Structural study of DNA duplex containing an
N-(2-deoxy-β-D-erythro-pentofuranosyl) formamide
frameshift by NMR and restrained molecular
dynamics

C. Maufrais, G. V. Fazakerley, J. Cadet¹ and Y. Boulard*¹

CEA, Département de Biologie Joliot Curie, Service de Biochimie et de Génétique Moléculaire, Bat 144,
CEA-Saclay, 91191 Gif-sur-Yvette Cedex, France and ¹CEA, Service de Chimie Inorganique et Biologique &
FRE2600, Département de Recherche Fondamentale sur la Matière Condensée, CEA-Grenoble, 38054 Grenoble
Cedex 9, France

Received July 8, 2003; Revised and Accepted September 1, 2003 PDB no. 1OSR

ABSTRACT

The presence of an N-(2-deoxy-β-D-erythro-pentofuranosyl) formamide (F) residue, a ring fragmentation
product of thymine, in a frameshift context in the sequence 5¢-d-(AGGACCACG)´d(CGTGGFTCCT)
has been studied by 1H and 31P nuclear magnetic resonance (NMR) and molecular dynamics. Two-
dimensional NMR studies show that the formamide residue, whether the cis or trans isomer, is rotated
out of the helix and that the bases on either side of the formamide residue in the sequence, G14 and
T16, are stacked over each other in a way similar to normal B-DNA. The cis and trans isomers were
observed in the ratio 3:2 in solution. Information extracted from 31P NMR data reveal a modification of
the phosphodiester backbone conformation at the extrahelical site, which is also observed during the
molecular dynamics simulations.

INTRODUCTION

The chemical modifications induced in cellular DNA exposed to the action of ionizing radiation have been shown to play an
important role in biological processes such as mutagenesis, carcinogenesis and cell lethality. Among the important
radiation-induced base lesions are the thymine ring fragmentation products, such as N-(2-deoxy-β-D-erythro-pentofuranosyl)
formamide, shown in Figure 1. DNA polymerase bypasses fragmentation base products like formamide with an estimated
frequency of 33 and 11% for the Klenow fragment and Taq DNA polymerase, respectively, and thus gives rise to a frameshift mutation (1). However, the formamide lesion also appears to be able to direct nucleotide incorporation, preferentially of guanine.

Frameshifts are caused by unpaired bases or bulges that result from recombination processes or from displacement of
bases during replication (2). Attempts to understand the molecular basis of frameshift mutation lead to numerous
investigations on the structural effects of unpaired nucleotides. All purine bulges studied by nuclear magnetic resonance (NMR) were shown to be intrahelical, having the unpaired base stacked with the neighboring bases in the oligonucleotide (3–9). The only reported example of an extrahelical purine bulge is a crystallographic study of an extrahelical adenine (10). Several duplexes containing single pyrimidine bulges investigated by NMR (11–16) were found to be extra- or intrahelical, depending on the temperature and sequence context. Duplexes with either an abasic site or an abasic analog lesion in a frameshift context were found to be extrahelical (17,18). (2-deoxy-β-D-erythro-pentofuranosyl) urea, another ring fragmentation product of thymine, in a frameshift situation can be either intra- or extrahelical (19). The extrahelical bulge can occur in two ways, including either a left-handed or a right-handed loop. The nature of the loop structure can be determined from unusual NMR interactions, observed in previous studies (17,19,20).

Formamide could be considered as an intermediate structure between a nucleobase and an abasic site. Part of the
coding information has been lost, but the pyrimidine remnant is still potentially able to form hydrogen bonds (Fig. 1). In a
previous paper, we reported a structural analysis by 1H NMR of a formamide residue incorporated opposite a guanine
in DNA (21). We observed two species, which corresponded to the two cis and trans isomers of formamide, that are present in
solution in dynamic equilibrium at a 3:2 ratio. The two isomers were found to be intrahelical and stabilized by hydrogen bonds.

In an attempt to understand why the formamide residue could give rise to a frameshift mutation, we report herein
results of studies on the structural and dynamic properties of an oligonucleotide with an unpaired formamide residue.

MATERIALS AND METHODS

NMR spectroscopy

Mixing a 9mer oligonucleotide with a 10mer oligonucleotide
containing the formamide residue gave rise to the duplex.
During the mixing procedure, the oligonucleotides were heated to 80°C and slowly cooled down to form the duplex. The sequence is: first strand, 5’ A1 G2 G3 A4 C5 C6 A7 C8 G9; second strand, 3’ T19 C18 C17 T16 F15 G14 G13 T12 G11 C10. The duplex was 2 mM single strand concentration dissolved in 10 mM phosphate buffer, 150 mM NaCl and 0.2 mM EDTA. The 3-(trimethylsilyl) propionate peak was used as the internal reference. NMR spectra were recorded on either a Bruker DRX500 or a DRX600 spectrometer and all the spectra were acquired in the phase sensitive mode (22). NOESY spectra were recorded with a mixing time of 40, 50, 60, 80, 100, 250 or 400 ms in D2O and 150 or 250 ms in H2O at 5, 7, 17 and 27°C. In D2O, the residual water resonance was presaturated during the relaxation and mixing delays. In H2O, the water signal was suppressed using the WATERGATE sequence (23). TOCSY experiments were recorded with a mixing time of either 40 or 80 ms. DQF-COSY were recorded with a protocol as in the previous study on the formamide opposite a guanine (21). For each isomer, we kept 226 distance constraints from NOE build-up curves and the torsion angles derived from the DQF experiments as constraints for the model constructions. The best energy refined structure of each isomer was hydrated with TIP3P water molecules (26) and, to obtain electroneutral hydrated models, 17 Na+ counterions were positioned with the LEaP AMBER module (27). The constructed system for the trans and cis isomers of formamide contained 3019 and 3047 water molecules, respectively, in a rectangular box of 50.0 × 58.3 × 46.8 Å3, with periodic boundary conditions applied to the solvent. The dielectric constant was taken to be equal to 1 (28). We applied the particle mesh Ewald method to treat long-range interactions and periodic boundary conditions. An additional cut-off of 10 Å has been used. During the 500 ps of production phase, torsion angles and 206 NMR distance constraints were applied. The conformations generated were analyzed and displayed on SGI workstations using the programs MORCAD (29) and MOLMOL (30). The structural properties were analyzed using the CURVES algorithm (31).

RESULTS AND DISCUSSION

Non-exchangeable protons

The melting temperature (Tm) of the duplex was determined by following the chemical shifts of the aromatic protons as a function of temperature. The Tm value, which is ~51°C, is slightly higher than that obtained for the F-G duplex (48°C).

The H6/H8/H2–H1’/H5 region of a NOESY spectrum of the complex (mixing time 400 ms, 17°C and pH 6.5) is shown in Figure 2. Resonance assignments were achieved following well-established procedures and as outlined for NMR studies of several other base modification-containing duplexes (19,21) using TOCSY and NOESY data sets. The sequential
assignment pathway along the two DNA strands is highlighted using 2-deoxyribose H1' (5.0–6.4 p.p.m.) to base H6/H8 (7.0–8.4 p.p.m.) dipolar correlations to identify neighboring nucleotides (Fig. 2). The pattern of NOE connectivities along the first strand is characteristic of a right-handed B-DNA helix, as illustrated by the A4H8–C5H5 interaction (peak A in Fig. 2), suggesting that no major disruption of the helix takes place between A4 and C5, the residues which could be influenced by the formamide residue on the opposite strand.

For the second strand, the characteristic connectivities for a right-handed B-DNA helix are observed from C10 to G14 and T16 to T19. Further, a direct interaction is observed between T16H6 and G14H1' (peak B in Fig. 2), which shows that the G14 and T16 bases are stacked over each other. This is confirmed in the H6/H8–H2'/H2'/CH3 region, where the T16H6–G14H2'/H2’ and the G14H8–T16CH3 interactions are observed (not shown). The direct T16H6–G14H1'/H2'/H2’ and G14H8–T16CH3 interactions illustrate that the formamide is positioned outside the helix.

After assignment of all the H6, H8 and H2 protons of the Watson–Crick base pairs, two unassigned cross-peaks remain in the H6/H8/H2–H1' region at 8.16 and 8.26 p.p.m. At this stage, only the F15H2/H1' cross-peaks of both the cis and trans isomers of the formamide are not assigned. They must therefore correspond to these interactions. In the 1-dimensional spectra integration of the H2 resonance at 8.26 p.p.m. relative to that of A7H8 shows that the former corresponds to 0.4 protons and thus to the minor species. In 2-dimensional spectra, the minor species is labeled with an exponent, 2. These resonances (8.16 and 8.26 p.p.m.) give rise only to NOEs with their own H1', H2'/H2’ and H3' protons (not shown).

We have carefully searched in other regions but no internucleotide interactions were observed between F15H2 of the two species and the adjacent bases. The H2 chemical shifts of the formamide are shifted ~1 p.p.m. downfield compared with those of the formamide stacked in the helix opposite a guanine in the same sequence context (21). The absence of internucleotide NOE interactions with H2, the downfield shift of the F15H2 protons and the direct interactions between G14 and T16 indicate that the formamide residue is extrahelical in both species. This requires the formation of a loop conformation in order that the two bases adjacent to the extrahelical base are stacked over each other.

There are two possibilities for bulging a base out of the helix; following the backbone it can be either a left-handed or a right-handed loop (17,19), as shown in Figure 3. The nature of the loop can be distinguished by searching for NOE interactions which can exist exclusively in one of the two conformations (20). We find the F15H4'–G14H3' and F15H3'–T16H5'/H5'' interactions unambiguously for both species, which demonstrates that the loops are left-handed (Table 1). Interactions which characterize a right-handed loop are absent. There are other interactions characteristic of a left-handed loop, however, they could not be assigned unambiguously due to overlap of the peaks.

NOESY spectra were recorded at short mixing times (40, 50, 60, 80 and 100 ms) to measure NOE build-up curves for determining proton–proton distances as previously described (21). The distances calculated between T16 and G14 for the duplex correspond to those typically observed between adjacent Watson–Crick base pairs in an overall B conformation.

**Sugar and base conformations**

The sugar conformations of the Watson–Crick residues were initially determined by comparing the intraresidue distances H6/H8–H2' with those of H6/H8–H3'. For all non-terminal sugars of the duplex, a predominantly C2'–endo conformation for the major and the minor species was observed. This is confirmed by measurement of the appropriate coupling

---

**Table 1.** Characteristic NOESY cross-peaks of a right-handed and a left-handed loop for the extrahelical nucleotide (n) and its neighbors

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Observed peak</th>
<th>Overlapping peak</th>
<th>Unobserved peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-handed loop</td>
<td>H2'/H2''(n–1)–H4''(n – 1)</td>
<td>Major, minor</td>
<td>Major, minor</td>
</tr>
<tr>
<td></td>
<td>H2'/H2''(n)–H3'(n – 1)</td>
<td>Major, minor</td>
<td>Major, minor</td>
</tr>
<tr>
<td>Left-handed loop</td>
<td>H4'(n)–H2'/H2''(n – 1)</td>
<td>Major, minor</td>
<td>Major, minor</td>
</tr>
<tr>
<td></td>
<td>H4'(n)–H3'(n – 1)</td>
<td>Major, minor</td>
<td>Major, minor</td>
</tr>
<tr>
<td></td>
<td>H3'(n)–H5'/H5''(n + 1)</td>
<td>Major, minor</td>
<td>Major, minor</td>
</tr>
<tr>
<td></td>
<td>H4'(n)–H5'/H5''(n + 1)</td>
<td>Major, minor</td>
<td>Major, minor</td>
</tr>
</tbody>
</table>

Major, major species; minor, minor species.
constants in DQF-COSY spectra (32) (see Supplementary Material, Fig. S1). These spectra also showed that the formamide sugar puckers for both isomers are in a predominantly C2'-endo conformation, in agreement with the intrasugar NOEs. The sugar/base orientations of the Watson–Crick residues were all anti by comparing the H6/H8–H1' and H6/H8–H2' distances. For both isomers of the formamide, the H2–H1', H2–H2'/H2'' and H2–H3' distances could be determined (21). They give an orientation in the syn range for both isomers. For example, the H2–H1' and H2–H2' distances for the cis isomer are in the range 2.1–2.6 and 4.0–4.5 Å, respectively, whereas the related distances are 3.5–4.0 and 2.1–2.6 Å for the trans isomer.

**Exchangeable protons**

One-dimensional NMR spectra were recorded between pH 5 and 8 and at different temperatures between 1 and 20°C. Apart from the expected line width changes, the spectra showed no significant differences. The best resolution was obtained at 5°C and pH 6.5. Under these conditions, we observed several resonances within the 12–14 p.p.m. range where thymine and guanine imino protons are expected and two relatively broad resonances between 8.82 and 9.05 p.p.m. (top of Fig. 4). Integration of these resonances shows the existence of two species in solution.

Figure 4 shows three regions of a NOESY spectrum recorded in 90% H2O and 10% D2O at 5°C with 250 ms mixing time. The T12NH resonance at 13.92 p.p.m., assigned from the intranucleotide interaction with its methyl group and the A7H2 proton, shows cross-peaks with two G imino protons. They correspond to G13 and G11, which show cross-peaks with the H5 and NH2 protons of C6 and C8, respectively. The cross-peak between the G13NH proton and an imino proton at 12.88 p.p.m. must correspond to G14NH; this received further support from the observation of interactions with the H5 and NH2 protons of C5 on the opposite strand. G14NH also interacts with the A4H2 proton and a T imino proton at 13.95 p.p.m., which must correspond to T16. We can assign the T16NH proton at 13.95 p.p.m. from interactions with its CH3 group and the H2 and NH2 protons of A4. The G14NH/T16NH and G14NH/A4H2 interactions are indicative of stacking of the C5-G14 and A4-T16 base pairs upon each other. From T16NH, the G3 imino proton at 12.71 p.p.m., which shows cross-peaks with C17H5 and C17NH2, can be assigned. Finally, the connectivities between G2 and G3 can be inferred from both the imino–imino and imino–amino interactions. No conditions where the T19 and G9 imino resonances of the terminal base pairs were resolved could be found, probably due to the occurrence of fast exchange with the solvent.

The amide protons of the isomers of F15 are not identified at this stage but the stacking of Watson–Crick base pairs A4–T16 and C5–G14 upon each other confirmed that F15 is extrahelical. The two resonances of exchangeable protons at 8.82 and 9.05 p.p.m. (Fig. 4C) are within the range of chemical shift of amide protons (33). They are assigned as the amide protons of the cis and trans isomers of formamide. In 1-dimensional spectra, the relative importance of the two latter resonances is in the ratio 3:2, with the major species at 9.05 p.p.m. (see top of Fig. 4). The two resonances at 8.82 and 9.05 p.p.m. show intraresidue interactions with the H2 (peaks C and D) and H1' (peaks E and F) protons of the minor and
major species of the formamide. The lack of additional cross-peaks from the two amide resonances indicates, in concordance with observations in D$_2$O, that each isomer of formamide is extrahelical. Comparing the relative intensity of the H1/H2 NOE cross-peak in the same NOESY spectrum of both species of the formamide gives us the side chain orientation (33). The H2/H1 cross-peak of the major species is more intense than that of the minor species after taking into account the 3:2 ratio. Thus the major species is the \textit{trans} isomer.

\section*{\textsuperscript{31}P NMR spectra}

Phosphorus NMR can potentially provide qualitative structural information on the phosphodiester backbone conformation. The dispersion of the \textsuperscript{31}P chemical shifts is related to the structure, sequence and position of the phosphates in a DNA duplex. Figure 5 shows 1-dimensional \textsuperscript{31}P spectra recorded at various temperatures. Most of the \textsuperscript{31}P peaks are well resolved and at 42°C we observe two peaks (~2.0 and ~1.96 p.p.m.) downfield of the others, which suggests that 2 of the 17 \textsuperscript{31}P atoms are in a different environment. At a lower temperature, 37°C, the peak at ~2.0 p.p.m. splits into two. On further lowering the temperature, a new downfield shifted peak appears (~2.37 p.p.m.) at 32°C. The strong temperature dependence of the chemical shift suggests that the \textsuperscript{31}P atom concerned is located on a terminal residue. Analysis of a 2-dimensional \textsuperscript{1}H/\textsuperscript{31}P heteronuclear single quantum correlation spectrum leads to the assignment of all the phosphorus resonances with possible interactions with both H3' (n – 1/n) and the H4' (n) sugar protons (5.34). Figure 6 shows the well-defined 2-dimensional \textsuperscript{1}H/\textsuperscript{31}P spectrum recorded at 42°C. The sequential assignment can be followed for each strand. The A1P-G2H3' and G2P-G3H3' cross-peaks are visible at lower contour plot levels. The G14 and F15 phosphorus atoms are clearly identified, but those corresponding to the \textit{cis} and \textit{trans} isomer duplexes are not resolved. A third resonance in this region, observed more clearly at 37°C (Fig. 5), must arise from one or both of these atoms of the minor species. However, due to its lower concentration and also to resonance overlap, it was not possible to assign it. The strongly temperature-dependent resonance observed at and below 37°C probably arises from C10, but due to the large line width and thus low sensitivity, this could not be confirmed.

The phosphorus atoms of G14 and F15 are shifted ~0.70 p.p.m. downfield from the center of those corresponding to the Watson–Crick base pairs. The \textsuperscript{31}P resonances correlate with the conformation of a phosphodiester in the BI conformation, in which the torsion angles $\epsilon$ and $\zeta$ are (t,g–), several p.p.m. upfield of the \textsuperscript{31}P signal in the BII (g–,t) conformation (35–37). A BI/BII equilibrium leads to an intermediate downfield shift for the \textsuperscript{31}P resonances (38). It is now admitted that the phosphate groups can be in equilibrium between the two BI and BII conformations but with different populations depending on the dinucleotide step and the flanking sequences. The downfield shifted resonances of the phosphorus atoms of G14 and F15, however, do not suggest a pure BII conformation, but they could imply equilibrium of the backbone structure between the BI and BII conformations or at least a modification of the backbone conformation due to the extrahelical base. Overall this is in agreement with a looped out structure.

Attempts were made to measure the \textsuperscript{31}P-\textsuperscript{1}H coupling constants to calculate the torsional angles $\theta$ (H3'-C3'-O-P), which are valuable constraints for molecular modeling. Unfortunately, $J_{p-\textit{H}}$ for the region of principal interest could not be measured accurately due to the line width. There is a very small chemical shift dispersion between the \textit{cis} and
trans forms resulting in a strong resonance overlap that precludes quantitative interpretation.

In addition to heteronuclear experiments, Blommers et al. (39) have demonstrated that experimental validation of the torsion angle $\epsilon$ in the $\mathrm{g}^\prime$ conformation could be checked by the E-COSY type appearance of the $\mathrm{H}_2^\prime$--$\mathrm{H}_3^\prime$ cross-peak in a DQF-COSY spectrum. The $\mathrm{H}_2^\prime$--$\mathrm{H}_3^\prime$ cross-peak patterns recorded in a DQF-COSY spectrum were analyzed (Fig. S1) without revealing such a typical pattern.

**Molecular modeling**

The initial structures for the MD calculations were constructed as described in Materials and Methods. The interresidue interactions observed in different regions of the NMR spectra show that the Watson–Crick base pairs adopt a right-handed B-DNA conformation with G14 and T16 stacked over each other. In all cases, the formamide remnant is outside the helix. In the left-handed looped out structure built from previous data (17,40), the plane of the 2-deoxyribose moiety of the formamide nucleoside is tangential to the helical cylinder and perpendicular to the helix axis on the major groove side. The formamide side chain cannot interact with the rest of the helix and is completely exposed to the solvent. The data from a refinement procedure conducted with A-DNA as the starting point rapidly converge under NMR data to a B-DNA structure.

We have first performed molecular simulations with implicit water because it allows us to rapidly test the stability of the different models. Futhermore, it is for a given amount of CPU time a more extensive sampling method compared to calculations with explicit water. For each of the isomers, two families of structures obtained from model building as described in Materials and Methods and also (see below) observed during MD runs with implicit water were used to initiate MD runs with explicit water. These conditions are a priori more accurate in testing the influence of water molecules on the structures.

For each of the isomers, the two initial structures were energy minimized and MD runs over 500 ps with explicit water and distance constraints derived from NOE build-up curves together with constraints on torsion angle $\delta$ were performed. The structures differ by the position of the extrahelical formamide. In the first structure, each of the isomers of F15 is completely exposed to the solvent (MD1t and MD1c). In the second structure, F15 lies in the major groove along the G14 base (MD2t and MD2c). In the MD1t runs for the trans isomer, the formamide completely exposed to the solvent maintains its position during the entire period of calculation. In contrast, in MD2t the formamide inclined towards the major groove at the start of the runs, becomes exposed to the solvent after 50 ps of phase production and maintains this position. For the cis isomer, results obtained for MD1c and MD2c are similar. The formamide explores a large conformational space. In some structures, the formamide is inclined towards the major groove close to G14, while in others it remains completely exposed to the solvent.

The difference between the two isomers is due to the possibility that the cis isomer could form a hydrogen bond with the backbone (Fig. 7b), whereas the trans isomer could not (Fig. 7a). As shown in detail in Figure 7c and d, the orientation of the cis isomer (Fig. 7d) allowed the formation of a hydrogen bond between the H2 proton and its phosphate group (F15H2–P14O), which stabilized the orientation of the formamide along the structure. The situation for the trans isomer is different because no hydrogen bond between the base and the backbone could be formed (Fig. 7c). In consequence, the formamide is completely exposed to the solvent. During the whole period of calculation the H2 proton of the trans isomer is always away from the duplex, in agreement with the absence of interresidue NOEs with this proton. It is now widely accepted that C–H⋯O contacts constitute electrostatically stabilized attractive interactions that can be considered as weak hydrogen bonds (41–43). The energy of the C–H⋯O interaction was calculated to be ~2 kcal mol$^{-1}$ (44). For both isomers, stacking of the A4–T16 and C5–G14 base pairs is unaffected by the presence of the formamide and the first strand remains close to B-DNA. Results for the trans isomer are quite similar to those of the cis isomer for the completely solvent exposed structure.

We will focus our analysis on the results for the cis isomer. Figure 7e shows a superposition of 25 structures taken during the MD simulation with implicit water. The overall structures as well as local features such as base stacking are well determined in the center of the duplex while the position of the formamide fluctuates during the MD runs. The conformational space explored by the formamide is shown in Figure 7f. The structure where the formamide lies in the major groove occurs only 34% of the time (Fig. 8a). The duplex presents a smooth rms curve during the 500 ps of phase production, as shown in Figure 8b, demonstrating that the structure is well stabilized over all the MD runs. The rms fluctuations around 1.58 ± 0.3 Å. Interestingly, during the MD runs the structures appear very consistent with the experimental data. The sum $F = \sum (r_{ij} - d_{ij})^2/d_{ij}$, where $r_{ij}$ represents the NMR distance measurements and $d_{ij}$ the proposed distance in a given structure, is the estimate of the overall agreement between computed and NMR data. The average of the sum for the whole duplex during the MD runs is 4.5 ± 0.3 Å; this decreases considerably when the terminal base pairs are eliminated, to 2.5 ± 0.2 Å, as shown Figure 8b. This difference is due principally to end fraying, as these base pairs are less well characterized. In Figure 8c, the hydrogen bonding F15H2–P14O illustrates that the formamide lies in the major groove for part of the time. Comparison of Figure 8b and c shows that movement of the formamide outside the helix does not disturb the global helix and that, whatever the formamide position, all structures are in good overall agreement with the NMR data. However, during the MD runs some of the NMR distances which characterize a left-handed loop vary with the formamide position. The F15H4′–G14H3′ distance varies from ~2.0 Å, when the formamide is completely exposed to the solvent, to 4.0 Å, when it lies in the major groove (Fig. 8d). Similar variations for both the F15H4′–G14H2′/2″ (2.8/2.5 to 4.7/3.2 Å) and F15H4′–T16H5′/H5″ distances (3.5/4.1 to 2.1/2.3 Å) were found; while some distances did not vary with the position of the formamide, such as F15H4′–T16H5′/5″ (not shown). The computed distances characteristic of a left-handed loop agree best with the NMR data when the formamide is completely exposed to the solvent. Populations of structures where the formamide lies in the major groove cannot be excluded, as calculated NMR distances are not very sensitive to this conformational change.
The looped-out structure deforms the phosphodiester backbone around the site of the extrahelical base, which is not in the classical B-form. From the analysis of an abasic frameshift structure four torsion angles differ from B-DNA angles when the abasic site is involved in a left-handed loop (17,40), $\varepsilon$ on the 5' side of the abasic site and $\beta$, $\gamma$ and $\zeta$ of the 2-deoxyribose
moiety of the abasic site. Similar variations are observed here during the MD runs. The torsion angles for the A4-C5 segment are shown in Table 2. The backbone angles $\beta$, $\gamma$, $\epsilon$ and $\zeta$ differ markedly at the G14-F15 and F15-T16 steps compared to the others. Moreover, angle $\xi$ of G14 and the $\beta$, $\gamma$ and $\epsilon$ angles of F15 fluctuate with the formamide position outside the helix (Fig. 8e and f). The two extreme values of these angles that correspond to the formamide completely exposed to the solvent or next to G14 are shown in Table 2. The $\epsilon$ and $\zeta$ angles of G14 are not characteristic of those found usually in the BI or BII conformations. Fluctuation of these angles suggests modification of the backbone conformation due to the extrahelical base. These unexpected phosphate backbone conformations are corroborated by the $^{31}$P experiments discussed above. Finally, all sugar puckers were found in the C2'-endo conformation and the glycosidic angle of formamide is in the syn range. However, fluctuations around the $\chi$ angle of formamide are important.

Two MD runs with explicit water molecules over 500 ps for the two extreme positions of the extrahelical formamide were performed to compare the stability of the models. In the first MD1 runs, F15 is completely exposed to the solvent. In the second MD2 runs, F15 at the beginning of the calculations lies in the major groove. The rms $\pm$ SD versus time evolution of the two duplexes follow stable trajectories after ~50 ps of phase production ($1.32 \pm 0.2$ Å for MD1 and $1.39 \pm 0.3$ Å for MD2) (Fig. 9a and b). During both MD runs, the base pairs at the center of the duplexes are well formed and the A4-T16 and C5-G14 base pairs remain stacked over each other. During MD1, the formamide never lies in the major groove, but is exposed to the solvent, the population of formamide being completely outside (Fig. 9c and e). Figure 9d and f shows the same parameters for MD2. During the first 100 ps of calculation of MD2, the formamide position varies between a position inclined towards the major groove next to G14 and a position completely exposed to the solvent. The F15H2±P14O distance varies from 3.2 Å, when the formamide lies in the major groove, to ~6.5 Å, when it is completely exposed to the solvent (Fig. 9d). After these 100 ps of phase production, the formamide remains exposed to the solvent and follows approximately the same trajectory as during MD1. The formamide lies in the major groove for only ~5% of the time during MD2 (Fig. 9f). Compared to the MD runs with explicit water, where a bimodal distribution was observed for the formamide position, a unimodal distribution was observed with explicit water. This difference could be explained by the fact that with explicit water the extrahelical formamide is better stabilized because it could form hydrogen bonds with
Table 2. Average torsion angles for the 500 ps MD runs with implicit water for the central d(A4-C5)-d(G14-F15-T16) segment of the duplex with the formamide in the cis conformation

<table>
<thead>
<tr>
<th>Base</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>ε</th>
<th>ζ</th>
<th>χ</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-DNA</td>
<td>-65</td>
<td>167</td>
<td>51</td>
<td>129</td>
<td>-157</td>
<td>-120</td>
<td>-103</td>
</tr>
<tr>
<td>Implicit water</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>-64 ± 9</td>
<td>168 ± 8</td>
<td>51 ± 9</td>
<td>146 ± 6</td>
<td>171 ± 6</td>
<td>-113 ± 9</td>
<td>-107 ± 9</td>
</tr>
<tr>
<td>C5</td>
<td>-71 ± 10</td>
<td>172 ± 5</td>
<td>57 ± 8</td>
<td>141 ± 6</td>
<td>172 ± 6</td>
<td>-108 ± 11</td>
<td>-109 ± 9</td>
</tr>
<tr>
<td>G14</td>
<td>-67 ± 10</td>
<td>161 ± 11</td>
<td>46 ± 10</td>
<td>159 ± 5</td>
<td>-82 ± 9</td>
<td>-95 ± 10</td>
<td>-90 ± 9</td>
</tr>
<tr>
<td>F15</td>
<td>70° ± 10</td>
<td>-89° ± 9</td>
<td>-79° ± 9</td>
<td>158 ± 8</td>
<td>-163 ± 10</td>
<td>80° ± 9</td>
<td>-63 ± 17</td>
</tr>
<tr>
<td>T16</td>
<td>-6° ± 8</td>
<td>169 ± 7</td>
<td>8° ± 8</td>
<td>160 ± 4</td>
<td>171 ± 6</td>
<td>-104 ± 9</td>
<td>-129 ± 8</td>
</tr>
<tr>
<td>Explicit water (MD1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>-64 ± 8</td>
<td>168 ± 9</td>
<td>48 ± 9</td>
<td>140 ± 7</td>
<td>173 ± 5</td>
<td>-101 ± 9</td>
<td>-114 ± 8</td>
</tr>
<tr>
<td>C5</td>
<td>-66 ± 9</td>
<td>172 ± 7</td>
<td>51 ± 9</td>
<td>139 ± 7</td>
<td>170 ± 8</td>
<td>-106 ± 12</td>
<td>-108 ± 8</td>
</tr>
<tr>
<td>G14</td>
<td>-65 ± 9</td>
<td>171 ± 7</td>
<td>52 ± 9</td>
<td>150 ± 7</td>
<td>-80 ± 11</td>
<td>-99 ± 12</td>
<td>-94 ± 7</td>
</tr>
<tr>
<td>F15</td>
<td>70° ± 12</td>
<td>-99° ± 15</td>
<td>-70° ± 12</td>
<td>157 ± 6</td>
<td>-160 ± 13</td>
<td>78° ± 15</td>
<td>-76 ± 18</td>
</tr>
<tr>
<td>T16</td>
<td>-96 ± 21</td>
<td>-115 ± 23</td>
<td>40 ± 14</td>
<td>150 ± 6</td>
<td>173 ± 5</td>
<td>-100 ± 9</td>
<td>-123 ± 8</td>
</tr>
<tr>
<td>Explicit water (MD2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>-65 ± 9</td>
<td>165 ± 9</td>
<td>44 ± 10</td>
<td>141 ± 7</td>
<td>173 ± 5</td>
<td>-101 ± 9</td>
<td>-115 ± 8</td>
</tr>
<tr>
<td>C5</td>
<td>-66 ± 9</td>
<td>171 ± 7</td>
<td>49 ± 9</td>
<td>141 ± 7</td>
<td>170 ± 7</td>
<td>-103 ± 11</td>
<td>-106 ± 8</td>
</tr>
<tr>
<td>G14</td>
<td>-68 ± 10</td>
<td>169 ± 8</td>
<td>49 ± 10</td>
<td>148 ± 7</td>
<td>-78 ± 11</td>
<td>-103 ± 14</td>
<td>-94 ± 8</td>
</tr>
<tr>
<td>F15</td>
<td>73° ± 10</td>
<td>-94° ± 15</td>
<td>-72° ± 16</td>
<td>157 ± 6</td>
<td>-159 ± 12</td>
<td>79° ± 13</td>
<td>-73 ± 18</td>
</tr>
<tr>
<td>T16</td>
<td>-88 ± 18</td>
<td>-122° ± 23</td>
<td>44 ± 12</td>
<td>149 ± 6</td>
<td>173 ± 5</td>
<td>-100 ± 9</td>
<td>-125 ± 8</td>
</tr>
</tbody>
</table>

Numbers in bold correspond to torsion angles when the formamide lies in the major groove.
Numbers in italic correspond to the variation of the ε angle during MD runs with explicit water.

aThe parameters for B-DNA are taken from Dickerson et al. (45).
bValues outside the normal range.

CONCLUSION

The presence of an extrahelical formamide lesion induces limited perturbations of the DNA backbone. NMR data show clearly that the duplex adopts a conformation very close to that of B-DNA with the formamide outside the helix. No conformational change on strand one is detected compared to B-DNA. For the modified strand, distortions of the sugar-phosphate backbone are limited to the lesion site. In NMR spectra, only the G14 and the F15 phosphorus atoms resonate outside the region usually observed for B-DNA duplexes. Molecular dynamics simulations show that a large number of conformations are accessible to the extrahelical formamide residue. 31P NMR spectra and computational studies reveal an equilibrium for the backbone conformation at the extrahelical site. The lack of conformational preference for the formamide backbone angles allows this structural freedom.

For a (2-deoxy-D-erythro-pentofuranosyl) urea residue in the same sequence context, Gervais et al. (19) found that the lesion could be intra- or extrahelical and these two conformations corresponded to the cis and trans isomers of the lesion, respectively. The intrahelical form (cis isomer) was stabilized by hydrogen bonds formed with a base above or below and on the opposite strand. For the extrahelical species (trans isomer), the looped-out conformation was right-handed, with the adjacent bases stacked over each other. The lesion is situated in the minor groove and could form both intra- and inter-strand hydrogen bonds. The unusual hydrogen bonding possibilities of the urea lesion could explain the major structural differences between this and the formamide duplex. Since the formamide residue cannot form hydrogen bonds with neighboring bases, which could stabilize it inside the duplex, only an extrahelical form was found for each isomer. On the other hand, the results obtained for the formamide in a frameshift situation are similar to those reported by Cuniasse et al. (17) and Lin et al. (18) for an abasic lesion, although no intersugar interactions were reported in the latter case. In both cases the
A basic lesion was extrahelical and the flanking base pairs were well stacked upon each other. Whatever the bulged out base, either formamide or an abasic lesion, the bases 5'¢ and 3'¢ to the lesion are well stacked, as if they belonged to consecutive nucleotides. A gain in stacking energy therefore brings the two bases closer together.

In DNA an extrahelical base is considered to play an important role in frameshift mutagenesis. If the base and the sugar are in the extrahelical conformation at the replication site, DNA polymerase may not incorporate a nucleotide opposite the lost base and cause a shift in the translational reading frame. The structural conformations obtained in this study suggest that the formamide lesion should exhibit mutagenic features similar to those of an abasic site.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS
We thank Marc Le Bret for the kind gift of the programs OCL and MORCAD. This work was, in part, supported by funds from the Ministère de l’Education Nationale et de l’Enseignement Supérieur, ACC-SV8.

REFERENCES


