Structures of Apurinic and Apyrimidinic Sites in Duplex DNAs*

Richard D. Beger and Philip H. Bolton‡

From the Department of Chemistry, Wesleyan University, Middletown, Connecticut 06459

Natural and exogenous processes can give rise to abasic sites with either a purine or pyrimidine as the base on the opposing strand. The solution state structures of the apyrimidinic DNA duplex, with Dₐ, indicating an abasic site,

\[ d(C₁G₂C₃G₄A₅D₆A₇C₈G₉CₐCₐC₁₁) \]

\[ d(GₐG₂G₉CₐCₐTₗAₕTₕG₉GₑGₓG₄Gₑ) \]

referred to as AD, and the apurinic DNA duplex with a dC₁₇, referred to as CD, have been determined. A particu-
larly striking difference is that the abasic site in CD is predominantly a β hemiacetal, whereas in AD the α and β forms are equally present. Hydrogen bonding with water by the abasic site and the base on the opposite strand appears to play a large role in determining the structure near the damaged site. Comparison of these structures with that of a duplex DNA containing a thy-
mine glycol at the same position as the abasic site and with that of a duplex DNA containing an abasic site in the middle of a curved DNA sequence offers some in-
sight into the common and distinct structural features of damaged DNA sites.

Damage to DNA bases can arise from a number of routes including oxidative stress, the action of various chemical agents, and by radiative processes (1–5). Common examples of base damages include the spontaneous deamination of cytosine to uracil, the oxidation of thymine to thymine glycol or urea, and the photochemical production of thymine dimer adducts. The first step in repair of damaged DNA in vivo is often the hydrolytic cleavage of the C-N bond between the sugar and the damaged or unusual base to generate an aldehydic abasic site that is sometimes referred to as an apurinic/apyrimidinic or AP₁ site (1–5).

The base opposite the abasic site is of interest because there are many routes by which an abasic site can be generated (2, 3, 6, 7). Deamination of a dC residue followed by the action of N-uracil DNA glycosylase leads to dG opposite the abasic site. Oxidation of dT to thymine glycol or urea and the subsequent action of a glycosylase lead to dA opposite an abasic site, whereas base damage to dA or dG can lead to having dC or dT opposite the abasic site. The base opposite the abasic site may have structural, dynamic, or other properties that affect the chemical reactivity of the duplex DNA, its recognition by proteins, or its interactions in subsequent repair reactions (8, 9). Also, it now appears that some polymerases may place all four bases opposite an abasic site and not only dA residues (10) leading to both apurinic and apyrimidinic sites. The structure of damaged DNA is sequence-dependent as is the structure of undamaged DNA, and the presence of an abasic site within a curved DNA sequence can have large, long range structural effects (11). Results on the structures of repair enzymes have led to the suggestion that the abasic site can be recognized directly (12).

Damaged DNA plays roles in controlling the cell cycle, and these roles are of considerable and growing interest because of their importance in apoptosis and carcinogenesis as noted in recent reviews (13–18). It now appears that there are “check-
points” for the presence of damaged DNA which need to be passed before a cell can go from G₁ to G₂ (18). The processes by which these checkpoints control the G₁ to G₂ passage are not yet known. It is also not known how many types of damaged DNA are recognized at these checkpoints nor how many differ-
ent procedures are used to recognize damaged DNA (19–22). Normal cells with damaged DNA are programmed to undergo apoptosis, and apparently only cells with damaged DNA and which avoid this step can become transformed into tumor cells (2, 16, 18, 23–26). Abasic sites are also poisons for topoisome-
rase, which may be important in apoptosis (27–29).

In many cells the predominant pathway to the formation of aldehydic abasic sites is the spontaneous deamination of cytosine to uracil (5, 6, 30–34). The enzyme N-uracil glycosylase excises uracil to form an aldehydic abasic site as shown in Scheme I.

For a typical *Escherichia coli* there are about 40–400 such events per cell division, and in a typical mammalian cell 4,000–
40,000 uracil are formed per cell division (2, 35, 36). The num-
ber of damaged sites in a particular genome depends on the state of the cell, and some damaged sites may not be repaired until the G₁ to G₂ passage occurs. The number of abasic sites in a “typical” human cell is not known because the rates of dam-
gage and repair are dependent on many factors. Ames and co-workers (37) have estimated that there are more than 10,000 damaged sites, of all types, per typical human cell at any given time. The presence of a direct relationship between DNA damage and cancer is known, but the number and types of damage to DNA which are required for transformation to occur are just now being determined.

The cleavage of the glycosidic bonds is catalyzed by DNA glycosylases, which were first identified by Lindahl and Nyberg

---

*This research was supported in part by American Cancer Society Grant NP-750. The 400-MHz NMR spectrometer was purchased with support from National Science Foundation Grant BIR 93-03077. The 500-MHz spectrometer was purchased with support from National Sci-
ence Foundation Grant BIR-95-12478 and by the Camille and Henry Dreyfus Foundation. The costs of publication of this article were de-
layed by the Camille and Henry Dreyfus Foundation Grant BIR-95-12478 and by the Camille and Henry Dreyfus Foundation.\‡ To whom correspondence should be addressed. Tel.: 860-685-2668; Fax: 860-685-2211; E-mail: pbolton@wesleyan.edu.

\* The abbreviations used are: AP, apurinic/apyrimidinic; AD, apyri-
midinic duplex; CD, apurinic duplex; NOESY, nuclear Overhauser ef-
fect spectroscopy; ROESY, rotating frame Overhauser effect spectro-
copy; TOCSY, total correlation spectroscopy; PECOSY, phased easy correlation spectroscopy; NOE, nuclear Overhauser effect; r.m.s., root mean square; DQ[COSY], double quantum filter correlation spectroscopy.
in 1974 (38), and many distinct classes of glycosylases are now known (6). Crystal structures of two uracil glycosylases have been determined (39, 40) as have the structure of uracil glycosylase inhibitor protein free and complexed to uracil glycosylase (41–43). The abasic site is not a chemically unique species but is an equilibrium mixture (44–47) of α- (I), and β- (II), hemiacetals that are 2-deoxy-D-erythro-pentofuranoses, of aldehyde (III), and of hydrated aldehyde (IV), as depicted below. The 3'-cleavage of the abasic site catalyzed by AP lyases is via a syn β-elimination reaction (6) (Scheme II).

Subsequent to the formation of the aldehydic abasic site the repair process can continue with the cleavage of both the 5'- and 3'-phosphodiester bonds. The 3'-cleavage reaction catalyzed by UV endonuclease V of bacteriophage T4 and by E. coli endonuclease III or the enzyme formamidopyrimidine-DNA glycosylase is a β-elimination that proceeds with syn stereochemistry, with abstraction of the 2'-pro-S proton, as discussed elsewhere (44–48). The 5'-cleavage is via a delta elimination when catalyzed by formamidopyrimidine-DNA glycosylase (40). After the abasic site is cleaved at both the 3'- and 5'-phosphodiester linkages the site is repaired via synthesis and ligation of the DNA (7).

There have been a number of investigations of the effects of unrepaired aldehydic abasic sites on replication. Studies on the effect of abasic sites on both in vivo and in vitro replication have been carried out on a variety of DNA polymerases (2, 5, 8, 49–51). Recent results indicate that DNA damage can affect DNA replication at sites remote from the site of damage (52) indicative of interactions of the polymerases with the DNA at positions remote from the site of damage.

The presence of an unrepaired aldehydic abasic site can lead to a stable mutation as well as having effects on transcription (2, 53, 54). The results to date suggest that the presence of an abasic site slows down but does not block transcription (53–55). The structural features of AD can be compared with those of a damaged DNA with thymine glycol (57) in the same position as the abasic site. The structural features of AD can also be compared with a DNA in which the abasic site is in the middle of a curved DNA stretch (11). These comparisons indicate what may be some of the common and distinct features of damaged DNAs.

**MATERIALS AND METHODS**

**NMR Procedures**—NMR data were acquired at 500 MHz on a Varian Inova and at 400 MHz on a Varian Unityplus spectrometer, at Wesleyan, and at 500 MHz on a Bruker 500 MHz DMI spectrometer at the University of Wisconsin, Madison, using methods described previously (11, 57–59) and modified as discussed below. All 31P experiments were obtained at 161 MHz using the 400-MHz Varian Unityplus. All of the Varian NMR data were processed using VNMR software, and all of the Bruker data were processed using Felix 95.0 software. All of the Varian two-dimensional data were obtained using States-Haberkorn and all of the Bruker data using TIPPI.

NOESY experiments on AD were run on the Bruker 500 at Madison with the sample in 2H2O at 25 °C with mixing times of 100 and 200 ms and a 1.6-σ equilibration delay. The spectral width in each dimension was 5,000 Hz. 512 t1 increments were acquired with 64 transients/t1 point and 1024 complex points in the t2 dimension. The data were processed with a sinebell apodization in both dimensions before 2048 × 2048 Fourier transformation. These spectra were used for NOESY cross-peak quantification.

NOESY experiments on CD were run on the Varian 500 with the sample in 2H2O at 15 °C with mixing times of 100 and 250 ms and a 1-σ equilibration delay. The spectral width in each dimension was 6,000 Hz. 448 t1 increments were acquired with 48 transients/t1 point and 4096 complex points in the t2 dimension. The data were processed with a Gaussian weighting in both dimensions before 4096 × 1024 Fourier transformation. These spectra were used for NOESY cross-peak quantification.

NOEY experiments on AD were run on the Varian 500 with the sample in 2H2O at 15 °C with mixing times of 100 and 250 ms and a 1-σ equilibration delay. The spectral width in each dimension was 6,000 Hz. 448 t1 increments were acquired with 48 transients/t1 point and 4096 complex points in the t2 dimension. The data were processed with a Gaussian weighting in both dimensions before 4096 × 1024 Fourier transformation was used. NOEY experiments, with watergate suppression of the water resonance, on CD in 90% H2O and 10% 2H2O, were run using the Varian 500 with the sample at 15, 5, and 1 °C with a mixing time of 50 ms and a 1-σ equilibration delay. The spectral width was 12,000 Hz in both dimensions, and the data were transformed using Gaussian shifted weighting functions into 4096 × 1024 real points.

**NOE cross-peak quantification**

**Quiet-NOEY experiments** (60) on CD and AD were run on the Varian 400 with the sample in 2H2O at 15 °C with a mixing time of 250
ms and a 1-s equilibration delay. In the middle of the mixing time a 180° shaped Gaussian pulse was applied on the aromatic region in one experiment and on the methyl region in a separate Quiet-NOESY experiment. The spectral width in each dimension was 5,000 Hz. 256 t1 increments were acquired with 48 transients/t1 point and 4096 complex points in the t2 dimension. The data were processed using Gaussian apodization in both dimensions before 4096 × 1024 Fourier transformation.

Band-selective TOCSY experiments (61) were run on the Varian 400 with the samples in D2O and at 15 °C for CD and 27 °C for AD with mixing times of 70 ms and a 1-s equilibration delay. Two 180° Gaussian shaped pulses were applied to the H3′ region during the spin echo before the TOCSY spin lock. The spectral width in each dimension was 5,000 Hz. 500 t1 increments were acquired with 48 transients/t1 increment and 4096 complex points in the t2 dimension. The J scale for coupling to 31P was set to 2 and to 3 in separate experiments. The data were processed with a Gaussian apodization in both dimensions before 4096 × 1024 Fourier transformation.

PECOSY spectra were acquired for the AD and CD samples at 400 MHz with 31P decoupling during the evolution period. The data were collected as 2,000 complex t1 points and 512 complex t2 points. Gaussian weighting was applied in both dimensions, and the data were zero filled to 2,000 × 2,000 real points.

TOCSY experiments were run on AD and CD with a mixing time of 80 ms with the samples in D2O at 400 MHz. The water resonance was presaturated during the 1.5-s delay period. The spectral width in both dimensions was 6,000 Hz. 2,048 complex points were collected in the t2 dimension and 256 complex points in t1. Gaussian weighting was applied before Fourier transformation into 2048 × 1024 real points.

A series of one-dimensional inversion recovery experiments was run on AD and CD to allow determination of the chemical shifts of the adenine H2 resonances on the basis of their relatively long T1 values. These spectra were acquired with the samples in D2O at 400 MHz. An equilibrium delay of 12 s was used, and the inversion recovery delay time was varied from 0.3 to 2.0 s in equal steps.

The Bruker 500 MHz spectrometer was used to obtain a 125-ms ROESY experiment on AD in D2O with the sample at 25 °C. A spectral width of 6,000 Hz was used in each dimension. The data were collected as 2,000 complex t1 points and 512 complex t2 points. Gaussian weighting was applied in both dimensions, and the data were zero filled to 2,000 × 2,000 real points.

Quantitation and Assignment of NOE Cross-peaks—The spectra were assigned using the procedures described previously for abasic site-containing DNAs (11, 58, 62). The standard sequential connectivities of B-form DNA could be used except at the site of the damage. In the α form the H1′ is spatially closer to the H2′ than the H2′. In the β form the H1′ is spatially closer to the H2′ than the H2′. AD NOE cross-peak volumes were quantified from the data obtained at 500 MHz at mixing times of 100 and 200 ms using FELIX 95.0 software. The NOE cross-peak volumes, of CD, were quantified in the data obtained at 500 MHz with mixing times of 100 and 250 ms using VNMRFIT software. For each assigned cross-peak the volume over a standard area was determined.

Structure Determination Procedures—Structure refinement by restrained molecular dynamics using X-PLOR 3.1 was performed as described previously (11, 57–59, 63) with the following modifications. The optimum weighting of the NOE constraints was found to be 60 kcal/mol, and the experimental volumes were used as the constraints. For AD, at each mixing time, 233 NOE volume constraints were used for the α and for the β structure, and the refinements of the two structures were carried out independently. During each structure calculation 34 dihedral constraints for the ribose (H1′-C1′-C2′-H2′ and H1′-C1′-C2′-H2′) were used as well as 20 dihedral constraints for the backbone (P-O3′-C3′-H3′) with a weighting of 40 kcal/mol.

For all four structures 28 hydrogen bonds were used for the 10 base pairs with a weighting of 40 kcal/mol, and 2 hydrogen bonds were used on the bound water protons to the abasic site hydroxyl and to the proton acceptors on the adenine or cytosine across from the abasic site with a weighting of 40 kcal/mol. The hydrogen bond proton-oxygen distance constraints were set to 1.86 ± 0.20 Å, and the nitrogen-proton distance constraints were set to 1.86 ± 0.20 Å. The nonbonded interaction cutoff was set to 11.5 Å. The distance over which the nonbonded interaction was switched from on to off was 9.5 to 10.5 Å. The distance cutoff for the Watson-Crick hydrogen bonding interactions was set to 7.5 Å, and the switching function was applied from 5.5 to 6.5 Å.

Starting structures were generated from canonical B-DNA by replac-
Fig. 2. The spectra shown are the experimental 500-MHz two-dimen- 
sional NOESY data set for AD ob-
tained with a 100-ms mixing time 
shown at the top. The region shown 
contains the H6/H8-H1' cross-peaks. The 
same region is also shown for the pre-
tained with a 100-ms mixing time 
sional NOESY data set for AD ob-
point were averaged for a 10-ps time frame to create a predicted spectrum. 
cutoff of 5.5 Å. Back-calculated cross-peak volumes for each 2-ps time 
form the opposing orientation was used. Starting structures were energy mini-
mized by 100 steps of Powell’s conjugate gradient minimization using 
X-PLOR. 
Relaxation matrix refinements were carried out at 300 K with a 
radial dielectric. The minimized structures thus obtained were then 
subjected to 100 steps of minimization using all of the constraints of the 
force field. This was followed by a relaxation matrix simulation and 200 
steps of conjugate gradient minimization. Each trajectory was run for 
200 ps, and each trajectory appeared to reach equilibrium after 100 ps. 
The structures from 140 to 148 ps were used to generate back-calcula-
ted spectra at 2-ps intervals, and these five spectra were averaged to 
compare with the experimental data as discussed below. 
The NOE cross-peak volumes for each of these structures were back-
calculated as described previously (11, 57–59, 63) separately using an 
overall correlation time of 5 ns, a leakage rate of 0.33 s\(^{-1}\), and a distance 
cutoff of 5.5 Å. Back-calculated cross-peak volumes for each 2-ps time 
point were averaged for a 10-ps time frame to create a predicted spectrum. 
Averaged predicted spectra were found to represent the experimental 
data more accurately than any single structure as discussed below. 
There were no missing or any extra peaks in the averaged, back-
calculated 100-ms NOESY spectra of AD or CD in the aromatic to H1' 
or aromatic to H2'/H2' spectral regions. There were no proton-proton or 
proton-phosphorus dihedral restraints that were violated from 140 ps to 
148 ps for any of the trajectories. For the middle 9 residues of each 
strand, the average r.m.s. deviation of the five structures from 140 ps to 
148 ps to the average of those five structures was 0.43 Å for AD \(\alpha\), 0.48 
Å for AD \(\beta\), 0.55 Å for CD \(\beta\) N3, and 0.43 Å for CD \(\beta\) O2. For all 11 pairs, the 
average r.m.s. deviation of the five structures from 140 ps to 148 ps 
to the beginning, canonical B-DNA structure was 2.46 Å for AD \(\beta\), 2.65 
Å for AD \(\alpha\), 3.06 Å for CD \(\beta\) N3, and 2.49 Å for CD \(\beta\) O2. The Q, R, and 
r.m.s. values (64) of the average of the back-calculated volumes for AD, 
which is the average of those predicted \(\alpha\) and \(\beta\) for the aromatic to H1' 
region of the 100-ms NOESY are 0.14, 0.29, and 0.18. The Q, R, and 
r.m.s. of the average back-calculated volumes for the CD \(\beta\) N3 and CD 
\(\beta\) O2 structures for the aromatic to H1' region of the 100-ms NOESY are 
0.15, 0.32, and 0.21. 
Surface accessibilities were modeled using the Connolly surface cal-
culation in Insight II as described previously (11, 57). Probe radii were 
incremented in 0.5 Å steps between 0.5 and 4.0 Å. The accessibilities 
were determined separately for the bases of dG\(_4\), dA\(_5\), dC\(_6\), dA\(_7\), dC\(_8\), 
dG\(_{12}\), dT\(_{16}\), dA\(_{17}\), dT\(_{18}\), and dC\(_{19}\) in all of the structures to allow their 
color coding. These surface areas were normalized to the surface acces-
sibilities of the free bases using the same probe radii (11, 57). 

RESULTS AND DISCUSSION 

The one-dimensional spectra of the imino and exchangeable protons and of the phosphorus nuclei of the two damaged DNA 
duplexes are shown in Fig. 1. The spectra of the two samples are quite distinct in each of these three regions. The spectra of 
the nonexchangeable protons indicate that many protons have differing chemical shifts depending on whether the residue 
opposite the abasic site is a dA or dC. The imino region of the 
two spectra are the most similar of the three regions shown. In 
both cases the net number of imino protons integrates to about 
10, indicating that all of the base pairs are present at 5°. The 
chemical shifts of the imino protons of the AT base pairs are to 
lower field for AD than for CD, and there is slightly better 
dispersion of the GC imino chemical shifts for the AD duplex 
than for the CD duplex. 
The \(31^P\) chemical shifts from the sites adjacent to the damaged 
site are different in AD and CD. The CD duplex has two 
signals that are significantly downfield of the main group, 
whereas the AD duplex has two that are only somewhat downfield 
of the main field. The most downfield \(31^P\) resonance in 
CD is from the phosphorus between residues 6 and 7, and the 
next most downfield one is from the phosphorus between resi-
dues 5 and 6. The two downfield resonances from AD are from 
the phosphorus sites between residues 6 and 7 and between 5 
and 6, but there was not sufficient spectral resolution to deter-
mine which was the one slightly more downfield. 
Thus, the one-dimensional data indicate that both damaged 
DNAs form duplexes, generally in the B-family with all possi-
brable base pairs present. However, the CD sample appears to 
exhibit the larger perturbation from B-form near the damaged 
site. The overall patterns of the chemical shifts are similar in 
both cases, but a sufficient number of differences suggests that
there are significant structural differences resulting from whether dC or dA is opposite the abasic site.

The proton assignments of AD have been presented previously (58), and many of these are indicated in Fig. 2. The same assignment strategy was used for CD and led to the conclusion that there are two β forms present rather than equal amount of α and β. The two β forms will be referred to as β N3 and β O2 because of the basis of their main structural difference. We have previously examined the AD duplex and a similar sequence context with both dA and dG opposite the abasic site as well as an abasic site in a curved DNA context, and essentially equal amounts of α and β had been observed in each of these cases (11, 58, 62).

The assignments of most of the CD resonances could be made using the standard interresidue and intraresidue NOEs supplemented by information from TOCSY and PE/DQ COSY experiments as discussed above. However, no appreciable signals from the damaged site in the α form could be detected from the CD sample at 15 °C. At higher temperatures small amounts of an α form could be detected, but these spectra were not analyzed in detail because under these conditions signals from three forms of CD were present. The results of NOESY experiments allowed clear identification of α and β forms on the basis of the H1'-H2'/H2' NOEs. In the β form the H1' is spatially closer to the H2' than the H1', and the H1'-H2' NOE is larger than the H1'-H2'. In the β form the H1' has a larger scalar coupling to the H2' than the H1'. The converse is the case for the α form. The lack of an α form of CD is consistent with results obtained quite some time ago on a 7-mer duplex with a dC opposite the abasic site, which indicated that essentially only one anomic form was present based on results on a sample with the 1°C labeled with 13C (44).

After the NOE volumes were quantified and the scalar couplings determined, this information was used to determine the structures of the two forms of each damaged DNA via restrained molecular dynamics. The predicted, back-calculated NOEs for each of the forms were then combined and compared with the experimental results as shown in Figs. 2 and 3. It is seen that the agreement between the experimental NOEs and those predicted by the structures of the DNAs is quite good in the regions shown and is equally good in the regions not shown. The main difference between the structure determination methods used here and those used previously is the inclusion of a water molecule at the damaged site during the restrained molecular dynamics. Our earlier results had shown that there is likely to be hydrogen bonding involving a water molecule at abasic sites in duplex DNAs (58), and the inclusion of the water molecule during the trajectories gave rise to structures that were in better agreement with the experimental results than could be obtained without the inclusion of the water molecules into the simulations.

The resulting structures of the AD duplex with the α and β forms of the abasic site are shown in Fig. 4. The overlay of five structures, obtained at 2-ps intervals, is shown at the top. The overlays show that the trajectories are stable over the 10-ps time interval with the r.m.s. deviation for the α structures at 0.48 Å and 0.4 Å for the β structures over this time period. The trajectories appeared to become stable at, or before, about 100 ps. The structures shown at the bottom of Fig. 4 are the average structure over the 10-ps period, and the expansion of the region near the damaged site shows the DNA backbone, including the sugar, in a thin line, the bases in a thicker line, and the water molecule in CPK mode. This representation emphasizes that in the β conformation a water molecule can be well situated for hydrogen bonding to both the abasic site 1'OH as well as to the hydrogen bond donor N3 of dA17. This is essentially what was observed when the structure was determined and the position of the water molecule examined afterward rather than being included in the structure determination (58). It appears that the presence of the water molecule allows the β form of AD to adopt a structure that is quite similar to that of B-form DNA at the damaged site.

The structure of the α form of AD is shown in Fig. 4. The optimum position of the water molecule is not as well suited for hydrogen bonding to the abasic site nor the dA in the position opposite the abasic site as in the β form. The position of the dA17 is much closer to the abasic site in the α form than is the case for the β form, and this is part of the structural differences
between the two forms.

The structural determination of the CD sample followed much the same route. The structures of the two forms were determined independently, the structures were each used to back-calculate results for the two structural forms, and these results were combined and compared with the experimental results as shown in Fig. 3. The variation for the CD case is that both structural forms have the abasic site in the \( \beta \) form. In one of the two structures the water molecule hydrogen bonds to the N3 of dC17, referred to as \( \beta \text{N3} \), and in the other the water molecule hydrogen bonds to the O2 of dC17, referred to as \( \beta \text{O2} \). The overlays of the structures at 2-ps time steps over a 10-ps time interval for both of these structures are shown in Fig. 5. The two structural forms are in slow exchange, on the NMR time scale, as indicated by the presence of distinct cross-peaks, for the two structural forms, which are separated by 10–20 Hz for the dT16 H6 to Me intraresidue cross-peak, as shown in Fig. 6. No cross-peaks that could be attributed to exchange between the two forms were observed in the NOESY, ROESY, or TOCSY data sets.

The overlays of the structures show that the trajectories are stable over the 10-ps time interval with the r.m.s. deviation for the \( \beta \text{N3} \) structures at 0.56 Å and 0.55 Å for the \( \beta \text{O2} \) structures over this time period. As shown at the bottom of Fig. 5 the prime difference between the two structures is the position of dC17 with the water molecule within hydrogen bonding distance of the 1’OH of the abasic site and with either the O2 or N3 of dC17.

Analysis of the original set of CD trajectories indicated that the experimental results are only consistent with the presence of two forms, both with a \( \beta \) abasic site, and that the dC17 residue was the site of conformational heterogeneity. Additional preliminary trajectories carried out without constraints on the position of the water molecule did not converge because the position of the water was found to have transitions between positions that allowed hydrogen bonding to the N3 or O2 of dC17. Because NOE and ROE connectivities between water and residues near the damaged site could be observed, the positions of the water molecules, in the two forms, were constrained as described above. Cross-peaks between water and the two forms of the H2 of dA7, the H2 of dA5, and the amino protons of dC17 were observed in NOESY spectra as well as in the ROESY spectrum shown in Fig. 6. These experimentally observed water-DNA contacts are consistent with the positions of the water molecules found in the trajectories of the \( \beta \text{N3} \) and \( \beta \text{O2} \) forms.

The structures of AD and CD were used as a basis to model the surface accessibilities of the residues of these damaged DNAs. As we have shown previously, the accessible surface area appears to be a good surrogate for “extrahelicity” as well as indicating which portions of the damaged DNAs are avail-
The surface accessibilities of most of the residues are the same in the cases of AD and CD with the exceptions of dT18 and dC19 as shown in Fig. 7. The residue dT18 has a significantly higher surface accessibility in both forms of CD than it does in either form of AD. The residue dC19 has the highest surface accessibility in the βN3 form of CD.

An overall view of the surface accessibilities of the damaged DNAs is given in Fig. 8 for the case of a 1.5 Å probe molecule. The surface accessibilities of the α and β forms of AD are shown, on the left, with the abasic site coded in red, the dA17 opposite the abasic site in blue, the dA5 and dA7 adjacent to the abasic site in yellow, and the dT16 and dT18 coded in green. The positions of major differences in surface accessibility are dA17, which is much more accessible from the major groove in the α form than in the β, and dA5 and dA7, which are more accessible from the minor groove in the α form than in the β. The surface accessibilities of CD are shown on the right side of Fig. 8 with βN3 on the top and βO2 on the bottom. The color coding is the same as for AD with the exception that it is dC17 in blue.

Thus, the structures of the two forms of CD are distinct from those of AD near the damaged site, and the extent of distortion from B-form DNA is more pronounced for CD. The least distorted structure is that of AD with the β form of the abasic site. The β form of the abasic site appears to be able to hydrogen bond to a water molecule that can also hydrogen bond to the N1 of dA17 on the opposing strand. This positioning of this water molecule allows the abasic site and dA17 to adopt positions analogous to those of B-form DNA.

When the abasic site is in the α form this hydrogen bonding arrangement apparently cannot be adopted. When the 1′OH of the abasic site is in the “down” position, associated with the α form, the arrangement of the abasic site, the water molecule, and dA17 is less favorable than in the β form in which the water molecule can effectively bridge the two strands. Thus, the most favorable arrangement that occurs in the AD α case is when the water molecule hydrogen bonds to both the 1′OH hydrogen of the abasic site and the N1 of dA17. In neither of these structural forms is the abasic site itself significantly exposed, consistent with the chemical stability of the abasic site in duplex DNA.
The origin of the structural differences between the CD and AD forms appears to be that dC is simply too small to hydrogen bond to a water molecule that is simultaneously hydrogen bonded to the abasic site while the dC is near the position it would occupy in B-form DNA. Thus, the structure near the abasic site is more distorted when dC is present than dA as evidenced both by the larger accessible surface and the absence of an α form in CD. The structures of the β O2 and β N3 forms have a water molecule hydrogen bonded to the dC residue. In the actual solution state there will be many water molecules interacting with the damaged sites of both AD and CD, and only the position of a single water molecule has been considered here.

The structure of a duplex DNA with a thymine glycol opposite dA in the same sequence context has also been determined (57). This structure is shown along with that of AD at the top of Fig. 9.
Fig. 9. The overall features of the AD and thymine glycol duplex DNAs are relatively similar near the damaged site in the undamaged strand. However, the thymine glycol residue has considerable extrahelicity, and the backbone is more distorted in the damaged strand of the thymine glycol DNA than is the case for AD.

The structure of a duplex DNA with an abasic site opposite dA has been determined in a curved dA tract DNA context (11), and the structures of dA and the dA tract-damaged DNA are shown in Fig. 9. The comparison shows that the structures of the two damaged DNAs are strikingly different. The structures of the α and β forms are considerably different in the two sequence contexts. This comparison shows that the structure of a DNA containing an abasic site is not determined solely by the type of damage but by the sequence context in which the damage appears.

It may have been the case that the structure of duplex DNA containing an abasic site is determined primarily by the “hole” left by the excised base. However, the results presented here indicate that the structure is determined, in part, by the base on the opposing strand, because the structures with dA and dC opposite the abasic site have distinct structural features. Comparison of the structural results on AD and on the abasic site in the curved context shows that the sequence context in which the damage occurs also has a role in determining the structure of damaged DNA. Comparison of the structures of damaged DNAs with the same sequence but with different types of damage, abasic site or thymine glycol, shows that the structure is dependent on the type of damage.

The surface accessibility of all of these structures have been determined. In each case the accessibility of at least some of the residues at and near the damaged site are considerably greater than those of undamaged DNAs. The structures of these damaged DNAs also all have alterations in the width and regularity of both the minor and major grooves as well as the ratio of the widths of the two grooves. None of these forms of damage appear to induce a large change in the radius of the DNAs at or near the damaged site.

The structures of only a small number of the naturally occurring damaged DNA sites have been determined. In addition to the structures discussed above the structures and other properties of duplex DNAs with 8-oxoguanine have been examined (65–68). Although some generalizations may be suggested by this small number of structures it is expected that a somewhat larger set of structures needs to be determined and the biological properties of the set determined. This allows a detailed examination of which structural features of damaged DNAs can be recognized as “generic” damaged DNA and which can be recognized as being associated with a specific type of damaged DNA.

REFERENCES

1. Beckman, K. B., and Ames, B. N. (1997) J. Biol. Chem. 272, 19633–19636
2. Freiberg, E. C., Walker, G. C., and Siede, W. (1995) DNA Repair and Mutagenesis, American Society for Microbiology, Washington, D. C.
3. Freiberg, E. C. (1985) DNA Repair, Freeman, New York.
4. Ward, J. F. (1991) in Physical and Chemical Mechanisms in Molecular Radiation Biology (Glass, W. A., and Varma, M. N., eds) pp. 403–421, Plenum Press, New York.
5. Demple, B., and Harrison, L. (1994) Annu. Rev. Biochem. 63, 915–948
6. Cunningham, R. P. (1997) J. Biol. Chem. 272, 183–189
7. Krokan, H. E., Standal, K., and Slupphaug, G. (1997) Biochem. J. 325, 1–16
8. Cuniase, P., Lazkarra, V. G., Guschlbauer, W., Kaplan, B. E., and Sowers, L. C. (1990) J. Mol. Biol. 213, 303–314
9. Bjørka, E., Buch, L., and Hansen, L. (1987) Biochem. J. 246, 1161–1166
10. Efrati, E., Tocco, G., Eritja, R., Wilson, S. H., and Goodman, M. F. (1997) J. Biol. Chem. 272, 2559–2569
11. Wang, K. Y., Parker, S. A., Goljer, I., and Bolton, P. H. (1997) Biochemistry 36, 11629–11639
