Synthesis, characterization and solution structure of tethered oligonucleotides containing an internal 3′-phosphoglycolate, 5′-phosphate gapped lesion

Hans-Dieter Junker¹, Silvia T. Hoehn¹, Richard C. Bunt¹, Vasilios Marathius¹, Jingyang Chen¹, Christopher J. Turner³ and JoAnne Stubbe¹,²,*

¹Department of Chemistry, ²Department of Biology and ³Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received August 13, 2002; Revised and Accepted October 24, 2002 PDB nos 1G5K, 1GJ1, 1N0O, 1N0K

ABSTRACT

Bleomycins (BLMs) are antitumor antibiotics that in the presence of iron and oxygen mediate DNA damage by 4′-hydrogen atom abstraction of pyrimidines 3′ to guanines. The resulting 4′-deoxyribose radicals can be trapped by O₂ and ultimately result in the formation of base-propanal and gapped DNA with 3′-phosphoglycolate (3′-PG) and 5′-phosphate (5′-P) ends. The role of this lesion in triggering double-strand cleavage of duplex DNA by a single BLM molecule and the mechanism by which this lesion is repaired in vivo remain unsolved problems. The structure of these lesions is an essential step in addressing both of these problems. Duplex DNAs (13mers containing tethered hexaethylene glycol linkers) with GTAC and GGCC cleavage sites have been synthesized in which gaps containing 3′-PG and 5′-P ends at the sites of BLM cleavage have been inserted. The former sequence represents a hot spot for double-strand cleavage, while the latter is a hot spot for single-strand cleavage. Analytical methods to characterize the lesioned products have been developed. These oligonucleotides have been examined using 2D NMR methods and molecular modeling. The studies reveal that the lesioned DNAs are B-form and the 3′-PG and 5′-P are extrahelical. The base opposite the gap and the base pairs adjacent to the gap remain well stacked in the DNA duplex. Titrations of the lesioned GGCC oligomer with HOO-CoBLM leads to a mixture of complexes, in contrast to results of a similar titration with the lesioned GTAC oligomer.

INTRODUCTION

The cytotoxic effects of a variety of antitumor antibiotics such as the bleomycins (BLMs) and enedynes are thought to be associated with their ability to bind to DNA and produce sequence-specific single-strand (ss) and double-strand (ds) breaks in the DNA by damaging the deoxyribose backbone. Even though these natural products differ in the details of their mechanism of action, they share the ability to produce a 4′-deoxyribose radical that can lead to two types of damage (Fig. 1) (1). If the radical is trapped with O₂ the DNA strand is cleaved directly, generating a gapped duplex containing 3′-phosphoglycolate (3′-PG) and 5′-phosphate (5′-P) ends and a base-propanal. Alternatively, under low O₂ tensions, a 4′-ketoabasic site with an intact DNA strand and a free nucleic acid base results (2–4). Our laboratory has focused recently on understanding the mechanism of BLM-mediated ds breaks and the role of the 3′-PG lesion in this process (5–7). Double-strand breaks are believed to be the major cause of cytotoxicity of the BLMs due to the difficulty of their repair (8–10). Our interest in the mechanism of action of these antitumor agents has led us to study the mechanisms of repair of the lesions generated by these natural products. This paper presents the first insight into the structures of duplex oligomeric DNAs containing a GXAC and GGXC sequence (where X is a gapped 3′-PG/5′-P lesion) using 2D NMR methods and molecular modeling. This study is the starting point for understanding the basis of recognition of damaged sugars and the mechanism of their repair.

Base excision repair (BER) is an important pathway in the repair of ss damage to the deoxyribose of DNA caused by BLM in the presence of its required cofactors Fe²⁺ and O₂ (11–13). In humans the apurinic/apyrimidinic endonuclease APE1 (also known as HAP1, Ref-1 or Apex) is the protein in the BER pathway that is responsible for the recognition and repair of abasic sites (14,15). This enzyme is also inductive in response to BLM (16,17) and can repair the 4′-ketoabasic sites and 3′-PG lesions generated by BLM (11,12,18). The efficiency of removal of the 4′-ketoabasic site in one sequence context is similar to abasic sites (13).

The 3′-PG lesion repair by APE1 in vitro is slow and depends on the location of the damage within the duplex, and thus the exact role this enzyme plays in its repair in vivo is not known. 3′-PGs are removed from oligonucleotide substrates with an efficiency that is 70–400-fold lower than removal of abasic sites (12,19). In mammalian cell extracts at least two additional 3′-phosphodiesterase activities have been detected, but not isolated due to their instability (20,21). Furthermore,
recent reports suggest that APE1 is one of a complex of enzymes involved in the BER pathway and in such a complex the kinetics could be very different (22).

The kinetics and mechanism by which activated BLM generates a gapped 3'-PG next to a 5'-P have been investigated in detail (Fig. 1). Activated BLM abstracts the H4' of a pyrimidine 3' to a guanine (recently reviewed in 2,3,23,24). O2 reacts with the resulting C4' radical which is reduced to a 4'-hydroperoxide. This species ultimately leads to strand scission generating a 3'-PG end with a $t_{1/2} = 2.5\pm5$ min at 4°C. Kinetic studies in one system reveal that 3'-PG production precedes base-propenal release and 5'-P formation which occur with a $t_{1/2}$ of 6.7 ± 0.3 min at 4°C (25). If these in vitro experiments are indicative of the reaction in vivo and the rates are sequence independent, then repair enzymes may encounter an intermediate in the strand scission pathway consisting of a 3'-PG adjacent to a fragment of deoxyribose with base still attached, the precursor to the 5'-P end (Fig. 1). However, most DNA damage does not result in ds lesions (ds:ss cleavage ratios are typically 1:10–1:20) (8,26,27) and the average kinetic parameters measured in vitro may not be representative of the lesions that result in ds breaks. Thus the actual lesion(s) that leads to ds cleavage and the lesions recognized and repaired in vivo have not been identified.

APE1 has been crystallized in the presence of an oligonucleotide containing a tetrahydrofuran (THF) analog of an abasic site (28). The binding pocket for the damaged deoxyribose is small and cannot accommodate a base. Consistent with this model, APE1 does not bind to nicked DNA containing a 3'-OH and a 5'-P mononucleotide (12). APE1 does, on the other hand, recognize, in addition to a modified deoxyribose, a space bracketed by duplex DNA (19). Thus structural data suggests that APE1 could recognize a 3'-PG and a gapped 5'.

Double-strand cleavage of DNA by BLM occurs only when a 3'-PG next to a 5'-P is generated initially at the ss cleavage site. Evidence suggests that a single molecule of activated BLM can cause both cleavage events without dissociation from the DNA. Based on our recent NMR studies with the hydroperoxide of cobalt BLM (HOO-CoBLM) bound to a hot spot for ds cleavage, a model involving the importance of the bithiazole tail of BLM in the reorganization of the BLM from one strand to the second has been proposed (7,29). The trigger that causes the BLM to reorganize, however, remains elusive. Since BLM-induced ds cleavage only occurs at 3'-PG ends, it is possible that either the 3'-PG next to the 5'-P, or a precursor thereof, triggers the reorganization of the bithiazole tail, which ultimately leads to ds cleavage.

Our laboratory has long been interested in the mechanism of BLM-mediated ss and ds cleavage of DNA. As a starting point to study ds cleavage catalyzed by BLM and the mechanism of its repair, we have synthesized gapped duplex DNAs with 3'-PG next to a 5'-P (denoted by X) in the sequence contexts d(CCAAAGXACTGGG) (1a) and d(CCAAGGXCTTGGG) (1b). We have previously shown that GTAC is a hot spot for BLM-induced ds cleavage and that GGCC is a hot spot for ss cleavage and not ds cleavage (6). To ensure the stability of the oligomeric duplex, hexaethylene glycol spacers have been used to connect C1 to G26 and G13 to C14 in the double hairpin of 1a and 1b (Fig. 2). In this paper we describe a general synthesis and characterization for PG lesions. The solution structure of 1a and 1b, established by 2D NMR experiments and molecular modeling, provides the first structural models of 3'-PG and 5'-P gapped lesions.
MATERIALS AND METHODS

Solid phase oligonucleotide synthesis

Compounds 5a and 5b (Fig. 3) were synthesized on the 1 μmol scale using deoxyribonucleotide phosphoramidites and reagents from Perkin Elmer and an ABI 392 DNA synthesizer (30,31). The synthesis of 5a is reported in detail and that of 5b followed a similar protocol. The 3-[(4¢,4-dimethoxytrityloxy)-glyceryl-1-succinoyl]-long chain alkylamino controlled pore glass columns (2; Fig. 3), the spacer 9-O-dimethoxytrityl-hexaethylene glycol 1-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite (3; Fig. 3), [3-(4,4¢-dimethoxytrityloxy)-2,2-dicarboxyethyl]propyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite (4) and the dicyanomimidazole solution (0.25 M) were purchased from Glen Research. The following changes from the standard automatic synthesis protocol of ABI were made: (i) synthesis was carried out using a 0.25 M dicyanomimidazole solution in place of the tetrazole activator; (ii) a coupling time of 325 s instead of 25 s was used in the synthesis cycle for 3 (Fig. 3); (iii) the coupling time using 4 was extended to 480 s and the capping procedure in this cycle was omitted.

Deprotection and reverse phase HPLC purification

5a was cleaved from the solid support (2 h, 20°C) and deprotected (8 h, 55°C) with 5 ml of concentrated (33%) aqueous ammonia. The results of 10 (1 μmol scale) syntheses were pooled and concentrated in vacuo. The resulting viscous oil was purified by reverse phase HPLC using a semi-preparative Alltech Econosil C18 column (10 × 250 mm) and a linear gradient of 10–50% acetonitrile in 0.1 M triethylammonium acetate (pH 7.3) over 25 min, followed by a linear gradient of 50–100% acetonitrile over 10 min (flow rate 3 ml/min). The retention time of the product was 20–23 min. Fractions containing this product were pooled, concentrated to a colorless gum, dissolved in 5 ml of 80% acetic acid and incubated at room temperature for 5 h. The acetic acid was then removed in vacuo and the residue was dissolved in 7 ml of concentrated aqueous ammonia and incubated for 90 min at room temperature. The ammonia was evaporated in vacuo to give a white solid, 6a. The residue was taken up in 4 ml of water, filtered and used directly in the next step. The isolated yield of 6a was 2.3 μmol, based on the calculated extinction coefficient (ε260 = 1.9 × 10^5/M/cm). The isolated yield of 6b was 1.3 μmol.

Oxidation of 6 (Fig. 3)

Conversion of 6a to 1a was carried out by a modification of the procedure of Urata and Akagi (32). Sodium periodate (0.2 M, 0.5 ml) was added at 0°C to 6a (2.3 μmol, 4 ml of water) and the mixture stirred for 2.5 h. The reaction was then quenched at 0°C by addition of methionine (0.2 M, 0.72 ml) and stirred for an additional 30 min. The pH was then adjusted to 5 by addition of NaH2PO4 (0.31 M, 1.6 ml) and acetic acid (20–50 μl). To this solution of 7a (2.3 μmol, 6.4 ml of water) was added NaClO2 (0.2 M, 90 μl) at room temperature (33). The oxidation reaction was monitored by ion exchange HPLC using a Dionex DNApac analytical ion exchange column (4 × 250 mm) using a linear gradient of 35–50% B over 30 min (flow rate 1.0 ml/min). Buffer A contained 25 mM sodium
acetate in 90% H₂O/10% acetonitrile (pH 6.0); buffer B contained buffer A and 1.0 M NaCl. Compounds 6a, 7a and 1a eluted with retention times of ~22.5, 23 and ~25 min, respectively. Compounds 6b, 7b and 1b eluted with retention times of 24.1, 24.7 and 26.9 min. The reaction was quenched after ~95% conversion (~3 h) with Na₂SO₃ (0.2 M, 90 µl) and the pH was adjusted to ~6.5 with 0.2 M Na₂HPO₄. The crude product (~75% pure) was dialyzed ( Spectrum dialysis tubing, mol. wt cut-off 1000 g/mol) against 2 l of 5 mM sodium phosphate (pH 6.9) buffer for 36 h at 4°C.

**Purification of 1a**

Compound 1a was purified by HPLC using a semi-preparative Dionex Nucleopac PA 100 column (9.0 × 250 mm). The same gradient as described above was applied (flow rate 4.0 ml/min). The product was dialyzed at 4°C against 2 × 2 l of 5 mM sodium phosphate (pH 6.9) buffer for 48 h. The isolated yield of 1a was 54% (1.2 µmol). The isolated yield of 1b, prepared by an identical procedure, was 0.72 µmol, 54%.

**Electrospray mass spectroscopy (ES-MS)**

Mass spectra were obtained in the MIT Biopolymer Laboratory using an ABI 140c high performance liquid chromatograph (HPLC) interfaced with a PE Sciex API365 mass spectrometer. The electrospray source was adjusted to a spray voltage of ~3.5 kV and a capillary temperature of 125±200°C. The product was desalted using a microdialysis chamber and 2.5 mM NaH₂PO₄ (pH 6.9) for 24 h. The solution was concentrated in vacuo and lyophilized several times from D₂O (99.9%; Cambridge Isotope Laboratories). Compound 1a was dissolved in 0.6 ml of D₂O (99.999%; Isotec) and annealed at 65°C, giving ~1.9 mM duplex 1a in 40 mM phosphate buffer (pH 6.8). Compound 1b was 1.4 mM.

**NMR experiments**

All NMR experiments were performed on custom-build 500, 600 or 750 MHz NMR spectrometers at the Francis Bitter Magnet Laboratory. The acquired data were transferred to a Silicon Graphics workstation and processed using Felix software, version 95 (Accelrys Inc.). ¹H and ¹³C chemical shifts were referenced to an internal standard, sodium 3-(trimethylsilyl)-1-propanesulfonate (TSP) at 0 p.p.m.

**PECOSY (flip angle mixing pulse of 36°), TOCSY (30, 60, 100, 150 and 250 ms mixing times) and NOESY (100, 200, 250 and 400 ms mixing times) experiments** were recorded at 750 MHz at 20°C in D₂O or at 4°C in 90% H₂O, 10% D₂O. Data sets of 4096 × 512 complex points were acquired with spectral widths of 8000 Hz in both dimensions and 32 scans per t₁ increment. During the relaxation delay period, a presaturation pulse of 2 s was used for solvent suppression. For the NOESY experiments in 90% H₂O/10% D₂O, a Watergate gradient pulse sequence (34) was used for water suppression. For Watergate experiments data sets with 4096 × 512 complex points were acquired with spectral widths of 15 000 Hz in both dimensions. For all experiments, spectra were zero filled to 4096 points in the t₁ dimension. The data were processed with a combination of exponential and Gaussian weighting functions. Baselines were corrected with a polynomial or an automatic baseline correction routine in t₂ when necessary.

³¹P GE-HSQC experiments (35) were carried out at 20°C and at 500 or 600 MHz. Data sets contained 1024 × 256 complex points with a spectral width of 6000 Hz in the proton channel and 600, 2000 and 5000 Hz in the ³¹P channel.
Experiments were conducted with a typical relaxation delay of 5 s and with d1 delay times of 10 and 20 ms. Sixteen to thirty-two scans for every t1 increment were used.

1H-13C GE-HSQC experiments (36) were acquired at 600 MHz at 20°C with 4096 × 512 complex points and a spectral width of 5000 Hz in the 1H dimension and 30 000 Hz in the 13C dimension. Transients (128) were recorded for each t1 increment. Long range 1H-13C couplings were detected by choosing dephasing–phasing delay times (τ) of 30 and 40 ms (37). BASHD-TOCSY (38) experiments were recorded with a mixing time of 80 ms and 128 transients. Data sets were collected at 4096 × 64 complex points with spectral widths of 100, 200 and 400 Hz for scaling factors of 3, 2 and 1 in the second dimension and 6000 Hz in the first dimension. All 31P spectra were referenced indirectly through the gyromagnetic ratio to trimethyl phosphate by external calibration on TSP (39).

Molecular modeling

Distance constraints. Distance constraints were derived from a 200 ms NOESY experiment at 750 MHz. Peak volumes were assigned by visual inspection using H2'-H2'' and/or cytosine H5-H6 NOEs as a guide for strong NOEs. NOEs were classified as strong, medium or weak with distances of 1.7–4.0, 2.0–5.0 and 3.0–6.0 Å, respectively. NOE volumes were also calculated with the peak pick protocol in Felix 95, 1.7±4.0, 2.0±5.0 and 3.0±6.0 Å, respectively. NOE volumes were classified as strong, medium or weak with distances of 1.7–4.0, 2.0–5.0 and 3.0–6.0 Å, respectively. During the molecular dynamics calculations, planarity constraints were included for the purine bases at 50 kcal mol⁻¹ Å⁻². The NOE force constant was 50 kcal mol⁻¹ Å⁻², while the dihedral angle force constant was 40 kcal mol⁻¹ rad⁻². These force constants were chosen to minimize NOE distance constraint and dihedral angle constraint violations. Non-bonded interactions had a cut-off of 11.5 Å. The non-bonded interaction was switched from on to off between 9.5 and 10.5 Å. During the molecular dynamics calculations, atomic coordinates were written to a trajectory file every 0.2 ps throughout the 100 ps molecular dynamics trajectory. The structure was averaged over the last 10 ps of the run, and the averaged structure was again subjected to 200 steps of conjugate gradient minimization. Ten averaged structures were calculated. Coordinates of the 10 averaged structures of 1a and 1b have been deposited in PDB. The coordinates for 1a are in PDB 1GSK and 1GJ1 (10 structures and one averaged minimized structure, respectively). The coordinates for 1b are deposited in PDB 1N0O and 1NOK (10 structures and one averaged minimized structure, respectively).

Back-calculations of NMR spectra. Back-calculations were performed in InsightII using the iterative relaxation matrix (IRMA) approach (43,44). The entire 200 ms NOESY spectrum at 750 MHz was back-calculated for the average of 10 structures of 1a where each structure resulted from a separate 100 ps molecular dynamics run. The spectra were calculated using the full relaxation matrix. Parameters for the IRMA calculations were set as follows: the rotational correlation time (τc) was 3 ns and T1 leakage was 10 s⁻¹.

RESULTS

Choice of the sequences

The sequence of the duplex 1a containing a PG lesion is shown in Figure 2A. Design considerations of 1a encompassed having a full turn of a DNA helix and a single BLM binding site (5'-GTAC-3'), located opposite the lesion. This context was chosen because we have previously demonstrated that 5'-GTAC-3' is a 'hot spot' for the generation of ds DNA breaks by a single BLM (6). This lesion in the presence of a metallo-BLM has allowed us to test our model for BLM reorganization during the ds cleavage process (7). Our initial
Solid support DNA synthesis

A prerequisite for structural NMR studies of PG containing lesions is the development of a reliable synthetic strategy for their large-scale synthesis using phosphoramidite chemistry. The design of the oligonucleotide containing the lesion requires the introduction of three modifications into the standard solid phase DNA synthesis protocol to synthesize the lesion precursor 5 (Fig. 3). Previous studies have developed a modified controlled pore glass solid support, allowing the incorporation of a glycerol unit at the 3' end of the oligonucleotide (Fig. 3). As noted above, incorporation of hexaethylene glycol phosphoramidite building blocks (Fig. 3, 3) allows the crosslinking of both DNA strand ends. The final step of the solid phase synthesis is the attachment of a 5'-phosphate precursor using 4 (Fig. 3) that leads to 5 (47). The length of the sequence and the incorporation of the three different modifications made the solid phase synthesis challenging. Initial attempts to synthesize 5a by an unmodified Applied Biosystems programmed DNA synthesis cycle resulted in the occurrence of failure sequences and a low overall yield. In particular, the coupling efficiency of the linkers varied from 80 to 95%. Therefore, a longer coupling time for the spacer was chosen, which improved the yield to 90–96%. Decreased coupling yields (<95%) were also observed for the last steps of the synthesis. Initially, longer reaction times (2–10×) and excess reagents (2×) were tried to enhance coupling efficiency. Evaluation of the crude deprotected oligonucleotides by ion exchange HPLC showed ultimately that increased amounts of side products resulted from acid-catalyzed depurination by the tetrazole activator. A significant improvement in the purity of the product was thus achieved by changing tetrazole to dicyanoimidazole. The advantages of this reagent are its reduced acidity (pKₐ 5.2) and its increased reactivity in comparison to tetrazole.

In the final step of the solid phase synthesis, the chemical phosphorylation reagent (4) was coupled to the oligonucleotide by omitting the capping steps in the programmed synthesis cycle and by using a coupling time of 6 min. The lesion precursor was then cleaved from the column and deprotected by standard methods (Fig. 3, step a). The presence of a dimethoxytrityl (DMT) group at the 5'-P end of the lesion allowed an efficient separation of the lesion precursor from failure sequences by reverse phase HPLC. In a final step, the 5'-P was deprotected by removal of the DMT group with acetic acid and subsequent ammonolysis (Fig. 3, step b). The isolated yield of 6a was ~23%.

Oxidation of 6

Compound 6 was converted to 1a by a two-step oxidation procedure (Fig. 3, steps c and d) (32,33). First, the terminal diol of the 3' glyceraldehyde was cleaved with sodium periodate to generate the 3'-glycol aldehyde 7 (>95% conversion). Second, 7 was oxidized to the acid using sodium chlorite. The different
retention times of 6, 7 and 1 on chromatography using an ion exchange HPLC system made it possible to monitor the course of the reaction (Supplementary Material, Fig. S1). The crude oxidation product (~75% purity by HPLC) was purified further by semi-preparative ion exchange HPLC to yield ~1.2 μmol (54%) of 1a. The purity of 1a was >95%, based on ion exchange and reverse phase HPLC chromatograms. The identity of the oligonucleotide was confirmed by ES-MS. Compound 1b was generated in similar overall yield and characterized by both HPLC and ES-MS.

Quantitation of the PG by enzymatic digestion

To characterize 1a further, an enzymatic digestion scheme was developed. This procedure allowed the detection and quantitation of the dG-3'-PG. 1a was digested with two enzymes: DNase and SVPDE. DNase cuts DNA randomly into small pieces with an average size of 3 nt. The resulting cleavage products are phosphorylated at the 5' end and contain free 3'-OH groups. SVPDE further digests these oligonucleotides into 5'-nucleotides. Nucleotides with a 5' modification like the PG or the hexaethylene glycol are not substrates for this enzyme under the chosen reaction conditions. The 3' modified nucleotides were separated from the 5'-nucleotides by ion exchange HPLC. Two peaks (A and B) in the chromatogram had significantly longer retention times than the 5'-nucleotides. Peak A (retention time 5.5 min) was desalted by reverse phase HPLC and analyzed by ES-MS. The molecular mass of 691.8 g/mol is in agreement with the chemical structure of 5'-phosphateoxynucleosine-3'-phospho-hexakis(ethylene glycol). Peak B (retention time 9.2 min) correlated with the synthesized reference compound, dG-3'-PG, and had an identical UV/vis spectrum.

To quantitate the amount of dG-3'-PG relative to the other nucleotides, the entire reaction mixture subsequent to treatment with DNase and SVPDE was incubated with alkaline phosphatase. This enzyme hydrolyzes the 5'-monophosphates, leaving nucleotides that were then characterized by C18 reverse phase ion pairing HPLC. The resulting chromatogram (Supplementary Material, Fig. S2) showed baseline resolution of the deoxynucleotides (dC, dG, T and dA), the dG-3'-PG and dG-3'-PG (ethylene glycol). The amounts of the nucleotides were compared to a reference mixture containing their expected ratios based on the sequence. The amount of all nucleotides was in agreement with the predicted structure; the dG-3'-PG accounted for 75% of the expected value.

Assignment of the proton chemical shifts

The base and sugar protons of 1a (and 1b) were assigned by standard 2D NMR methods (48). The sugar spin systems were identified by through bond correlations, measured by TOCSY experiments with mixing times of 60 and 90 ms. The 5' and 6' hydrogens were assigned based on TOCSY (150 ms mixing time) and NOESY (250 ms mixing time) experiments. Table S1 (Supplementary Material) contains the chemical shifts of most of the protons of 1a. Table S2 (Supplementary Material) contains the chemical shifts of most of the protons of 1b.

The 250 ms NOESY experiment at 20°C (750 MHz). (A) Solid lines indicate the NOE connectivities from C1 to G6 (black) and from A8 to G13 (red). The following NOE cross-peaks are labeled by an arrow: A, A5 H1' to G6 H8; B, A8 H1' to C9 H6; C, A8 H8 to C9 H5. (B) Solid lines indicate the NOE connectivities from A17 to T24. Red lines indicate the NOE connectivities from the residues T19 to C21 that are located opposite the lesion. The following NOE cross-peaks are labeled: A, T19 H1' to A20 H8; B, A20 H1' to C21 H6; C, A20 H8 to C21 H5; D, A5 H2 to T23 H1'.

Sequential NOE connectivities can also be traced in the lesion containing strand from A1 to G6 and from A8 to G13 (Fig. 5A). In 1a it was not possible to detect any NOE interactions between G6 and A8, suggesting that there is no collapse of the gap between these two bases. The NOESY data allow a description of the structural features of the nucleotides flanking the lesion as well. Sequential NOEs between the sugar hydrogens (H1', H2' and H2") of A5 and H8 of G6 indicate that the 3' PG modification of G6 does not significantly alter the stacking of G6 with A5. For example, a strong NOE is detected between A5 H1' and G6 H8 (Fig. 5A, arrow A). The only structural perturbation of B-form DNA on this side of the lesion is apparent from an unusually strong interstrand NOE between A5 H2 and T23 H1' (Fig. 5B, arrow D).

The influence of the lesion is also apparent at the 5'-P side. A8 H8 is significantly downfield shifted (~0.3 p.p.m.), outside...
the envelope of the base protons (Fig. 5A). Despite this perturbation, NOEs between A8 H1’ and C9 H6 (Fig. 5A, arrow B) and between A8 H8 and C9 H5 (Fig. 5A, arrow C) indicate stacking interactions between these residues.

NOE connectivities can also be traced along the intact strand from C14 to G26. Figure 5B shows a portion of the base to H1’ connectivities from A17 to G25. The NOESY data indicate that there are also minimal structural perturbations of the residues in the intact strand opposite to the lesion. There are no interruptions of the sequential NOE connectivities from G18 to C21 (Fig. 5B). NOE cross-peaks between A20 H1’ and C21 H6 (Fig. 5B, arrow B) and between A20 H8 and C21 H5 (Fig. 5B, arrow C) are indicative of stacking interactions between the adenine A20 and the cytosine C21. The only apparent perturbation is the decreased intensity of T19 H1’ and A20 H8 (Fig. 5B, arrow A).

Assignment of the exchangable protons (49)

The Watergate NOESY data at 5°C show 12 distinct signals in the imino region from 12–14 p.p.m. (Fig. 4B) and 10 of these 12 protons show NOE interactions allowing chemical shift assignments. The imino protons belonging to base pairs G6 C21 and A8 T19 broadened in the 1D spectra (Fig. 4B) and are not detectable in the 2D spectra. The broad signal at 12.8 p.p.m. has been assigned to the G6 imino proton based on our ability to assign hydrogen bonding of the amino protons of C21 within this base pair. The two amino protons (7.86 and 6.99, Supplementary Material, Table S1) have been assigned based on their NOEs to their own C21 H5. However, NOEs between these protons are only one third of the intensity of the NOEs between the amino groups of the other cytosines, indicative of exchange. Furthermore, the chemical shift of the hydrogen bonded proton is upfield shifted 0.5–0.7 p.p.m. relative to the unperturbed base pairs, also indicative of enhanced exchange. Increasing the temperature of the sample resulted in duplex melting from the lesion towards the linkers. At 25°C the imino protons of the base pairs G18 C9 and A5 T23 showed significantly reduced peak intensities in comparison to the signals of the other imino protons that are closer to the linker.

**1H-31P NMR GE-HSQC measurements**

31P chemical shifts of the DNA phosphodiesters can be a sensitive indicator of an altered DNA conformation (50,51). 1H-31P GE-HSQC experiments were therefore undertaken and revealed a narrow dispersion of signals (~2.4 to ~3.3 p.p.m.) including that associated with the 3′-PG, ruling out significant structural backbone perturbations of the intact strand and of the 3′-PG. Altered 31P chemical shifts are, however, associated with the linker connections at the 5′-ends of both strands of 1a and the 5′-P end at A8 (45). The chemical shifts of the phosphates associated with the linkers are downfield shifted by ~1 p.p.m. relative to generic phosphodiesters, in agreement with the literature (45). The 5′-P end at A8 of 1a is observed at ~0.3 p.p.m., typical for monophosphates (25). This assignment is based on a cross-peak of the H4’ of A8 and this P. Detection of this 5′-P, due to its altered relaxation properties, required a change in the delay time (t_a) from 10 to 20 ms in the GE-HSQC experiment (Supplementary Material, Fig. S3).

**Assignment of the PG protons**

Assignment of the methylene protons of the PG lesion is required to detect interactions with its environment. A 13C-HSQC experiment was used to make a preliminary assignment of the PG protons and their carbon chemical shift to 4.23 and 67.5 p.p.m., respectively. The low spectral dispersion of the H5′/H5″ deoxyribose protons prevented unambiguous assignment of PG protons and therefore a long range GE-HSQC experiment was used as further confirmation. This experiment involves polarization transfer through three bonds from the methylene protons of the PG to its carboxylate carbon. A cross-peak was observed between protons at 4.21 p.p.m. and the phosphoglycolate at 178.9 p.p.m. (Fig. 6A). The proton chemical shift agrees with the assignment made using the HSQC experiment and published values for model compounds containing a 3′-PG (52,53). For a comparison of signal intensities the coupling between the methyl groups of the thymine and the C6, C5 and C4 carbons of the pyrimidine ring are also shown (Fig. 6B). The assignment of the
ments support the results from the 31P chemical shifts that (Supplementary Material, Table S4). Thus these measures
phosphates, giving rise to $e$ (measurements allowed the calculation of 17 dihedral angles 
ration was determined by BASHD-TOCSY experiments. These 
conformation. Finally, the phosphodiester backbone confor-
methylene protons of the PG allowed careful examination of 
the NOESY spectra to look for interactions with other residues 
of the DNA. Unfortunately, all of the NOEs could be assigned 
to interactions between 5'/5'' protons and other hydrogens. 
The lack of detectable interactions indicates flexibility in its 
conformation.

Sugar and backbone conformations
The PECOSY data allowed the determination of most (41/52) 
of the H1' to H2'/H2'' coupling constants of the deoxyribose 
of 1a (Supplementary Material, Table S3). The values for $J_{1',2'}$ 
were between 7.8 and 9.8 Hz, indicative of an S-type sugar 
conformation and the 2' endo sugar pucker found in B-form 
DNA. The coupling constants of the nucleotides A8 ($J_{1',2'} = 
7.8$ Hz and $J_{1',2''} = 5.9$ Hz) and G6 ($J_{1',2'} = 9.8$ Hz and $J_{1',2''} = 
5.8$ Hz) flanking the lesion are also consistent with a B-like 
conformation. Finally, the phosphodiester backbone confor-
modation was determined by BASHD-TOCSY experiments. These 
measurements allowed the calculation of 17 dihedral angles ($e$) 
between the 3'-hydrogens and their corresponding phosphates, 
giving rise to $e$ values between 150° and 172° (Supplementary 
Material, Table S4). Thus these measurements support the results from the 31P chemical shifts that there are no significant structural perturbations of the DNA backbone. The only exception is the dihedral angle associated with the P linked to the glycolate in which the value of $e$ is 142°.

Molecular dynamics runs
The molecular modeling used 548 experimentally determined 
distance constraints and dihedral angle constraints for the 
backbone and the sugar conformations. Since the presence of 
the hexaethylene glycol spacers allowed assignment of all 12 
imino protons, hydrogen bonds between base pairs were 
constrained for all residues, including the ones at the 
oligonucleotide ends. Ten separate molecular dynamics runs 
starting from B-form DNA were performed to give 10 separate 
structures. The overlaid structures resulting from these runs 
(Fig. 7) show good agreement. The nucleotide opposite to and 
the nucleotides adjacent to the PG lesion exhibit a regular 
B-form conformation. The PG is extrahelical in all 10 
structures. The 3'-PG and the 5'-P at A8 are disordered since 
there are no distance constraints limiting their position. Their 
flexibility is shown in Figure 8. The distance between the 5'-P 
and the 3'-PG phosphate is 8 Å, in comparison with a value of 
6.5 Å for B-form DNA. Electrostatic repulsion between the 
negatively charged PG and the 5'-P may be the cause of this 
increased distance. Table 1 summarizes the results of the 
molecular modeling.

Back-calculating structures
The entire 200 ms NOESY spectrum for 1a was back-
calculated using the IRMA protocol in InsightII. An ensemble 
of 10 structures was back-calculated (Supplementary 
Material, Fig. S4). All of the experimentally observed peaks 
are observed in the back-calculated spectrum with intensities 
similar to those observed experimentally.

Titration of 1b with HOO-CoBLM
The structure of 1b was obtained from 762 NOE-derived 
distance restraints and 47 dihedral angle restraints [15 from 
BASHD-TOCSY experiments (H3'-31P) and 22 from PE-
COSY] using the protocols described for 1a. The resulting 
structural model is very similar to 1a (Supplementary 
Material, Fig. S5).

DISCUSSION
The long range goals of our structural studies on 1a and 1b are 
to understand the mechanism of BLM-mediated ds cleavage 
and to understand how both ss and ds DNA lesions containing

Table 1. Summary of molecular modeling of 1a

<table>
<thead>
<tr>
<th>Metric</th>
<th>10 structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>r.m.s. deviation to average structure for all atoms (Å)</td>
<td>1.00</td>
</tr>
<tr>
<td>r.m.s. deviation to average structure for the PG (Å)</td>
<td>1.69</td>
</tr>
<tr>
<td>Average NOE violation (Å)</td>
<td>0.037</td>
</tr>
<tr>
<td>Average dihedral angle violation (°)</td>
<td>0.986</td>
</tr>
<tr>
<td>Average number of NOE violations &gt;0.5 Å</td>
<td>0</td>
</tr>
<tr>
<td>Average number of dihedral angle violations &gt;5°</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 8. Conformational disorder in the 3'-PG and the 5'-P ends of 1a.
Four out of 10 structures from Figure 6 are shown. Each structure is shown 
in a different color. The average of 10 structures is shown for the strand 
opposite the lesion (T19–C21, in black).
PG moieties can be recognized and repaired. Since the GTAC sequence is a hot spot for ds cleavage (6) and the GGCC sequence is a hot spot for ss cleavage with little ds cleavage, a 3'-PG lesion next to a 5'-P was studied in each sequence context. It was hoped that a comparison of the structural models for these two lesions might provide insight into both processes.

Information from 2D NMR spectroscopy has been used to provide the first structural models of gapped lesions containing a 3'-PG and a 5'-P end (Fig. 7 and Supplementary Material, Fig. S5). All of the modeled structures for 1a and 1b show the PG as extrahelical and the 5'-P disordered with the average distance between two phosphates adjacent to the gap greater than in B-form DNA. Furthermore, the 1D NMR of the oligomers (Fig. 4) in water indicate that the imino protons adjacent to the gap exchange more rapidly than the remaining imino protons in the duplexes, suggesting flexibility. Both duplex structures are B-form with the base opposite the gap and bases adjacent to the gap well stacked.

The similarities in structure between 1a and 1b by themselves provide little insight into why the intact duplex of 1a provides an excellent ds cleavage site for BLM while the duplex of 1b does not. Our previous studies have shown that titration of 1a with HOOCoblM led to a single complex whose structure was readily determined (29). A similar titration of 1b with HOOCoblM was carried out in this study in an effort to understand the basis for the difference in behavior of these two oligomers with respect to ds cleavage. The results with 1b were not easily interpretable. Multiple binding modes were apparent precluding detailed chemical shift assignments (data not shown). The results of the 1b titrations are similar to those observed by titration of lesioned GXAC, where X is an abasic site (instead of the 3'-PG/5'-P of 1a) with HOOCoblM. The abasic site lesion does not lead to ds cleavage and the titration with HOOCoblM also leads to multiple complexes (29). Our structural model of 1a complexed with HOOCoblM (29) suggests that one excellent ss cleavage site appropriately spaced on each strand of duplex DNA is essential for efficient ds cleavage. However, both the nature of the sequence context and the nature of the lesion appear to be important in BLM-mediated ds cleavage in ways not yet understood.

Despite the prevalence of gapped DNA duplexes generated by numerous DNA damaging processes little information is available about their structures and the basis of their recognition by DNA repair enzymes. Recent NMR studies and molecular dynamics simulations have focused on gapped duplexes with 3' and 5' OH ends opposite guanine and adenine (54,55). In both cases the NMR data revealed a single major species in which the DNA was B-like and the A and G opposite the gap well stacked in the helix. The results from molecular dynamics simulations indicated in both cases several families of structures, with straight and bent DNA conformations. These studies indicate, as do ours on 1a and 1b, that duplex DNA is, on average, unperturbed from B-form, but flexible in the region of the gap.

Recent structures of two types of endonucleases, APE1 and Endo IV, in complex with the same THF (an analog of an abasic site)-containing oligonucleotide support the importance of flexibility in recognition of substrate and provide insight into additional factors that might influence the recognition process (14,28). In both structures the deoxyribose analog is flipped out of the helix and into a specific binding pocket on the enzyme (14,28,56). In both cases the DNA is bent (35° and 90°, respectively) away from the lesion and in the case of APE1 a comparison between the Cβ backbone of free and oligonucleotide-bound structure reveal a r.m.s.d. of ~0.7 Å. These data support the protein’s ability to recognize flexibility in the substrate and suggest the importance of a protrusion from the duplex phosphodiester backbone in the recognition process. In both structures the size of the pocket that accommodates the lesion is small.

While there are no structures of oligonucleotides containing gapped 3'-PG, 5'-P lesions bound to repair enzymes, both APE1 and Endo IV can repair this lesion. Thus, as suggested by the THF-APE1 structure, both flexibility and the 3'-PG projection from the helix in 1a and 1b may be important in the recognition of this type of lesion by these repair enzymes. The PG lesion is also small enough to fit into the enzyme’s binding pocket (14), as can 3'-phosphate and 3'-phosphoglycoaldehyde lesions also repaired by these enzymes (15,57). Thus the charge of the PG does not appear to be crucial for recognition by APE1. Nevertheless, in vitro the recognition and repair of abasic sites appears to be more efficient than the recognition of PG lesions (11,12). It is unclear at this point what elements contribute to the lower repair efficiency and if this decreased efficiency is also observed in vivo. The detailed mechanism of repair of different lesions and which step(s) is rate limiting remain to be established.

In addition to substrate flexibility and extrahelicity, APE1 appears to recognize a space bracketed by duplex DNA, suggested by its ability to bind tightly to DNA containing a single-nucleotide gap with a 3'-OH end and a 5'-P end, but not to nicked DNA with the same ends (12). This putative space bracketed by duplex DNA is seen in our structural model for the PG lesion and is thus consistent with this indicated requirement for recognition (Fig. 6).

Many of the studies on the repair of PG lesions have been carried out on oligonucleotide substrates. However, PG lesions have also been ligated into specific vectors to test ss and ds break repair in cell extracts and intact cells (58–60). When APE1 is neutralized by polyclonal antibodies, repair of ionizing-radiation-induced ss damage such as 3'-PG ends still occurs in partially purified HeLa cell nuclear extracts (60). Thus under these conditions other phosphodiesterases are responsible for the repair of these lesions. Nevertheless, when the amount of APE1 is reduced in vitro by APE1 antisense RNA expression, hypersensitivity of the cells results with respect to oxidizing agents that can generate PG lesions (61,62). Furthermore, recent studies with cells derived from germ cell tumors with overexpressed APE1 suggest that the cells are more resistant to BLM, providing further support that this enzyme plays an important role in repairing the damage created by the drug (17). As indicated in Figure 1, however, the 4'-ketobasic site similar to the abasic site is also produced by BLM and might be the predominant lesion repaired by APE1.

The details of PG lesion repair in vivo remain to be elucidated, and the importance of understanding the structure of this lesion is evident. The ds DNA breaks induced by BLM all contain PG lesions, and PG lesions at ds DNA breaks are poor substrates for APE1 (12), the only human
phosphodiesterase capable of removing this lesion characterized to date (11–13,18,20,21). The ability of this protein to interact with other proteins in the BER repairosome could, however, alter the rate of repair in ways that remain to be determined. The cytotoxicity of BLM is thought to be related to its ability to cause ds DNA breaks (63). Thus understanding the structure of a PG lesion is an important step toward understanding the BLM-induced ds cleavage mechanism (29) and the mechanism of repair of the lesions generated by the drug.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM34454 to J.S. The NMR facility is supported by NIH grant RR-00995.

REFERENCES


