Published online 19 November 2007

Nucleic Acids Research, 2008, Vol. 36, No. 1 ***253–262***
doi:10.1093/nar/gkm622

DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics

Jingyang Chen[1], François-Yves Dupradeau[2], David A. Case[2], Christopher J. Turner[3] and JoAnne Stubbe[1,4,*]

[1]Department of Chemistry Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, [2]Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, [3]Francis Bitter Magnet Laboratory and [4]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received June 7, 2007; Revised July 24, 2007; Accepted July 30, 2007

ABSTRACT

Abasic sites are common DNA lesions resulting from spontaneous depurination and excision of damaged nucleobases by DNA repair enzymes. However, the influence of the local sequence context on the structure of the abasic site and ultimately, its recognition and repair, remains elusive. In the present study, duplex DNAs with three different bases (G, C or T) opposite an abasic site have been synthesized in the same sequence context (5′-CCA AAG_6 XA_8C CGG G-3′, where X denotes the abasic site) and characterized by 2D NMR spectroscopy. Studies on a duplex DNA with an A opposite the abasic site in the same sequence has recently been reported [Chen,J., Dupradeau,F.-Y., Case,D.A., Turner,C.J. and Stubbe,J. (2007) Nuclear magnetic resonance structural studies and molecular modeling of duplex DNA containing normal and 4′-oxidized abasic sites. *Biochemistry*, 46, 3096–3107]. Molecular modeling based on NMR-derived distance and dihedral angle restraints and molecular dynamics calculations have been applied to determine structural models and conformational flexibility of each duplex. The results indicate that all four duplexes adopt an overall B-form conformation with each unpaired base stacked between adjacent bases intrahelically. The conformation around the abasic site is more perturbed when the base opposite to the lesion is a pyrimidine (C or T) than a purine (G or A). In both the former cases, the neighboring base pairs (G6-C21 and A8-T19) are closer to each other than those in B-form DNA. Molecular dynamics simulations reveal that transient H-bond interactions between the unpaired pyrimidine (C20 or T20) and the base 3′ to the abasic site play an important role in perturbing the local conformation. These results provide structural insight into the dynamics of abasic sites that are intrinsically modulated by the bases opposite the abasic site.

INTRODUCTION

The integrity of the genome is under constant threat of endogenous and exogenous DNA damaging agents, which produce a variety of DNA damage products. The abasic sites are common DNA lesions. Lindahl has estimated that there is a constant level of ~10 000 abasic sites for a typical human cell (1). Most of the abasic sites are believed to result directly from spontaneous depurination (1), or indirectly from deamination of cytosine to uracil, which is then eliminated by uracil glycosylases (2). Abasic sites also result from hydrolysis of oxidized or alkylated bases by lesion-specific glycosylases (3–5).

The abasic sites in DNA are chemically labile lesions. In solution, they exist predominantly as a mixture of ring-closed α and β hemiacetals with a minor amount of ring-opened aldehyde and aldehyde hydrate (<1%) (6,7). The aldehyde form is susceptible to base-catalyzed β-elimination that leads to a DNA strand break (8). The half-life of an abasic site in oligomeric duplex DNA has been measured by our group and the Sheppard group in five different sequence contexts and ranges from 200 to 900 h under physiological conditions (pH 7.5, 150 mM NaCl at 37°C) (9,10).

*To whom correspondence should be addressed. Tel: +1 617 253 1814; Fax: +1 617 258 7247; Email: stubbe@mit.edu
Present address:
Jingyang Chen, Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge St. Boston, MA 02114, USA.
Francois-Yves Dupradeau, DMAG, EA 3901, Faculté de Pharmacie, Université de Picardie Jules Verne, 1-3, rue des Louvels, 80037 Amiens Cedex 1, France.

Unrepaired abasic sites are mutagenic (11). Studies on abasic sites within a template for DNA replication suggest that a number of DNA polymerases (Pol I and III from *Escherichia coli*, Pol α from *Drosophila* and T4 DNA polymerase) preferentially incorporate dA opposite an abasic site (12–14). During translesion DNA synthesis catalyzed by Y-family DNA polymerases, abasic sites also cause frameshifts in the replication product (15–18).

In vivo, the abasic sites are thought to be repaired via the base excision repair pathway, which in humans initially involves apurinic/apyrimindic endonuclease 1 (Ape1) (19,20). Crystallographic studies on Ape1 complexed with a duplex oligonucleotide containing a tetrahydofuran (THF) abasic site analog reveal bending (35°) of the DNA helical axis, extrahelical flipping of THF relative to the phosphodiester backbone and penetration of protein side chains into the regions adjacent to the abasic site (21). Similar observations have been made in human uracil–DNA glycosylase and *E. coli* Endonuclease IV (22,23). Whether the conformation of the abasic site in duplex DNA in solution is inherently perturbed or whether the repair enzymes actively alter DNA structures in the region of the lesion remains unclear.

Solution structural studies by 2D NMR methods and molecular modeling on oligonucleotides containing site-specifically incorporated abasic sites suggest that the conformation of DNA in solution is different from that bound to the repair enzymes determined crystallographically (22,23). Studies from our group using 2D NMR spectroscopic methods on d(5′-CCA AAG_6 $\mathbf{X}A_8$C TGG G)-3′, where **X** denotes the abasic site, indicate that while the deoxyribose moiety is partially extrahelical, the overall structure is predominantly B-form (24). Because the unpaired and flanking bases are proposed to participate in the interaction with repair enzymes (21–23), elucidating their roles in determining the local conformation of the abasic site is important to understanding the basis of recognition of abasic sites by the apurinic/apyrimidinic endonucleases. Thus far, no systematic efforts have been reported on the influence of flanking or unpaired bases on the conformation of abasic sites.

In the present study, duplex DNAs containing an abasic site with the three different unpaired bases (G, C or T, Figure 1) in the same sequence context have been synthesized and characterized by 2D NMR spectroscopy. Molecular modeling and molecular dynamics calculations have been applied to study the influence of this base on the local conformation of the lesioned DNA. These results are compared with those from our previous study on the abasic site with an unpaired A in the same sequence (25). The results indicate that all four duplexes adopt overall B-form conformations and that the base opposite the abasic site is intrahelical. The conformation around the abasic site is more perturbed when the unpaired base is a pyrimidine (C or T) than a purine (A or G). In the former cases, the neighboring base pairs (G6-C21 and A8-T19) are closer to each other than those in B-form DNA. Molecular dynamics simulations reveal that the perturbation results from transient H-bond interactions between the unpaired pyrimidine (C20 or T20) and A8, 3′ to the abasic site. These results provide the structural insight into how the bases opposite the abasic site modulate its dynamics.

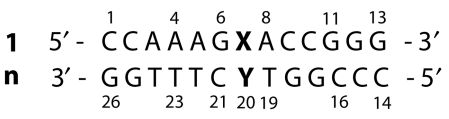

Figure 1. Sequences of the duplex DNA constructs containing a single abasic site (**X**) with four different bases (**Y**). $n = 2$, **Y** = G, duplex is Ab/G; $n = 3$, **Y** = C, duplex is Ab/C; $n = 4$, **Y** = T, duplex is Ab/T. Ab/A (**Y** = A) was described in (25).

EXPERIMENTAL SECTION

Materials

The oligonucleotides were synthesized on a 10-μmol scale by Invitrogen Inc. *E. coli* uracil–DNA glycosylase (UDG) was purified from expression plasmid pET21a-UDG kindly supplied by Dr James Stivers, Johns Hopkins University, as previously described (26). All other chemicals were purchased from Sigma-Aldrich Co.

Preparation of NMR samples

The single-stranded oligonucleotide containing an abasic site, d(CCA AAG **X**AC CGG G) (**1**) was synthesized by treating d(CCA AAG UAC CGG G) with UDG as previously described (24,25). Compound **1** was annealed in a 1:1 ratio with different templates (**2**, **3** or **4**, Figure 1) to generate three duplex DNAs (referred to as Ab/G, Ab/C and Ab/T, Figure 1). Duplex DNA formation of Ab/G, Ab/C and Ab/T was confirmed by anion-exchange HPLC analysis as previously described (25) (Supplementary Figure S1). The final NMR samples contained Ab/G (2.7 mM), Ab/C (2.5 mM) or Ab/T (2.7 mM) in 325 μl of 10 mM sodium phosphate, 0.2 mM EDTA, pH 6.5 in D_2O or 10% D_2O/90% H_2O (v/v).

NMR experiments

The 2D-NMR experiments were performed on a custom-built 591 MHz spectrometer at the Francis Bitter Magnet Laboratory. Spectra acquisition (TOCSY, NOESY, E-COSY, Watergate-NOESY, ^{1}H- ^{31}P HSQC and H3′-selective ^{31}P-^{1}H COSY) and data processing have been described in detail previously (24,25).

NMR structure determination and refinement

Distance restraints were derived from the volumes of the cross peaks in NOESY spectra (200 ms mixing time). The peak volumes were calculated with the peak picking protocol in the Felix 2001 software package. Volumes of cross peaks between H2′ and H2″ (1.772 Å) or between H5 and H6 (2.460 Å) of cytosine were used to calibrate the proton–proton distances. The cross peak intensities were classified as weak, medium or strong with the distance restraints of 1.5–3.0, 2.0–5.0 or 3.0–6.0 Å, respectively. Alternatively, a more rigorous approach of iterative relaxation matrix analysis was applied to estimate the distance restraints using the program MARDIGRAS (27,28). MARDIGRAS calculations were carried out on

Table 1. Summary of pairwise all-atom RMSD, NMR restraints and NMR violations of the final 10 structures of Ab/G, Ab/C and Ab/T

	Ab/G		Ab/C		Ab/T	
	α	β	α	β	α	β
Pairwise RMSD (Å)						
Starting structures	4.54	4.54	4.55	4.55	4.55	4.55
Final 10 structures	1.55	1.45	1.53	1.41	1.62	1.73
Number of NMR restraints						
NOE	455	455	478	479	452	453
Dihedral angle	58	58	63	63	62	61
Violation of restraints						
NOE violation (Å)	0.071	0.107	0.112	0.089	0.092	0.110
Dihedral angle violation (°)	2.27	2.31	2.39	2.38	2.36	2.38

the NOESY spectra with a correlation time of 2.0 ns. For each NOESY cross peak, an upper and a lower distance restraint was calculated. The final distance restraints derived from MARDIGRAS calculations agreed well with the manual classification as described above.

Dihedral angle restraints of the deoxyribose moieties were derived from the coupling constants measured in E-COSY experiments. Except for the abasic site, the coupling constants of H1′-H2′ and H1′-H2″ were consistent with a Southern (C2′-*endo*) conformation. A summary of NMR restraints for Ab/G, Ab/C and Ab/T is presented in Table 1.

All structural calculations were performed on an SGI Altix 3700 LINUX server with 128 Intel Itanium 1.3 GHz processors at The Scripps Research Institute (TSRI) using the SANDER module of AMBER 8.0 (29). The starting model structures were built using the Biopolymer module of Insight 2000 (Accelrys Inc.). Three starting structures were constructed, including an A-form, a B-form and a third form with arbitrary helical parameters. The all-atom root-mean- square deviation (RMSD) for the three structures was ~4.5 Å. The Hawkins–Cramer–Truhlar pairwise generalized-Born implicit solvent model (IGB = 1) was applied to simulate the solvent dielectric constant (29). The initial model structures were energy-minimized by 500 steps of conjugated gradient minimization followed by 500 steps of steepest descent minimization to remove high-energy interactions introduced during the manual model building. The simulated annealing procedure was carried out by heating the structure from 300 to 900 K in 5 ps followed by cooling back to 300 K over the next 15 ps. During the cooling process, weak temperature coupling (slow cooling) was applied for the first 13 ps followed by strong temperature coupling (rapid cooling) for the last 2 ps (30). NMR restraints were enforced during the annealing process with an increasing force constant from 5 to 50 kcal/mol over the first 3 ps and was maintained at 50 kcal/mol for the rest of the calculation (30). The annealed structure was then subjected to 100-ps constant temperature (300 K) molecular dynamics with NMR restraints. Ten structures for each starting structure (A-, B- and the third form) were calculated by sampling conformations during the last 10 ps of the molecular dynamics at a 1-ps interval. The resulting 30 structures have an averaged RMSD of <2.0 Å, which indicates a structural convergence after refinement. Ten final structures were chosen from the 30 structures based on the fewest NMR violations to represent the ensemble of the averaged structure and showed no preference relative to the starting model. The statistical analysis of the final 10 structures is summarized in Table 1. The NMR restraint files, the 10 selected structures, and the averaged structures for each abasic site duplex have been deposited in Nucleic Acid Database (NDB entry no. 2O7W, 2O7X, 2O7Y, 2O7Z, 2O80 and 2O82).

Molecular dynamics simulations using an explicit solvent model

The molecular dynamics simulations were performed at a constant temperature (300 K) using the software AMBER 8.0 following the protocol used by Dupradeau *et al*. (31) The averaged structures were first solvated using the TIP3P solvent model with approximately 6000 water molecules in a truncated octahedral box of solvent with a thickness of 10 Å. Twenty-four sodium counter ions were included in the calculations to neutralize the system. The molecular dynamics calculations were carried out on a 3-ns time scale. The conformation was sampled every 1 ps, and was analyzed using the PTRAJ module available in the Amber 8.0 package (29).

RESULTS

Chemical shift assignments of non-exchangeable protons

Standard TOCSY (60 and 110 mixing time) and NOESY (100, 200 and 400 ms mixing time) experiments were performed to assign the non-exchangeable protons of each duplex (32). The proton chemical shifts associated with the deoxyriboses were assigned using TOCSY experiments. The NOE connectivities between base protons (H6 of pyrimidines, H8 of purines) and deoxyribose protons (H1′, H2′/H2′, H3′, H4′ and most of H5′/5″) were established by analyzing the NOESY spectra (200 ms mixing time) to confirm the assignments. NOE connectivities between the H5 and H6 protons of cytosines and between the methyl and H6 protons of thymidines were also used to assist the assignments of NOE signals in overlapping regions.

For Ab/G, the NOE connectivities in the base-to-H1′ region are easily detected due to dispersion of signals (Supplementary Figure S2). In **1** of Ab/G (Figure 1), the NOE is disrupted between G6 and A8 (Supplementary Figure S2, indicated by ×). In **2** of Ab/G, the NOE connectivities are observed from T19 to C21, indicating that the unpaired base (G20) adopts an intrahelical conformation as is observed in regular B-form DNA. Similar results have previously been obtained for Ab/A (25).

For Ab/C and Ab/T, the NOE connectivities between the H1′ protons and the base protons are shown in Supplementary Figures S3 and S5, respectively. The assignments of a number of base-H1′ NOE signals

(C20 to T24 of Ab/C, G18 to T24 of Ab/T) were more challenging due to poor signal dispersion. Confirmation of these assignments was accomplished by NOESY and TOCSY spectra involving the methyl protons of thymidines and/or H5/H6 protons of cytosines (Supplementary Figure S4 for Ab/C and S6 for Ab/T). Surprisingly, in contrast with Ab/G and Ab/A, a weak NOE interactions was detected between G6 H1′ and A8 H8 for both Ab/C and Ab/T (Arrows in Supplementary Figures S3 and S5).

Chemical shift Assignments of exchangeable protons

The exchangeable protons (imino protons of guanines and amino protons of cytosines) of each duplex are indicative of Watson–Crick base pair interactions and global duplex conformation. These protons were assigned by Watergate-NOESY experiments at 4°C in 10% D_2O/H_2O (v/v). In the previous studies of Ab/A, 8 out of total 12 imino protons were detected by this method (25). The remaining four imino protons in Ab/A (C1-G26, G13-C14, G6-C21, and A8-T19) could not be detected in either 1D or 2D spectra, consistent with weak H-bond interactions of the 5′ and 3′ ends or adjacent to the abasic site (25).

Supplementary Figure S7 shows the cross peaks of the T-imino to A-H2 proton and G-imino to C-amino proton with the overlaid 1D spectrum of each duplex. Besides the eight imino signals previously detected in Ab/A, additional imino protons possibly associated with T19 in Ab/G and Ab/C (red arrows, Supplementary Figure S7A and B) and G6 in Ab/T (solid red line, Supplementary Figure S7C) suggest subtle differences in the base-pair interactions around the abasic site among these duplexes.

In contrast to the imino protons, the amino protons of all cytosines are readily assigned based on their strong cross peaks with their own C-H5 and C-H6. An example of the assignment strategy is shown in Supplementary Figure S8 for Ab/G. The chemical shift assignments of non-exchangeable and exchangeable protons are summarized in Supplementary Table S1 (Ab/G), S2 (Ab/C) and S3 (Ab/T) and were deposited in BioMagResBank (BMRB entry 15223, 15224, 15227, 15228, 15238 and 15239).

Chemical shift assignments of the protons associated with the abasic site deoxyribose

Assignments of the protons associated with the abasic site deoxyribose are essential to distinguish between its α and β configurations and to identifying their interactions with neighboring residues. The chemical shifts of the H1′, H2′/H2″, and H3′ protons associated with the abasic site deoxyribose were assigned using the 60 ms TOCSY spectra (Figure 2 and Supplementary Table S1–S3). In each duplex (Ab/G, Ab/C and Ab/T), these protons are shifted upfield relative to the deoxyribose protons found in B-form DNA (H1′, 5.0–5.2 p.p.m. versus 5.3–6.5 p.p.m.; H2′/2′, 1.8–2.0 p.p.m. versus 2.0–3.0 p.p.m.; H3′, 4.5–4.6 p.p.m. versus 4.7–5.0 p.p.m.). The H4′ and H5′/5″ protons could not be assigned due to spectral crowding in those regions.

The sugar protons at each abasic site exist as two sets of signals (Figure 2, solid red lines), consistent with α and β

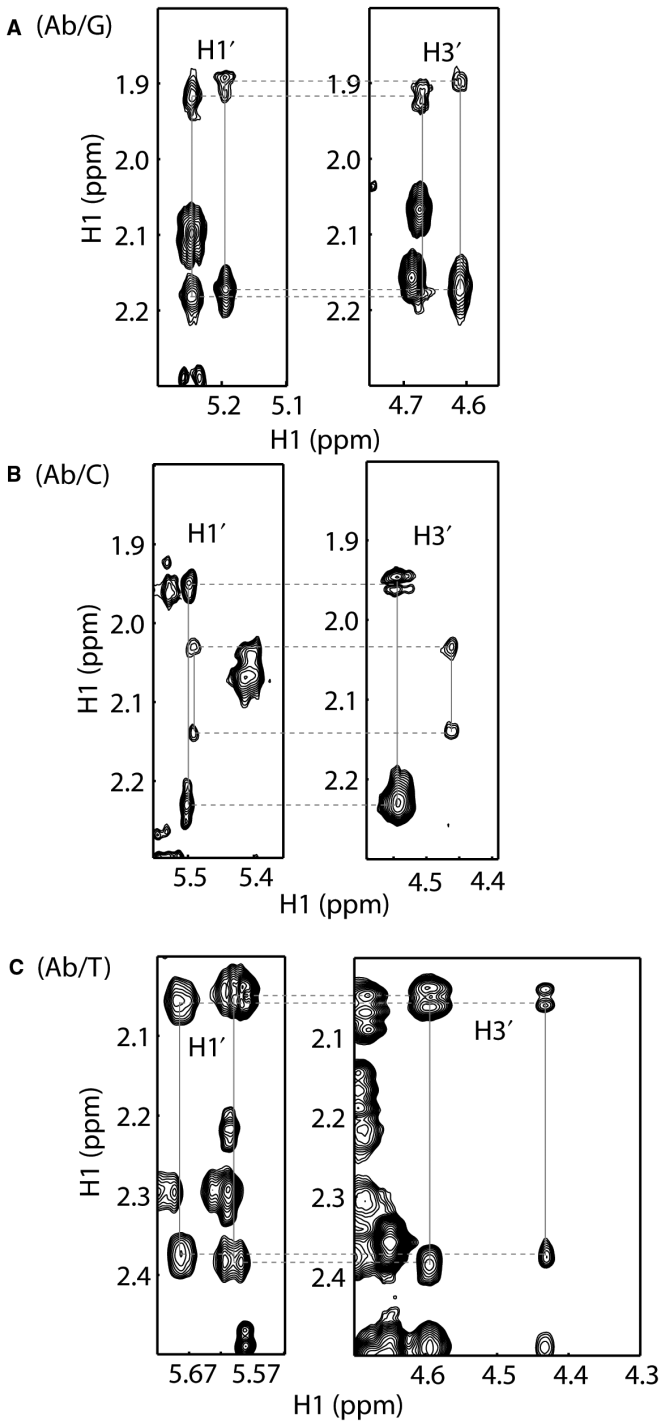

Figure 2. Chemical shift assignments of protons associated with the abasic site deoxyribose by TOCSY experiments (60 ms mixing time). For each oligonucleotide, the through-bond connectivities between H1′ and H2′/2″ and between H3′ and H2′/2″ are indicated by the solid red lines. The dashed red lines indicate the connectivities between H1′ and H3′ through the same H2′/2″ protons. A, Ab/G; B, Ab/C; C, Ab/T.

configurations. The assignment of each anomer was based on the relative intensities of the NOE signals (200 ms mixing time) between H1′ and H2′/2″ (24,25). For the α anomer, the H1′ is closer to H2′ than to H2″, producing a more intense cross peak. The intensity pattern is reversed

for the β anomer. The NOESY spectra containing the abasic site proton regions are shown in Supplementary Figures S9–S11 with the connectivity of H2′ and H2″ indicated by solid red lines. The ratio of α:β anomers for each duplex is approximately 60:40 based on signal intensities.

Sugar conformation

E-COSY experiments were used in an effort to measure the coupling constants of the deoxyriboses to define sugar conformations. For all the residues except the abasic site, the values of $J_{H1', H2'}$ ranged from 8 to 10 Hz and $J_{H1', H2''}$ from 5 to 6.5 Hz, characteristic of the 2′-*endo* sugar pucker (S-type) observed in B-form DNA. The coupling constants for the abasic sites proved to be problematic as only the signals associated with Ab/C [and Ab/A in (25)] were of sufficient intensity to detect(Supplementary Figure S12). For both Ab/C and Ab/A, the coupling constants ($J_{H1', H2'} = 4.3$ Hz and $J_{H1', H2''} < 3$ Hz for the α anomer and $J_{H1', H2'} = 5.6$ Hz and $J_{H1', H2''} = 4.3$ Hz for the β anomer) are consistent with a near-N-type (O1′-endo) conformation for the α anomer and a near-S-type (C1′-exo) conformation for the β anomer (25).

Phosphodiester backbone conformation

In order to characterize the conformation of the phosphodiester backbone, the dihedral angles of H3′-C3′-O3′-P have been determined by measuring the $J_{H3', P}$ coupling constants using H3′-selective ^{31}P-^{1}H COSY experiments (33,34). Due to the small dispersion of H3′ and ^{31}P chemical shifts, not all cross peaks were well resolved. For Ab/G and Ab/A, all ^{31}P chemical shifts were between −3.8 and −5.0 p.p.m. and all the $J_{H3'-P}$ values were within 6–8 Hz, consistent with those expected for a generic B-form DNA.

For Ab/C and Ab/T, the majority of the ^{31}P chemical shifts and the $J_{H3', P}$ coupling constants were also consistent with B-form DNA. However, with Ab/C, three ^{31}P signals were shifted downfield between −3.1 p.p.m. and −3.4 p.p.m. (cross peaks C, D and E in Supplementary Figure S13) and were assigned to nucleotides adjacent to the abasic sites (G6, A8 and T19). Assignments were made based on the H3′ and H4′ chemical shifts obtained from the ^{31}P-^{1}H HSQC experiments. With Ab/T, two ^{31}P signals were shifted downfield between −3.2 p.p.m. and −3.5 p.p.m. (cross peak C and D in Supplementary Figure S14). The ^{31}P-^{1}H HSQC experiments revealed these two signals to be associated with G6 and T19 adjacent to the abasic sites. Thus, both Ab/C and Ab/T appear to have perturbed backbone conformations relative to B-form DNA.

The Karplus equation (^{3}J(HCOP) = 15.3 $\cos^2\phi$ − 6.1 $\cos\phi$ + 1.6) was used to derive dihedral angle constraints for H3′-C3′-O3′-P from the experimentally determined coupling constants ($J_{H3'-P}$). The H3′-C3′-O3′-P dihedral angles for most residues were constrained to be –40° ± 25° as in B-form DNA. For G6, A8 and T19 in Ab/C and G6 and T19 in Ab/T, despite aberrant ^{31}P signals, the coupling constants fell in the normal 6–8 Hz range. However, the downfield shifted ^{31}P chemical shifts suggested perturbed dihedral angles. Thus, all four possible solutions to the Karplus equation (±40° and ±115°) were tested in separate structural calculations. The solution that consistently gave the lowest energy was used in the final structural calculations. The final H3′-C3′-O3′-P dihedral angle constraints for Ab/C were set to 40° ± 25° (G6), −40° ± 25° (A8), and −40° ± 25° (T19) and for Ab/T, these were 40° ± 25° (G6) and −40° ± 25° (T19).

NOE interactions involving the protons associated with the abasic site and adjacent bases

Our previous molecular modeling studies of the abasic site conformation in Ab/A (Figure 1) indicated that the NOE patterns between H2′/H2″ and H3′ of this site and the base proton (A8H8) 3′ to this site, was informative about sugar conformation relative to the DNA backbone (25). The medium NOEs observed between H3′ and A8H8 and weak NOEs between H2′/2″ and A8H8 indicated a partial extrahelicity of the deoxyribose (25).

The NOE interactions between the abasic site in Ab/G, Ab/C or Ab/T and their neighboring bases therefore were determined (Table 2). The abasic site in Ab/G shows a similar pattern to Ab/A with both anomers, except that the NOE between 3′H and A8H8 is weak. In the case of both Ab/C and Ab/T, the crosspeaks associated with 3′H and A8H8 are all medium, while those associate with the H2′/H2″ protons are variable and anomer specific (Table 2).

For the Ab/C and Ab/T constructs, several additional NOEs were detected between G6 and A8, the two bases adjacent to the abasic site. For Ab/C, weak NOE signals were observed between A8-H8 and G6-H1′ for both anomers (indicated by the dashed box in Supplementary Figure S3) and between A8-H8 and G6-H2″. For Ab/T, weak NOE signals were also observed between A8-H8 and G6-H1′ for each anomer (arrows in Supplementary

Table 2. Summary of the NOE interactions between protons around the abasic sites

Ab/G (α)		Ab/G(β)		Ab/T (α)		Ab/T (β)		Ab/C (α)		Ab/C(β)	
A8H8-**X**H3′	w	A8H8-**X**H3′	w	A8H8-**X**H3′	m	A8H8-**X**H3′	m	A8H8-**X**H3′	m	A8H8-**X**H3′	m
A8H8-**X**H2′	w	A8H8-**X**H2′	w	A8H8-**X**H2′	w			A8H8-**X**H2′	w		
A8H8-**X**H2″	w	A8H8-**X**H2″	w					A8H8-**X**H2″	w	A8H8-**X**H2″	w
				A8H8-G6H1′	w	A8H8-G6H1′	w	A8H8-G6H1′	w	A8H8-G6H1′	w
								A8H8-G6H2″	w	A8H8-G6H2″	w
										G6H4′-AbH2′	w
										G6H4′-AbH2″	w

(m, medium; w, weak; **X**, abasic site).

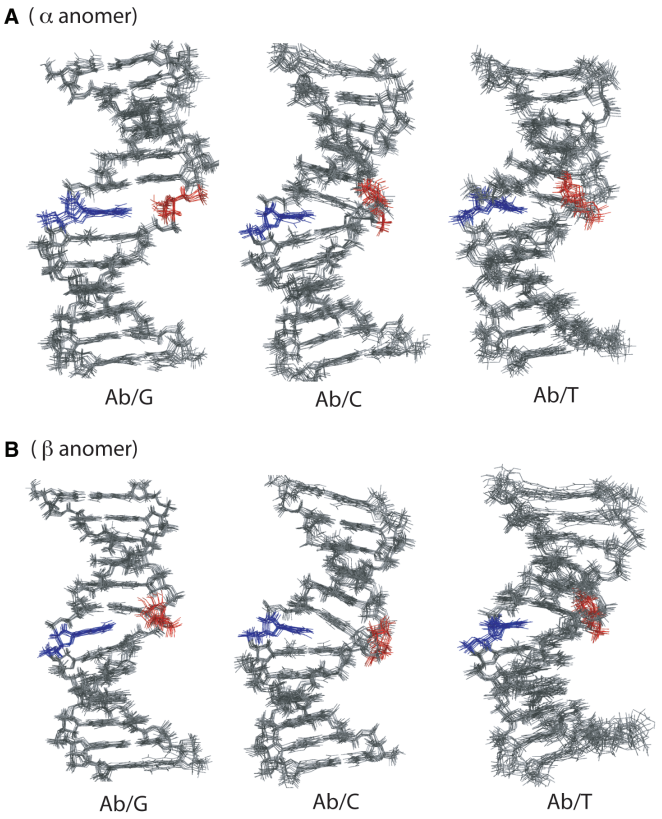

Figure 3. The overlay of 10 final structures of Ab/G, Ab/C and Ab/T. The abasic site is colored in red and the unpaired base is colored in blue. (**A**) Structures with the α anomer at the abasic site. (**B**) Structures with the β anomer at the abasic site.

Figure S5), however, no NOE signals were observed between A8H8 and G6-H2″. These NOEs and the unusual ^{31}P chemical shifts described above indicate that the abasic sites of Ab/C and Ab/T adopt different conformations relative to Ab/A and Ab/G.

Averaged structures of Ab/G, Ab/C and Ab/T

Ten structures of the α and β anomers of Ab/G, Ab/C and Ab/T were calculated from three starting structures using the experimentally determined distance and dihedral angle constraints. None of the 10 structures contained NOE violations >0.5 Å or dihedral angle violations >5°. The RMSDs of the final 10 structures are between 1.5 and 1.7 Å (Table 1). An overlay of the 10 structures for each construct is shown in Figure 3 and reveals that they are B-form with the unpaired bases (colored in blue) stacked intrahelically between flanking base pairs. A variety of conformations are observed for the abasic site deoxyribose (colored in red, Figure 3) and the bases adjacent to it (G6 and A8).

Abasic site conformation

The averaged structure of Ab/G shows that in the α anomer, the abasic site adopts an intrahelical conformation and in the β anomer, it is partially extrahelical (Figure 4). In the final 10 structures for the α anomer, 2 of them show that the 1′-OH group is within the H-bonding

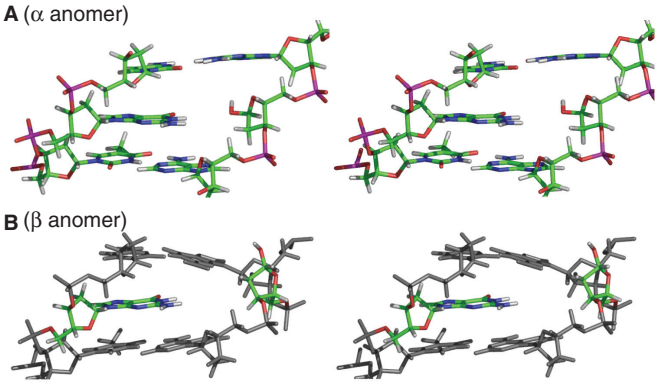

Figure 4. A stereo view of the conformation of the abasic site region in the averaged structure of Ab/G. (**A**) α anomer. (**B**) β anomer. In B, only the abasic site and the unpaired bases are colored for clarity.

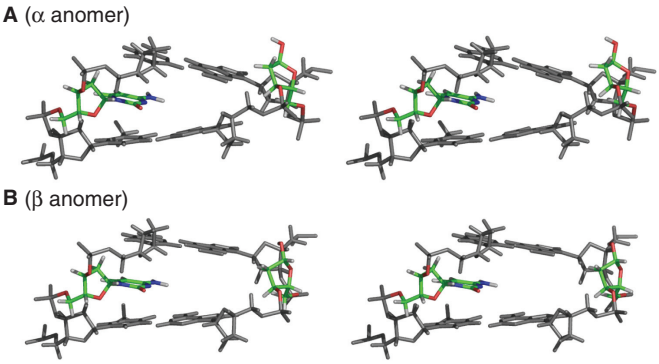

Figure 5. A stereo view of the conformation of the abasic site region in the averaged structure of Ab/C. (**A**) α anomer. (**B**) β anomer. The abasic site deoxyribose and the unpaired base are colored for clarity.

distance from the imino group of the opposite G20. Since no imino proton signals were detected for G20, this putative H-bonding interaction, if it exists, must be transient. The partially extrahelical conformation for the Ab/G β anomer is very similar to that of Ab/A determined in our earlier study (25).

Figures 5 and 6 show the averaged abasic site conformation in Ab/C and Ab/T. In all structural models, the abasic site is extrahelical. In addition, the backbone conformation around the abasic site is perturbed so that G6 and A8 are closer to each other than in normal B-form DNA. These results are consistent with and defined by the NOE interactions between A8H8 and several sugar protons of G6 in both Ab/C and Ab/T (Table 2).

Molecular dynamics simulations of abasic-site-containing duplexes

Molecular dynamics simulations were carried out to characterize the conformational flexibility in the abasic site region of all constructs. For Ab/G, an overlay of 10 snapshots representing 20 ps of molecular dynamics simulations for the α and β anomers of Ab/G are shown in Supplementary Figure 15A and B. The results indicate

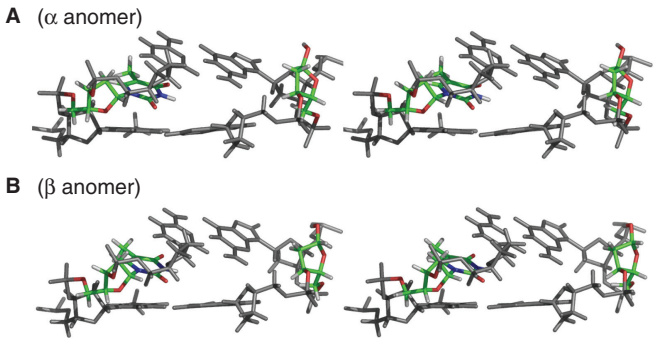

Figure 6. The stereo view of the conformation of the abasic site region in the averaged structure of Ab/T. (**A**) α anomer. (**B**) β anomer. The abasic site deoxyribose and the unpaired base are colored for clarity.

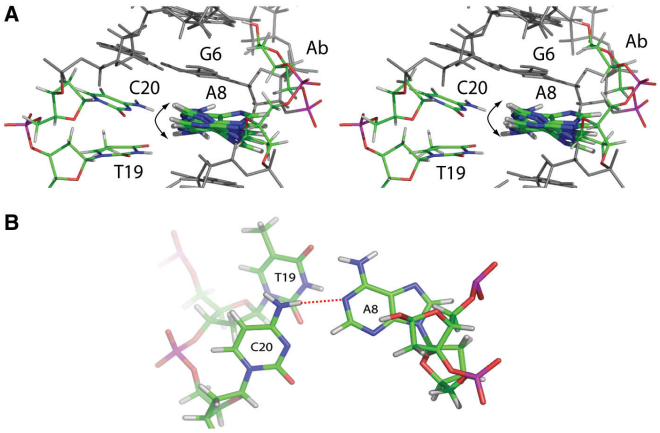

Figure 7. Molecular dynamics of Ab/C (α anomer) suggesting transient H-bonding interactions between C20 and A8 formation. (**A**) A stereo view of a snapshot of the conformation at the abasic site. For the conformation of A8 (in stick), five snapshots representing 10 ps molecular dynamics simulations are overlaid to show the conformational flexibility of A8 (indicated by the double-headed arrow). (**B**) A snapshot of the conformation showing the transient H-bonding interactions (red dashed line) between A8 and C20. The conformation of T19, the expected base-pair partner of A8, is also shown.

that the abasic site deoxyribose adopts a range of conformations from intrahelical to extrahelical. The extrahelical motion of the abasic site deoxyribose occurs without significant distortion of the conformations of adjacent residues, which is illustrated by the snapshot of the α anomer of Ab/G with an extrahelical abasic site (Supplementary Figure 15C). A similar extrahelical motion of the abasic site has also been observed with Ab/A during the simulations (25).

The molecular dynamics simulations of Ab/C reveal that the abasic site deoxyribose remains partially extrahelical with flexibility largely reflected by the motion associated with A8. Figure 7A shows that in the α anomer, A8 (in stick) adopts a range of conformations (indicated by the double-headed arrows) that are capable of forming H-bonding interactions with both T19 (the expected base-pair partner) and C20 (the base opposite the abasic site). Figure 7B shows the snapshot with the transient

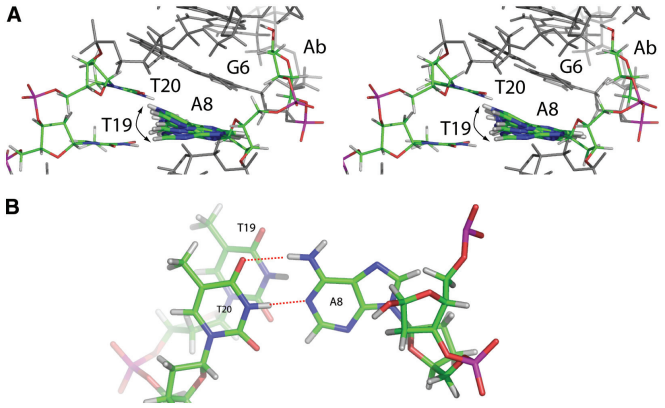

Figure 8. Molecular dynamics of Ab/T (α anomer) suggesting transient H-bonding interactions between T20 and A8. (**A**) A stereo view of a snapshot of the conformation at the abasic site. For the conformation of A8 (in stick), five snapshots representing 10 ps molecular dynamics simulations are overlaid to show the conformational flexibility of A8 (indicated by the double-headed arrow). (**B**) A snapshot of the conformation showing the transient H-bonding interactions (red dashed lines) between A8 and T20. The conformation of T19, the expected base-pair partner of A8, is also shown.

H-bonding interaction between A8 and C20 (viewed from the top along the helical axis). Quantitative analysis of the distance (red trace, Supplementary Figure S16) and angle (blue trace, Supplementary Figure S16) between the H-bond donor of C20 and the accepter of A8 during simulation suggests that the optimal H-bond interaction occurs every ~15 ps (black triangular dot, Supplementary Figure S16). Similar results were obtained for the β anomer.

The molecular dynamics simulations of Ab/T also reveal that the abasic site deoxyribose remains partially extrahelical with the flexibility also dominated by the motion associated with A8 3′ for both anomers. Figure 8A shows that in the α anomer, A8 (in stick) adopts a range of conformations (indicated by the double-headed arrows) that are capable of forming H-bonding interactions with both T19 (its expected base-pair partner) and T20 (the base opposite the abasic site). Figure 8B shows the snapshot with the transient H-bonding interaction between A8 and T20 (viewed from the top along the helical axis). Quantitative analysis suggests that the optimal H-bond interaction occurs every 8 ps between T20 and A8, more frequently than Ab/C (Supplementary Figure S17). This observation can be explained by the stronger interaction (2 H-bonds) between T20 and A8 than that in Ab/C (1 H-bond between C20 and A8 in Ab/C).

DISCUSSION

The first structural models of natural abasic sites in duplex DNA were reported in the 1990s by the Bolton group. They examined the duplex DNA of 5′-GAXAC-3′, where the base opposite the abasic site (**X**) was either a pyrimidine (C) or a purine (A) (35,36). Unfortunately, ambiguous chemical shift assignments, contradictory reports of the anomeric distribution, and the absence of

Table 3. Summary of previous 2D-NMR studies on duplex DNA containing abasic sites (**X**)

Sequence context	Anomeric ratio (α:β)	Conformation of abasic site	Conformation of unpaired base	References
5′-CGCGA**X**ACGCC-3′ 3′-GCGCT**A**TGCGG-5′	~50:50	α and β: intrahelical	α and β: intrahelical	(36)
5′-CGCGA**X**ACGCC-3′ 3′-GCGCT**C**TGCGG-5′	0:100	Intrahelical	intrahelical	(36)
5′-CGCGA**X**ACGCC-3′ 3′-GCGCT**A**TGCGG-5′	~50:50	α: extrahelical β: intrahelical	α and β: intrahelical	(35)
5′-CGCATT**X**TTGCG-3′ 3′-GCGTAA**A**AACGC-5′	~45:55	α: extrahelical β: intrahelical	α and β: intrahelical	(37)
5′-CCAAAG**X**ACTGGG-3′ 3′-GGTTTC**A**TGACCC-5′	~60:40	α and β: extrahelical	α and β: intrahelical	(24)
5′-CCAAAG**X**ACCGGG-3′ 3′-GGTTTC**A**TGGCCC-5′	~60:40	α and β: extrahelical	α and β: intrahelical	(25)

NOEs between basic sites and neighboring bases, limited the conclusions that could be drawn from these studies. Later structural studies by the Bolton group and our group in several sequence contexts suggest that the overall duplex is B-form with the unpaired base adopting an intrahelical conformation (see Table 3 for the summary of structural features in the abasic site region observed in the previous studies) (24,25,37). However, the limited NOEs observed between abasic sites and nucleotides in their environment led to limited generalizations that can be made about their conformation, anomeric distribution and flexibility in different sequence contexts.

In the present study, we extend our previous structural studies on the 5′-G**X**AC-3′/3′-C**A**TG-5′ sequence to all four unpaired bases (G, T, A and C) in the same sequence. The structural modeling based on 2D NMR analysis indicates that all four duplexes are B-form with the unpaired bases being intrahelical. In addition, for all four abasic site duplexes, a mixture of α and β anomers (60:40) were observed, which suggests that the anomeric distribution is not perturbed by the identity of unpaired bases.

The present studies on Ab/A, Ab/G, Ab/C and Ab/T have provided insight into the influence of the unpaired base on the structure of the 5′-G**X**AC-3′ duplex. Compared with Ab/A and Ab/G, the structural models of Ab/C and Ab/T indicate unusually close proximity between bases adjacent to the abasic site (G6 and A8) as a result of several weak NOEs between G6 and A8 (Table 2). Furthermore, for Ab/T and Ab/C, molecular dynamics calculations reveal transient H-bonding interactions between A8, 3′ to the abasic site, and the pyrimidine opposite the lesion. Examples of transient H-bonding interactions between an unpaired base and an adjacent base normally paired with another base have been documented (38). Computer modeling of a THF analog with C opposite the THF in the 5′-TC**F**TA-3′ sequence without NMR restraints revealed a similar transient H-bonding interaction between T, 3′ to THF, and the unpaired C during simulation (38). Such transient H-bonding interactions were not observed during the simulation of THF with an unpaired G in the same sequence context (38). Additional studies are required to establish whether this structural feature exists for all abasic sites with unpaired pyrimidines or whether it is dependent on the sequence context adjacent to the abasic site as well.

Transient H-bonding interactions between the base 3′ to the abasic site and the unpaired pyrimidine raise the issue of their biological importance. H-bonding interactions between the incoming nucleotide and bases adjacent to the abasic sites have been proposed to be responsible for frameshift mutations in translesion DNA replication by Y-family DNA polymerases (15–18). Structures of Dpo4 (a Y-family DNA polymerase from *Sulfolobus solfataricus*) have been solved in a complex with a primer/THF-containing-template (39). In one structure, the incoming nucleotide (dCTP, the would-be base opposite THF) formed base pair interactions with G, 5′ to THF, which would lead to a –1 frameshift. In the structure, the THF moiety was partially extrahelical to accommodate the structural requirement of the base pair interaction between the incoming nucleotide and the base 5′ to THF. Although this conformation was observed in a complex with the DNA polymerase, our results suggest that the inherent structural flexibility of the abasic site could lead to the transient base pair interaction (40).

Our structural models of abasic sites with four unpaired bases in 5′-G**X**AC-3′ represent the first step toward understanding the influence of the sequence context on anomeric distribution, conformation and flexibility of the abasic site. To investigate the influence of neighboring bases on the conformation of the abasic site, we also attempted to apply 2D NMR spectroscopy to obtain structural information of the abasic site in the sequence of 5′-CCAAGG**X**CAGGG-3′/3′-GGTTCC**G**GTCCC-5′. The 5′-GG**X**C-3′/3′-CC**G**G-5′ sequence was chosen for direct comparison to our previous structural studies of the 3′-phosphoglycolate/5′-phosphate lesion in the same sequence (41). In contrast with the 5′-G**X**AC-3′ sequences, we could not assign most protons associated with the neighboring bases (G6 and G17-C19, see Supplementary Table S4 for partial proton assignment). No NOE signals were detected for these residues in the region showing connectivity between base protons and H1′ (Supplementary Figure S18). In addition, no signals were detected for the abasic site protons by either NOESY or TOCSY. However, a mixture of anomers with a ratio of 60:40 was suggested based on two NOE

signals associated with H5 and H6 of C8 (Cross peak A and B, Supplementary Figure S18). The inability to detect NMR signals in the abasic site region in the 5′-GG**X**C-3′/3′-CC**G**G-5′ sequence underlines the challenge of obtaining structural models using the 2D NMR spectroscopy. Samples with ^{13}C or ^{15}N labeling should facilitate the assignments of residues and yield richer information on NMR restraints in the abasic site region in future studies.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work is supported by NIH Grant GM 34454 to J.S. and GM 45811 to D.A.C. The NMR facility is supported by NIH Grant RR-00995. F.-Y.D. was funded by the French 'Ministère de l'Education Nationale et de la Recherche'. Funding to pay the Open Access publication charges for this article was provided by NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
2. Lindahl,T. (1980) Uracil-DNA glycosylase from *Escherichia coli*. *Methods Enzymol.*, **65**, 284–290.
3. Nash,H.M., Bruner,S.D., Scharer,O.D., Kawate,T., Addona,T.A., Spooner,E., Lane,W.S. and Verdine,G.L. (1996) Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. *Curr. Biol.*, **6**, 968–980.
4. Dizdaroglu,M., Karahalil,B., Senturker,S., Buckley,T.J. and Roldan-Arjona,T. (1999) Excision of products of oxidative DNA base damage by human NTH1 protein. *Biochemistry*, **38**, 243–246.
5. Wyatt,M.D., Allan,J.M., Lau,A.Y., Ellenberger,T.E. and Samson,L.D. (1999) 3-methyladenine DNA glycosylases: structure, function, and biological importance. *Bioessays*, **21**, 668–676.
6. Manoharan,M., Ransom,S.C., Mazumder,A., Gerlt,J.A., Wilde,J.A., Withka,J.A. and Bolton,P.H. (1988) The characterization of abasic sites in DNA heteroduplexes by site specific labeling with ^{13}C. *J. Am. Chem. Soc.*, **110**, 1620–1622.
7. Wilde,J.A., Bolton,P.H., Mazumder,A., Manoharan,M. and Gerlt,J.A. (1989) Characterization of the equilibrating forms of the aldehydic abasic site in duplex DNA by ^{17}O NMR. *J. Am. Chem. Soc.*, **111**, 1894–1896.
8. Bailly,V. and Verly,W.G. (1988) Possible roles of beta-elimination and delta-elimination reactions in the repair of DNA containing AP (apurinic/apyrimidinic) sites in mammalian cells. *Biochem. J.*, **253**, 553–559.
9. Chen,J. and Stubbe,J. (2004) Synthesis and characterization of oligonucleotides containing a 4′-keto abasic site. *Biochemistry*, **43**, 5278–5286.
10. Zheng,Y. and Sheppard,T.L. (2004) Half-life and DNA strand scission products of 2-deoxyribonolactone oxidative DNA damage lesions. *Chem. Res. Toxicol.*, **17**, 197–207.
11. Loeb,L.A. and Preston,B.D. (1986) Mutagenesis by apurinic/apyrimidinic sites. *Annu. Rev. Genet.*, **20**, 201–230.
12. Boiteux,S. and Laval,J. (1982) Coding properties of poly(deoxycytidylic acid) templates containing uracil or apyrimidinic sites: in vitro modulation of mutagenesis by deoxyribonucleic acid repair enzymes. *Biochemistry*, **21**, 6746–6751.
13. Sagher,D. and Strauss,B. (1983) Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: uniqueness of adenine nucleotides. *Biochemistry*, **22**, 4518–4526.
14. Randall,S.K., Eritja,R., Kaplan,B.E., Petruska,J. and Goodman,M.F. (1987) Nucleotide insertion kinetics opposite abasic lesions in DNA. *J. Biol. Chem.*, **262**, 6864–6870.
15. Boudsocq,F., Iwai,S., Hanaoka,F. and Woodgate,R. (2001) *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. *Nucleic Acids Res.*, **29**, 4607–4616.
16. Kokoska,R.J., Bebenek,K., Boudsocq,F., Woodgate,R. and Kunkel,T.A. (2002) Low fidelity DNA synthesis by a Y family DNA polymerase due to misalignment in the active site. *J. Biol. Chem.*, **277**, 19633–19638.
17. Kokoska,R.J., McCulloch,S.D. and Kunkel,T.A. (2003) The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase η and *Sulfolobus solfataricus* Dpo4. *J. Biol. Chem.*, **278**, 50537–50545.
18. Potapova,O., Grindley,N.D. and Joyce,C.M. (2002) The mutational specificity of the Dbh lesion bypass polymerase and its implications. *J. Biol. Chem.*, **277**, 28157–28166.
19. Xu,Y.J., Kim,E.Y. and Demple,B. (1998) Excision of C-4′-oxidized deoxyribose lesions from double-stranded DNA by human apurinic/apyrimidinic endonuclease (Ape1 protein) and DNA polymerase β. *J. Biol. Chem.*, **273**, 28837–28844.
20. Masuda,Y., Bennett,R.A. and Demple,B. (1998) Dynamics of the interaction of human apurinic endonuclease (Ape1) with its substrate and product. *J. Biol. Chem.*, **273**, 30352–30359.
21. Mol,C.D., Izumi,T., Mitra,S. and Tainer,J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair and coordination. *Nature*, **403**, 451–456.
22. Parikh,S.S., Mol,C.D., Slupphaug,G., Bharati,S., Krokan,H.E. and Tainer,J.A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.*, **17**, 5214–5226.
23. Hosfield,D.J., Guan,Y., Haas,B.J., Cunningham,R.P. and Tainer,J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397–408.
24. Hoehn,S.J., Turner,C.J. and Stubbe,J. (2001) Solution structure of an oligonucleotide containing an abasic site: evidence for an unusual deoxyribose conformation. *Nucleic Acids Res.*, **29**, 3413–3423.
25. Chen,J., Dupradeau,F.-Y., Case,D.A., Turner,C.J. and Stubbe,J. (2007) Nuclear magnetic resonance structural studies and molecular modeling of duplex DNA containing normal and 4′-oxidized abasic sites. *Biochemistry*, **46**, 3096–3107.
26. Drohat,A.C., Jagadeesh,J., Ferguson,E. and Stivers,J.T. (1999) Role of electrophilic and general base catalysis in the mechanism of *Escherichia coli* uracil DNA glycosylase. *Biochemistry*, **38**, 11866–11875.
27. Borgias,B.A. and James,T.L. (1989) Two-dimensional nuclear Overhauser effect: complete relaxation matrix analysis. *Methods Enzymol.*, **176**, 169–183.
28. James,T.L., Borgias,B.A., Bianucci,A.M. and Zhou,N. (1990) Determination of DNA and protein structures in solution via complete relaxation matrix analysis of 2D NOE spectra. *Basic Life Sci.*, **56**, 135–154.
29. Case,D.A., Darden,T.E., Cheatham,T.E. III, Simmerling,C.L., Wang,J., Duke,R.E., Luo,R., Merz,K.M., Wang,B. *et al.* (2004) *Amber 8*. University of California, San Fransisco, San Fransisco, USA.
30. Smith,J.A., GomezPaloma,L., Case,D.A. and Chazin,W.J. (1996) Molecular dynamics docking driven by NMR-derived restraints to determine the structure of the calicheamicin gamma(I)(1) oligosaccharide domain complexed to duplex DNA. *Magn. Reson. Chem.*, **34**, S147–S155.
31. Dupradeau,F.Y., Case,D.A., Yu,C., Jimenez,R. and Romesberg,F.E. (2005) Differential solvation and tautomer stability of a model base pair within the minor and major grooves of DNA. *J. Am. Chem. Soc.*, **127**, 15612–15617.
32. Wijmenga,S.S., Mooren,M.W. and Hilbers,C.W. (1993) In Roberts,G.C.K. (ed), *NMR of Macromolecules: a Practical Approach*. Oxford University Press, Oxford, UK, pp. 217–283.
33. Bax,A. and Lerner,L.M. (1988) Measurement of ^{1}H-^{1}H coupling-constants in DNA fragments by 2D NMR. *J. Magn. Reson.*, **130**, 296–299.

34. Sklenar,V. and Bax,A. (1987) Measurement of H-1-P-31 NMR coupling-constants in double-stranded DNA fragments. *J. Am. Chem. Soc.*, **109**, 7525–7526.
35. Goljer,I., Kumar,S. and Bolton,P.H. (1995) Refined solution structure of a DNA heteroduplex containing an aldehydic abasic site. *J. Biol. Chem.*, **270**, 22980–22987.
36. Beger,R.D. and Bolton,P.H. (1998) Structures of apurinic and apyrimidinic sites in duplex DNAs. *J. Biol. Chem.*, **273**, 15565–15573.
37. Wang,K.Y., Parker,S.A., Goljer,I. and Bolton,P.H. (1997) Solution structure of a duplex DNA with an abasic site in a dA tract. *Biochemistry*, **36**, 11629–11639.
38. Barsky,D., Foloppe,N., Ahmadia,S., Wilson,D.M.III and MacKerell,A.D. Jr (2000) New insights into the structure of abasic DNA from molecular dynamics simulations. *Nucleic Acids Res.*, **28**, 2613–2626.
39. Ling,H., Boudsocq,F., Woodgate,R. and Yang,W. (2004) Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts. *Mol. Cell*, **13**, 751–762.
40. Wilson,D.M.III, Takeshita,M., Grollman,A.P. and Demple,B. (1995) Incision activity of human apurinic endonuclease (APE) at abasic site analogs in DNA. *J. Biol. Chem.*, **270**, 16002–16007.
41. Junker,H.D., Hoehn,S.T., Bunt,R.C., Marathius,V., Chen,J., Turner,C.J. and Stubbe,J. (2002) Synthesis, characterization and solution structure of tethered oligonucleotides containing an internal 3′-phosphoglycolate, 5′-phosphate gapped lesion. *Nucleic Acids Res.*, **30**, 5497–5508.