Abasic Sites in Duplex DNA: Molecular Modeling of Sequence-Dependent Effects on Conformation

Leila Ayadi,* Christian Coulombeau,* and Richard Lavery#

*Laboratoire d'Etudes Dynamiques et Structurales de la Sélectivité, Laboratoire de Chimie Bioorganique, Université Joseph Fourier, F-38041 Grenoble, Cedex 9, France; and #Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, Paris 75005, France

ABSTRACT Molecular modeling calculations using JUnction Minimization of Nucleic Acids (JUMNA) have been used to study sequence effects on the conformation of abasic sites within duplex DNA. We have considered lesions leading to all possible unpaired bases (X), adenine, guanine, cytosine, or thymine contained within two distinct sequence contexts, CXC and GXG. Calculations were carried out on DNA 11-mers using extensive conformational search techniques to locate the most stable abasic conformations and using Poisson-Boltzmann corrected electrostatics to account for solvation effects. The results, which are in very good agreement with available experimental data, point to strong sequence effects on both the position of the unpaired base (intra or extrahelical) and on the overall curvature induced by the abasic lesion. For CXC, unpaired purines are found to lie within the helix, while unpaired pyrimidines are either extrahelical or in equilibrium between the intra and extrahelical forms. For GXG, all unpaired bases lead to intrahelical forms, but with marked, sequence-dependent differences in induced curvature.

INTRODUCTION

The loss of a purine or pyrimidine base, resulting in the formation of an apurinic or apyrimidinic site (AP site), is the most frequent lesion that occurs in DNA. Such AP sites can be formed by spontaneous hydrolysis of the N-glycosidic bond (Lindahl and Nyberg, 1972), by chemical modifications of nucleic bases, or by physical agents (UV or γ radiation) (Lindahl, 1993). They are also generated enzymatically as intermediates in the course of the repair of modified or abnormal bases (Lindahl, 1993; Loeb and Preston, 1986; Wallace, 1988; Demple and Harrison, 1994).

AP sites are both cytotoxic and highly promutagenic due to a lack of coding information (Loeb and Preston, 1986; Boiteux and Laval, 1982). The repair of AP sites is thus essential to cell viability. AP sites are specifically repaired by a base excision repair pathway (Demple and Harrison, 1994; Barzilay and Hickson, 1995; Doetsch, 1995). This mechanism is initiated by the recognition of the AP site by an AP endonuclease, which cleaves the phosphodiester backbone 5′ to the lesion, leaving a 3′-hydroxyl nucleotide and deoxyribose 5′-phosphate as termini. The repair is completed by deoxyribose-phosphodiesterase, a DNA polymerase that replaces the missing nucleotide and a DNA ligase that joins the phosphodiester backbone.

It is remarked that an abasic site is not a chemically unique species, but rather an equilibrium of four forms: α- and β-hemiacetals (2-deoxy-D-*erythro*-pento-furanoses), aldehyde, and hydrated aldehyde. The aldehydic form, however, represents $<1\%$ of the mixture (Manoharan et al., 1988; Wilde et al., 1989). In many studies (Millican et al., 1984; Takeshita et al., 1987), this intrinsic instability of the abasic site is avoided using a 3-hydroxy-3 (hydroxymethyl) tetrahydrofuran analog, and this choice has also been made in the present modeling study. The reader is also referred to a recent NMR study of abasic bulges, which was also carried out with a tetrahydrofuran analog (Lin et al., 1998).

How AP endonucleases recognize and specifically bind to AP sites is still unknown, mainly due to the lack of structural information on the perturbations introduced by the site. It is thus uncertain which features of abasic sites stimulate repair. Solving this problem clearly requires information on how variable the conformation of an AP site is as a function of the missing base and of its sequence context.

Available solution studies on duplexes containing unpaired purines, flanked by a variety of neighboring basepairs, show that these bases remain stacked within the helix and that the double helix conserves a regular right-handed geometry (Cuniasse et al., 1987, 1990; Kalnik et al., 1988, 1989). However, it has been difficult to characterize the detailed consequences of such AP sites.

By using NMR spectroscopy, Cuniasse et al. (1990) have shown that an unpaired thymine flanked by two cytosines (facing an abasic site analog) is in equilibrium between intra and extrahelical positions and sensitive to temperature effects. By using a combination of NMR spectroscopy and restrained molecular dynamics (rMD) simulations, Coppel et al. (1997) found that when the same unpaired base is flanked by two guanines an intrahelical conformation with a

Received for publication 2 April 1999 and in final form 11 August 1999.

Address reprint requests to Dr. Christian Coulombeau, LEDSS, Laboratoire de Chimie Bioorganique, UMR CNRS 5616, Université Joseph Fourier, Grenoble 1, BP 53X, F-38041 Grenoble, Cedex 9, France. Tel.:33-4-76-51-44-34; Fax: 33-4-76-51-43-82; E-mail: Christian.Coulombeau@ujf-grenoble.fr.

The structures discussed in this article are available from Prof. Christian Coulombeau in the form of PDB files.

0006-3495/99/12/3218/09 $2.00

moderate kink is stabilized. It is remarked that a moderate 20° kink is also found for this sequence with a ribonolactone abasic site (Drogue et al., unpublished results). An NMR/rMD study of this lesion (Jourdan et al., 1999) led to only 10° of kinking, but it should be noted that the kink angle cannot be determined directly from the NMR data and is difficult to define with precision.

Other evidence of sequence effects comes from thermodynamic studies of AP sites (Shida et al., 1994; Ide et al., 1995), which have underlined either the impact of the unpaired base or, in a recent comprehensive study of 13-mers containing tetrahydrofuran analog sites, the impact of the basepairs neighboring the AP site (Gelfrand et al., 1998). These contrasting, but partial, results clearly point to the need for further investigations.

We have already used internal coordinate molecular mechanics techniques to study the base opening pathway in an oligomer containing an abasic site (Ayadi et al., 1997). This study detected the experimentally observed extrahelical minimum and confirmed its stability with respect to an intrahelical conformation. In the present work we extend our modeling to all possible abasic sites (that is, all possible unpaired bases facing a tetrahydrofuran AP site analog) situated between two CG or two GC flanking pairs. This implies calculations on eight different abasic duplexes. In all cases, we have used extensive search procedures available within our modeling program to ensure that the most stable conformations are located. Solvent effects are taken into account using Poisson-Boltzmann (PB) electrostatics, which have already shown to be effective in obtaining a correct ordering of the stability of nucleic acid conformers in aqueous solution (Zakrzewska et al., 1996; Zacharias and Sklenar, 1997, 1999). It is remarked that the approach used here can successfully predict sequence-dependent curvature in both natural (Sanghani et al., 1996a) and locally modified (Sanghani et al., 1996b) DNAs, although we take into account neither explicit counterions nor explicit water molecules.

The calculations we have carried out involve two series of AP sites described by the abbreviations CXCa and GXGa. Here X indicates the unpaired base and the subscript indicates the presence of an AP site, O being the analog of the abasic site. The sequences and the numbering used throughout this article are listed below.

(a) Unpaired bases flanked by CG basepairs:

$$5'\text{-}G_1\ C_2\ G_3\ T_4\ G_5\ O_6\ G_7\ T_8\ G_9\ C_{10}G_{11}\text{-}3'$$
$$3'\text{-}C_{12}G_{13}C_{14}A_{15}C_{16}X_{17}C_{18}A_{19}C_{20}G_{21}C_{22}\text{-}5'$$

CACa	X=A
CGCa	G
CTCa	T
CCCa	C

(b) Unpaired bases flanked by GC basepairs:

$$5'\text{-}C_1\ G_2\ C_3\ A_4\ C_5\ O_6\ C_7\ A_8\ C_9\ G_{10}C_{11}\text{-}3'$$
$$3'\text{-}G_{12}C_{13}G_{14}T_{15}G_{16}X_{17}G_{18}T_{19}G_{20}C_{21}G_{22}\text{-}5'$$

GAGa	X=A
GGGa	G
GTGa	T
GCGa	C

METHODOLOGY

All the structures studied were energy-optimized with the JUnction Minimization of Nucleic Acids (JUMNA) algorithm, which has been the subject of a number of previous publications (see, for example, Lavery, 1988; Lavery et al., 1995). JUMNA models DNA by breaking each strand down into a series of 3′-monophosphate nucleotides. The nucleotides are positioned in space with respect to a common helical axis system using three translational variables (Xdisp, Ydisp, rise) and three rotational variables (inclination, tip, twist), while internal coordinates (dihedral angles and valence angles within sugar rings and along the phosphodiester backbone) are used to describe the internal flexibility of each nucleotide. All bond lengths are kept fixed. Sugar ring closure (C4′–O4′ bonds) and internucleotide junctions closure (C5′–O5′ bonds) are ensured by harmonic energy penalty functions. This hybrid internal/helicoidal representation enables efficient energy minimization and allows a versatile control over the local and global characteristics of the nucleic acid conformation.

Initial optimization for each sequence studied used B-DNA fiber coordinates (Arnott et al., 1980). These conformations were then refined in two steps. The first step consisted of applying various distortions and then freely reminimizing. In the present study, the distortions included the helical twist (which was varied between 24 and 45°) and the rise (which was varied between 2.5 and 4.3 Å). Sugar puckers were also searched using the phase angle with ranges of 100–178° for purines and 85–145° for pyrimidines, coupled with amplitude variations of 35–47°.

Two other types of deformation were also imposed: 1) superhelical constraints within JUMNA (Sanghani et al., 1996a, b) were used to vary the curvature of the sequences investigated by inserting them within repeating polymers having 10-nucleotide repeats. After curvature, the resulting conformations were used as starting conformations for reconstructing symmetry-unconstrained oligomers; 2) base opening constraints were used to generate extrahelical conformations by changing the angle of the appropriate glycosidic bond with respect to the basepair C1′–C1′ axis (after projection into the plane perpendicular to the local helical axis) (Bernet et al., 1997).

The second step of the procedure involved using all the minimal energy conformations located for a given sequence as starting points for all other sequences investigated. This “crossing-over” was repeated until no new conformations were located after energy minimization. This lengthy procedure has the advantage of thoroughly scanning conformational space and of ensuring that the final conformations are the most stable minima for each sequence, independently of the starting conformations employed.

The JUMNA program uses the Flex force field, which includes electrostatic and Lennard-Jones terms between nonbonded atoms (including an angular dependent hydrogen bonding contribution), in addition to dihedral and valence angle contributions. To deal with environmental effects, counterions are modeled by reducing the net charge on each phosphate to −0.5e, while solvent electrostatic damping is modeled by introducing a sigmoidal distance dependence of the dielectric function (Hingerty et al., 1985). Although this simple approach is rapid and quite successful in reproducing DNA conformation in solution (Fritsch and Westhof, 1990; Poncin et al., 1992), it does not treat the modifications in solvation enthalpy that can occur during major conformational changes. Earlier studies have shown that these effects can be treated more correctly using PB electrostatics (Zakrzewska et al., 1996; Bernet et al., 1997; Zacharias and Sklenar, 1997, 1999). PB calculations were consequently performed here using the DelPhi program (Gilson et al., 1987; Gilson and Honig, 1988; Sharp and Honig, 1990; DelPhi, *Biosym/MSI*, 9685 Scranton Road, San Diego, CA 9121-2777), in three “focusing” steps using successively smaller boxes around the DNA oligomer. In the first of these runs, 30% of the box is filled by the oligomer and Coulomb boundary conditions are used. For the

following runs, 60 and 90% of the box is filled. Solute polarizability is mimicked by using an internal dielectric constant of 2, while the external dielectric constant is set to 80. A probe sphere of 1.05 Å was used to define the solvent-accessible surface. The resulting electrostatic energy (Coulomb and reaction field) was then used in place of the electrostatic energy calculated with the Flex force field to derive a corrected total conformational energy.

The conformational parameters used to describe the minimal energy conformations follow the Cambridge convention for nucleic acids (Dickerson et al., 1989) and were calculated using the CURVES algorithm (Lavery and Sklenar, 1988, 1989, 1996). The results presented below include both local variables (describing the relative position of successive basepairs steps) and global variables (which relate the bases to an overall helical axis). Molecular structures were displayed using Insight II (Insight II 97.0, *Biosym/MSI*, 9685 Scranton Road, San Diego, CA 9121-2777).

RESULTS

We begin by discussing the stability and the overall structure of the abasic site-containing oligomers before turning to the details of the conformational perturbations induced by the abasic site as a function of sequence. Following the procedures described above, we searched for the most stable conformations of each abasic oligomer using a large number of starting structures. The total energy for each conformation was then corrected for solvent contributions by using PB electrostatics. The resulting energies were classed and all conformations having energies >15 kcal/mol above the most stable conformation for each sequence were discarded. The results are shown in Table 1.

These stable conformations have been classified according to their global conformational features. The intrahelical conformations were found to differ by their global curvature angle and were classified into three families: (I) slightly curved, with an angle of curvature varying from 0 to 20°; (II) moderately curved, with angles varying from 20 to 40°; and (III) strongly curved, with angles varying from 40 to 60°. Extrahelical conformations were classified as a function of the position of the extrahelical base: (IV) with the unpaired base in the major groove and (V) with the unpaired base in the minor groove.

TABLE 1 Relative energies (kcal/mol) of the abasic sites as a function of sequence and conformation after PB electrostatic corrections

	State				
Sequence	I	II	III	IV	V
CAC	**0.0**	**0.7**	3.0	13.9	9.4
CGC	11.6	**0.0**	5.9	7.0	9.1
CTC	**0.7**	—	4.1	1.6	**0.0**
CCC	8.5	5.7	5.6	7.1	**0.0**
GAG	**0.0**	1.9	4.1	5.7	6.6
GGG	**0.0**	1.6	7.0	6.3	7.3
GTG	4.4	**0.0**	7.7	11.1	—
GCG	6.5	2.8	**0.0**	—	—

States I–III have the unpaired base stacked into the helix with increasing degrees of curvature, while states IV and V have the unpaired base unstacked into the major and minor grooves, respectively. The states with energies within 1 kcal/mol of the minimum are shown in bold type.

We begin by looking at the CXCa series, where the unpaired base falls between two pyrimidines (C, C) and the abasic site between two purines (G, G). This series shows a clear difference between unpaired purines, which prefer to be stacked within the helix and unpaired pyrimidines, which favor extrahelical conformations. There are, however, considerable differences within these categories as a function of sequence. For CACa, two conformations with either slight ($CACa_I$, family I) or moderate ($CACa_{II}$, family II) curvature clearly have acceptable energies (separated by only 0.7 kcal/mol), although the most stable conformation belongs to family I. In contrast, CGCa only exhibits a single low energy conformation with moderate curvature.

For the unpaired pyrimidines, although the most stable conformation has the unpaired base in the minor groove for both CTCa and CCCa, the former sequence is characterized by a second stable conformation $CTCa_I$ with an intrahelical base and slight curvature. This conformation is only 0.7 kcal/mol less stable than the extrahelical conformation.

Turning to the GXGa series, where the unpaired base lies between two purines (G, G) and the abasic site between two pyrimidines (C, C), we see that there is a clear preference for intrahelical conformations, although, once again, there are sequence-related effects. Both unpaired purine sequences, GAGa and GGGa, show only slight curvature in their energetically preferred states. In contrast, the unpaired pyrimidines show either moderate or strong curvature for GTGa and GCGa, respectively. In all of these cases the energy differences with the other families of intrahelical conformation and with the extrahelical conformations are relatively large. It is remarked that for both (GTGa, GCGa) sequences no minor groove extrahelical conformations were found within our imposed 15 kcal/mol energy range and, for the latter sequence, no major groove extrahelical conformation satisfied this criterion.

If we consider that conformations within 1 kcal/mol of the most stable conformation are likely to be accessible to the abasic oligomers in solution (bold entries in Table 1), then we can summarize our findings by noting that in the CXCa series, unpaired purines will be intrahelical, while unpaired pyrimidines will be either extrahelical (for CCCa) or in equilibrium between intra and extrahelical conformations (CTCa). These conclusions are in agreement with experimental observations. For adenine between two cytosines (CACa), Cuniasse et al. (1987) has derived two conformations using NMR/rMD methods where the unpaired base (and the abasic linkage) lie inside the helix. These two structures are close in stability and both agree with the NMR data. They differ mainly in the puckering of the abasic sugar. The same authors found that an unpaired guanine flanked by two cytosines (CGCa) was also in an intrahelical position, although this base showed less stacking than that found with unpaired adenine. Within the same sequence context, unpaired thymine (CTCa) showed a temperature-dependent equilibrium between intrahelical and extrahelical forms, while unpaired cytosine (CCCa) was found to be exclusively extrahelical.

For the GXGa series, the same energy criterion of 1 kcal/mol suggests that all unpaired bases will be intrahelical, although the degree of curvature will depend on their nature. Experimental results concerning this series are only available for the case of GTGa, where Coppel et al. (1997) have used NMR/rMD methods to show that the unpaired thymine stacks within the helix and leads to a kink of ~30°. This is in very good agreement with our findings and increases our confidence in the prediction of intrahelical conformations for the remaining members of the series.

Before ending the discussion of stability, it is worth stressing that the PB correction of the electrostatic energies is an essential step in obtaining reliable predictions. Although the overall range of the uncorrected relative energies shown in Table 2 is similar to that of the corrected energies, the preferred conformations show a number of important differences. In disagreement with the experimental results mentioned above, these values suggest that CGCa should be in equilibrium between intra and extrahelical conformations. However, again in disagreement with experiment, they do not support and intra/extrahelical equilibrium in the case of CTCa. The uncorrected results also generally favor intrahelical states with more significant curvature than those found after PB correction. All these changes can be traced to the fact that exposing the hydrophobic faces of a base leads to an energy penalty with the PB correction, but not with a simple distance-dependent dielectric function.

Conformational details

The most stable conformations of each sequence are illustrated in Fig. 1. A number of general remarks can be made concerning the structural consequences of an abasic site. First, structural deformations are largely restricted to the level of the abasic site itself and to the two neighboring basepair steps. In the case of the intrahelical conformations, curvature is thus a consequence of local kinking, which can occur on either side of the unpaired base or, for moderately or strongly kinked conformations, at the following or preceding basepair steps. The details as a function of sequence are given in Table 3.

TABLE 2 Relative energies (kcal/mol) of the abasic sites as a function of sequence and conformation before PB electrostatic corrections

	State				
Sequence	I	II	III	IV	V
CAC	1.3	3.6	**0.0**	21.2	9.2
CGC	14.4	**0.7**	8.4	3.4	**0.0**
CTC	8.5	—	7.5	4.6	**0.0**
CCC	19.0	14.0	13.0	9.6	**0.0**
GAG	1.8	1.1	**0.0**	**0.4**	9.1
GGG	2.6	**0.0**	5.8	5.1	6.9
GTG	4.8	**0.0**	8.1	13.9	—
GCG	10.5	3.8	**0.0**	—	—

Conformational states I–V are as defined in Table 1. The states with energies within 1 kcal/mol of the minimum are shown in bold type.

The overall angle of curvature (Table 3, *top*) varies from ~10° to 20° for the least-curved intrahelical conformations and for the stable extrahelical conformations, to >50° in the case of the most strongly distorted intrahelical sequence, GCGa. The direction of curvature nevertheless is always toward the major groove at the level of the abasic site, although in the GXGa series this curvature is somewhat biased in the direction of the strand containing the abasic site. For the extrahelical conformations, extrusion of the unpaired base is accompanied by extrusion of the abasic backbone linkage and the reestablishment of stacking interactions between the basepairs on either side of the abasic site. Such conformations thus generally show only moderate kinking or overall curvature (18° for CTCa and 17° for CCCa). In some cases, the extruded base and backbones of the extrahelical conformations are stabilized by additional hydrogen bonds with the rest of the double helix, as in the case of the most stable CTCa conformation, which shows a T(HN3)-G9(O3′) base-backbone interaction and of CCCa, which shows a C(H1N4)-T8(O2) base-base interaction.

As shown in Table 3, *bottom*, the number of significant kinks increases as the overall curvature of the oligomer becomes more important. These kinks always occur at one of the steps adjacent to the unpaired base, but they may also spread out to next-neighboring steps on the 3′-side (as in case of GTGa) or on the 5′-side (as in the case of GCGa).

A CURVES (Lavery and Sklenar, 1996) analysis of the stable abasic oligomer conformations shows that, despite local deformations at the abasic site, all the oligomers remain within the B-DNA family with average basepair Xdisp values between −1.9 and −1.2 Å, average basepair inclinations between −5° and 8° (for all but certain terminal basepairs), and average rise and twist of 3.1–4.1 Å and 31–43°, respectively. The impact of the abasic site is mainly seen in the rotational intra-basepair parameters, buckle and propeller. Changes with respect to the range of parameters found for normal repeating sequence oligomers (Poncin et al., 1992) occur notably for propeller values (33–45°) for the pairs on the 3′-side of the abasic site in the optimal GAGa, GGGa, and GTGa conformations. It is remarked that large propeller values were also found in the NMR/rMD structures of a GTGa site (Coppel et al., 1997). These values correspond to a shift of the C7 residue into the space generated by the abasic site and can be accompanied by the formation of inter-basepair hydrogen bonds involving C7 and the unpaired base on the opposing strand. This is true for both the GGGa and GTGa (which show characteristic large negative buckles 3′ to the abasic site), but not for GAGa. Such interactions also occur for the strongly kinked GCGa conformation, but without the creation of large propeller values.

An analysis of the backbone conformations also shows that large changes are restricted to the vicinity of the abasic site. The most significant changes occur in α (C3′-O3′-P-O5′) for the extrahelical conformations of the unpaired base and the abasic nucleotide preferred by the CTCa and CCCa

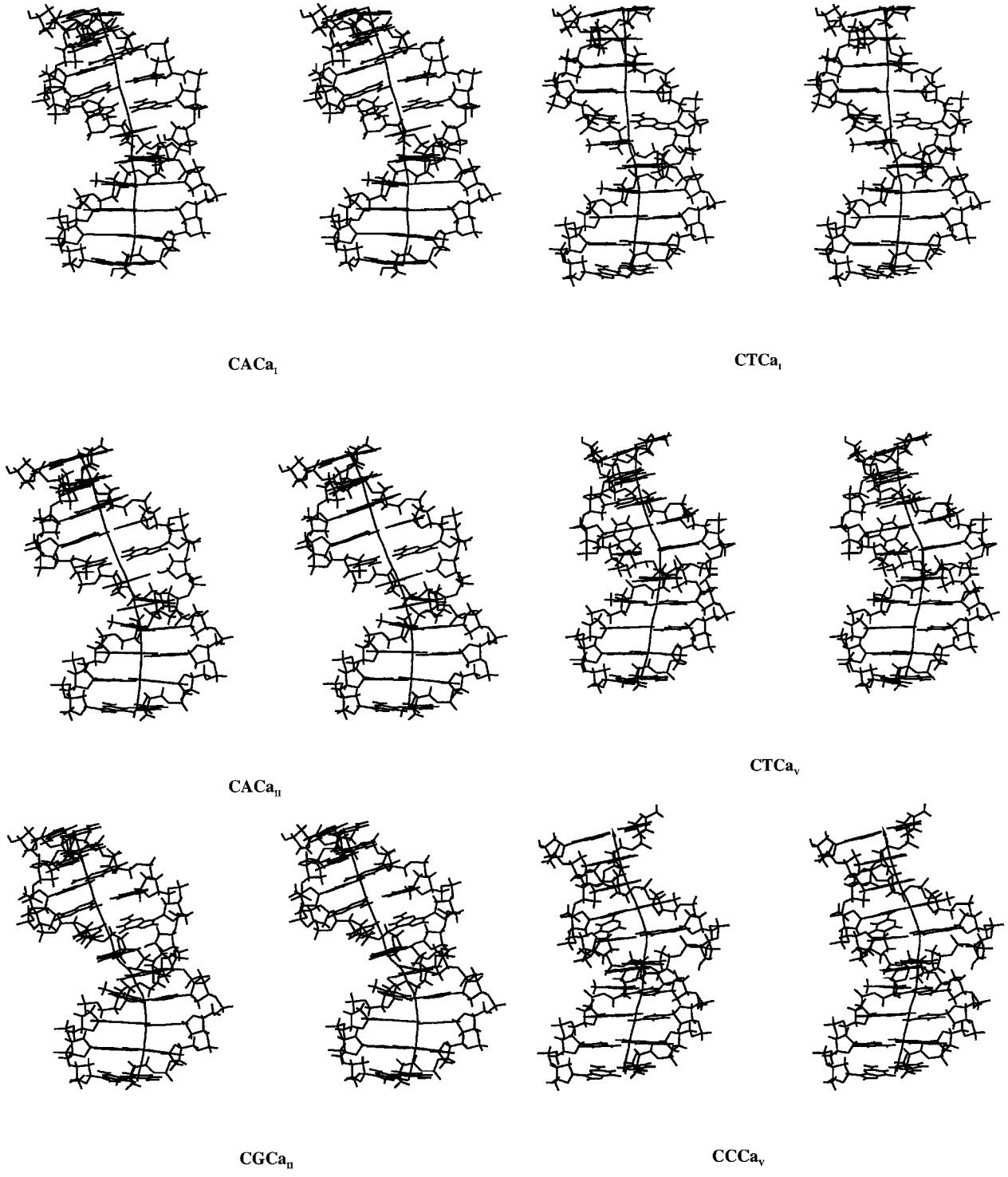

FIGURE 1 Stereoviews of the most stable conformations of the abasic oligomers belonging to the CXCa and GXGa series. Conformations within 1 kcal/mol of the optimal state are illustrated (see Table 1). In each case the unpaired base is at the center of the strand closest to the viewer, and the major groove at the center of the oligomer lies on the left-hand side of the figure.

oligomers. Note that these changes, which seem to be directly coupled to swinging out the unpaired base, are not accompanied by changes in γ as in the well-known "crankshaft" states. Of the other backbone torsions ζ is the most variable, but changes in its value are sequence-related and not restricted to residues at or around the abasic site. It is also remarked that all glycosidic angles remain in the *anti* domain, including those of the extrahelical unpaired bases. Most sugar puckers remain in the canonical C2′-endo conformation; however, C3′-endo puckers are found in both strands, 5′ to the abasic level of (GTGa, GGGa) and in the unpaired base strand of CACa. A number of O4′-endo

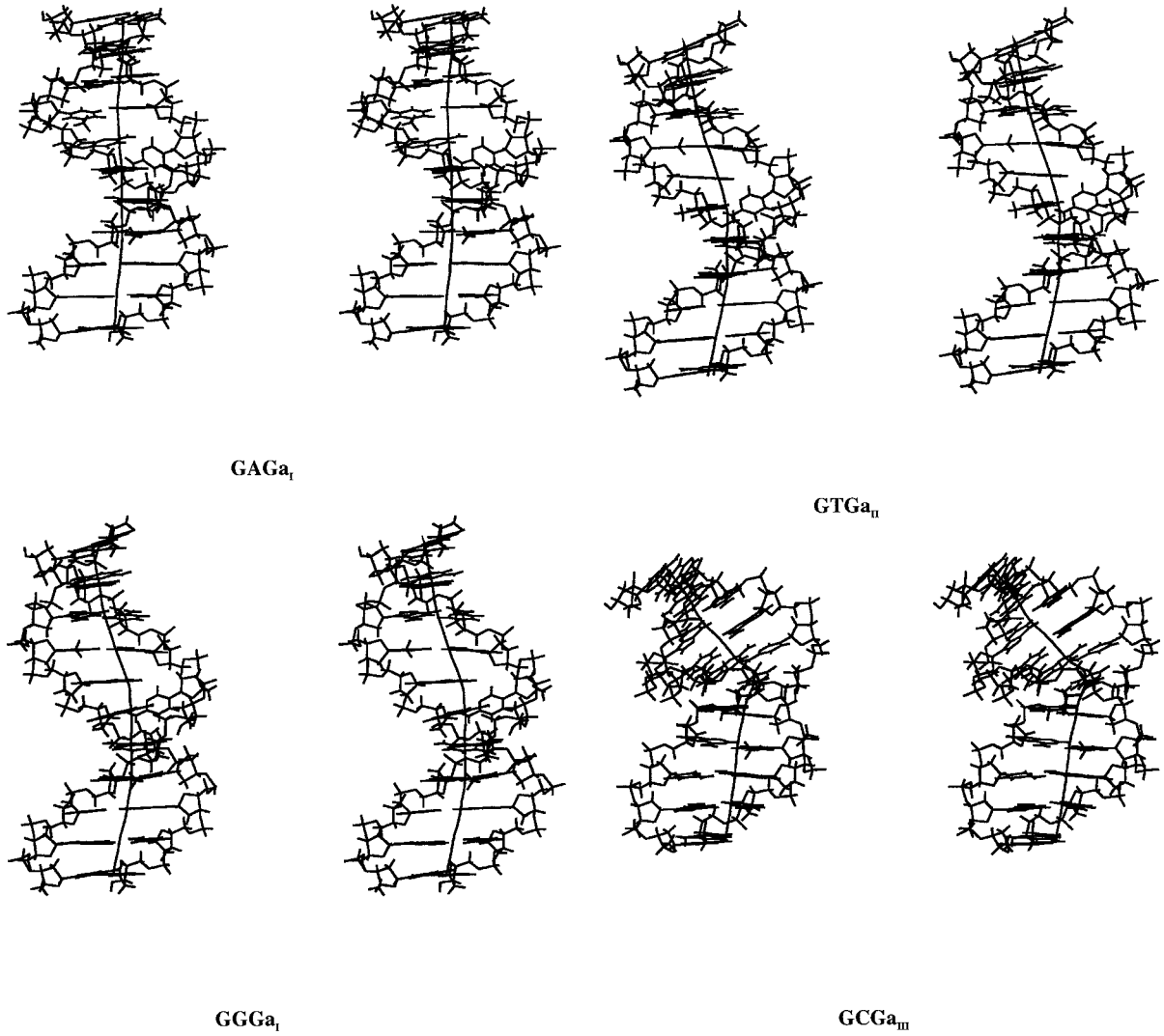

FIGURE 1 Continued.

conformations are also seen one or two residues away from the abasic sites in both the GXGa and CXCa series.

Base-stacking

The components of the base-stacking interactions for both the unpaired base facing the abasic site and the basepairs flanking the abasic site are presented in Table 4 and defined in Fig. 2. An analysis of these data for the oligomers currently studied shows that the sequence-dependent stacking patterns are globally very like those found in earlier modeling of helically symmetric repeating-sequence DNA (Poncin et al., 1992). Several effects of the overall conformation of the abasic sites are also clearly visible. In the case of the extrahelical conformations of CTCa and CCCa there is significant stacking between the neighboring basepair (components 5 + 6 + 7 + 8 in Table 4) but, naturally, little stacking for the expelled base. For all the intrahelical conformations, the stacking of the basepairs flanking the abasic site is rather weak, although the interaction between the two cytosines on either side of the abasic site (component 6 in Table 4) in the GXGa conformations is significant in all but the most strongly curved case (GCGa). In this same family, GAGa stands out for the poor stacking of its unpaired base due, in this case, principally to the deformation of the basepairs flanking the abasic site rather than to the overall deformation of the duplex.

DISCUSSION AND CONCLUSIONS

By using a careful molecular mechanics approach combining controlled helical deformations and combinatorial backbone conformational searching, we have been able to locate the most stable conformations of DNA 11-mers containing single abasic sites (or, more precisely, their tetrahydrofuran analogs) in two different sequence contexts, generating an unpaired base between two cytosines CXCa or between two guanines GXGa. We have also tried to account reasonably for solvation effects by using an electrostatic term obtained by solution of the PB equation.

TABLE 3 Conformational details of the stable abasic oligomer conformations

Sequence	State I	State II	State III	State IV	State V
	Overall angle of curvature (°)				
CAC	11	28			
CGC		32			
CTC	15				18
CCC					17
GAG	19				
GGG	17				
GTG		26			
GCG			52		
	Kink sites				
CAC	—	G5pO6			
CGC		G5pO6			
		O6pG7			
CTC	—				—
CCC					—
GAG	—				
GGG	C5pO6				
	O6pC7				
GTG		C5pO6			
		O6pC7			
		C7pA8			
GCG			A4pC5		
			C5pO6		
			O6pC7		

Conformational states I–V are as defined in Table 1.

The results show that whatever the nature of the unpaired base, and for both the sequence contexts studied, the abasic site has only a local impact on the conformation of the DNA double helix. With only rare exceptions, the parameters of the phosphodiester backbones remain in the domain expected for the B-family DNA and, similarly, the abasic lesion does not appear to be linked to any characteristic change in sugar pucker. However, it is found that sequence effects clearly govern the location of the unpaired base (intra or extrahelical) and the extent of induced axial curvature.

Curvature, which is the most important global perturbation introduced by the abasic site, occurs toward the major groove at the level of the abasic site. The degree of curvature depends both on the nature of the abasic site and of the flanking basepairs sequences. For the CXCa series, an unpaired purine is found to be intrahelical, leading to slight or moderate overall curvature. In agreement with the experimental results of Cuniasse et al. (1990), poorer stacking is found for CGCa than for CACa. For an unpaired pyrimidine we find moderately curved conformations with either an extrahelical unpaired base, in the case of CCCa, or an equilibrium between extra and intrahelical forms, in the case of CTCa. This is again in accord with experimental results. It is also confirmed that extrahelical unpaired bases lie in the grooves, eventually forming interactions with neighboring bases or phosphate residues (particularly in the case of the less soluble cytosine residue) (Joshua-Tor et al., 1988; Miller et al., 1988; Morden et al., 1990). Extrahelical conformations, which are created principally by changes in the backbone dihedral α (in contrast to earlier assumptions; Letellier et al., 1991; Cuniasse et al., 1989) lead to stacking between the basepairs on either side of the abasic site and thus to only slight curvature.

For the GXGa series, all unpaired bases remain stacked within the double helix, but sequence strongly influences overall curvature. Curvature is larger with unpaired pyrimidines than with unpaired purines. For the unpaired purines, a movement of the C7 residue on the 3′-side of the abasic site into the abasic cavity is seen and this effect was also seen in the case of GTGa. The only experimental data in this series concern the latter sequence, where NMR/rMD and our results agree in assigning a moderately curved intrahelical conformation (Ayadi et al., 1999), but disagree in the extent to which the unpaired base moves into the abasic cavity. The GCGa sequence is unique in the present investigation in showing a very marked overall curvature (52°). These marked sequence effects can clearly be related to base-stacking around the abasic site, with poor stacking (or, conversely, higher solubility) of the unpaired base and strong stacking of the bases neighboring the abasic site contributing to a partial displacement of the unpaired base and enhanced axis curvature.

TABLE 4 Stacking energies

Sequence	Component 1	2	3	4	1 + 2 + 3 + 4	5	6	7	8	5 + 6 + 7 + 8
$CACa_I$	−2.5	−4.7	−0.5	−6.8	−14.5	0.1	0.2	0.2	−0.6	−0.1
$CACa_{II}$	−3.3	−4.9	−0.5	−6.9	−15.6	0.0	0.6	0.2	−0.4	0.4
$CGCa_{II}$	−6.0	−9.1	−2.8	−9.5	−27.4	−0.1	1.1	0.3	−0.5	0.8
$CTCa_I$	−3.0	−0.4	−2.9	−4.6	−10.9	−0.1	−0.3	0.5	−0.3	−0.2
$CTCa_V$	0.3	−0.5	−0.2	−2.3	−2.7	1.1	−3.0	−2.7	−2.5	−7.1
$CCCa_V$	−0.1	−0.1	0.1	−0.1	−0.2	0.7	−2.2	−0.9	−2.3	−4.7
$GAGa_I$	−2.6	−0.3	−6.7	0.0	−9.6	0.2	−3.8	−0.3	−0.6	−4.5
$GGGa_I$	−1.2	−1.7	1.2	−19.4	−21.1	0.0	−7.0	−0.6	−0.2	−7.8
$GTGa_{II}$	−2.7	−1.2	−5.0	−10.2	−19.1	−0.1	−7.6	−0.6	−0.2	−8.5
$GCGa_{III}$	−1.9	−7.9	−7.8	1.3	−16.3	−1.0	0.2	0.0	0.2	−0.6

Components 1–4: stacking between the unpaired base and the flanking basepairs. Components 5–8: stacking between the flanking basepairs (see Fig. 3).

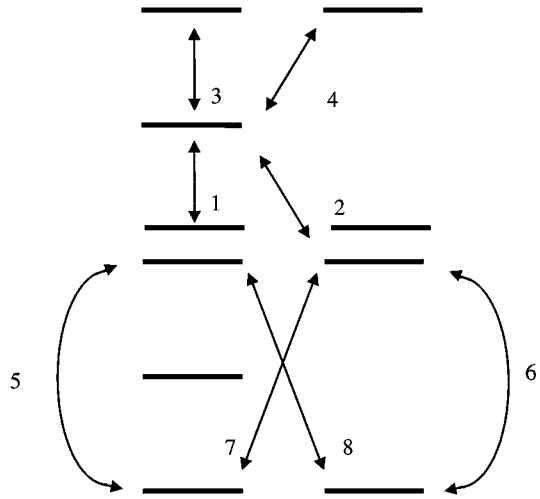

FIGURE 2 Schematic diagram of stacking interactions used in Table 4. The interactions 1–4 (*top*) involve the unpaired base, while the interactions 5–8 (*bottom*) involve the basepairs flanking the abasic site. Stacking energies were calculated as the sum of the nonbonded interactions (Lennard-Jones and electrostatic) between the corresponding base atoms.

In line with the suggestions of Gelfrand et al. (1998) we also see several possible structures for the cavity created by an abasic lesion. When the unpaired base stacks well within the double helix (CACa, CGCa, and intrahelical CTCa), the cavity largely remains open. In contrast, for extrahelical conformations, the gap closes by stacking between the neighboring pairs. Intermediate cases occur when the unpaired base partially shifts into the cavity, but our modeling points rather to a deformation of the 3′-side neighboring basepair (seen with GTGa, GAGa, and GGGa). We should, however, be cautious that this effect may be due to the absence of explicit water molecules in our modeling.

We lastly turn to the biological implications of this study and to the mechanisms by which AP endonucleases may recognize AP sites. Apart from the presence of a particular ring structure for the damaged nucleotide (Hare et al., 1983) it seems clear that this mechanism needs to be able to deal with changes linked to the AP site targeted and to its sequence context. Thus, in certain cases, a clear abasic cavity or an extrahelical unpaired base (Mol et al., 1995; Gorman et al., 1997; Shida et al., 1996) may be the trigger, but in others our results would support the idea of a discriminating role for axis curvature derived from sharp kinks at or close to the AP site (Miaskiewicz et al., 1996; Greger and Kemper, 1998). Such interpretations of binding specificity remain essentially static, but we should not ignore the possible effects of abasic sites on the dynamics of the double helix and we are currently attempting to extend our modeling to answer such questions.

The authors thank P. Vatton and M. L. Dheu-Andries for valuable technical assistance.

We gratefully acknowledge financial support from the Association pour la Recherche sur le Cancer (ARC) and the Ligue Nationale Contre le Cancer (LNFCC).

REFERENCES

Arnott, S., R. Chandrasekaran, D. L. Birdsall, A. G. W. Leslie, and R. L. Ratliff. 1980. Left-handed DNA helices. *Nature*. 283:743–745.

Ayadi, L., Ce. Coulombeau, C. Coulombeau, and R. Lavery. 1997. Conformational analysis of DNA abasic sites dynamics. *In* Biological Macromolecular Dynamics. S. Cusack, editor. Adenine Press, New York. 83–87.

Ayadi, L., M. Jourdan, C. Coulombeau, J. Garcia, J. Lhomme, and R. Lavery. 1999. Experimental and theoretical studies of the conformational perturbations induced by an abasic site. *J. Biomol. Struct. Dyam*. In press.

Barzilay, G., and I. D. Hickson. 1995. Structure and function of apurinic/apyrimidinic endonucleases. *Bioessays*. 17:713–719.

Bernet, J., K. Zakrzewska, and R. Lavery. 1997. Modelling basepair opening: the role of helical twist. *J. Mol. Struct*. 398:473–482.

Boiteux, S., and J. Laval. 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.

Coppel, Y., N. Berthet, C. Coulombeau, Ce. Coulombeau, J. Garcia, and J. Lhomme. 1997. Solution conformation of an abasic DNA undecamer duplex d(CGCACXCACGC). d(GCGTGTGTGCG): the unpaired thymine stacks inside the helix. *Biochemistry*. 36:4817–4830.

Cuniasse, P., G. V. Fazakerley, W. Guschlbauer, B. Kaplan, and L. C. Sowers. 1990. The abasic site as a challenge to DNA polymerase. A nuclear magnetic resonance study of G, C and T opposite a model abasic site. *J. Mol. Biol*. 213:303–314.

Cuniasse, P., L. C. Sowers, R. Eritja, B. Kaplan, M. F. Goodman, J. A. H. Cognet, M. Lebret, W. Guschlbauer, and G. V. Fazakerley. 1987. An abasic site in DNA. Solution conformation determined by proton NMR and molecular mechanics calculations. *Nucleic Acids Res*. 15: 8003–8022.

Cuniasse, P., L. C. Sowers, R. Eritja, B. Kaplan, M. F. Goodman, J. A. H. Cognet, M. Lebret, W. Guschlbauer, and G. V. Fazakerley. 1989. Abasic frameshift in DNA. Solution conformation determined by proton NMR and molecular mechanics calculations. *Biochemistry*. 28:2018–2026.

Demple, B., and L. Harrison. 1994. Repair of oxidative damage to DNA: enzymology and biology. *Annu. Rev. Biochem.* 63:915–948.

Dickerson, R. E., M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C.-S. Tung, E. von Kitzing, A. H.-J. Wang, and V. B. Zurkin. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787–791.

Doetsch, P. W. 1995. What's old is new: an alternative DNA excision repair pathway. *Trends Biochem. Sci*. 20:384–386.

Fritsch, V., and E. Westhof. 1990. Minimization and molecular dynamics of Z-DNA modified by acetylaminefluorene. *In* Modelling of Molecular Structures and Properties. J. L. Rivail, editor. Elsevier, Amsterdam. 627–634.

Gelfrand, C. A., G. E. Plum, A. P. Grollman, F. Johnson, and K. J. Breslauer. 1998. Thermodynamic consequences of an abasic lesion in duplex DNA are strongly depend on base sequence. *Biochemistry*. 37:7321–7327.

Gilson, M. K., and B. Honig. 1988. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies and conformational analysis. *Proteins: Struct., Funct., Genet.* 4:7–18.

Gilson, M. K., K. A. Sharp, and B. Honig. 1987. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem*. 9:327–335.

Gorman, M. A., S. Morera, D. G. Rothwell, E. de La Fortelle, C. D. Mol, J. A. Tainer, I. D. Hickson, and P. S. Freemont. 1997. The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *EMBO J*. 16:6548–6558.

Greger, B., and B. Kemper. 1998. An apyrimidinic site kinks DNA and triggers incision by endonuclease VII of phage T4. *Nucleic Acids Res*. 26:4432–4438.

Hare, D. R., D. E. Wemmer, S.-H. Chou, G. Drobny, and B. R. Reid. 1983. Assignment of the non-exchangeable proton resonances of d(C-G-C-G-

A-A-T-T-C-G-C-G) using two-dimensional nuclear magnetic resonance methods. *J. Mol. Biol.* 171:319–336.

Hingerty, B. E., R. H. Ritchie, T. L. Ferrell, and J. E. Turner. 1985. Dielectric effects in biopolymers: the theory of ionic saturation revisited. *Biopolymers*. 24:427–439.

Ide, H., H. Shimizu, Y. Kimura, S. Sakamoto, K. Makino, M. Glackin, S. S. Wallace, H. Nakamuta, M. Sasaki, and N. Sugimoto. 1995. α-Deoxyadenosine, a major anoxic radiolysis product of adenine in DNA, is a substrate for *Escherichia coli* endonuclease IV. *Biochemistry*. 34: 6947–6955.

Joshua-Tor, L., D. Rabinovich, H. Hope, F. Frolow, E. Appella, and J. L. Sussman. 1988. The three-dimensional structure of a DNA duplex containing looped-out bases. *Nature*. 334:82–84.

Jourdan, M., J. Garcia, E. Defrancq, M. Kotera, and J. Lhomme. 1999. The 2′-désoxyribonolactone lesion in DNA: refined solution structure determined by NMR and molecular modeling. *Biochemistry*. 38:3985–3995.

Kalnik, M. W., C. N. Chang, A. P. Grollman, and D. J. Patel. 1989. NMR studies of abasic sites in DNA duplexes: deoxyadenosine stacks into the helix opposite acyclic lesions. *Biochemistry*. 28:3373–3383.

Kalnik, M. W., C. N. Chang, F. Johnson, A. P. Grollman, and D. J. Patel. 1988. NMR studies of abasic sites in DNA duplexes: deoxyadenosine stacks into the helix opposite the cyclic analog of 2-deoxyribose. *Biochemistry*. 27:924–931.

Lavery, R. 1988. Junctions and bends in nucleic acids: a new theoretical modelling approach. *In* Structure and Expression. W. K. Olson, R. H. Sarma, M. H. Sarma, and M. Sundaralingam, editors. Adenine Press, New York. 121–211.

Lavery, R., and H. Sklenar. 1988. The definition of generalized helical parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* 6:655–667.

Lavery, R., and H. Sklenar. 1996. CURVES 5.1. Helical analysis of irregular nucleic acids. Laboratory for Theoretical Biochemistry, CNRS, Paris.

Lavery, R., K. Zakrzewska, and H. Sklenar. 1995. JUMNA (JUnction Minimisation of Nucleic Acids). *Comp. Phys. Commun*. 91:135–158.

Letellier, R., E. Taillandier, J. R. Bertrand, and C. Malvy. 1991. Molecular modelling of 9-aminoellipticine interactions with abasic oligonucleotides. *J. Biomol. Struct. Dyn.* 9:579–597.

Lin, Z., K.-N. Hung, A. P. Grollman, and C. de Los Santos. 1998. Solution structure of duplex DNA containing an extrahelical abasic site analog determined by NMR spectroscopy and molecular dynamics. *Nucleic Acids Res.* 26:2385–2391.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature*. 362:709–715.

Lindahl, T., and B. Nyberg. 1972. Rate of depurination of native deoxyribonucleic acid. *Biochemistry*. 11:3610–3618.

Loeb, L., and B. Preston. 1986. Mutagenesis by apurinic/apyrimidinic sites. *Annu. Rev. Genet*. 20:201–230.

Manoharan, M., S. C. Ransom, A. Mazumder, J. A. Gerlt, J. A. Wilde, J. M. Withka, and P. H. Bolton. 1988. The characterization of abasic sites in DNA heteroduplexes by site-specific labeling with ^{13}C. *J. Am. Chem. Soc*. 110:1620–1622.

Miaskiewicz, K., J. Miller, M. Cooney, and R. Osman. 1996. Computational simulations of DNA distortions by a cis, syn-cyclobutane thymine dimer lesion. *J. Am. Chem. Soc*. 118:9156–9163.

Miller, M., R. W. Harrison, A. Wlodawer, E. Appella, and J. L. Sussman. 1988. Crystal structure of 15-mer DNA duplex containing unpaired bases *Nature*. 334:85–86.

Millican, T. A., G. A. Mock, M. A. Chauncey, T. P. Patel, M. A. Eaton, J. Gunning, S. D. Cutbush, S. Neidele, and J. Mann. 1984. Synthesis and biophysical studies of short oligodeoxynucleotides with novel modifications: a possible approach to the problem of mixed base oligodeoxynucleotide synthesis. *Nucleic Acids Res*. 12:7435–7453.

Mol, C. D., C. F. Kuo, M. M. Thayer, R. P. Cunningham, and J. A. Tainer. 1995. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature*. 374:381–386.

Morden, K. M., B. M. Gunn, and K. Maskos. 1990. NMR studies of a deoxyribodecanucleotide containing an extrahelical thymidine surrounded by an oligo(dA).oligo(dT) tract. *Biochemistry*. 29:8835–8845.

Poncin, M., B. Hartmann, and R. Lavery. 1992. Conformational sub-states in B-DNA. *J. Mol. Biol*. 226:775–794.

Sanghani, S. R., K. Zakrzewska, S. C. Harvey, and R. Lavery. 1996a. Molecular modelling of (A4T4NN)n and (T4A4NN)n: sequence elements responsible for curvature. *Nucleic Acids Res.* 24:1632–1637.

Sanghani, S. R., K. Zakrzewska, and R. Lavery. 1996b. Modeling DNA bending by phosphate neutralization. *In* Biomolecular Structure and Dynamics, Proceedings of the 9th Conversations. R. H. Sarma and M. H. Sarma, editors. Adenine Press, New York. 267–278.

Sharp, K. A., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem*. 19:301–332.

Shida, T., M. Arakawa, and J. Sekiguchi. 1994. Properties of DNA duplexes containing an abasic site: effect of nucleotide residue of the opposite part of a 2-deoxyribosylformamide residue. *Nucleosides Nucleotides*. 13:1319–1326.

Shida, T., M. Noda, and J. Sekiguchi. 1996. Cleavage of single- and double-stranded DNAs containing an abasic residue by *Escherichia coli* exonuclease III (AP endonuclease VI). *Nucleic Acids Res.* 24: 4572–4576.

Takeshita, M., C.-N. Chang, F. Johnson, S. Will, and A. P. Grollman. 1987. Oligodeoxynucleotides containing synthetic abasic sites. Model substrates for DNA polymerases and apurinic/apyrimidinic endonucleases. *J. Biol. Chem.* 262:10171–10179.

Wallace, S. S. 1988. AP endonucleases and DNA glycosylases that recognize oxidative DNA damage. *Environ. Mol. Mutagen.* 12:431–477.

Wilde, J. A., P. H. Bolton, A. Mazumder, M. Manoharan, and J. A. Gerlt. 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.

Zacharias, M., and H. Sklenar. 1997. Analysis of the stability of looped-out and stacked-in conformations of an adenine bulge in DNA using a continuum model for solvent and ions. *Biophys. J.* 73:2990–3003.

Zacharias, M., and H. Sklenar. 1999. Conformational analysis of single-base bulges in A-form DNA and RNA using a hierarchical approach and energetic evaluation with a continuum solvent model. *J. Mol. Biol.* 289:261–275.

Zakrzewska, K., A. Madami, and R. Lavery. 1996. Poisson-Boltzmann calculations for nucleic acids and nucleic acids complexes. *Chem. Phys.* 204:263–269.