Structure of a Duplex DNA Containing a Thymine Glycol Residue in Solution*

(Received for publication, October 18, 1996, and in revised form, January 19, 1997)

Hsiang Chuan Kung and Philip H. Bolton‡

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

Oxidative stress, ionizing radiation, and other events can induce the oxidation of the thymine in DNA to thymine glycol. The presence of thymine glycol can have significant biological consequences, and there are specific repair enzymes for thymine glycol in a wide range of organisms. The structure of a duplex DNA containing a single thymine glycol (5,6-dihydroxy-5,6-dihydrothymidine) has been determined by the combined use of NMR and restrained molecular dynamics. The duplex of d(CGCGAATGCGCG) paired with d(G25C22G20, C19T18A17G16) with Tg indicating thymine glycol, has been used for these studies. The structure shows that the thymine glycol induces a significant, localized structural change with the thymine glycol largely extrahelical. This structural information is consistent with the biological consequences of thymine glycol in DNA. This structure is compared with that of a DNA duplex with an abasic site in the same sequence context.

Damage to DNA can occur by the spontaneous deamination of cytosine to uracil and through the action of alkylating agents, oxidants, drugs, and toxins, ionizing radiation, and other modes of action (1–3). The exposure of DNA in cells bound to proteins as a solid or free in solution to ionizing radiation or oxidative stress can lead to the conversion of thymine to thymine glycol (5,6-dihydroxy-5,6-dihydrothymidine). Approximately 10–20% of the damage to DNA induced by ionizing radiation, including that used to treat tumors, is the result of thymine base oxidation and fragmentation. These products can also be produced by oxidative stress. In addition, the importance of damaged DNA as a control on the cell cycle is increasingly becoming a focus of research efforts so that the effects of damaged thymines on the structures, stability, dynamics, and interactions of DNA are of interest.

It has been known for several decades that ionizing radiation can stop the reproduction of cells as well as kill cells. These activities form the basis for using radiation in chemotherapy, and the research in this area has been frequently reviewed (1–13). Since the mid-1950s there have been studies on the effects of ionizing radiation on the chemical integrity of DNA (4, 5, 7, 9, 14). Ionizing radiation can induce a number of types of damage to DNA including single and double strand breaks, oxidation of the purine and pyrimidine bases, and cross-linking of DNA with proteins. Radiation-generated oxidants are thought to react with thymine to lead to the formation of thymine glycol, thymine peroxide, thymine hydroperoxide, and other oxidized forms of the base. Some of these oxidized forms of thymine subsequently react further to form urea. Some of the same types of damage also occur to DNA during oxidative stress (5, 15, 16).

The damage to the thymine bases in DNA is of special interest because the major forms of damaged thymine are known; thymine is the most easily oxidized base, and many of the biological consequences of damaged thymines are known. Thus, the damaged thymines offer the opportunity to correlate the changes in the physical properties of DNA caused by ionizing radiation or oxidative stress with the biological effects.

Thymine glycol in DNA can be excised in vitro by Escherichia coli endonuclease III or other enzymes, which liberates the thymine glycol and subsequently carries out a β-elimination reaction to cleave the 3’ phosphodiester as was first shown by Demple and Linn (17, 18). We have shown that endo III cleaves the 3’ phosphodiester of abasic sites via a syn β-elimination reaction (17–20). A mammalian homolog of endo III has recently been found (21). dX174 containing thymine glycol are inactivated in E. coli hosts deficient in endo III much more so than in wild type hosts, indicating that endo III is most likely involved in thymine glycol repair in vivo (22). dX174 containing thymine glycol are also inactivated in E. coli hosts deficient in exonuclease III and endo IV much more so than in wild type hosts, indicating that class II apurinic endonuclease activity is needed for thymine glycol repair (22). Thymine glycols can apparently also be repaired through the ultraviolet induced SOS repair mechanism, which is more commonly associated with thymine dimer repair. Thus, there may be as many as three routes to thymine glycol repair: endo III; exonuclease III and endo IV; and SOS (13). These results indicate that thymine glycol sites can be repaired in vivo in organisms ranging from E. coli to mammals.

The presence of thymine glycol in DNA can have profound consequences on DNA replication (9, 13, 23–27). The presence of thymine glycol is a block to replication (23–28). Studies of the effects of thymine glycol on the replication of M13 in vitro using polymerase I, E. coli DNA polymerase I, and T4 polymerases indicate that thymine glycol stops replication either one residue before or at the site of damage. Examination of the replication of M13, prepared with a single thymine glycol, indicated that thymine glycol is a weak mutagen about 0.3% of the time and primarily acts as a replication block (23).

The presence of thymine glycol may also induce structural changes at the initiation and termination points leading to alterations in the relative amounts of the termination bands of DNA synthesis (22). These effects of thymine glycol may be due to...
to the interactions of the double stranded DNA containing the damaged site with the polymerase.

The presence of a pyrimidine to the damaged thymine seems to enhance the probability of reading through a thymine glycol site in replication more so than a purine, indicating a sequence-dependent effect (22). The base to a thymine glycol is apparently also important in determining the extent of the block to replication (28). There is also evidence that dA is the residue placed opposite thymine glycol by DNA polymerase (9, 13, 23).

Taken together these results indicate that thymine glycol sites can be repaired in vivo as well as in vitro and that if there is no repair, replication can be blocked and thymine glycol is mildly mutagenic. These molecular biology studies also suggest that the sequence context may be important and that the presence of thymine glycol might have structural consequences. Because thymine glycol can not be planar, unlike the normal DNA bases, structural consequences are not unexpected. It is likely that the presence of thymine glycol has effects on transcription, regulation, and DNA packaging as well.

The studies described below on the physical properties of DNA containing thymine glycol are aimed at determining how the changes in DNA structure, dynamics, and stability can be related to the biological consequences of thymine glycol in DNA. These results on thymine glycol will also be compared with those from our ongoing studies of the structures of DNAs containing abasic sites.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation**—DNA containing thymine glycol was prepared by oxidation of the parent ssDNA with 0.1 mM KMnO₄. The oxidation was carried out in a 300-ml plastic jar containing 20 ml of 0.2 M KH₂PO₄ at pH 8.6 and 50 OD₂₆₀ of d(C¹G²C³G⁴A⁵T⁶A⁷C⁸G⁹C¹⁰C¹¹). The mixture was stirred with a magnetic stirrer for 20 min in an ice bath. With the sample at 4 °C, it was treated with 8 ml of 0.1 M KMnO₄ (Aldrich) for 5 min. The reaction was quenched by the addition of 0.5 ml of allyl alcohol (Aldrich), which converts the MnO₄ into MnO₂. The sample was kept at 4 °C for at least 1 h to allow the MnO₂ to completely precipitate. The reaction mixture was then centrifuged to remove the MnO₂, and the supernatant containing the products was diluted to 400 ml with distilled water and was desalted with the use of a Waters C-18 cartridge. The DNA was eluted from the C-18 column with 5 ml of 60% CH₃CN/H₂O in three steps (2 x 3 ml and 1 x 1 ml).

**Purification of the Oxidized DNA**—The ssDNA containing thymine glycol was purified by HPLC on a semi-preparative reversed-phase PRP-1 column (Hewlett-Packard) and eluted (elution time is 18 min) with a 2–25% gradient of 50% acetonitrile, 25 mM phosphate buffer at pH 7.0 mixed with 25 mM phosphate buffer at pH 7.0. A typical chromatogram is shown in Fig. 1. The column flow rate was 2 ml/min, and detection was at 254 nm. The other peaks produced by oxidation were not identified and are likely to contain oxidized guanine, adenosine, or other oxidation products. The purified single strand containing thymine glycol was collected, diluted, and desalted using a C-18 column. The isolated yield was typically about 24%.

**Characterization of the Thymine Glycol in the Single Stranded DNA**—To confirm that the DNA prepared as described above contains a single thymine glycol and no other modifications, further analysis was carried out. A sample of 1 OD of the ssDNA containing thymine glycol was incubated with 2 units of venom phosphodiesterase, 4.1 units of alkaline phosphatase, 1 x Tris, 1 x MgCl₂ at pH 8 and 37 °C for 24 h. The parent strand was also digested using the same conditions except that the incubation time needed for complete digestion was only 1 h. Both DNAs were completely digested to nucleosides by this procedure, and the nucleoside mixtures were analyzed by HPLC.

The HPLC was performed using a PRP-1 column with a 3–18% gradient of 50% acetonitrile, 25 mM phosphate buffer at pH 7.0 mixed with 25 mM phosphate buffer at pH 7.0. The column flow rate was 2
non-exchangeable protons

inimo region

\[ 31P \]

\[ 8 \quad 10 \quad 12 \quad 14 \quad 16 \quad 18 \quad 20 \quad 22 \quad 24 \quad 26 \quad 28 \quad 30 \quad 32 \quad 34 \quad 36 \quad 38 \quad 40 \quad 42 \quad 44 \quad 46 \quad 48 \quad 50 \quad 52 \quad 54 \quad 56 \quad 58 \quad 60 \quad 62 \quad 64 \quad 66 \quad 68 \quad 70 \quad 72 \quad 74 \quad 76 \quad 78 \quad 80 \quad 82 \quad 84 \quad 86 \quad 88 \quad 90 \quad 92 \quad 94 \quad 96 \quad 98 \quad 100 \]

The spectra shown are the DNA duplex of \( \text{d(CGCGA'TgA'CGCGC'15)} \) paired with \( \text{d(GG}^{15} \text{CGT'T}\text{T}^{15} \text{TACG}) \), with Tg indicating thymine glycol. The top spectrum is the 400-MHz proton NMR spectrum of the sample in \( 2H_2O \) solution. The middle spectrum is the 400-MHz spectrum of the imino proton region obtained with the sample in 90% \( H_2O/10\% 2H_2O \). The bottom spectrum is the 161-MHz \( 31P \) spectrum of the duplex obtained with proton decoupling.

\( R \) (5,6)- to (5,6)-stereoisomers obtained in the nucleoside oxidation (30). It is not obvious to us how a 5'-phosphate controls the stereochemistry of the permanganate oxidation.

The heteroduplex was formed by mixing equimolar quantities based on the extinction coefficients of the two strands and by monitoring the titration of the single strand containing the thymine glycol with its complementary strand via the one-dimensional NMR spectrum. The duplex was lyophilized several times in \( 2H_2O \) and dissolved in 0.5 ml of 99.96% \( 2H_2O \). The purified duplex was studied at 1–1.5 mm concentration in pH 7.0 buffer containing 10 mM sodium phosphate, 100 mM sodium chloride, and 0.05 mM EDTA in 99.96% \( 2H_2O \). For experiments involving the exchangeable imino protons, the duplex was lyophilized and dissolved in 90% \( H_2O/10\% 2H_2O \).

**NMR Procedures—Two-dimensional NOESY, total correlation spectroscopy, and phased easy correlation spectroscopy spectra were collected at 400 and 750 MHz. The 400-MHz data were obtained using a Varian Unityplus spectrometer at Wesleyan, and the 750-MHz spectra were obtained using a Bruker DMX 750 spectrometer at the University of Wisconsin at Madison. The Varian NMR results were processed using VNMR software, and FELIX 95.0 software was used for the Bruker data.**

NOESY experiments in \( 2H_2O \) were carried out at mixing times of 150 and 250 ms with a 1 s equilibrium delay with presaturation of the water resonance, using the Bruker 750 MHz with a spectral width of 11904.8 Hz in each dimension. For each mixing time, 512 \( \tau_1 \) increments were acquired with 64 scans for each increment of the evolution time. The \( F_1 \) dimension was zero-filled to 2000, and the data were processed with sinebell apodization in each dimension prior to 4000 \( \times \) 2000 Fourier transformation. These data were used for quantification of the NOESY cross-peaks.

The total correlation spectroscopy spectra were obtained at 750 MHz and 25 °C with a mixing time of 60 ms and a delay time of 1 s between acquisitions. In this experiment, 512 \( \tau_1 \) increments were collected with
The starting structure was generated from a canonical B-form DNA in INSIGHT II by modifying the thymine residue in parent strand to thymine glycol and then minimized by DISCOVER 3.0. Because there are four possible thymine glycol structural forms that are the two chair forms and two boat forms, a MNDO minimization using INSIGHT II and molecular dynamics using X-PLOR were performed on thymine glycol. The potential was determined from MNDO, and the molecular dynamics showed that the free thymine glycol underwent rapid interconversion between the boat and chair forms. This indicated that the MNDO-generated potential was a reasonable one for thymine glycol. This potential for thymine glycol was then inserted to the topology file of the duplex DNA.

The energy of the starting structure was minimized in 100 steps of Powell’s conjugate gradient minimization using X-PLOR. The relaxation matrix refinements were carried out in真空 at 300 K. These were further minimized using the force field with all restraints for 100 steps of minimization and then subjected to a 60 ps relaxation matrix simulation followed by 200 steps of conjugate gradient energy minimization. The structures were saved at every 2-ps interval. The root mean square deviation over the 50–60 ps portion of the trajectory is 0.212 Å.

The NOE cross-peak volumes for the structures every 2 ps were back calculated using an overall correlation time of 5 ns, a leakage rate of 0.33 s⁻¹, and a distance cut-off of 5.5 Å. The NMR R, Q, and root mean square values (35) for the averaged structures are 0.10, 0.05, and 0.12, which are slightly better than the 0.11, 0.05, and 0.12 obtained for the 60 ps structure. The R, Q, and root mean square values all declined about 10–20% from 10 to 50 ps, at which time they stabilized. The structures and experimental data have been submitted to the Brookhaven data bank.

RESULTS AND DISCUSSION

The basic properties of the damaged DNA duplex could be determined from the one-dimensional spectra shown in Fig. 2. The number of signals in the one-dimensional spectrum of the nonexchangeable protons indicates that the duplex adopts one structure on the NMR time scale. The intensity of the imino region of the spectrum indicates that there are 10 base pairs present in the duplex. The range of ³¹P chemical shifts indicates that there are not any phosphodiester with highly unusual conformations.

Assignment of the NMR Spectrum of the DNA Duplex Containing Thymine Glycol—The spectra were assigned by the application of sequential assignment procedures used for B-form DNA (36, 37). A portion of a 750-MHz NOESY spectrum is shown in Fig. 3 with many of the assignments of the H6/H8-H1' region indicated. The H6/H8-H2' and H2' regions of a 750-MHz NOESY spectrum are shown in Fig. 4, and many of the H6/H8-H2' assignments are indicated. The assignments of the thy-
FIG. 5. Stereoview showing the superposition of the structures from 50 to 60 ps in the trajectory viewed into major groove along with the calculated global helix axis.
FIG. 6. Stereoview showing the superposition of the structures from 50 to 60 ps in the trajectory viewed into minor groove along with the calculated global helix axis.
FIG. 7. Refined structures of the duplex DNA containing thymine glycol and the α and β forms of the abasic site. The view is looking into major groove. The central region containing the damaged site is shown on an expanded scale on the bottom.

FIG. 8. Refined structures of the duplex DNA containing thymine glycol and the α and β forms of the abasic site. The view is looking into minor groove. The central region containing the damaged site is shown on an expanded scale on the bottom.
The sequential 3'-base to 5'-deoxyribose connectivities associated with B-form DNA were observed for residues 1–5 and 6–11 of the thymine glycol containing strand. There is a clear break in the inter-residue NOE connectivities found between the thymine glycol at position 6 and the A at position 5. The sequential 3'-base to 5'-deoxyribose connectivities associated with B-form DNA were observed for all of the residues of the complementary strand.

The NMR data presented to this point indicate that the thymine glycol containing DNA can be described as being close to an normal B-form DNA structure as monitored by the imino proton, the $^{31}$P chemical shifts, and the proton-proton NOEs except at the A5-Tg6 junction. At this base step the NOE connectivities are not consistent with B-form DNA.

The structure of the thymine glycol containing duplex was determined by the combination of molecular dynamics with experimental NOE and dihedral angle constraints. The range of structures consistent with the experimental data are shown in Figs. 5 and 6. These structures are those obtained at 2-ps intervals in the molecular dynamics trajectory between 50 and 60 ps. These structures indicate the range of structures consistent with the data and not the range of structures that the molecule would sample over a 10-ps time period. The experimental NOE data and that predicted from these structures are shown in Figs. 3 and 4. The comparison shows that the agreement between the predicted and experimental data is very good. The predicted data are the averages over the structures.
Effects of Thymine Glycol on Duplex DNA Structure

9235

obtained from 50 to 60 ps. The average over the predicted structures offers better agreement with the experimental data than any of the individual structures.

The structure of duplex DNA containing an abasic site in the same position as this DNA has a thymine glycol has been previously determined. Figs. 7 and 8 show the structures of the duplex DNAs that have the α and β forms of the abasic site, as well as the structure of the duplex DNA containing the thymine glycol. The comparison shows that the structures that are found depend on which damaged site is present. The damaged DNAs containing the α and β forms of the abasic site are distinct from one another as well as from the thymine glycol containing duplex DNA. The thymine glycol is not accommodated into the normal DNA stacking.

The "Extrahelicity" of the Thymine Glycol Residue—Extrahelical bases have been previously observed in DNA duplexes by crystallography (38, 39) and by NMR methods (40–45). The thymine glycol residue in this DNA appears to be extrahelical because it is not in the central stack of the DNA. However, there is no generally agreed upon definition of extrahelicity.

To move toward a quantitative measure of extrahelicity, we have examined the surface area of DNA residues as a function of the radius of the probe molecule. It was thought that extrahelical residues would have a larger surface area for probe molecules about the size of a water molecule and that the percentage of the surface area accessible to the probe molecule might provide a means to quantify the extent of extrahelicity.

The surface areas of nucleic acid residues in canonical structures as a function of probe molecule radius has been previously examined for other reasons.

The percentage of the surface area of the thymine glycol residue may be considered to be approximately half extrahelical.

To move toward a quantitative measure of extrahelicity, we have examined the surface area of DNA residues as a function of the radius of the probe molecule. It was thought that extrahelical residues would have a larger surface area for probe molecules about the size of a water molecule and that the percentage of the surface area accessible to the probe molecule might provide a means to quantify the extent of extrahelicity.

The surface areas of nucleic acid residues in canonical structures as a function of probe molecule radius has been previously examined for other reasons.

The percentage of the surface area of the thymine glycol residue may be considered to be approximately half extrahelical.

To move toward a quantitative measure of extrahelicity, we have examined the surface area of DNA residues as a function of the radius of the probe molecule. It was thought that extrahelical residues would have a larger surface area for probe molecules about the size of a water molecule and that the percentage of the surface area accessible to the probe molecule might provide a means to quantify the extent of extrahelicity.

The surface areas of nucleic acid residues in canonical structures as a function of probe molecule radius has been previously examined for other reasons.

The percentage of the surface area of the thymine glycol residue may be considered to be approximately half extrahelical. These NMR results on thymine glycol containing DNA indicate that the presence of thymine glycol induces a significant and highly localized alteration in the structure of the DNA. The base of the thymine glycol is largely accessible to the solvent and other molecules. It is likely that this structural perturbation due to the presence of thymine glycol effects the recognition of the DNA by proteins and may also effect the packaging of DNA in the replication complex, in nucleosomes, in viruses, and in other contexts. The distortions that accompany the presence of the thymine glycol may provide a readily recognizable target for repair enzymes as well as the direction interaction with the thymine glycol.

The structural perturbation due to thymine glycol is quite different from that found for DNA, which contains aldehydic abasic sites. The comparison of the structures of duplex DNAs of the same sequence that have abasic site damage or thymine glycol shows that the structure of the damaged DNA is dependent on the nature of the damaged site. Thus, the partial repair of a thymine glycol site to an abasic site will reduce the structural consequences of the damage to DNA.

Acknowledgments—The 750-MHz spectra were obtained with the assistance of Dr. Fris V. Abidagwa using the NMR facility at the University of Wisconsin at Madison, which is supported by National Institutes of Health Grant RR02301 with equipment purchased with support from the University of Wisconsin (NSF DMB-8415048 and NIH RR02781) and the United States Department of Agriculture.

REFERENCES

44. Hare, D., Shapiro, L., and Patel, D. J. (1986) Biochemistry 25, 7456–7464
Structure of a Duplex DNA Containing a Thymine Glycol Residue in Solution
Hsiang Chuan Kung and Philip H. Bolton

doi: 10.1074/jbc.272.14.9227

Access the most updated version of this article at http://www.jbc.org/content/272/14/9227

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 42 references, 6 of which can be accessed free at http://www.jbc.org/content/272/14/9227.full.html#ref-list-1