Characterization of the Effects of a Thymine Glycol Residue on the Structure, Dynamics, and Stability of Duplex DNA by NMR*

(Received for publication, March 17, 1993, and in revised form, May 12, 1993)

Jung Yie Kao, Igor Goljer, Thu A. Phan, and Philip H. Bolton‡
From the Chemistry Department, Wesleyan University, Middletown, Connecticut 06459

A duplex DNA containing a single thymine glycol (5,6-dihydroxy-5,6-dihydrothymidine) has been studied by NMR and other methods. Oxidative stress, ionizing radiation, and other causes can induce the oxidation of thymine to thymine glycol. The presence of thymine glycol is known to have significant biological consequences, and there are repair enzymes for thymine glycol in a wide range of organisms. These studies have been carried out on the DNA duplex of d(C1G2A3A4G5Tg6C7A8GBC9O) paired with d(G5A7G8T9C10A11G12T13C14G15G16), with Tg indicating thymine glycol. The presence of thymine glycol lowers the thermal stability of duplex DNA. The NMR results indicate that thymine glycol induces a large, localized structural change in duplex DNA with the thymine glycol base being extrahelical as well as the opposing base on the complementary strand. This structural information is consistent with the biological consequences of thymine glycol in DNA.

For an organism to survive and to successfully reproduce, there must be means by which its DNA can be repaired. Damage to DNA can occur by the spontaneous deamination of cytosine to uracil and through the action of alkylating agents, oxidants, drugs, toxins, and ionizing radiation (1-3). The exposure of DNA in cells, bound to proteins as a solid or free in solution, to ionizing radiation or to oxidative stress leads to the conversion of thymine to thymine glycol (5,6-dihydroxy-5,6-dihydrothymidine) as shown in Fig. 1. Ames and co-workers (4) have shown that the average human cell repairs about 320 thymine glycol sites/day based on the examination of the thymine glycol content of urine. Since ~10-20% of the damage to DNA by ionizing radiation is thymine base oxidation and fragmentation, and these products are also produced by oxidative stress, the effects of damaged thymines of the structures, stabilities, dynamics, and interactions of DNA is of interest.

It has been known for several decades that ionizing radiation can kill or stop the reproduction of cells. Since the mid-1950s there have been studies of the effects of ionizing radia-

* This research was supported in part by Grant DE-FG02-93ER61586 from the Department of Energy. The 600-MHz spectra were obtained with the assistance of Dr. Frits Abildgaard using the NMR facility at the University of Wisconsin, Madison, which is supported by National Institutes of Health Grant RR02301 with equipment purchased with support from the University of Wisconsin, National Science Foundation Grant DMB-8415048, National Institutes of Health Grant RR02781, and the United States Department of Agriculture. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed.

**FIG. 1. The oxidation of thymine to thymine glycol is depicted. The KMnO₄ oxidation of thymine yields only cis thymine glycol, and the 5R,6S cis stereoisomer is shown. The 5R,6S stereoisomer is the predominant product formed in the oxidation of ssDNA.**

The main questions concerning the biological consequences of thymine base damage are as to the effects of such damage on DNA function and whether the damage can be repaired. 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 (19, 20). We have recently shown that
Effects of Thymine Glycol on Duplex DNA Structure

**FIG. 2.** The HPLC traces that illustrate the purification of thymine glycol-containing DNA, as well as the characterization of the thymine glycol-containing DNA, are shown. In panel a, the top trace is of the unmodified T11 ssDNA strand, the middle trace is of the oxidation products of T11, and the bottom trace is of the purified Tg11 ssDNA. In panel b, the top HPLC trace is of the mixture of nucleosides formed by the digestion of T11, and the bottom trace is of the mixture of nucleosides formed by the digestion of Tg11. In panel c, the top HPLC trace is of the thymine glycol nucleosides formed by oxidation of the nucleoside thymidine, and the bottom HPLC trace is of the thymine glycol nucleosides, isolated from the digest of Tg11, formed in the oxidation of ssDNA.

endo III cleaves the 3' phosphodiester of abasic sites by a syn β-elimination reaction (21, 22). Wallace and co-workers (23) have shown that φX174 containing thymine glycol are inactivated in E. coli hosts deficient in endo1 III much more so than in wild type hosts, indicating that endo III is most likely involved in thymine glycol repair in vivo. They have also shown that φX174 containing thymine glycol are inactivated in E. coli hosts deficient in exo 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 (23). Thymine glycols can apparently also be repaired through the

---

1 The abbreviations used are: endo, endonuclease; ss, single-stranded; HPLC, high performance liquid chromatography; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; DQCOSY, double quantum filtered correlation spectroscopy.

**FIG. 3.** The 400-MHz proton NMR spectra of d(C,GzC,A,G,Tg,C7A,GeCl~Cl~), with Tg indicating thymine glycol, and d(C,GzC,A,G,Tx,C7A,GeC6C11) are shown. The position of the H6 resonance of the T6 resonance is indicated.

**FIG. 4.** The spectra shown are the DNA duplex of d(C,GzC,A,G,Tg,C7A,GeC6C11) paired with d(GzzCzGzTz,CzGzGzGzC11), with Tg indicating thymine glycol. The top spectrum is the 600-MHz proton NMR spectrum of the sample in 'H2O solution. The middle spectrum is the 400-MHz spectrum of the imino proton region obtained with the sample in 90% H2O/10% 2H2O. The bottom spectrum is the 161-MHz 31P spectrum of the duplex obtained with proton decoupling.

uv-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, exo III, and endo IV SOS (10). These results indicate that thymine glycol sites can be repaired in vivo, at least in E. coli.

The presence of thymine glycol in DNA has profound
Effects of Thymine Glycol on Duplex DNA Structure

FIG. 5. The spectra shown are taken from a 600-MHz two-dimensional NOESY data set obtained with a 250-ms mixing time. The region in panel A contains the H6/H8–H1' cross-peaks used to make sequential assignments with the connectivities of the d(GpCpCpTpCpApGpTpGp(Gp) strand indicated. The region in panel B contains the H6/H8–H1' cross-peaks to make sequential assignments with the connectivities of the d(CpCpCpApGpTpCpAp Cp(Cp)Cp(Cp)) strand indicated.

Figures 5 and 6 illustrate the spectra obtained from a 600-MHz two-dimensional NOESY data set. The spectra were acquired with a 250-ms mixing time. The region in panel A contains the H6/H8–H1' cross-peaks used to make sequential assignments with the connectivities of the indicated strands. The region in panel B contains the H6/H8–H1' cross-peaks to make sequential assignments with the connectivities of the indicated strands.

consequences on DNA replication (8, 10, 24–28). Several groups have shown that the presence of thymine glycol leads to a block of replication (23–28). The presence of thymine glycol blocks replication in vivo (23). 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 1 residue before or at the site of damage and 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 (23).

Furthermore, the presence of a pyrimidine 5' to the damaged thymine seems to allow the reading through of a thymine glycol site in replication more so than a 5' purine, indicating a sequence-dependent effect (23). Other studies have indicated that the base that is 3' to a thymine glycol is also important in determining the extent of the block to replication (28). There is also evidence that thymine glycol predominately codes for dA (8, 10, 28). 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 (28).

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, that 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. Since thymine glycol cannot 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 be compared with those from our ongoing studies of abasic sites in DNA.

EXPERIMENTAL PROCEDURES

Preparation of Thymine Glycol-containing DNA—The oxidation was carried out in a 300-ml plastic jar equipped with a stir bar. The reaction mixture consisted of 50 \( A_{260} \) of d(CpCpCpApGpTpCpAp Cp(Cp)Cp(Cp)) in 20 ml of 0.2 M KHPO₄ at pH 8.6. The reaction mixture was equilibrated to 4 °C, and then 8 ml of 0.1 M KMnO₄ (Aldrich) was added, while the solution was vigorously stirred. The oxidation reaction proceeded for 5 min and was then quenched by the addition of 0.5 ml of allyl alcohol (Aldrich). The reaction mixture was kept at 4 °C for at least 30 min to allow the MnO₂ to completely precipitate. The reaction mixture was then centrifuged to remove the MnO₂, the solution containing the products was diluted to 400 ml with distilled water, and the solution was
Effects of Thymine Glycol on Duplex DNA Structure

desalted with the use of a Waters C18 cartridge. The DNA was eluted from the C18 column with 5 ml of 60% CH3CN/H2O in three steps (2 × 2 ml and 1 × 1 ml).

The ssDNA containing thymine glycol was purified by HPLC using a Hewlett-Packard PRP-1 column with a CH3CN/phosphate buffer, and a typical set of chromatograms are shown in Fig. 2a. The top trace is that of T11, the middle trace is of oxidized T11, and the bottom trace is that of purified Tg11. The numerous small peaks, observed in the middle trace, produced by oxidation were not identified and are likely to contain oxidized guanine and adenosine. The ssDNA containing thymine glycol peak was collected, diluted, and desalted using a C18 column. The isolated yield of thymine glycol-containing DNA was typically 13 Åw, about 25%, and three runs of this procedure were typically needed to prepare sufficient material for a single NMR sample.

We had originally attempted to prepare thymine glycol-containing DNA using osmium oxidation. A large number of variations on the published procedures, reviewed by Demple and Levin (10), were attempted, as well as running the oxidation in mixed solvents, to enhance the solubility of osmium tetroxide. We were unable to find any T in the Tg11 sample, and only the expected nucleosides were detected. The thymine glycol nucleoside elutes well before C and does not have sufficient absorbance at 254 nm to be detected. The thymine glycol nucleoside can be detected at 220 nm.

Permanganate oxidation of thymine has been shown to yield only the two cis isomers (31). The determination of the stereochemistry of the cis thymine glycol produced by the oxidation of single-stranded DNA was determined by HPLC and confirmed by NMR. Thymine glycol nucleosides were prepared from Tg11 by the digestion described above and chromatographed on a Whatman ODS-3 semipreparative column, 10 mm × 50 cm, with a mobile phase of 3% acetonitrile, 25 mM phosphate buffer at pH 7 mixed with 25 mM phosphate buffer at pH 7.0. The column flow rate was 2 ml/min, and the nucleosides were detected at 264 nm. There was no evidence for any T in the Tg11 sample, and only the expected nucleosides were detected. The thymine glycol nucleoside elutes well before C and does not have sufficient absorbance at 254 nm to be detected. The thymine glycol nucleoside can be detected at 220 nm.

Characterisation of the Thymine Glycol in the Single-stranded DNA—To unambiguously show that the DNA prepared and purified as described above contains a single thymine glycol, further analysis was needed, with the details given below. The DNA was digested to nucleosides followed by chromatography to determine the A:G:C:T base composition. The results showed that the A:G:C ratios were the same as in the parent strand and that no T was present in Tg11. The thymine glycol nucleosides were isolated by HPLC from the digestion mixture of Tg11 and rechromatographed to resolve the two cis stereoisomers. The chemical identity of each of the stereoisomers of the thymine glycol nucleoside were confirmed by obtaining one- and two-dimensional NMR data on the purified nucleosides.

The ssDNA containing thymine glycol was digested to nucleosides and analyzed by HPLC (16, 30, 46, 47). Typical chromatograms are presented in Fig. 2b with the top trace of the nucleosides obtained from the digestion of T11 and the bottom trace of the nucleosides obtained from the digestion of Tg11. The A, G, C nucleosides obtained by digestion of Tg11 were found to have the same elution times and ratios as those of T11. The nucleosides were produced by the digestion of T11 with venom phosphodiesterase I and alkaline phosphatase in tris buffer at pH 8.0. The order of elution of nucleosides is C, G, T, A, and the elution times were 12.9, 14.8, 16.2, and 19.9, respectively. The HPLC was run using a PRP-1 column with a 3–18% gradient of 50% acetonitrile, 25 mM phosphate buffer at pH 7 mixed with 25 mM phosphate buffer at pH 7.0. The column flow rate was 2 ml/min, and the nucleosides were detected at 264 nm. There was no evidence for any T in the Tg11 sample, and only the expected nucleosides were detected. The thymine glycol nucleoside elutes well before C and does not have sufficient absorbance at 254 nm to be detected. The thymine glycol nucleoside can be detected at 220 nm.

Permanganate oxidation of thymine has been shown to yield only the two cis isomers (31). The determination of the stereochemistry of the cis thymine glycol produced by the oxidation of single-stranded DNA was determined by HPLC and confirmed by NMR. Thymine glycol nucleosides were prepared from Tg11 by the digestion described above and chromatographed on a Whatman ODS-3 semipreparative column, 10 mm × 50 cm, with a mobile phase of 3% acetonitrile in water with a flow rate of 2 ml/min. The elution time of the 5S,5R stereoisomer is 18.0 min, and that of the 6S,5R stereoisomer is 17.0 min. A typical chromatogram, with detection at 220 nm, is the bottom

![Diagram](image-url)

**Fig. 7.** The spectrum shown is a region of the 600-MHz TOCSY spectrum of the DNA duplex containing thymine glycol obtained with a mixing time of 250 ms. The lines shown connect the H2' and H2'' resonances of the indicated residues. The F2 chemical shift is that of the H1' proton of the same residue.

**Table I**

Chemical shift assignments of the DNA duplex of d(C6G5C4A3G2T1C10G9C8C7C6) paired with d(G7C6T5C4A3G2T1C10G9C8C7C6T1), with Tg2 indicating thymine glycol

<table>
<thead>
<tr>
<th></th>
<th>H8/H6</th>
<th>H5/H2</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H3'</th>
<th>H4'</th>
<th>H5'</th>
<th>H5''</th>
<th>CH5</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7.63</td>
<td>5.71</td>
<td>5.71</td>
<td>1.97</td>
<td>2.38</td>
<td>4.81</td>
<td>3.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>7.95</td>
<td>5.86</td>
<td>2.27</td>
<td>2.73</td>
<td>4.95</td>
<td>4.33</td>
<td>4.14</td>
<td>4.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>7.39</td>
<td>5.43</td>
<td>5.48</td>
<td>2.02</td>
<td>2.34</td>
<td>4.82</td>
<td>4.14</td>
<td>3.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>8.19</td>
<td>6.07</td>
<td>2.68</td>
<td>2.91</td>
<td>5.02</td>
<td>4.39</td>
<td>4.15</td>
<td>4.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G5</td>
<td>7.57</td>
<td>5.90</td>
<td>2.72</td>
<td>2.85</td>
<td>4.90</td>
<td>4.34</td>
<td>3.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T6</td>
<td>4.77</td>
<td>5.21</td>
<td>1.95</td>
<td>2.16</td>
<td>4.63</td>
<td>3.88</td>
<td>3.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>7.57</td>
<td>5.59</td>
<td>5.48</td>
<td>2.02</td>
<td>2.34</td>
<td>4.79</td>
<td>4.11</td>
<td>3.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>8.17</td>
<td>5.96</td>
<td>2.71</td>
<td>2.85</td>
<td>4.99</td>
<td>4.22</td>
<td>4.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G9</td>
<td>6.88</td>
<td>5.75</td>
<td>2.51</td>
<td>2.60</td>
<td>4.93</td>
<td>4.35</td>
<td>4.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>7.37</td>
<td>5.29</td>
<td>5.95</td>
<td>2.08</td>
<td>2.42</td>
<td>4.73</td>
<td>4.39</td>
<td>4.13</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>7.63</td>
<td>5.89</td>
<td>6.21</td>
<td>2.22</td>
<td>2.21</td>
<td>4.52</td>
<td>3.98</td>
<td>4.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td>7.86</td>
<td>5.77</td>
<td>2.63</td>
<td>2.80</td>
<td>4.97</td>
<td>4.75</td>
<td>4.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C13</td>
<td>7.43</td>
<td>5.30</td>
<td>5.95</td>
<td>2.04</td>
<td>2.47</td>
<td>4.72</td>
<td>4.40</td>
<td>4.22</td>
<td>4.14</td>
<td></td>
</tr>
<tr>
<td>T15</td>
<td>7.28</td>
<td>5.71</td>
<td>1.96</td>
<td>2.36</td>
<td>4.76</td>
<td>4.22</td>
<td>4.08</td>
<td>4.07</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>G16</td>
<td>7.89</td>
<td>5.74</td>
<td>2.64</td>
<td>2.80</td>
<td>4.83</td>
<td>4.34</td>
<td>4.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A17</td>
<td>7.83</td>
<td>6.02</td>
<td>2.58</td>
<td>2.68</td>
<td>4.76</td>
<td>4.33</td>
<td>4.15</td>
<td>4.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C18</td>
<td>7.23</td>
<td>5.09</td>
<td>5.59</td>
<td>1.91</td>
<td>2.62</td>
<td>4.62</td>
<td>4.07</td>
<td>3.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T19</td>
<td>7.36</td>
<td>5.75</td>
<td>2.45</td>
<td>2.62</td>
<td>4.93</td>
<td>4.34</td>
<td>4.23</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G20</td>
<td>7.89</td>
<td>5.85</td>
<td>1.89</td>
<td>2.32</td>
<td>4.96</td>
<td>4.33</td>
<td>4.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21</td>
<td>7.34</td>
<td>5.42</td>
<td>5.72</td>
<td>1.98</td>
<td>2.39</td>
<td>4.68</td>
<td>4.08</td>
<td>4.04</td>
<td>3.95</td>
<td></td>
</tr>
<tr>
<td>G22</td>
<td>7.94</td>
<td>6.14</td>
<td>2.60</td>
<td>2.32</td>
<td>4.65</td>
<td>4.15</td>
<td>4.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE II

<table>
<thead>
<tr>
<th>Chemical shift assignments of the DNA duplex of d(C,GzA,TzCzA,GzC,TzCzGzA) paired with d(GzC,TzGzA,GzC,TzCzGzA)</th>
<th>H5'/H6</th>
<th>H5/H2</th>
<th>H1'</th>
<th>H2'</th>
<th>H2''</th>
<th>H5''</th>
<th>H3'</th>
<th>H4'</th>
<th>H3''</th>
<th>H4''</th>
<th>CH3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7.78</td>
<td>6.05</td>
<td>6.04</td>
<td>2.06</td>
<td>2.48</td>
<td>4.85</td>
<td>3.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>8.09</td>
<td>5.92</td>
<td>2.79</td>
<td>2.86</td>
<td>5.12</td>
<td>4.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>7.48</td>
<td>5.57</td>
<td>5.72</td>
<td>2.14</td>
<td>2.50</td>
<td>5.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>8.35</td>
<td>6.22</td>
<td>2.85</td>
<td>3.03</td>
<td>5.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G5</td>
<td>7.62</td>
<td>5.97</td>
<td>2.48</td>
<td>2.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T6</td>
<td>7.32</td>
<td>6.10</td>
<td>2.08</td>
<td>2.48</td>
<td>4.96</td>
<td>4.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>7.62</td>
<td>5.75</td>
<td>2.08</td>
<td>2.43</td>
<td>4.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>8.33</td>
<td>6.13</td>
<td>2.84</td>
<td>3.01</td>
<td>5.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G9</td>
<td>7.82</td>
<td>5.92</td>
<td>2.63</td>
<td>2.68</td>
<td>5.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>7.52</td>
<td>5.47</td>
<td>6.12</td>
<td>2.23</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>7.81</td>
<td>5.89</td>
<td>6.38</td>
<td>2.29</td>
<td>2.55</td>
<td>4.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td>8.02</td>
<td>5.85</td>
<td>2.71</td>
<td>2.68</td>
<td>5.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G13</td>
<td>8.00</td>
<td>6.12</td>
<td>2.42</td>
<td>2.89</td>
<td>5.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C14</td>
<td>7.98</td>
<td>5.46</td>
<td>6.15</td>
<td>2.16</td>
<td>2.62</td>
<td>5.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T15</td>
<td>7.61</td>
<td>5.81</td>
<td>2.13</td>
<td>2.50</td>
<td>5.00</td>
<td>4.24</td>
<td>1.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G16</td>
<td>8.09</td>
<td>5.66</td>
<td>2.82</td>
<td>2.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A17</td>
<td>8.31</td>
<td>6.34</td>
<td>2.82</td>
<td>3.01</td>
<td>5.16</td>
<td>4.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C18</td>
<td>7.37</td>
<td>5.32</td>
<td>5.88</td>
<td>2.04</td>
<td>2.82</td>
<td>4.82</td>
<td>4.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T19</td>
<td>7.40</td>
<td>5.86</td>
<td>2.18</td>
<td>2.57</td>
<td>4.82</td>
<td>4.19</td>
<td>1.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G20</td>
<td>8.02</td>
<td>5.86</td>
<td>2.18</td>
<td>2.57</td>
<td>5.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21</td>
<td>7.49</td>
<td>5.57</td>
<td>5.92</td>
<td>2.04</td>
<td>2.48</td>
<td>4.94</td>
<td>4.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G22</td>
<td>8.08</td>
<td>6.29</td>
<td>2.74</td>
<td>2.52</td>
<td>5.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 8.** The spectra shown are slices taken from a 600-MHz NOESY data set on the thymine glycol-containing DNA duplex obtained with a mixing time of 250 ms. The five slices were extracted at the chemical shifts of the indicated thymine glycol protons. The vertical dashed lines are at the chemical shifts of the indicated thymine glycol and C7 protons.

For comparison, the results obtained on the ratio of stereoisomers formed in the permanganate oxidation of thymine nucleoside, under the same conditions as those used for the ssDNA, are also shown in the top trace of Fig. 2c. These results show that the thymine glycol in Tg11 is the 5R,6S stereoisomer shown in Fig. 1. The ratio of 5R,6S to 5S,6R is at least 5 to 1 in Tg11, whereas the ratio is about unity when the oxidation occurs on a nucleoside. Since the ratio of the two stereoisomers produced by oxidation of thymine is close to unity, whereas that produced by the oxidation of ssDNA is at least 5 to 1, there is partial stereochemical control over the oxidation in ssDNA.

The identification of each of the two stereoisomers was confirmed one- and two-dimensional NMR. In particular, the NOEs between the H6 and H1' protons were found to be in agreement with those previously reported (29) for each stereoisomer.

DQCOSY data (32) were obtained on Tg1 at 400 and 600 MHz. The NMR methods—Two-dimensional NOESY, TOCSY, and DQCOSY data (32) were obtained on Tg1 at 400 and 600 MHz. The 400-MHz spectra were obtained using spectrometers at Wesleyan and the 600-MHz spectra using an AM 600 at the University of Wisconsin, Madison.

**RESULTS**

**Characterization of ssDNA Containing Thymine Glycol—**

The 400-MHz proton NMR spectrum of the resulting single

**FIG. 9.** The spectra shown are slices taken from a 600-MHz NOESY data set on the thymine glycol-containing DNA duplex obtained with a mixing time of 250 ms. The three slices were extracted at the chemical shifts of the indicated A17 protons. The vertical dashed lines are at the chemical shifts of the indicated A17 protons.
strand of d(CGCGCGTGGGTGCATCGC), with Tg indicating thymine glycol, is shown in Fig. 3. The NMR results show that the thymine is oxidized as indicated by the chemical shift of the Tg methyl resonance. The proton NMR spectrum of the parent strand is also shown for comparison. The spectra of the two samples are essentially identical except for the resonances of the oxidized site.

**NMR Characterization of DNA Duplex Containing Thymine Glycol**—The duplex DNA was prepared by the preparation of the one-to-one mixture of d(CGCGCGTGGGTGCATCGC) and d(GCGCGCGGTGCGGAAGCTGG). This duplex will be referred to as Tg1. Fig. 4 contains the proton NMR spectrum of Tg1 in 2H2O. This spectrum shows that a duplex is formed since the chemical shifts of both strands in the mixture are distinct from those of the individual strands. The 31P spectrum in Fig. 4 suggests that the backbone of the duplex DNA is mostly regular, since the chemical shifts are in the general region associated with B-DNA. The imino proton spectrum of Tg1 in Fig. 4 shows that the duplex is base-paired, since an integration of the imino proton spectrum indicates that Tg1 has 10 Watson-Crick base pairs. The imino spectrum was obtained as a function of temperature, and these results (not shown) indicated an NMR melting temperature of Tg1 is about 40 °C. The optical melting temperature of Tg1 is 45 °C, determined at a concentration of 0.7 A260, and that of the analogous, undamaged DNA duplex is 52 °C. The optical melting curve of the Tg1 duplex exhibits lower cooperativity than that of the undamaged DNA duplex.

Taken together, these results show that duplex DNA containing thymine glycol can be prepared at high, NMR level purity and that duplex DNA containing a thymine glycol can be formed with the complementary strand. In addition, the NMR properties of the thymine-glycol-containing duplex are in the general range of B-DNA. The more detailed characterization of the structural features of Tg1 are described below.

**Assignment of the NMR Spectrum of the DNA Duplex Containing Thymine Glycol**—The spectra were assigned by the application of the sequential assignment procedures used for B-form DNA (33, 34). A portion of a 600-MHz NOE spectrum is shown in Fig. 5, and the H6/H8/H1' and the sequential assignment pathways of both strands are indicated. The H6/H8/H2' and H2* regions of a 600-MHz NOE spectrum is shown in Fig. 6. The H6/H8/H2' sequential assignments of both strands are indicated. A portion of the 600-MHz TOCSY spectrum of Tg1 is also shown in Fig. 7 with the assignments of some of the H1'-H2', H1'-H2' cross-peaks indicated, including that of the thymine glycol residue. The assignments of the Tg1 duplex and the corresponding undamaged DNA are listed in Tables I and II.

The sequential 5'-base to 3'-deoxyribose connectivities associated with B-form DNA were observed for residues 1-5 and 7-11 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 C at position 7.

The sequential 5'-base to 3'-deoxyribose connectivities associated with B-form DNA were observed for residues 22-18 and 16-12 of the complementary strand. There is a clear break in the inter residue NOE connectivities found between A17 and the G at position 16. A17 is the residue opposite the thymine glycol.

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 31P chemical shifts, and the proton-proton NOEs except at the C7-Tg6 and A17-G16 junctions. At these two base steps, the NOE connectivities are not consistent with B-form DNA.

The NOEs of Tg6 have been examined in detail, and Fig. 8 shows traces taken from a NOESY data set at the chemical shifts of the CH3, H6, H1', H2', and H2'' protons of the thymine glycol. The are several points about the NOEs of Tg6 that are of interest. The CH3 has, at most, a very weak NOE to H6. The CH3-H6 distance in thymine glycol is only somewhat greater than in thymine. Thus, the lack of this NOE suggests that the Tg is undergoing significant motion on the nanosecond time scale. Similarly, the intraresidue H6', H6-H2', and H6-H2'' NOEs are also very weak. The H1'-H2' and H1'-H2'' NOEs are at about the expected intensity. The only interresidue NOE that has been observed is a very weak one between the H6 of C7 and the H1' of Tg6.

These results on the NOEs of the thymine glycol indicate that the thymine glycol is most likely not stacked in the DNA duplex and that the base portion, at least, is extrahelical and may be disordered. The weak intraresidue H6-deoxyribose NOEs of Tg6 can arise from significant internal motion, disorder, and/or the extrahelical structure. The deoxyribose does not appear to be disordered.

Extrahelical bases have been previously observed in DNA duplexes that contain unpaired residues due to single-base insertions. Extrahelical purines have been investigated by crystallography (35-39). Extrahelical pyrimidines have also been detected by NMR methods (40-43). The NMR properties of the Tg6 and A17 are similar to those of the previously studied extrahelical residues in having low intensity sequential NOEs along with intraresidue NOEs consistent with an anti-conformation about the glycosidic bond.

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 and that of the opposing residue are both extrahelical. 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 nucleosomes, in viruses, and in other contexts. The extrahelical thymine glycol may provide a readily recognizable target for repair enzymes. In addition, since the biological effects of thymine glycol are apparently sequence-dependent, the structural effects may also depend on the sequence.

The structural perturbation due to thymine glycol is in marked contrast to that found for DNA, which contains aldehydic abasic sites. We have previously shown that the abasic sites produced by the removal of a base by a glycosylase induce relatively small structural perturbations in DNA (44, 45). Thus, the partial repair of a thymine glycol site to an abasic site will reduce the structural consequences of the damage to DNA.

The NMR and other results presented here show that the presence of thymine glycol has a large structural effect on DNA. Future studies are underway to examine the contacts of the extrahelical bases with water as well as to characterize the structure of the DNA at the damaged site.
REFERENCES