base pairs. Results of these studies confirm the predominance of the Watson−Crick anti:anti conformation for 8oxoG:C and the Hoogsteen syn:anti conformation for 8oxoG:A. The preferred conformation does not depend on the base pair geometry used to initiate the simulation. For 8oxoG:A, we observed multiple anti→syn transition events following two qualitatively different pathways. One pathway involves extrusion of 8oxoG into the major groove, followed by rotation of the glycosidic bond and reformation of the base pair. The alternate pathway did not involve extrusion of 8-oxoG; instead, the AT pair 5′ to 8oxoG opens, accommodating rotation of the lesion inside the duplex. Reproducible results were obtained with a continuum solvent model; however, the anti→syn transition was not observed in twelve explicit solvation simulations with durations of up to 20 ns, despite the fact that simulations using anti showed structural instability and those using syn were stable. This observation suggests that these simulations are too short to sample the relevant transitions. Thus, these simulations present another example of the utility of treating solvation as a continuum, providing better-converged simulations and reduced computational requirements that facilitate the use of multiple simulations to examine the statistical nature of alternate transition pathways.

To gain additional insight into the nature of these transitions, we complemented our nonequilibrium simulations with replica exchange molecular dynamics, providing a detailed study of the thermodynamic properties of this system. We mapped free energy landscapes for several sequences with the 8oxoG glycosidic and nucleotide extrusion angles as two reaction coordinates. The resulting landscapes were consistent with results of the standard MD experiments, and in each case, the lowest free energy state matched available experimental data. The anti conformation of G in the G:C pair was approximately 4 kcal/mol lower in free energy than the syn. This difference was reduced to ∼3 kcal/mol for 8oxoG:C. The 8oxoG:C pair also showed broadening of the energy minima and a reduced barrier to base pair opening as compared to G:C. For 8oxoG:A, the trend was reversed, with the syn conformation for the 8oxoG base ∼3 kcal/mol lower in free energy than the anti.

The combination of transitions sampled during nonequilibrium dynamics and thermodynamic data from replica exchange provides novel insights into the dynamic behavior of this system, showing how this behavior is affected by chemical modifications generated in oxidatively damaged DNA. We note, however, that these landscapes were obtained using a modified replica exchange approach incorporating the Locally Enhanced Sampling approximation. The absolute free energy values using this method do not correspond to what would be obtained using standard MD simulations at the same temperature. Thus the landscapes provide only a qualitative view of the landscape;

however in the present case our conclusions are supported by the consistency between the landscapes, our data from standard molecular dynamics, and published experimental observations on the relative structure and stability of the mismatches.

Finally, the effect of the flanking base pairs on rotation of the glycosidic bond was investigated with multiple simulations of 8oxoG:A, with three alternate sequences used for the flanking base pairs. Variations in the opening rates are consistent with experimentally observed increased stability for GC compared to AT pairs. Moreover, the anti→syn transition probability and transition path were highly dependent on sequence context. Flanking pairs with reduced rates of opening for flanking base pairs also had reduced tendency for extrusion of 8oxoG during rotation about its glycosidic bond.

The use of a continuum solvent model in this study enabled a large number of simulations, providing reproducible observation of spontaneous conformational transitions along alternate pathways, for several mismatches and with variation of the flanking base pair sequence. No comparable studies using explicit solvent have been reported. However, it is important to keep in mind the potential limitations of continuum solvent models. For example, umbrella sampling studies[58] of base pair opening in explicit solvent located states with partially separated bases bridged by well-ordered water molecules with increased residency times. Continued improvement in simulation methods and computational resources will enable future validation of the calculations reported here against more detailed (and presumably more accurate) techniques.

Even when experiments were performed with continuum solvation, some barriers remained inaccessible on the multi-nanosecond time scale at 300 K. Thus, we were unable to directly observe spontaneous syn→anti transitions in constant temperature simulations of 8oxoG:A, although both directions were sampled in our replica exchange simulations. While syn 8oxoG is stable in solution, the modified base adopts an anti conformation when bound to MutY; thus, our simulations model the *reverse* of the process that presumably occurs during formation of the complex. We are performing similar studies for the complex to gain insight into how the preferred structures and transition pathways observed for free DNA are affected by the presence of the enzyme.

Acknowledgment. Support for this project was provided by NIH GM6167803 (C.S.), CA047995 (CDS), and CA17395 (A.P.G.) and by the National Computational Science Alliance Grant MCA02N028 (C.S.), which provided computational resources at NCSA. C.S. is a Cottrell Scholar of Research Corporation.

Supporting Information Available: Complete ref 22. This material is available free of charge via the Internet at http://pubs.acs.org.

JA052542S