Conformations of Nicked and Gapped DNA Structures by NMR and Molecular Dynamic Simulations in Water†

Claus Roll, Christophe Ketterlé, Valérie Faibis, G. Victor Fazakerley, and Yves Boulard*

CEA, Service de Biochimie et de Génétique Moléculaire, Bât. 142, Département de Biologie Cellulaire et Moléculaire, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France

Received September 25, 1997; Revised Manuscript Received December 15, 1997

ABSTRACT: We have analyzed and compared the molecular structures and dynamics of DNA duplexes containing a nick or a gap of one nucleotide where the base in front of the gap is a guanine. The continuous strand has the sequence 5′(CAGAGTCXCTGGCTC) where the residue X is absent for the nick, 14-mer, and where it is a G residue for the gap. Duplexes were formed with the two corresponding 7-mers. Neither of these is phosphorylated adjacent at the nick site, but it is a good model for a single strand break. For the nick structure, the quantitative NMR data show that the global conformation is very close to canonical B-form DNA, but it displays enhanced local flexibility. For the gap structure, we observe only one species in which the extra G is well stacked into the helix. The two half-helices around this residue also show a B-form conformation. As with the nick duplex, the adjacent G imino protons show enhanced exchange with solvent. The gap does not close completely. Using distance constraints, MD calculations show that the nick conformation is very close to a duplex with no lesion but is indeed more flexible in the central part. The gapped structure shows two families of conformations. One is close to B-DNA, the other is significantly kinked at the gap which reduces the size of the cavity. We observe a spine of hydration within the cavities, similar, but of different geometry in the two cases.

A gap of one nucleotide in DNA may occur mainly during one of two different processes. The first one is the irreversible chemical modification of the DNA double helix. DNA is damaged by hydrolysis or by oxidation during the normal cellular metabolism *(1, 2)*. A DNA double helix with a gap of one nucleotide is an intermediate in the major repair mechanism of this type of damage. The second process which induces the formation of a gap of one nucleotide is the ionizing radiolysis of H_2O. One of the products of the radiolysis, the hydroxyl radical, attacks DNA at different sites. Nicks or gaps of one nucleotide are formed after the attack at the phosphate backbone *(3)*.

Base excision repair is the major pathway involved in the repair of such lesions. The single nucleotide containing the damaged base is specifically recognized. First, the damaged base is excised, then the phosphate backbone is incised 5′ of the resulting abasic site. Finally, the abasic residue is excised *(2, 4−7)*. The DNA containing a gap of one nucleotide thus occurs during this repair mechanism as an intermediate. The gap-filling step requires a DNA-polymerase, DNA-polymerase I in *Escherichia coli (2)* and DNA-polymerase β in mammalians *(8, 9)*, which links the newly inserted nucleotide at the 3′-OH position of the gap. The resulting nick is closed by a DNA-ligase.

According to the current model of the nucleotide insertion step, derived from crystallographic data *(10−12)*, all the molecules require a particular spatial positioning *(13)*. Two nuclear proteins, HU in *E. coli* and poly(ADP-ribose) polymerase (PARP) in mammalians, each of which possess different functions, have increased affinities for DNA-containing gaps *(14, 15)*. These two proteins are able to bend DNA double helices *(16, 17)*. They are believed to participate in this repair mechanism. Further, electrophoretic studies have shown that DNA with a gap has a greatly increased flexibility compared to DNA without any lesion *(18)*. A common pattern of DNA-containing unpaired bases, as we have shown in previous papers for DNA duplexes containing abasic sites *(19−21)* and others for oligonucleotides containing bulges of one unpaired base *(22−24)* with structural analysis by NMR,[1] is that, generally, purines are stacked inside the helix and may induce a kink at the position of the unpaired base.

This report is the first of a series of studies of duplexes containing in front of the gap one of the four bases. Neither structural studies derived from NMR nor X-ray diffraction analysis have been published on gapped DNA. Here, we report the structural and dynamic analysis by NMR and molecular dynamics calculations of a duplex of 14 base pairs with an extra guanine in front of the gap. We included in this series the study of a duplex of 14 base pairs, containing a model nick: the 5′ phosphate at the single-strand break is removed. The structural differences with the phosphorylated

† This work was, in part, financed by a grant from the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche, ACC-SV8.

* To whom correspondence should be adressed. Tel: +(33) 1 69088087. Fax: +(33) 1 69084712. E-mail: ybou@matthieu.saclay.cea.fr.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; 1D, one-dimensional; 2D, two-dimensional; TOCSY, total correlation spectroscopy; MD, molecular dynamics; PME, particle mesh Ewald; rms, root-mean-square.

S0006-2960(97)02377-5 CCC: $15.00
Published on Web 02/28/1998

nick are only minor *(25, 26)*. The study of these different duplexes will help us to understand the influence of relaxing constraints in a DNA double helix on the phosphate backbone in the nicked duplex and additionally on the base stacking in the gapped duplex.

The structure and the dynamic behavior of a duplex containing a single nucleotide gap are important parameters in determining the repair efficiency. DNA−protein recognition, complex formation, and the reaction mechanism depend on these parameters.

EXPERIMENTAL PROCEDURES

NMR Spectroscopy. All the single-stranded oligonucleotides were synthesized by Eurogentec. The nick- and gap-containing duplexes were formed by mixing two single-stranded oligonucleotide 7-mers with, respectivly, a single-stranded 14-mer or a 15-mer oligonucleotide. During the mixing procedure, the oligonucleotides were heated at 80 °C and then slowly cooled to form the duplex. The duplexes were 4 mM in concentration dissolved in 10 mM phosphate buffer, 150 mM NaCl, and 0.2 mM EDTA. As internal reference, the tetramethylammonium chloride peak at 3.18 ppm was used. NMR spectra were recorded on Bruker AMX500 or AMX600 spectrometers. NOESY and TOCSY spectra were recorded and treated as described previously *(27)*. The mixing times of the NOESY spectra were of 40, 55, 70, 85, 100, and 400 ms in D_2O and 250 ms in H_2O. In D_2O, the residual water resonance was presaturated during the relaxation and mixing delays. In H_2O, the water signal was suppressed with the WATERGATE sequence *(28)*. TOCSY spectra were recorded with a 60 or 120 ms mixing time. For the nicked and the gapped duplexes, NMR distance constraints were obtained as described in earlier studies *(19, 20)*. For each type of interaction [H6/H8(*n*)-H1′(*n*), H6/H8(*n*)-H1′(*n* − 1)...], we derived internal reference distances from the parts of the duplex (excluding terminal base pairs) which are clearly in a classical B form conformation and applied them to the central region.

The only difference in DNA sequence between the nick- and gap-containing samples is that we added a supplementary nucleotide in the center of the uninterrupted strand in the latter case. Therefore, to simplify comparisons between the two duplexes, we used the following nomenclature:

```
     1  2  3  4  5  6  7     9  10 11 12 13 14 15
5'   C  A  G  A  G  T  C     C  T  G  G  C  T  C  3'
3'   G  T  C  T  C  A  G     G  A  C  C  G  A  G  5'
    29 28 27 26 25 24 23    22 21 20 19 18 17 16

     1  2  3  4  5  6  7  8  9  10 11 12 13 14 15
5'   C  A  G  A  G  T  C  G  C  T  G  G  C  T  C  3'
3'   G  T  C  T  C  A  G     G  A  C  C  G  A  G  5'
    29 28 27 26 25 24 23    22 21 20 19 18 17 16
```

and we use gapG to designate the gapped structure.

Molecular Modeling. Nick, gapG, and a reference 15-mer duplex models have been built in B-DNA form *(29)*. The reference 15-mer duplex has a C residue named C22′ in front of the G8 base in order to have continuity of the sugar−phosphate backbone. A set of conformations was generated for each structure by varying the roll angles in the central base pair region, i.e., (i) between C7•G23 and C9•G22 base pairs for the nick, (ii) between C7•G23 and G8 and G8 and C9•G22 for the gap, and (iii) between C7•G23 and G8•C22′ and G8•C22′ and C9•G22 for the reference 15-mer duplex. Each structure was minimized with the SANDER module of AMBER *(30)*, using the Cornell et al.'s force field *(31)*, in two phases. Three types of constraints were applied during the first minimizations: torsion angles, NMR distances, and weak reinforcements (5 kcal mol^{-1} $Å^{-2}$) of the hydrogen bonds of terminal base pairs to avoid end fraying. The torsion angle constraints were applied on δ angles (C5′−C4′−C3′−O3′) with a value of 144° and a constraint force of 50 kcal mol^{-1} deg^{-2}, forcing all the sugar puckers into a C2′-*endo* conformation, as observed by NMR. All internucleotide NMR distance constraints were applied with a constant force of 5 kcal mol^{-1} $Å^{-2}$. For the reference 15-mer duplex, the first minimizations were calculated with incorporating the NOE constraints found for the gapG. The second minimizations were performed for all the systems without any constraints in order to relax the obtained conformations. All minimizations were stopped when the rms energy gradient was less than 0.1 kcal mol^{-1} $Å^{-1}$. Starting from the best energy refined structures of nick, gapG, and the reference 15-mer B-DNA, we have hydrated the DNA duplexes with TIP/3P water molecules *(32)* and positionned Na^+ counterions with the LEaP AMBER module, to obtain electroneutral hydrated models. The sizes of the water boxes are 41.9 Å × 40.3 Å × 69.6 Å, 41.9 Å × 40.4 Å × 73.9 Å, and 39.8 Å × 41.6 Å × 73.3 Å, respectively, for the nick, gapG, and the reference 15-mer B-DNA. These boxes contain 2400−2600 water molecules. The dielectric constant was taken equal to 1 *(33)*. We have applied the particle mesh Ewald method to treat long-range interactions and periodic boundary conditions. An additional cutoff of 10 Å has been used. The dynamics protocol consists of the following steps. First, we minimized the hydrated structures by applying 200 steps of the conjugate gradient method with SANDER, to relax close contacts. Second, we prepared the hydrated models during 40 ps following the protocol described in ref *34*. Then, 500 ps of MD production on each system was performed. During this production phase the same three types of constraints described above for the first minimizations stage were applied. We defined an angle θ as follows. We have defined two-half helices from each side of the nick, the gap, or the central G•C base pair of the reference B-DNA, and built the two helical axes with OCL *(35)* according to ref *36*. The conformations generated were analyzed and displayed on SGI workstations using the program MORCAD *(37)*.

RESULTS

Nick Structure. The melting temperature of the duplex was determined by following the chemical shifts of the methyl groups as a function of temperature. For this duplex, we were unable to determine the exact value because of superposition of these peaks at temperatures above 60 °C. The melting curves up to this temperature indicate that we have not yet reached the T_m.

Exchangeable Protons. Figure 1 shows the assignment of the exchangeable protons. The regions shown are part of a NOESY spectrum recorded in 90% H_2O and 10% D_2O with a mixing time of 250 ms at 2 °C. The upper part

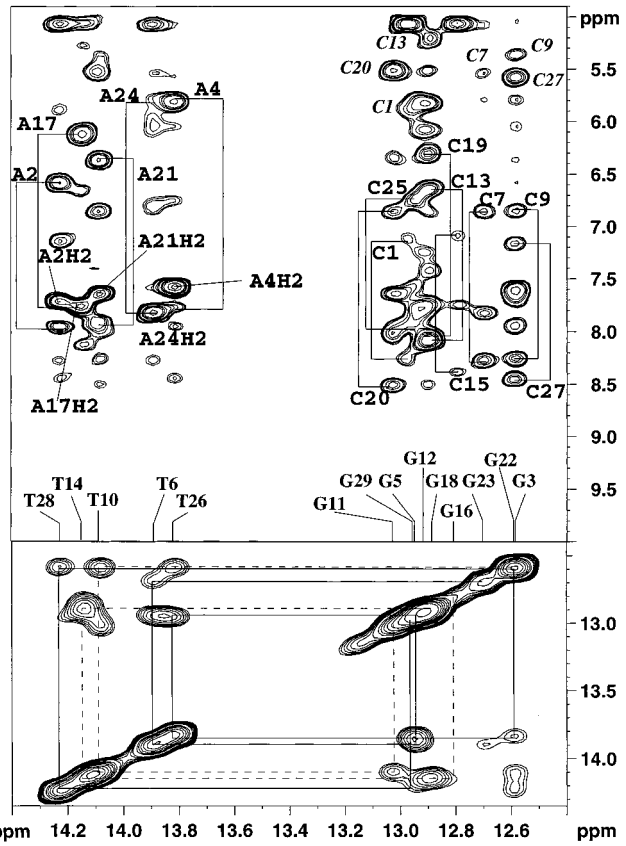

FIGURE 1: Expanded contour plot of the imino-amino/H2/H5 (upper part) and imino−imino (lower part) regions of the NOESY spectrum (250 ms mixing time) of the duplex containing the nick recorded in 90% H_2O and 10% D_2O at 2 °C. Imino-CH5 interactions are labeled in italic print.

represents the interactions between the imino and amino/H2/H5 protons, the lower part the imino−imino interactions. One can observe the characteristic intra- and interbase interactions for the exchangeable protons of an oligonucleotide in a B-DNA conformation from one end of the duplex to the other. In the lower part, the peaks, that are linked with a solid line, show the imino−imino interactions between the base pairs C1·G29 to C7·G23, those linked with broken lines, show the interactions between the base pairs from C15·G16 to C13·G18 and from G11·C20 to C9·G22. The assignments shown in Figure 1 are confirmed by NOEs with CH5 protons. The imino proton of G12 can now be identified from the interactions with the H5 protons of the cytosines C20 and C13 of the neighboring base pairs. The H5 protons of C19 and C25 have chemical shifts close to the water resonance frequency. Therefore, imino-H5 cross-peaks overlap with imino-H_2O exchange cross-peaks.

Thus, the base pairs from C1·G29 to T6·A24 and from C15·G16 to T10·A21 show the characteristic interactions of Watson−Crick base pairs in a standard B form conformation. The interactions in the center of the duplex are observed at the G22N1H and G23N1H resonance frequencies that are identified, respectively, by the interactions with C9H5 and C7H5. The two imino protons G22N1H and G23N1H also show the normal base pair interactions on each side of the nick. Unfortunately, due to poor spectral dispersion we cannot observe a NOE between these protons. However, G23N1H shows a cross-peak with C9H5 in a spectrum plotted at lower levels.

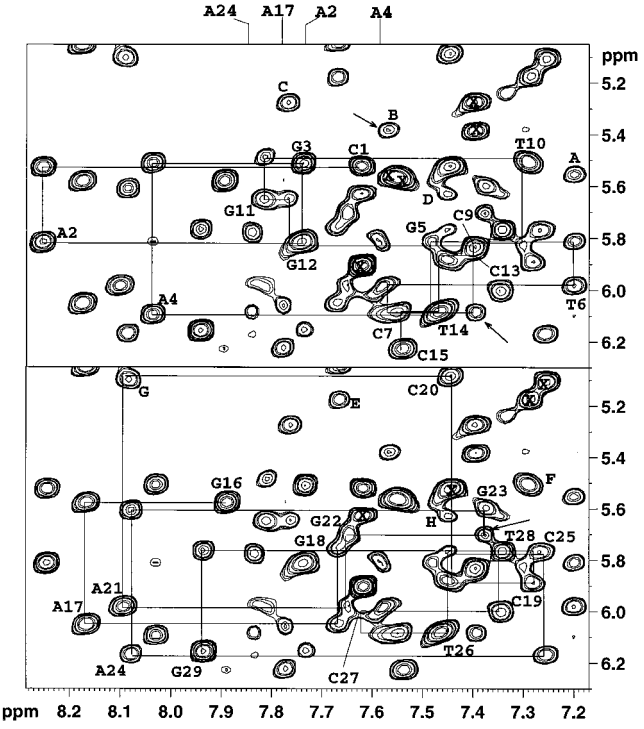

FIGURE 2: Expanded contour plot of the H6/H8−H1′ (upper and lower parts) region of the NOESY spectrum (400 ms mixing time) of the duplex containing the nick, recorded in D_2O at 13 °C. Cross-peaks marked with an X correspond to CH5−CH6 interactions. Cross-peaks A−H correspond to H6/H8 (*n*)-CH5 (*n* + 1) interactions. Cross-peaks marked with an arrow correspond to C7−C9 (upper part) and G22−G23 interactions (lower part).

The exchange behavior of the imino protons was studied by following the evolution of the line widths of the resonances in 1D spectra when increasing the temperature as they reflect the exchange rate with bulk solvent. The line broadening as a function of temperature is most rapid for the terminal imino protons due to fraying of the helix and intermediate for the central imino protons G22 and G23. The slowest evolution is observed for the imino protons in the centers of the two half-helices. We can attribute the intermediately increasing exchange with water protons in the central region of the duplex to more frequent base-pair opening and higher flexibility and thus greater accessibility of water to these imino protons.

Nonexchangeable Protons. The assignment of the non-exchangeable protons was obtained by analyzing a NOESY spectrum recorded with a 400 ms mixing time and a TOCSY spectrum with a 60 ms mixing time. The upper and lower parts of Figure 2 show the region corresponding to the H6/H8−H1′ interactions in the NOESY spectrum at 13 °C. We can follow the H6/H8−H1′ connectivities predicted for a right-handed B-DNA helix on each strand: C1 to C15 (upper part), G16 to G22 and G23 to G29 (lower part).

In particular, we are interested in the interactions in the duplex center between C7 and C9 and on the other strand between the guanines G22 and G23, flanking the nick. The cross-peaks indicated with arrows correspond to internucleotide interactions C7−C9 and G22-G23. In the upper part of Figure 2, peak B corresponds to the C7H6−C9H5 interaction and the other cross-peak indicated with an arrow in Figure 2 to the C7H1′−C9H6 interaction. In the lower part, the cross-peak indicated with an arrow corresponds to

the G22H1′−G23H8 interaction. These results indicate that C7 and C9 are stacked one on the top of each other as well as G22 on G23. The H6/H8(n)−H2′/H2″(n − 1) interactions, data not shown, between C9 and C7 and their neighbors confirm this interpretation. The H6/H8(n)−H2′/H2″(n − 1) interactions between G22 and G23 cannot be unambiguously confirmed because of poor spectral resolution. Further the C7H6−C9H6 and G22H8−G23H8 interactions observed in the matrix columns at the respective resonance frequencies, data not shown, confirm that the bases in the center of the duplex are stacked one on top of the other.

These assignments are confirmed in the other regions of the spectrum. All the connectivities predicted for B-DNA are observed in this duplex. The only exception is for C27, which shows very weak interactions with the H1′ peaks. However, all the expected cross-peaks for this residue are well observed in the H6−H2′/H2″/CH3 region. The sugar conformations were determined by comparing the relative intensities of the cross-peaks H6/H8−H2′ and H6/H8−H3′ in a NOESY spectrum recorded with a 60 ms mixing time. We observe for all the sugars a predominantly C2′-*endo* conformation as the H6/H8-intraresidue NOE with the H2′ proton is very much stronger than with the H3′ proton whereas the reverse would be expected for a C3′-*endo* conformation *(20)*. In the same spectrum, we observe that the sugar/base orientations are all anti by comparing the intensities of the H6/H8−H1′ and H6/H8−H2′ cross-peaks. Those with H1′ are weak at short mixing times, whereas they would be similar to CH5−CH6 cross-peaks for a syn conformation *(38)*.

gapG Duplex. We have recorded 1D spectra of the oligonucleotide as a function of pH and we do not observe changes in the spectrum, indicating a structural transition of the oligonucleotide. Whatever the physicochemical conditions of the sample, we did not find evidence for the presence of an equilibrium between different species in solution. We chose to study the gapG sequence at pH 6.5 and 22 °C as the 1D spectrum presented good resolution.

Nonexchangeable Protons. In this section, we will focalize our description around the gap, as the NMR results for the other base pairs of the oligonucleotide were the same as for the nick structure. Figure 3 shows the H8/H6−H1′/H5 region of a NOESY spectrum recorded in D_2O at 22 °C. We can follow the characteristic B-DNA connectivities all along the first strand from C1 to C15. In the center, we are able to identify connectivities between G8 and C7 and also between C9 and G8, indicating that G8 is well stacked, as in B-DNA, inside the helix. In the same manner, all the expected cross-peaks characteristic of a B-DNA structure were observed, for the other two strands, G16 to G22 and G23 to G29. But the interaction between G23H8 and G22H1′ was not observed showing that these two protons are too far apart to give a NOE effect. We have carefully searched in all the regions of all the NOESY spectra for cross-peaks between G22 and G23 but we have not been able to identify one. For the base pairs on either side of G8, T6·A24:G5·C25, and C9·G22:T10·A21, NOE interactions are observed in NOESY spectra recorded with short mixing times which are typical of a classical B-DNA. These demonstrate that the duplex forms a normal B-DNA structure with G8 stacked normally between C7 and C9 but with a gap in front of G8 that does not allow an NOE effect between the nucleotides G22 and G23.

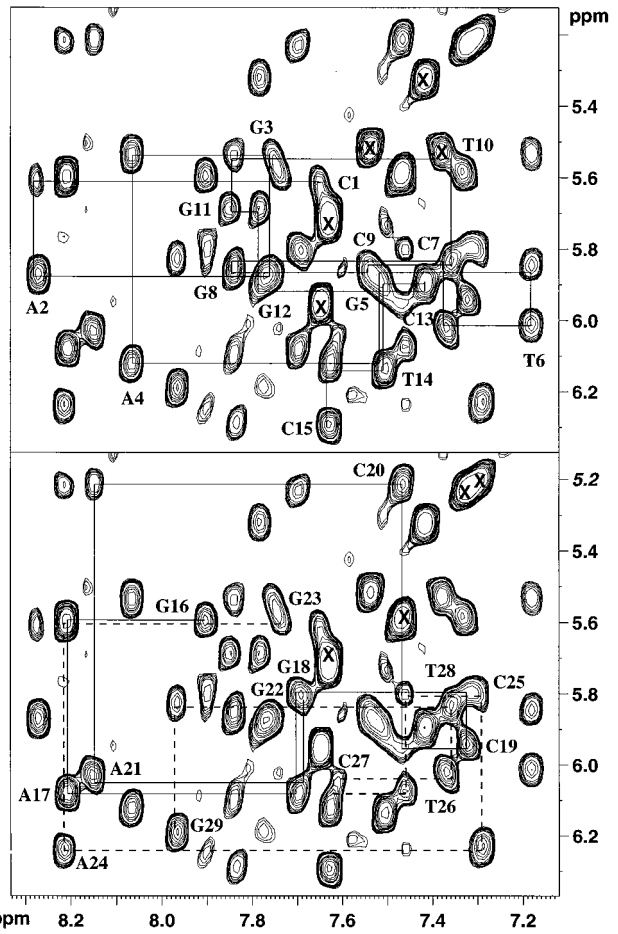

FIGURE 3: Expanded contour plot of the H6/H8−H1′ region of the NOESY spectrum (400 ms mixing time) of the gapG duplex in D_2O at 22 °C and pH 6.

Exchangeable Protons. All the expected exchangeable protons of the Watson−Crick base pairs were identified in H_2O as presented in Figure 4. In the lower part, the imino−imino connectivities of thymine and guanine residues observed in a NOESY spectrum at 9 °C are shown. The G22 and G23 imino protons are well observed but unfortunately overlap at 12.72 ppm, the resonance integrates for two protons in the 1D spectrum. As the two amino protons of C7 and C9 are well separated by ca. 0.9 and 1.3 ppm, respectively (see Figure 4, upper part), this showed that the two G·C base pairs around the gap are well formed with three hydrogen bonds. All the imino protons presented in the 12.3−14.5 ppm region have been assigned. We did not observe, in the range 14.5−10.0 ppm in neither 1D or 2D spectra recorded at different temperatures, an additional resonance that we could attribute to the imino proton of G8. This indicates that the imino proton of G8 exchanges rapidly with the solvent.

Part of the 1D spectrum recorded as a function of temperature at pH 6.5 is shown in Figure 5. As expected, the line widths of the imino protons of the terminal base pairs G16·C15 and C1·G29 broaden rapidly on increasing the temperature as a result of exchange with the solvent. At ca. 16 °C, they have disappeared, whereas the two imino protons of G22 and G23 which could also be considered as terminal are well observed. Given that the two resonances

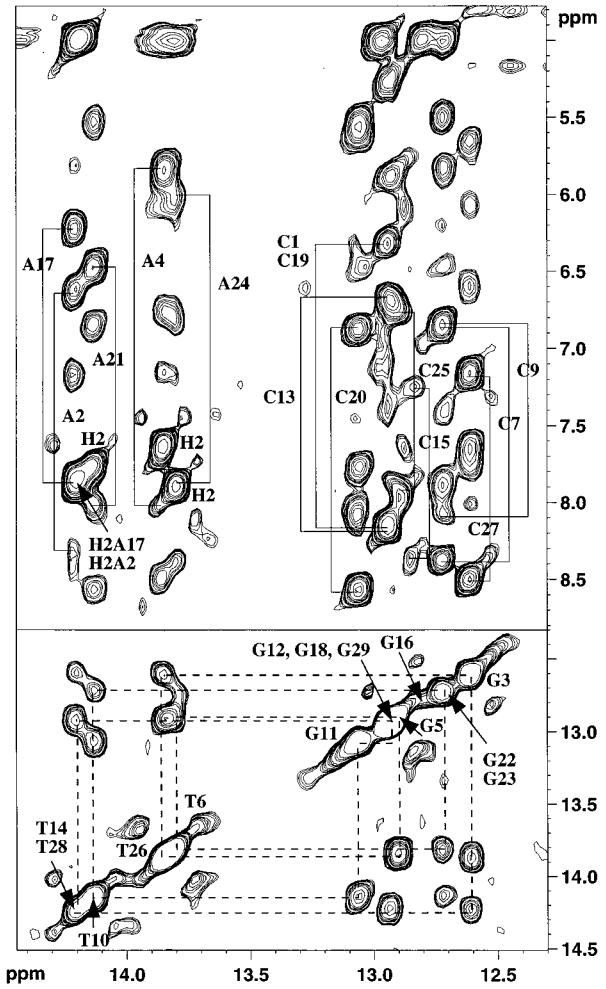

FIGURE 4: Expanded contour plot of the imino−amino/H2/H5 (upper part) and imino−imino (lower part) regions of the NOESY spectrum (250 ms mixing time) of the duplex containing the gapG recorded in 90% H_2O and 10% D_2O at 9 °C.

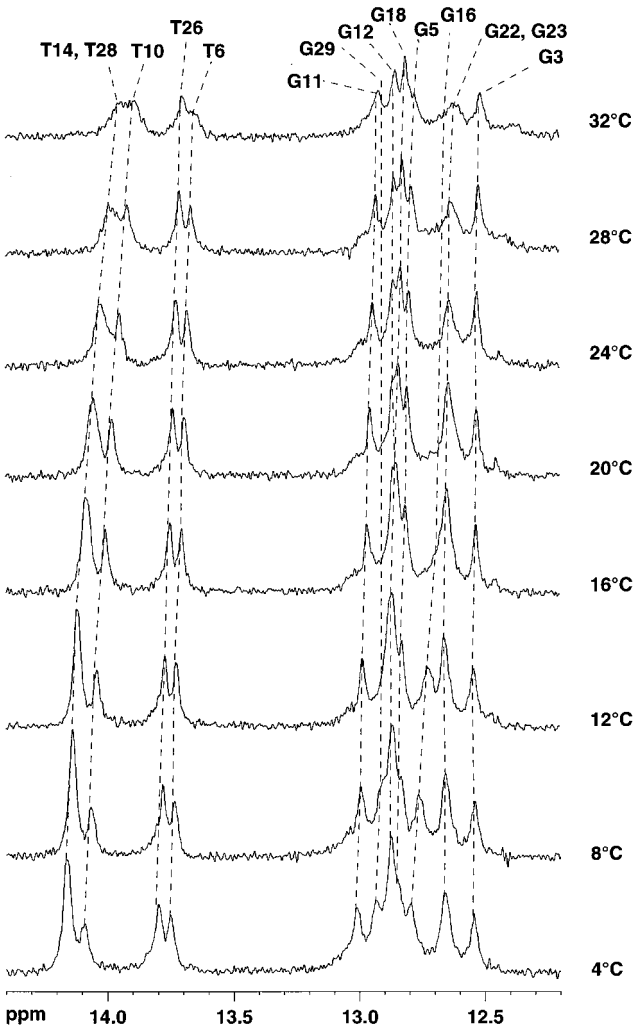

FIGURE 5: Temperature dependence of the imino protons of the gapG duplex.

overlap, we were not able to determine the exchange rate of one relative to the other. However, from the line width it is clear that exchange is much slower than for the terminal base pairs and not much faster than for imino protons well embeded in the helix. Despite solvent accessibility to the gap site, as we do not observe the G8 imino proton, this has little effect upon the dynamic behavior of the imino protons of the adjacent base pairs. Clearly, base pair fraying is much less important than for the terminal G·C base pairs.

The NMR data give us the following characteristics. Both duplexes correspond globally to B-DNA. Even in the center, we observe the characteristic interproton distances and sugar puckers. However, we observe an increased flexibility and solvent accessibility for both duplexes in the central region from the imino proton exchange. We have determined 126 interproton distances for the nicked duplex and 110 for the gapped duplex. We estimate that the error in the distance measurements is ±15%.

Molecular Modeling Studies. During the 500 ps PME MD simulation, the nicked DNA presents a smooth rms curve, showing that the structure is well stabilized from 50 ps to the end of the run, Figure 6a. The rms fluctuates around 1.7 ± 0.2 Å. The rms curve for gapG is also quite well stabilized and fluctuates around 2.2 ± 0.4 Å. It presents a sharp peak which reaches 3.7 Å around 270 ps, Figure 6b, which corresponds to a bent conformation of gapG, Figure 6d. When we sample the θ angles for the nick and gapG structures along the dynamics, Figure 6, panels c and d, we obtain the population histograms shown in Figure 6, panels e and f, respectively. The nick presents one homogeneous Gaussian population centered around 17°. The averaged value calculated on the reference B-DNA is 15° in calculations with the distance constraints for the two half-helices derived from the gapG structure, and a similar value is found without any distance constraints. The plot for gapG, Figure 6f, presents two distinct populations centered around 28° and 11°, which correspond respectively to two distinct structures: a bent one and an approximately straight one (named the straight conformation in the following text), well characterized in the time ranges 260−330 and 360−420 ps, respectively.

The agreement with the NMR data and the stability of these models is shown in Figure 7. In the nicked duplex, for only four out of the 126 distances determined by NMR do we observe a difference of more than 15% between the NMR derived distances and those of the minimized structure (Figure 7a). For the minimized gapped structure, eight out of the 110 experimental distances differ from the minimized structure distances of more than 15% (Figure 7b). We observe similar time-averaged results during the MD runs for these systems. For example, we have analyzed for the

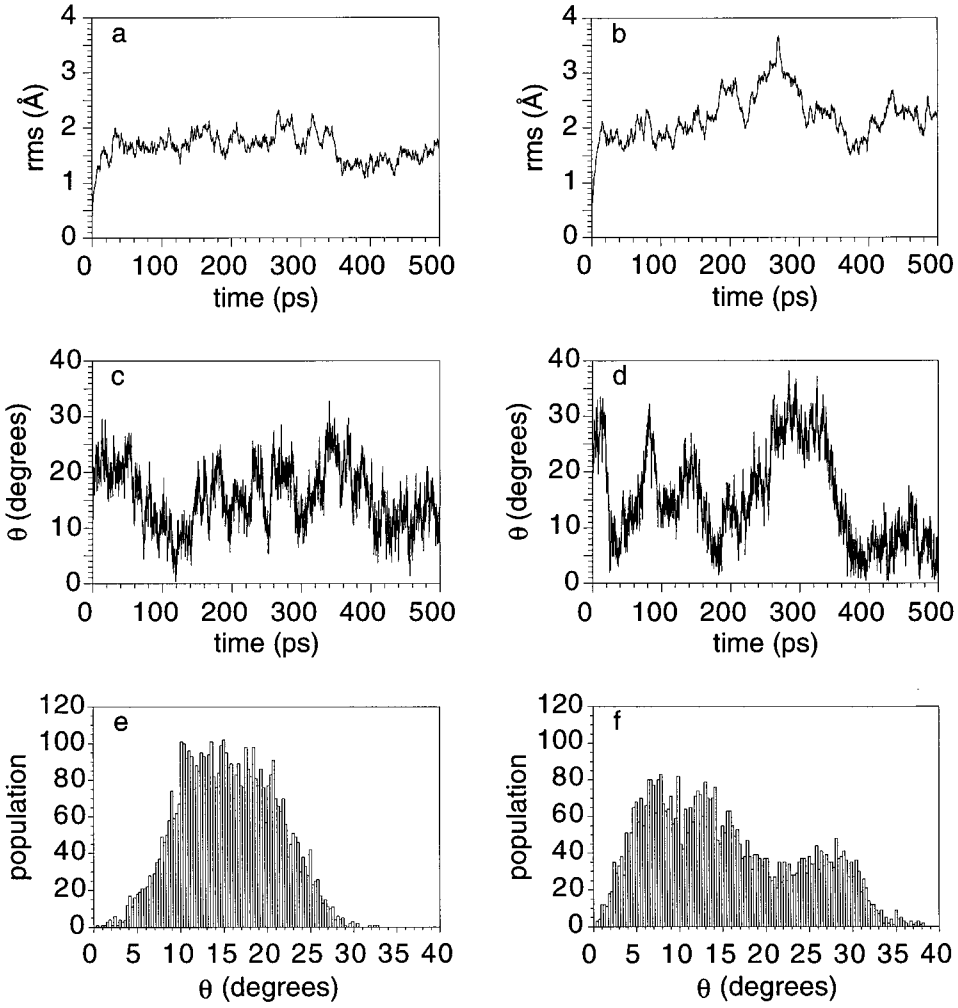

FIGURE 6: Plots versus time of (a) rms plot for the nick, (b) rms plot for gapG, (c) θ angle plot for the nick, and (d) θ angle plot for gapG. Plots of the population of conformations displaying a particular θ angle versus θ during the 500 ps MD runs for (e) the nick and (f) gapG.

gapG system two snapshots taken during the MD runs which correspond to the least and most bent structures that we have observed. In Figure 7c, we observe for the straight structure that 14 out of the 110 distances differ by more than 15%. For the bent structure, we observe similar results (Figure 7d). These show that the initial structures of the MD runs of the two duplexes fit well the NMR-derived data and that the two models of gapG observed during the MD runs are very close to the NMR observations. Average fluctuations around the distance constraints and their maximum violation depend on their type (inter- or intra-residue) and their length as described in a previous study *(39)*.

In Figure 7, panels e−g, we show the evolution of energy terms for the global system (solute and solvent) during the MD runs for the nicked structure. Figure 7e shows the variation of the total energy during the 500 ps phase production of the MD run. We observe a small decrease during the first 150 ps which is largely determined by that of the electrostatic energy of the molecule as shown in Figure 7f. This slight decrease may be due to the prolongation of the equilibration phase for such large systems where a pivot point of fluctuation exists. However, as shown in Figure 7g, both energies are very stable during the period from 150 to 500 ps, which illustrates the quality of the MD runs. Furthermore, it should be noted that these calculations were carried out at constant temperature as shown in Figure 7h and not with constant energy. We observe very similar results for the energy terms in the gapG system including the initial decrease in energy followed by stabilization. We note that the energy terms are mostly dominated by the solvent, but the rms curves shown in Figure 6 were only calculated for the DNA and demonstrate that the calculated structures do not drift during the MD runs. All the structures described here were derived from the period 150−500 ps where the energy of the systems are stable.

The best minimized conformations for the gapG is shown in Figure 8 and also two structures, straight and bent, observed in the MD runs. Note the difference between a zig-zagged helicoidal axis for the straight conformation of gapG, compared to the straight helicoidal axis for the minimized structure of gapG, and the particularly bent conformation of gapG, which considerably differs from the two other structures. To characterize the curvature or kinking of the structures, we have plotted the tilt, roll, and twist parameters between C7·G23 and C9·G22 base pairs for the nick, the reference 15-mer B-DNA with C22′ in front of G8, giving a duplex without any lesion and the gapG, Figure 9. The gapG Figure 9c shows slightly larger fluctuations for the three wedge parameters than for those of the reference B-DNA or for the nick. We observe during gapG MD runs, progressively, an overtwisting between C7·G23 and G8 and an undertwisting between G8 and C9·G22, which shows that

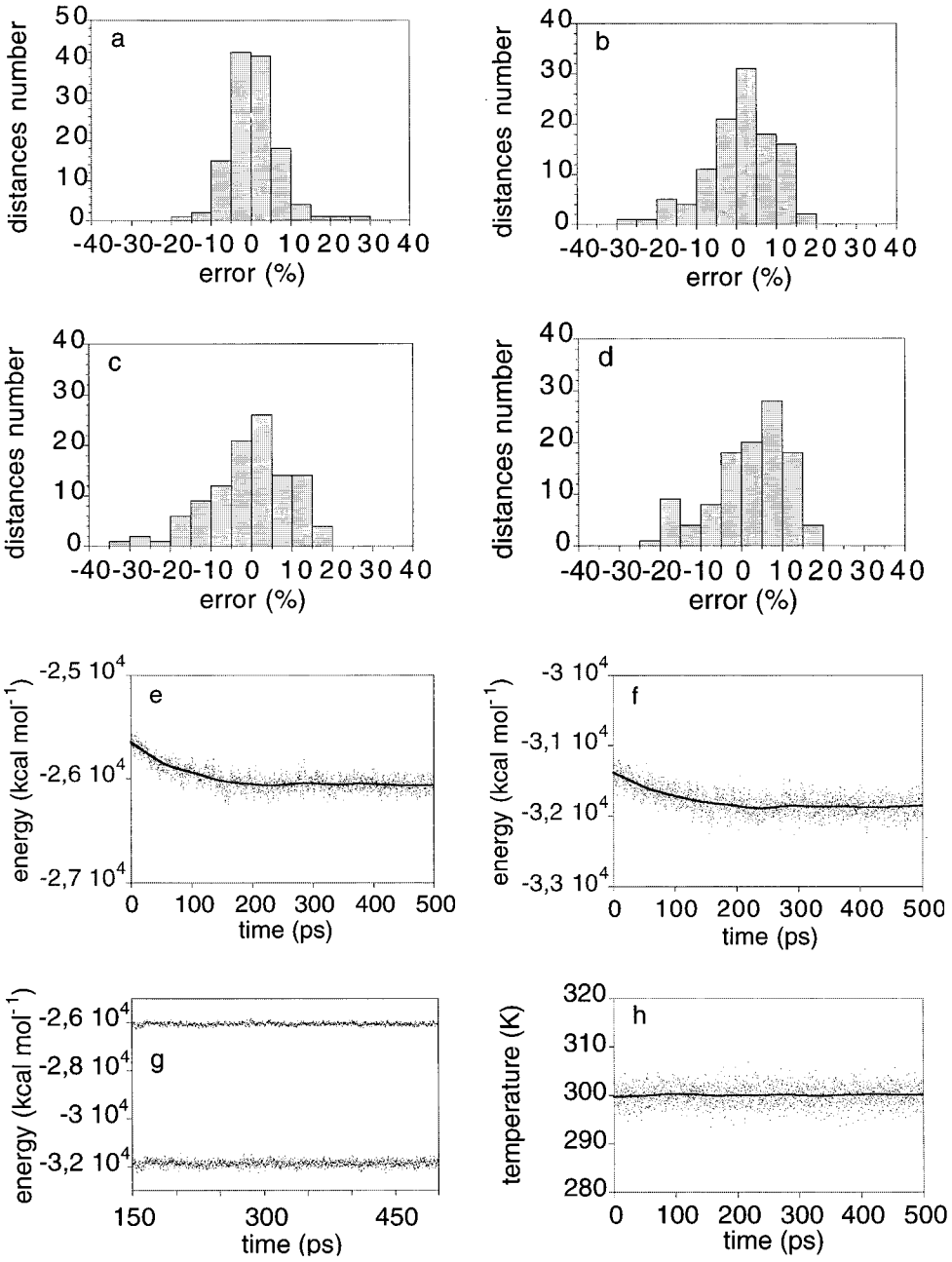

FIGURE 7: Percentage error between NMR distances and calculated distances for (a) the minimized nick structure and (b) the minimized gapG. Panels c and d show corresponding plots for respectively the straight and bent structures of gapG observed in the MD runs. Panel e shows the variation during MD runs of the nick for the total energy, panel f for the electrostatic energy, and panel h for the temperature. A zoom of the two energy terms in panels e and f is shown in panels g between 150 ps to the end of the simulation.

G8 stacks preferentially over the cytosine in the 3′ direction. The fluctuations of the twist parameters are anti-correlated with the fluctuations of the roll parameters, as previously observed *(40)*, but this is less so for the tilt. In going from the minimized conformation to the bent conformation of the gapG, the roll angle between C7·G23 and G8 progressively increases, whereas the roll angle between G8 and C9·G22 decreases with the same amplitude. The tilt angles on each side of G8 are higher for the bent structure than for the straight one. For the straight conformation, the average values for the roll and tilt angles are close to 0°.

We have computed the minor groove width for the nick, the reference 15-mer B-DNA and the gapG structures by measuring the distance between phosphorus atoms, on the unbroken strand, flanking the C7·G23 and C9·G22 region. The average minor groove width value calculated for all the structures has a value of 5.5 Å, which is close to the normal value of 5.7 Å for B-DNA *(41)*. For all the structures, we observe an increase of the average minor groove width value for the three central base pairs of the sequence that we have studied.

For gapG, it is noteworthy that the minor groove width in the G8−C9 region is broad (up to 8 Å) for the bent conformation and particularly narrow (around 3.5 Å) for the straight conformation. This is well correlated with the variation of θ shown in Figure 6a and suggests that the transition between the bent and the straight structures is accompanied by a major change in the local width of the minor groove.

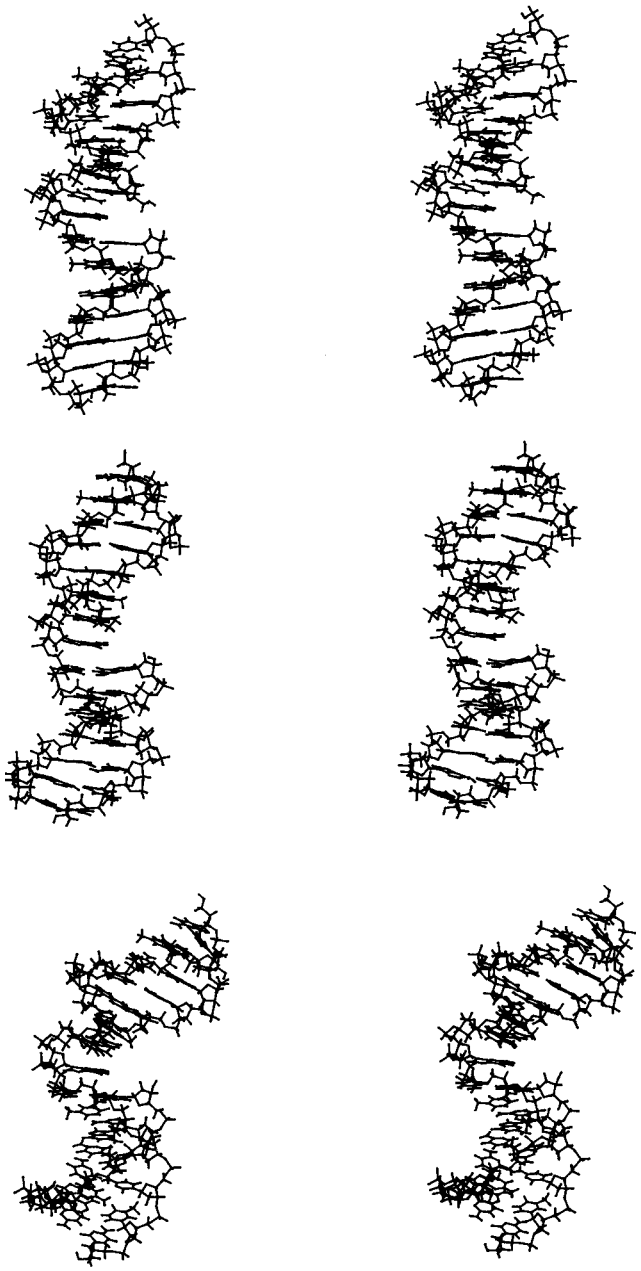

FIGURE 8: Stereoscopic views of gapG: after minimization (upper), a snapshot of the straight structure taken during the 500 ps MD run (middle) and a snapshot of the bent structure (lower).

Water Organization. The best way to verify the quality of a molecular dynamics simulation with explicit solvent molecules is to calculate the averaged self-diffusion coefficient on all water molecules. Explicit TIP/3P water molecules have a self-diffusion coefficient of 1.0×10^{-9} m^2 s^{-1}, which corresponds to the sixth of the coefficient of the slope of the rectilinear plot of diffusion averaged on all the water molecules, Figure 10a thick line. This is of the same order of the experimental value for bulk water and from theoretical studies *(42)*.

In front of the nick, there is no particular structuration of the water molecules. To describe the behavior of water molecules in and around the gap, we have compared them with water molecules in the minor groove and also with bulk solvent. The first essentially involves the minor groove spine of hydration which presents a low diffusion constant profile during MD runs. Two examples are shown in Figure 10b, compared to the second case where solvent molecules are free to diffuse, Figure 10a thin line. In our models of gapG, there is a cavity in front of G8, which opens into the major groove. This cavity is sufficiently wide to allow some water molecules to enter. It is hydrated mainly by the establishment of hydrogen bonds between water oxygen atoms and amino protons of C7 and G22 and both the amino and imino protons of G8. Figure 10c shows the diffusion curves of two water molecules observed in this cavity for the bent conformation. Note that a water molecule situated at the extremity of the cavity, Figure 10c thin line, which is the most exposed to the bulk solvent, can locally diffuse, reaching another state during about 30 ps, but can then return to its initial position. In contrast, the other water molecule which is in the center of the cavity is trapped, Figure 10c thick line, as for a water molecule in the minor groove. The residence time of this second water molecule in the cavity is longer compared to the first. During 500 ps, water molecules in the center of the cavity exchange 1−4 times with bulk solvent, which corresponds to a lifetime of 100−300 ps, whereas those at the extremity of the cavity exchange about 5 times more often. The longest lifetimes for water molecules in the gap are similar to those of water molecules tightly bound in the spine of hydration in the minor groove. Figure 10d shows similar curves for water molecules belonging to the cavity for a gapG in a straight conformation. The two types of water molecules as described above can also be observed, one which is fixed in the cavity and a second which can diffuse. The main difference is that, since the cavity is larger for the straight structure, when the water molecule begins to diffuse out of the cavity, it cannot return to its initial position because the exchange with another water molecule from the bulk solvent takes place.

Figure 11a shows the spine of hydration in the major groove in front of G8, for the bent conformation, which adopts a S shape. We can point out a water molecule belonging to the spine of hydration in the major groove which is involved in hydrogen bonds with both G8 and G22 amino groups. The same S shape is observed for the straight structure, Figure 11b, but since the cavity is larger, there are two hydration layers. During the MD runs, when the gapG bends, the cavity narrows ejecting water molecules into the major groove.

DISCUSSION

In earlier NMR studies, undammaged DNA duplexes were compared with duplexes containing a nick *(25, 26)*. The conclusion that there are only minimal differences in structure was based on chemical shift comparisons or differences in distances determined with NOE build-up curves. No structural calculations were reported for the duplexes containing the nick. An X-ray diffraction analysis of a DNA duplex containing a nick lead to the same conclusions. The DNA duplex is in a B conformation with only slight distortions at the lesion site. There is a hydrogen bond between the two hydroxyl groups adjacent to the nick and a narrow minor groove. The minor groove width may be sequence dependent as the nick was placed in the center of four A·T base pairs *(43)*. For gapped DNA, no structural studies have been published.

Further studies on nicked or gapped DNA duplexes have reported the curvature and the flexibility as observed by gel

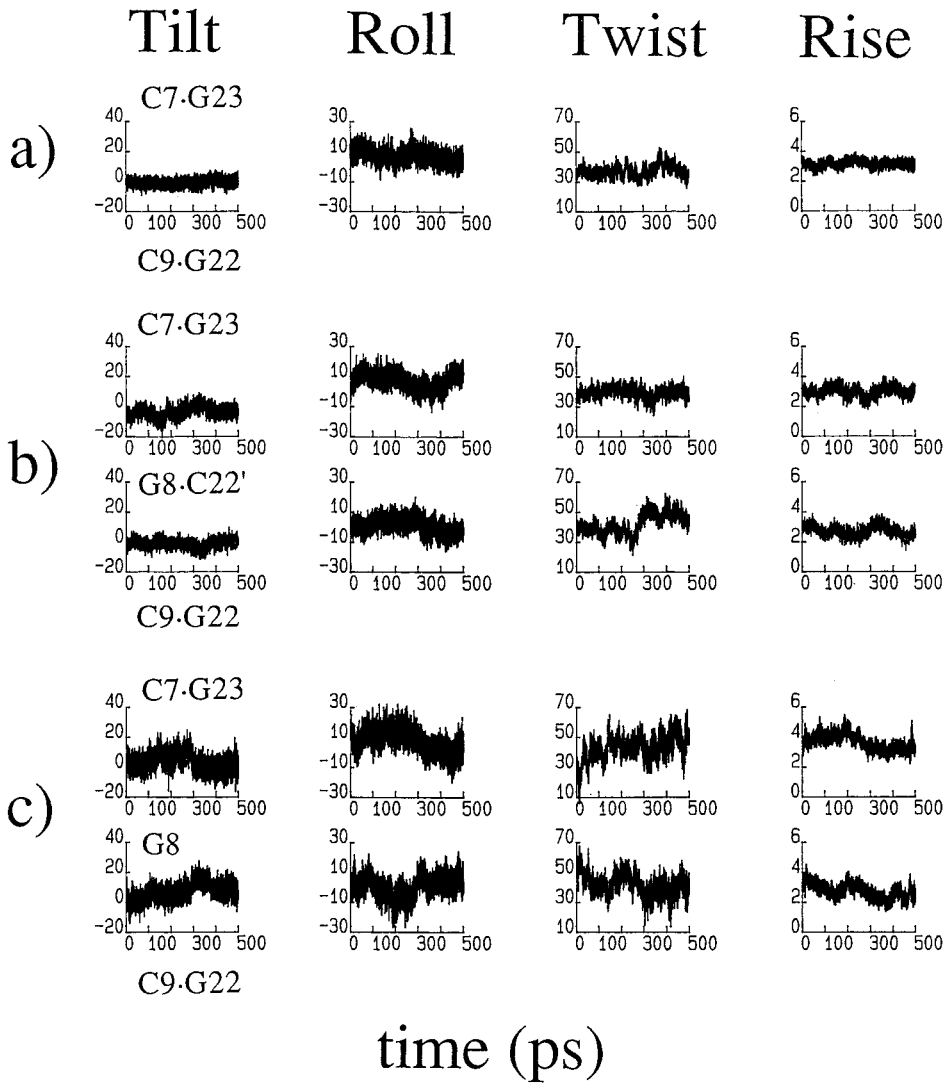

FIGURE 9: Wedge parameters, tilt, roll, twist (in degrees), and rise (in angstroms) versus time (in picoseconds) for the central bases of (a) the nick, (b) the reference 15-mer B-DNA, and (c) the gapG.

electrophoresis or electron microscopy, but with different conclusions. DNA duplexes containing nicks displayed nearly the same flexibility as a comparable duplex without any lesion *(18)* but gaps induce increased flexibility. In another report *(16)*, the authors compared a duplex containing a central nick to a duplex of the same sequence and without any lesion. The authors concluded that the nicked duplex is kinked and that this duplex possesses an increased flexibility compared to the control sample.

Previous simulations with explicit simulation of water and counterions but without PME were limited to a relatively short time scale (ca. 100 ps) and they display highly distorted structures *(44)* or instability such as frequent base-pair opening *(45)*. The treatment of the long-range electrostatic energy terms was done with cutoff techniques. Recent studies using the new force field parametrization for nucleic acids by Cornell et al. *(31)* with explicit representation of water and counterions and with PME approach suggest that the applied computational method is accurate for DNA structures *(46−49)*. For example, it has been shown by Cheatham and Kollman *(48)* that this molecular dynamics simulation protocol well simulates the differences in structure between A-RNA, B-DNA, or DNA:RNA hybrid duplexes. In a recent paper, Duan et al. *(49)* observed a good agreement between their simulation of the Dickerson dodecamer duplex and the X-ray structure for the spine of hydratation in the minor groove. These results show that the PME method is both efficient and accurate for the evaluation of long-range electrostatic interactions, in particular for DNA being studied by molecular dynamics simulations.

Our minimized structures obtained from NMR data were used to initiate MD runs. The starting models for calculations were in good agreement with NMR constraints. For the nick, the two G•C base pairs on either side of the nick are well stacked as shown by NMR and during MD calculations. In contrast, for the gapG, we do not observe NMR interactions between G22 and G23 but for the other strand we observe characteristic B-DNA connectivities between C9, G8, and C7. These results are well reproduced during PME MD runs. These 500 ps PME MD simulations present low rms with no drifts which demonstrates that the calculated structures are stable. Moreover, the expected minor groove water spine of hydration is observed.

The MD runs revealed one homogeneous bent population for the nick, around 15°, as for the duplex with no lesion, and two populations for the gapG, one bent, centered around 28°, and a straight one centered around 11°. We have obtained the same results in two 500 ps MD runs starting

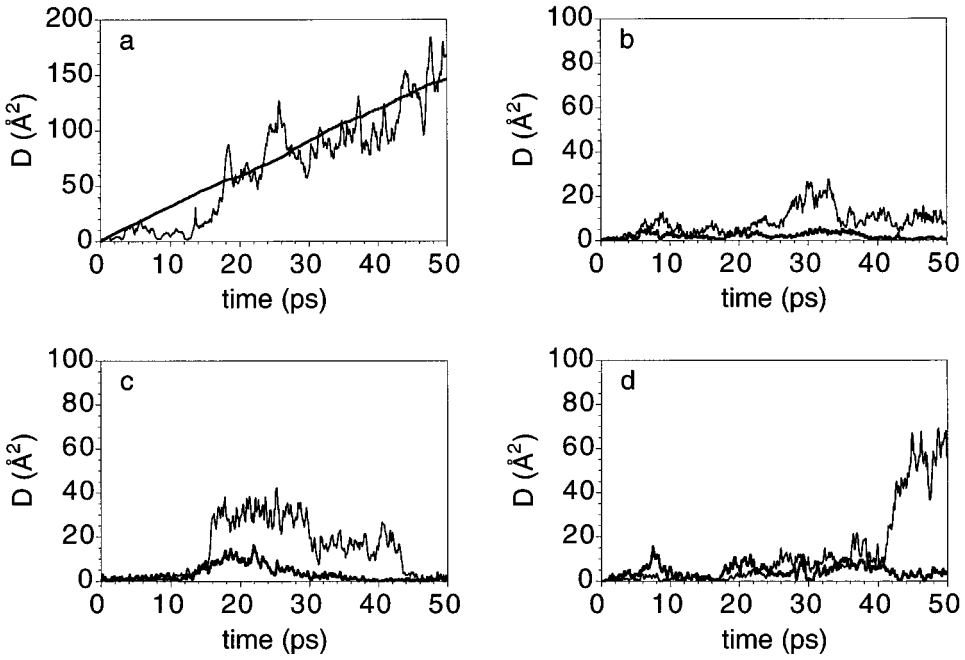

FIGURE 10: Plots of the diffusion of water molecules versus time for (a) a bulk water molecule (thin line) and the average of all the bulk water molecules (thick line), (b) two water molecules belonging to the minor groove, (c) two water molecules belonging to the cavity of the bent gapG conformation and (d) two water molecules belonging to the cavity of the straight gapG conformation. For panels c and d, a water molecule at the limit of the cavity (thin line) and a water molecule in the center of the cavity (thick line).

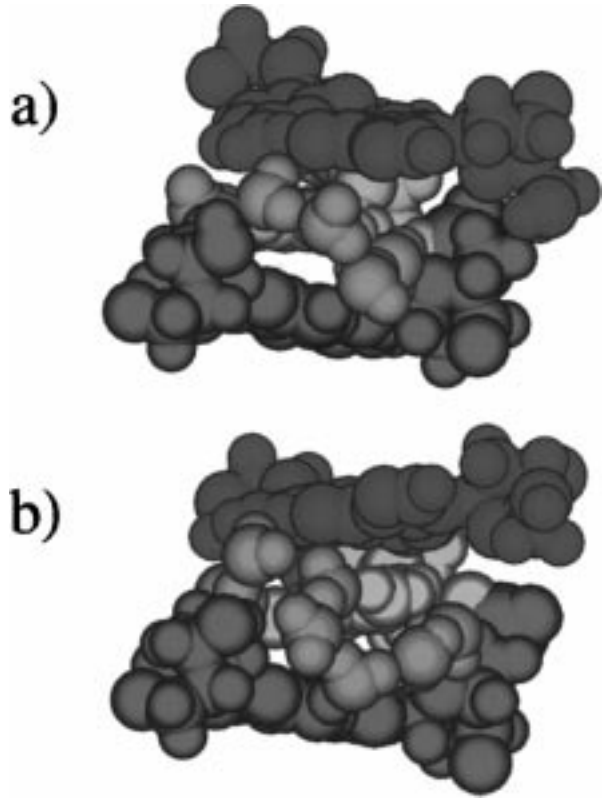

FIGURE 11: Hydrated structures of the central bases of gapG for (a) the bent conformation and (b) the straight conformation. The hydration spine of water molecules is shown in blue and the unpaired G residue in yellow.

from different initial conformations of gapG. We are confident in these results as they are reproducible and give rise to stable structures. Also, they are in agreement with other studies on duplexes containing nicks, bulges, or abasic sites by other techniques *(16, 24, 50−52)*. NMR is unable to provide any information about curvature, as this would require measuring very long interproton distances, which is impossible.

We have analyzed the curvature of the gapG duplex, and it appears that it is a complex movement which depends on various structural parameters such as local wedge angles (roll and twist principally) and, consequently, the minor groove width. The complexity of this motion is due to the presence of a strand break associated with a gap. This motion is a combination of a scissor movement around a pivot base, which corresponds to G8 in this system, and a rotational movement along the helicoidal axis centered on the pivot base. The consequence is that the two- half-helices can move independently, one with regard to the other. The average structure of gapG could be more bent than the nick, due to the presence of a cavity in front of G8, which allows more flexibility. The wedge parameters show fluctuations around the gapG, which are more pronounced than those of the nick and the reference B-DNA 15-mer.

We have searched to identify, during MD runs, the organization of water molecules around the gap and the nick. The hydration of the nick region does not differ from an unbroken B-DNA, except that the missing phosphate residue inhibits the presence of a cone of hydration in this part *(53)*. This is in good agreement with our NMR observations as the imino protons of G22 and G23 are well observed in H_2O, which means that they are not especially exposed to the solvent. The cavity present in the gap structure is hydrated in our models. The imino proton of G8 exchanges rapidly with the solvent, and thus, it is not implicated in a strong hydrogen bond. We cannot distinguish whether imino proton exchange occurs with solvent in the cavity or with bulk solvent as in Watson−Crick base pairs *(54)*. In contrast, the imino protons of the two adjacent G·C base pairs are well identified in 1D NMR spectrum. This shows that they exchange slowly with the solvent and that they must be

protected and engaged in hydrogen bonds. When we increased the temperature, we observed that G22 and G23 exchange in an intermediate rate between that of the terminal base pairs and well-protected imino protons. This NMR result is in good agreement with our models for gapG since we observe a local flexibility around the gap which is higher than that for the Watson−Crick base pairs but lower than the fraying at the ends of the duplex.

Water participates in DNA structure (*48*), and it is well observed in this study. We have pointed out a difference in the water organization in the cavity if the gapG structure is bent or straight but the precise form is influenced by the width of the cavity and also the direction of curvature. For the bent structure, the cavity is narrow and the S shape of the spine of hydration becomes flater compared to the straight structure.

Moreover, the diffusion curves of these water molecules reveal a third class of water molecules with diffusion fingerprints close to those belonging to the minor groove spine of hydration. The best trapped water molecules in the cavity of gapG have residence times in the range 100−300 ps, which is an upper evaluation of the residence time of water molecules involved in minor groove spines of hydration, but rather less than other types of water molecules trapped for instance, by amidinium groups of netropsin compounds (*34*).

CONCLUSION

We find by NMR that the two half-helices in the gapG duplex are in the B conformation and that there is an increased flexibility in the center of the duplex. For correct reincorporation of the excised base by polymerase β in a gap structure, the base on the undamaged strand must be intrahelical to provide the coding information. This is what we have observed for the gapG duplex. Further, no major conformational change is observed relative to a normal B-DNA. However, relative to the nicked duplex, which is virtually identical in its conformation to a duplex without any lesion, we observe for the gapG duplex a similar family of structures but also a second one in which there is a more pronounced kink at the gap site during the MD runs. The solvent occupation of the cavity is significantly modified, as well as the width of the minor groove, in passing from a straight to a bent structure.

We are not aware of any biochemical studies on the repair of single gaps in the presence of the proteins necessary for filling the gap. The repair mechanism is a highly complex process. The questions that we are trying to address are (1) what might be the signal such that the lesion is recognized and (2) what substrate conformation is necessary? The above results show that the lesion site shows enhanced flexibility and may be more kinked than the parent duplex.

We are currently studying, in the same sequence context, the three other Watson−Crick bases at the X position. From these we should be able to determine whether, at least in the initial steps, the residue X is preferentially intrahelical as for gapG, extrahelical, or in equilibrum between the two.

ACKNOWLEDGMENT

We are most grateful to Dr. M. Le Bret for providing us with the programs MORCAD and OCL.

REFERENCES

1. Lindahl, T. (1993) *Nature 362*, 709−715.
2. Dianov, G., and Lindahl, T. (1994) *Curr. Biol. 4*, 1069−1076.
3. von Sonntag, C. (1991) in *Physical and Chemical Mechanisms in Molecular Radiation Biology* (Glass, W. A., and Varna, M. N., Eds.) 287−321, Plenum Press, New York.
4. Boiteux, S., O'Connor, T. R., Lederer, F., Gouyette, A., and Laval, J. (1990) *J. Biol. Chem. 265*, 3916−3922.
5. Dianov, G., Sedgwick, B., Daly, G., Olsson, M., Lovett, S., and Lindahl, T. (1994) *Nucleic Acids Res. 22*, 993−998.
6. Müller, E., Boiteux, S., Cunningham, R. P., and Epe, B. (1990) *Nucleic Acids Res. 18*, 5969−5973.
7. Wallace, S. S. (1994) *Int. J. Rad. Biol. 66*, 579−589.
8. Singhal, R. K., Prassad, R., and Wilson, S. H. (1995) *J. Biol. Chem. 270*, 949−957.
9. Sobol, R. W., Horton, J. K., Kühn, R., Gu, H., Singhal, R. K., Prasad, R., Rajewsky, K., and Wilson, S. H. (1996) *Nature 379*, 183−186.
10. Pelletier, H., Sawaya, S. H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science 264*, 1891−1903.
11. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science 264*, 1930−1935.
12. Steitz, T. A., Smerdon, S. J., Jager, J., and Joyce, C. M. (1994) *Science 266*, 2022−2025.
13. Joyce, C. M., and Steitz, T. A. (1995) *J. Bacteriol. 177*, 6321−6329.
14. Castaing, B., Zelwer, C., Laval, J., and Boiteux, S. (1995) *J. Biol. Chem. 270*, 10291−10296.
15. Lindahl, T., Satoh, M. S., Poirier, G. G., and Klungland, A. (1995) *Trends Biochem. Sci. 20*, 405−411.
16. Le Cam, E., Fack, F., Ménissier de Murcia, J., Cognet, J. A. H., Barbin, A., Sarantoglou, V., Revet, B., Delain, E., and de Murcia, G. (1994) *J. Mol. Biol. 235*, 1062−1071.
17. Hodges-Garcia, Y., Hagerman, P. J., and Pettijohn, D. E. (1989) *J. Biol. Chem. 264*, 14621−14623.
18. Mills, J. B., Cooper, J. P., and Hagerman, P. J. (1994) *Biochemistry 33*, 1797−1803.
19. Cuniasse, P., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Le Bret, M., Guschlbauer, W., and Fazakerley, G. V. (1987) *Nucleic Acids Res. 15*, 8003−8022.
20. Cuniasse, P., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Le Bret, M., and Fazakerley, G. V. (1989) *Biochemistry 28*, 2018−2026.
21. Cuniasse, P., Fazakerley, G. V., Guschlbauer, W., Kaplan, B., and Sowers, L. C. (1990) *J. Mol. Biol. 213*, 303−314.
22. Hare, D., Shapiro, L., and Patel, D. J. (1986) *Biochemistry 25*, 7456−7464.
23. Kalnik, M. W., Norman, D. G., Swann, P. F., and Patel, D. J. (1989) *J. Biol. Chem. 264*, 3702−3712.
24. Rosen, M. A., Live, D., and Patel, D. J. (1992) *Biochemistry 31*, 4004−4014.
25. Pieters, J. M. L., Mans, R. M. W., van den Elst, H., van der Marel, G. A., van Boom, J. H., and Altona, C. (1989) *Nucleic Acids Res. 17*, 4551−4565.
26. Snowdon-Ifft, E. A., and Wemmer, D. E. (1990) *Biochemistry 29*, 6017−6025.
27. Faibis, V., Cognet, J. A. H., Boulard, Y., Sowers, L. C., and Fazakerley, G. V. (1996) *Biochemistry 35*, 14452−14464.
28. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
29. Arnott, S., Campbell Smith, P. J., and Chandrasekaran, R. (1976) in *Atomic coordinates and molecular conformations for DNA-DNA, RNA-DNA and DNA-RNA helices* (Fasman, G. D., Ed.) Vol. 2, CRC Press, Cleveland, OH.
30. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1995) *AMBER 4.1*, University of California, San Francisco.
31. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, G. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
32. Jorgensen, W. L., Chandrasekhar, J., Madura, J., Impeyer, R. W., and Klein, M. L. (1983) *J. Chem. Phys. 79*, 926.

33. Singh, U. C., Weiner, S. J., and Kollman, P. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 755−759.
34. Ketterlé, C., Gabarro-Arpa, J., Ouali, M., Bouziane, M., Auclair, C., Helissey, P., Giorgi-Renault, S., and Le Bret, M. (1996) *J. Biomol. Struct. Dyn. 13*, 963−977.
35. Gabarro-Arpa, J., Cognet, J. A. H., and Le Bret, M. (1992) *J. Mol. Graph. 10*, 166−173.
36. Rosenberg, J. M., Seeman, N. C., Day, R. O., and Rich, A. (1976) *Biochem. Biophys. Res. Commun. 69*, 979−987.
37. Le Bret, M., Gabarro-Arpa, J., Gilbert, J. C., and Lemarechal, C. (1991) *J. Chim. Phys. 88*, 2489−2496.
38. Patel, D. J., Koslowski, S. A., Nordheim, A., and Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 1413−1417.
39. Cognet, J. A. H., Boulard, Y. and Fazakerley, G. V. (1995) *J. Mol. Biol. 246*, 209−226.
40. Babcock, M. S., and Olson, W. K. (1994) *J. Mol. Biol. 237*, 98−124.
41. Saenger, W. (1984) *Principles of Nucleic Acid Structures*, Springer-Verlag, New York.
42. Weerasinghe, S., Smith, P. A., Mohan, V., Cheng, Y.-K., and Pettitt, B. M. (1995) *J. Am. Chem. Soc. 117*, 2147−2158.
43. Aymami, J., Coll, M., van der Marel, G. A., and van Boom, J. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2526−2530.
44. Miaskiewicz, K., Osman, R., and Weinstein, H. (1993) *J. Am. Chem. Soc. 115*, 1526−1537.
45. Beveridge, D. L., and Ravishanker, G. (1994) *Curr. Opin. Struct. Biol. 4*, 246−255.
46. Miaskiewicz, K., Miller, J., Cooney, M., and Osman, R. (1996) *J. Am. Chem. Soc. 118*, 9156−9163.
47. de Souza, N. O., and Ornstein, R. L. (1997) *Biophys. J. 72*, 2395−2397.
48. Cheatham, T. E., and Kollman, P. A. (1997) *J. Am. Chem. Soc. 119*, 4805−4825.
49. Duan, Y., Wilkosz, P., Crowley, M., and Rosenberg, J. M. (1997) *J. Mol. Biol. 272*, 553−572.
50. Coppel, Y., Berthet, N., Coulombeau, C., Coulombeau, C., Garcia, J., and Lhomme, J. (1997) *Biochemistry 36*, 4817−4830.
51. Rice, J. A., and Crothers, D. M. (1989) *Biochemistry 28*, 4512−4516.
52. Woodson, S. A., and Crothers, D. M. (1988) *Biochemistry 27*, 3130−3141.
53. Subramanian, P. S., and Beveridge, D. L. (1989) *J. Biomol. Struct. Dyn. 6*, 1093−1122.
54. Leroy, J. L., Kochoyan, M., Huynh-Dinh, T., and Guéron, M. (1988) *J. Mol. Biol. 200*, 223−238.

BI972377W