VOLUME 114, NUMBER 11
MAY 20, 1992

JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

Interstrand Cross-Linking of Duplex DNA by Nitrous Acid: Covalent Structure of the dG-to-dG Cross-Link at the Sequence 5′-CG

James J. Kirchner, Snorri Th. Sigurdsson, and Paul B. Hopkins*

Contribution from the Department of Chemistry, University of Washington, Seattle, Washington 98195. *Received November 12, 1991*

Abstract: It has previously been shown in synthetic oligodeoxynucleotides that the interstrand cross-linking reaction of nitrous acid with duplex DNA preferentially forms thermally- and base-stable links between deoxyguanosine residues at the duplex sequence 5′-CG. The covalent nucleus of this linkage is shown herein to result from the presence of a residue in which the original deoxyguanosyl residues on opposite strands are cross-linked through a single N^2 atom common to both [N^2-(2-deoxyinosyl)deoxyguanosine residue]. Variation of nitrous acid concentration, reaction time, and temperature established conditions under which up to a several percent yield of cross-linked oligodeoxynucleotide was obtained. Evidence for the covalent structure of the cross-link reported herein includes comparison of spectroscopic properties (mass spectrum, ultraviolet spectra at three pH values, and ^{1}H NMR spectrum) of N^2-(2-deoxyinosyl)deoxyguanosine isolated by enzymatic hydrolysis of cross-linked oligodeoxynucleotides to those of the same substance and its derivatives previously isolated from nitrous acid-treated calf thymus DNA. Further evidence in favor of cross-linking through N^2 is reported: substitution of deoxyinosine, which lacks an N^2 amino group, in place of deoxyguanosine at both sites in the duplex sequence 5′-CG abolishes dG-to-dG (dI-to-dI) cross-linking.

Introduction

Chemical agents which are capable of covalently altering DNA and the mechanisms by which they do so are of current widespread interest. Among these substances are numerous antibiotics, antitumor agents, carcinogens, and mutagens. One such agent is nitrous acid, which has long been known to convert the amino functional groups in DNA to carbonyl groups.[1] This reaction is of historical significance in molecular biology, having served as the first reliable chemical method for altering the genetic material with which a host cell was subsequently transformed.[2] Less well-known is the fact that nitrous acid is also a DNA interstrand cross-linking agent.[3,4] Although in random sequence DNA the extent of the interstrand cross-linking reaction is less than that of deamination, it is nevertheless significant: It has been estimated that for every four deaminations of deoxyguanosine, the most readily deaminated of the DNA components,[5] one interstrand cross-link is formed.[4] Interstrand cross-links are generally believed to be toxic, if not lethal to cells. The occurrence and consequences of deamination and cross-linking reactions of DNA with nitrous acid are of all the more interest given that there is a considerable dietary and environmental exposure of humans in our culture to oxides of nitrogen which can initiate these reactions.[6] The bioregulatory agent nitric oxide,[7] which is formed

(1) (a) Strecker, A. *Ann.* **1861**, *118*, 116. (b) Kossel, A. *Ber.* **1865**, *85*, 1928. (c) Kossel, A.; Steudel, H. *Z. Physiol. Chem.* **1903**, *37*, 377. (d) Schuster, H.; Schramm, G. *Z. Naturforschung* **1958**, *13b*, 697. See, also: Shapiro, R.; Pohl, S. H. *Biochemistry* **1968**, *7*, 448.

(2) Zimmerman, F. K. *Mutat. Res.* **1977**, *39*, 127.

(3) (a) Geiduschek, E. P. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 950. (b) Becker, E. F. *Biochim. et Biophys. Acta* **1967**, *142*, 238.

(4) Becker, E. F., Jr.; Zimmerman, F. K.; Geiduschek, E. P. *J. Mol. Biol.* **1964**, *8*, 377.

(5) (a) Schuster, H. *Z. Naturforsch.* **1960**, *15b*, 298. (b) Litman, R. *M. Chim. Phys.* **1961**, *58*, 997. (c) Shapiro, R.; Yamaguchi, H. *Biochim. Biophys. Acta* **1972**, *281*, 501.

(6) (a) Committee on Nitrite and Alternative Curing Agents in Food. *The Health Effects of Nitrate, Nitrite, and N-Nitroso Compounds*; National Academy Press: Washington, DC, 1981. (b) WHO Task Group on Environmental Health Criteria for Nitrates, Nitrites, and *N*-Nitroso Compounds. *Environmental Health Criteria 5, Nitrates, Nitrites, and N-Nitroso Compounds*; World Health Organization: Geneva, 1978. (c) *Handbook of Food Additives*, 2nd ed.; Furia, T. E., Ed.; CRC Press: Cleveland, 1972; pp 150–155. (d) Wakabayashi, K. In *Mutation and the Environment, Part E: Environmental Genotoxicity, Risk, and Modulation*; Mendelsohn, M. L., Albertini, R. J., Eds.; Wiley-Liss: New York, 1990; pp 107–116.

0002-7863/92/1514-4021$03.00/0

Table I. Interstrand Cross-Linking of DNAs with HONO

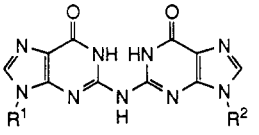
descriptor	DNA	yield (%) cross-link[a]	T_m[b] (°C)
TGCA	5'AATATAATTGCAATTAT TATTAACGTTAATATAA	0.15	40 (34)[c]
TCGA	5'AATATAATTCGAATTAT TATTAAGCTTAATATAA	0.60	40 (34)[c]
CCGG	5'AATATAATCCGGATTAT TATTAGGCCTAATATAA	2.4	42 (38)[c]
ATCCGGAA	5'AATATAATCCGGAATAT	0.07	20
CCGI	5'AATATAATCCGIATTAT TATTAIGCCTAATATAA	0.60	32
CCIG	5'AATATAATCCIGATTAT TATTAGICCTAATATAA	0.01	31
$(CG)_6$	5'CGCGCGCGCGCG GCGCGCGCGCGC		

[a] 0.5 o.d. 5'-^{32}P-radiolabeled DNA, 500 mM $NaNO_2$, pH 4.15, 300 mM NaOAc/HOAc buffer, 25 °C, 100 min, 563 mM total [Na^+]. [b] [Duplex DNA] ca. 30 μM, pH 4.15, 300 mM NaOAc/HOAc, 563 mM total [Na^+]. [c] Same as *b*, except [Na^+] = 63 mM.

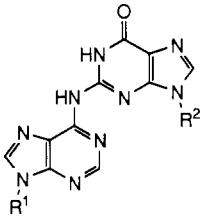
in significant quantities in many cell types, provides yet another initiator of this reaction cascade, having been demonstrated in vitro to lead to DNA deamination.[8] By analogy it is reasonable to expect that nitric oxide will lead to DNA interstrand cross-linking.

This paper details the elucidation of the covalent structure of the thermally- and base-stable interstrand cross-links formed in DNA exposed to nitrous acid. In 1977, Shapiro et al. provided the first candidate structures for the nucleus of this interstrand cross-link.[9-11] Calf thymus DNA was incubated with nitrous acid, and the product was processed with steps including enzymatic hydrolysis and chromatographies. Of those substances isolated, the most relevant here are **1a** (0.06% yield) and **2** (0.01% yield). It was suggested[9] that **1a** might arise by diazotization of a guanyl amino group and attack at the attached carbon, C2, by a second guanyl amino group, with loss of nitrogen.[12] From that work, it was not possible to establish whether **1a** and **2** had arisen from interstrand, intrastrand, or interhelical cross-linking. These substances might even have had their origin in reactions of denatured regions of the calf thymus DNA preparations studied. Using a panel of synthetic oligodeoxynucleotides, we recently investigated an hypothesis implicit in the structure of the more abundant of these two products, **1a**, that deoxyguanosine residues on opposite strands of duplex DNA might be cross-linked with nitrous acid.[13-15] Those studies revealed that heat- and base-stable interstrand linkages from dG-to-dG were in fact formed in highest yield when two deoxyguanosine residues were present in adjacent base pairs and on opposite strands. The duplex sequence 5'-CG was found to cross-link with a 4–25-fold preference over 5'-GC.[13] A molecular mechanics study suggested that the structural nucleus of **1a** might be accommodated at this preferentially cross-linked sequence with minimal structural reorganization.[14]

The central structural question addressed here is whether the dG-to-dG, thermally- and base-stable cross-links formed preferentially at 5'-CG in synthetic oligodeoxynucleotide duplexes are the result of covalent linkage as in **1a**. Described herein are experiments which conclusively demonstrate that the covalent linkage we observe in synthetic interstrand cross-linked oligodeoxynucleotides is the same as that observed by Shapiro et al. in calf thymus DNA and assigned structure **1a**. The data herein lend further support to the formulation of this substance as **1a**.



1a R^1=R^2= 2'-deoxy-ß-D-ribofuranosyl
1b R^1= 2'-deoxy-ß-D-ribofuranosyl; R^2=H
1c R^1=R^2= H



2 R^1=R^2= 2'-deoxy-ß-D-ribofuranosyl

Results and Discussion

Stability of Duplex DNAs. We verified at the outset that the duplex/single strand equilibria of the oligodeoxynucleotides studied herein did in fact reside predominantly on the side of duplex under the conditions of sodium concentration and pH (4.15) used for cross-linking reactions. This was accomplished by determination of the thermal denaturation profiles monitored at 260 nm (Table I). The oligodeoxynucleotides TGCA and TCGA (see Table I) whose cross-linking efficiencies are compared below exhibited very similar melting profiles, with, for our purposes, essentially identical T_m's (half maximal hyperchromicity using baseline correction at high and low temperature extremes[16]) of 40 °C at 563 mM sodium ion concentration, pH 4.15. The oligodeoxynucleotides CCGI and CCIG (see Table I) whose cross-linking reactions are compared below exhibited the similar T_m values of 32 and 31 °C, respectively, at this same pH and sodium ion concentration. Finally, the fully self-complementary oligodeoxynucleotide CCGG (see Table I) melted at 42 °C, some 22 °C higher than the partially self-complementary oligodeoxynucleotide ATCCGGAA, which in the duplex form is predicted to contain two dA–dA mismatched base pairs. The latter, mismatched DNA was specifically chosen as a predominantly single strand analogue of the duplex CCGG DNA.

Optimization of the Yield of Cross-Linked DNA. Initial attempts to cross-link synthetic oligodeoxynucleotides afforded an

(7) Ignaro, L. J. *Ann. Rev. Pharmacol.* **1990**, *30*, 535. See, also: Schuman, E. M.; Madison, D. V. *Science* **1991**, *254*, 1503.
(8) Wink, P. A.; Kasprzak, K. S.; Maragos, C. M.; Elespuru, R. K.; Misra, M.; Dunams, T. M.; Cebula, T. A.; Koch, W. H.; Andrews, A. W.; Allen, J. S.; Keefer, L. K. *Science* **1991**, *254*, 1001.
(9) Shapiro, R.; Dubelman, S.; Feinberg, A. M.; Crain, P. F.; McCloskey, J. A. *J. Am. Chem. Soc.* **1977**, *99*, 302.
(10) Dubelman, S.; Shapiro, R. *Nucleic Acids Res.* **1977**, *4*, 1815.
(11) Dubelman, S. *Analysis of Damaged DNA: A Study of DNA Cross-Linking by Nitrous Acid*; Ph.D. Thesis, New York University, October, 1975.
(12) Zollinger, H. *Acc. Chem. Res.* **1973**, *6*, 335.
(13) Kirchner, J. J.; Hopkins, P. B. *J. Am. Chem. Soc.* **1991**, *113*, 4681.
(14) Kirchner, J. J.; Solomon, M. S.; Hopkins, P. B. In *Structure and Function: Proceedings of the Seventh Conversation in Biomolecular Stereodynamics*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Albany, NY, 1992.
(15) Kirchner, J. J. *Structure and Sequence Specificity of the Nitrous Acid-Induced DNA Interstrand Cross-Link*; Ph.D. Thesis, University of Washington, December, 1991.
(16) Albergo, D. D.; Marky, L. A.; Breslauer, K. J.; Turner, D. H. *Biochemistry* **1981**, *20*, 1409.

ca. 0.1% yield of cross-linked DNA, as measured by Cerenkov counting of bands excised from polyacrylamide gel. Because synthetic DNA is available at modest cost in micromolar quantities and cross-linked DNA is easily separated from residual single strands, this yield was satisfactory for nucleotide resolution studies of sequence specificity, which require only subnanomolar quantities of ^{32}P-labeled, cross-linked DNA.[17-19] Because characterization of the cross-link at atomic resolution would ultimately require quantities at least an order of magnitude larger than this, we explored the impact of concentration of nitrous acid, reaction time, and temperature on the yield of cross-linking.

The self-complementary DNAs TGCA and TCGA (Table I) were independently incubated under a variety of conditions with acidic, aqueous sodium nitrite. The resulting reaction mixtures were ethanol precipitated and analyzed for dG-to-dG cross-linked DNA[20] by denaturing polyacrylamide gel electrophoresis (DPAGE). Quantitation by phosphorimagery afforded the results shown in Figure 1.

The yield of dG-to-dG cross-linked TCGA exposed to sodium nitrite at pH 4.15, 24 °C, for 100 min rose with a roughly first order dependence on sodium nitrite concentration in the range of 12.5–100 mM (Figure 1). Above 100 mM (250 and 500 mM), the yield continued to rise, but with an order in sodium nitrite less than one. Consistent with our previous observations, the yield of dG-to-dG cross-linked TCGA was greater than that in TGCA at all sodium nitrite concentrations studied. First order dependences of diazotization reaction rate[21] and deamination of dC[22] on nitrite concentration have previously been observed under similar experimental conditions.

The yield of dG-to-dG cross-link in TCGA (Table I) exposed to 500 mM sodium nitrite at pH 4.15, 24 °C, likewise rose with time out to the longest time point measured, 5000 min, reaching almost 6% at that time; the yield rose linearly with time from 10 to 1000 min (Figure 1).

The temperature dependence of the yield of dG-to-dG cross-linked product arising from exposure of TCGA and TGCA (Table I) to 500 mM sodium nitrite at pH 4.15 for 100 min was measured between 10 and 65 °C (Figure 1). The yields rose initially with increasing temperature but then fell sharply, being maximal at 25 °C for both TCGA and TGCA. At least qualitatively, this is consistent with an increase in the rate constant for cross-linking with increasing temperature, followed at higher temperatures by an offsetting loss of duplex substrate as the DNA is melted. Some further support for this hypothesis was found in the relative efficiency of cross-linking of CCGG and ATCCGGAA (Table I) at 25 °C. As described above, at the ionic strength of these cross-linking reactions, the former melts well above 25 °C while the latter, which contains two mismatched, dA-to-dA pairs, melts well below 25 °C (see Table I). At this same temperature, the former DNA cross-links with over 30-fold higher efficiency than

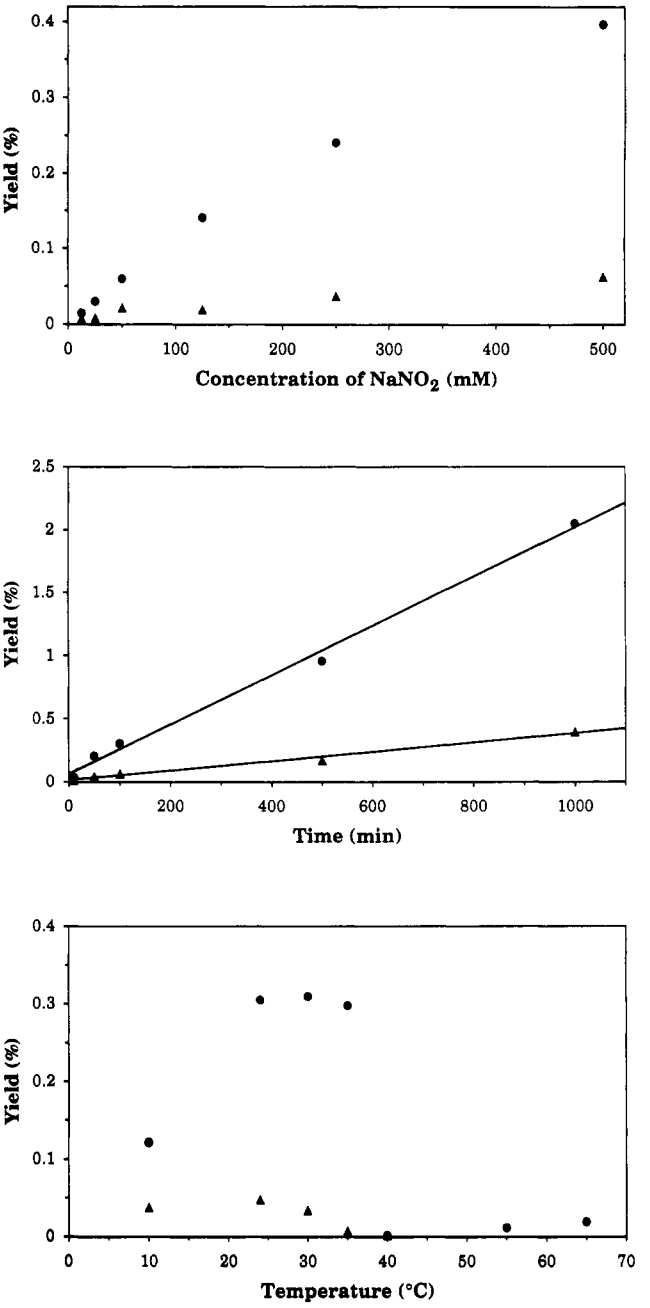


Figure 1. Yield of dG-to-dG interstrand cross-linked DNA as determined by denaturing PAGE as a function of sodium nitrite concentration (upper), time (middle), and temperature (bottom): ●, TCGA; ▲, TGCA (see Table I). See experimental section for further details.

the latter. Regardless of the origin of the effect, for the purpose of yield elevation, there appeared little advantage to use the reaction temperatures different from ambient.

It is noteworthy that the yields of dG-to-dG cross-linked product in both TCGA and TGCA rise with increasing nitrite concentration, with time, and, to a point, with temperature. In all cases, however, the yield rises more rapidly for TCGA. We have seen no evidence for reversion of cross-linked to single strand DNA. Together, these observations suggest that the previously reported preference for dG-to-dG cross-linking at the sequence 5′-CG over 5′-GC is kinetic in origin. Which mechanistic step(s) is responsible remains undefined.

Armed with these empirical observations concerning the yield of cross-linked product, we proceeded to identify the responsible covalent lesion.

Covalent Structure of the Cross-Link. Shapiro's isolation of substance **1a** from nitrous acid-treated calf thymus DNA provides an obvious candidate for the structural nucleus of the interstrand,

(17) Weidner, M. F.; Millard, J. T.; Hopkins, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 9270.
(18) Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. Th.; Raucher, S. *Tetrahedron* **1991**, *47*, 2475.
(19) Millard, J. T.; Weidner, M. F.; Kirchner, J. J.; Ribeiro, S.; Hopkins, P. B. *Nucleic Acids Res.* **1991**, *19*, 1885.
(20) We have observed that all but the most selective interstrand cross-linking reagents react appreciably to cross-link the ends of synthetic DNA duplexes. The resulting "end"-cross-linked duplexes exhibit a higher mobility on denaturing PAGE than the more centrally (in this case dG-to-dG) linked duplexes and are thus readily distinguished.[19] For a given reagent and end-sequence, the extent of end-cross-linking is independent of the nucleotide sequence at the center of the duplex. In this paper, all yields refer to the least mobile electrophoretic product, shown previously to be dG-to-dG cross-linked at 5′-CG sequences.[13,14]
(21) (a) Edwards, J. O.; Abbott, J. R.; Ellison, H. R.; Nyberg, J. *J. Phys. Chem.* **1959**, *63*, 359. (b) Stedman, G. *J. Chem. Soc.* **1960**, 1702.
(22) Frankel, A. D.; Duncan, B. K.; Hartman, P. E. *J. Bacteriol.* **1980**, *142*, 335.
(23) For example: (a) Norman, D.; Live, D.; Sastry, M.; Lipman, R.; Hingerty, B. E.; Tomasz, M.; Broyde, S.; Patel, D. *J. Biochemistry* **1990**, *29*, 2861. (b) Tomic, M. T.; Wemmer, D. E.; Kim, S.-H. *Science* **1987**, *238*, 1722.
(24) Maniatis, T.; Fritsch, D. G.; Sambrook, J. *Molecular Cloning: A Laboratory Approach*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1981.

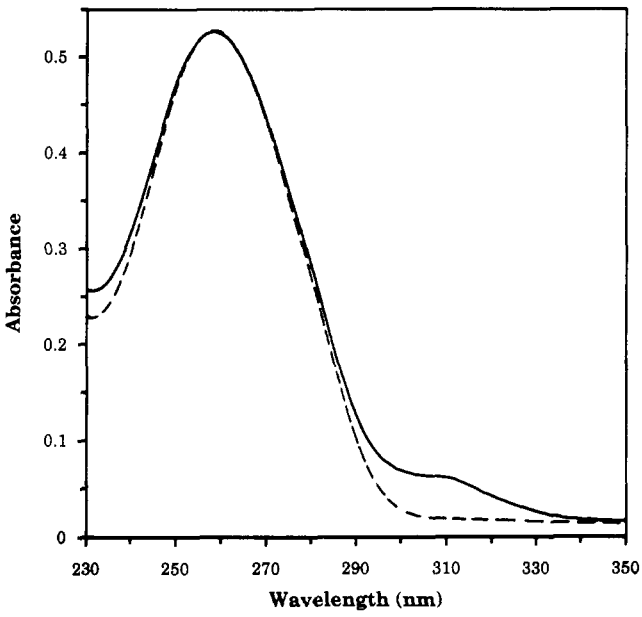


Figure 2. Ultraviolet spectra of interstrand cross-linked DNA. CCGG (Table I) cross-linked dG-to-dG at 5′-CG (—) and uncross-linked (– –).

dG-to-dG cross-link in duplex DNA.[9-11] The mechanistically rational proposal that this connectivity arises by diazotization of the exocyclic amino group of one dG residue, followed by displacement of dinitrogen with the exocyclic amino group of a neighboring dG residue on the opposite strand serving as a nucleophile, necessitates the presence of these exocyclic amino groups for cross-linking. To test this hypothesis, two DNAs, CCGI and CCIG (see Table I) containing the deoxyinosine residue (which bears a hydrogen rather than an amino group at C2) in place of deoxyguanosine were prepared and exposed to nitrous acid. Consistent with the requirement for the amino function at the site of cross-linking, exposure to nitrous acid followed by DPAGE analysis revealed that CCGI cross-linked ca. 60-fold more efficiently than CCIG (Table I). In other words, a dI-to-dI cross-link is not formed to an appreciable extent at the sequence (5′-CI)·(5′-CI).

The UV spectrum of the substance **1a** isolated from nitrous acid-treated calf thymus DNA has been reported[9-11] and exhibits appreciable absorbance of ultraviolet radiation at wavelengths longer than 300 nm, where the common residues of DNA do not absorb significantly. This suggested that if the nucleus **1a** was present in interstrand cross-linked oligodeoxynucleotides, it might be detectable by UV spectroscopy. The DNA CCGG was treated with nitrous acid, and the dG-to-dG cross-linked product was excised and eluted from a denaturing gel. The resulting material was desalted by reverse phase chromatography. The UV spectrum of the resulting material in H_2O displayed spectral density out to 340 nm (Figure 2). This result is again consistent with the extended chromophore of **1a**.

The strongest evidence that **1a** is the nucleus of the interstrand dG-to-dG cross-link came from enzymatic hydrolysis of the sugar-phosphate backbone of cross-linked DNA and, ultimately, isolation and spectroscopic characterization of **1a** itself. Nitrous acid-cross-linked CCGG isolated as described above was exposed to snake venom phosphodiesterase and calf intestinal alkaline phosphatase. HPLC analysis (Figure 3) of the hydrolysate with detection at 260 nm revealed, in order of elution, dC, dI (presumably from deamination of dA), dG, dT, dA, and finally a strongly retained substance which we attribute to the nucleus of the cross-link. Simultaneous detection at 310 nm, where **1a** is known to absorb, revealed a greatly enhanced absorbance of this last peak relative to all those mentioned previously (Figure 3). Quantitation of the released nucleosides using ϵ_{260} revealed a ratio of 2.0(dC):1.1(dG):6.5(dT):7.2(dA+dI), one dG residue short of the calculated (and experimentally verified) starting single strand ratios of 2(dC):2(dG):6(dT):7(dA). Assuming that each mole

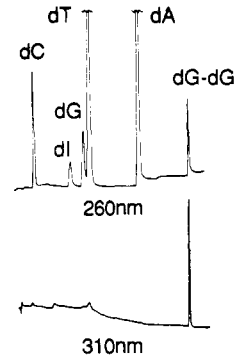


Figure 3. HPLC analysis of enzymatic hydrolysate of interstrand cross-linked CCGG (Table I): upper, detection at 260 nm; lower, detection at 310 nm. (Retention times increase to right.)

of cross-linked duplex contributes one mole of the last-eluting substance, a molar extinction coefficient of 16 800 M^{-1} cm^{-1} at pH 7 can be estimated.

The UV spectra of **1a** isolated by Shapiro et al. from calf thymus DNA vary greatly with pH with respect to line shape and λ_{max} values, and as such serve as a fingerprint for this substance. We repeated these measurements on the strongly retained substance isolated from cross-linked oligodeoxynucleotides. Because sequence specificity was not of foremost concern, this sample was prepared using the synthetic DNA $(CG)_6$ (see Table I). Not only did this sequence afford an almost 10-fold enhancement in the weight percentage of 5′-CG sequences relative to CCGG, it lacks dA residues which would likely limit column loading during HPLC purification of **1a**. Accordingly, synthetic $(CG)_6$ was treated with 1.0 M $NaNO_2$ at pH 4.15 (300 mM NaOAc/HOAc), 25 °C, for 43 h. Enzymatic hydrolysis as described above, followed by HPLC purification afforded ca. 0.3 o.d. (5% yield based on total 5′-CG sequences and the ϵ_{260} reported above) units of the strongly retained substance, which was in two HPLC solvent systems indistinguishable from that obtained from hydrolysis of cross-linked CCGG. The resulting spectra strongly suggest the identity of the substances isolated from calf thymus and synthetic DNA (Figure 4).

The mass spectrum of the strongly retained substance isolated from hydrolyzed CCGG was determined using electrospray ionization (Figure 5) in positive ion mode. Adducts of a single proton, sodium ion, and potassium ion were found at masses corresponding to the formula of **1a**, as were ions corresponding to fragmentation products **1b** and **1c**. Shapiro et al. have previously reported the same fragmentation pattern for the pertrimethylsilylated derivative **1a** isolated from calf thymus DNA.[9] They have noted that these data nicely accommodate structure **1a** and preclude linkage through the deoxyribose moieties. Subsequently, the electrospray mass spectrum of the relatively more easily accessible, strongly retained substance isolated from $(CG)_6$ was determined. That spectrum was entirely confirmatory of the above conclusions.

The proton NMR spectrum of the strongly retained substance was determined at 500 MHz in D_2O using a 7-o.d. sample isolated from cross-linking of 320 o.d. of $(CG)_6$ essentially as described above. The spectrum was fully consistent with the structure **1a**. The number and location (chemical shifts) of resonances matched those reported by Shapiro et al.[9,11] obtained from a spectrum recorded at 100 MHz. The relatively increased sensitivity and resolution of the current measurement provide further support for the C_2 symmetry of **1a**, revealing a single set of resonances for the two deoxyribose residues and two aromatic nonexchangeable hydrogens.

Conclusion

We have previously demonstrated that the thermally- and base-stable cross-links which are formed on exposure of highly polymerized DNA of undefined sequence (e.g., calf thymus DNA) to acidic sodium nitrite can also be formed in the better-defined setting of synthetic oligodeoxynucleotide duplexes.[13-15] Using a panel of synthetic DNAs, we showed that these stable cross-links

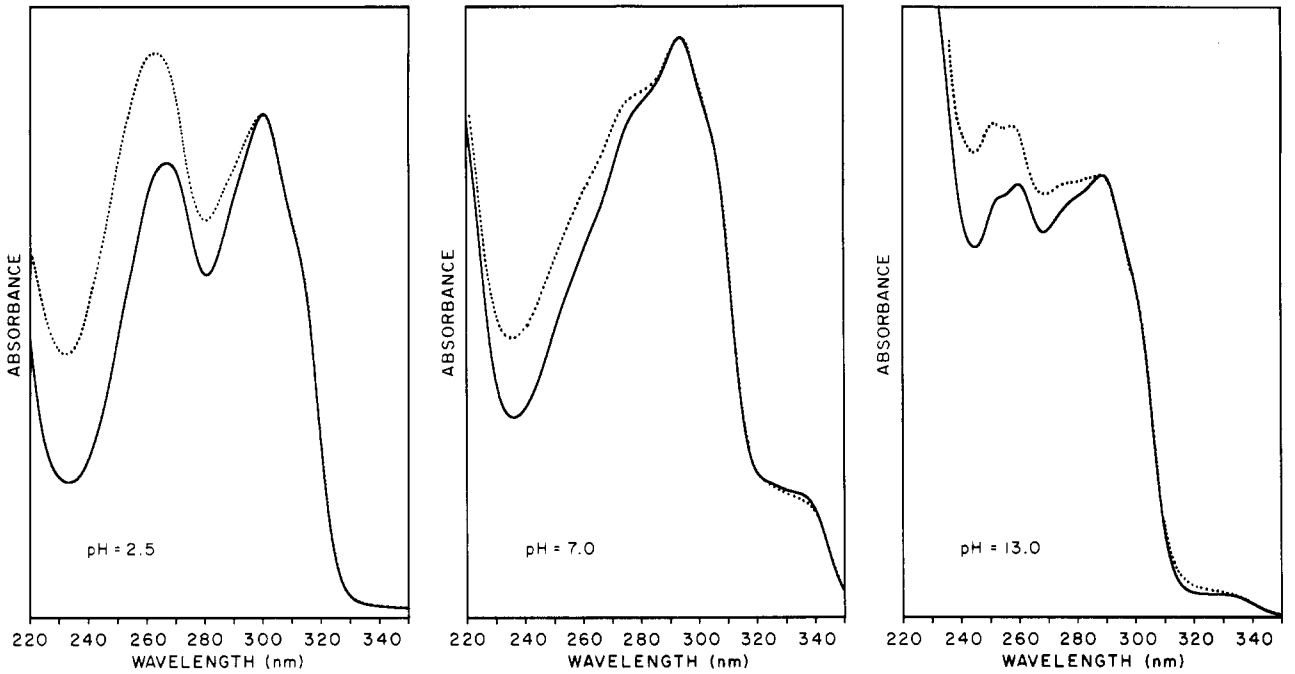
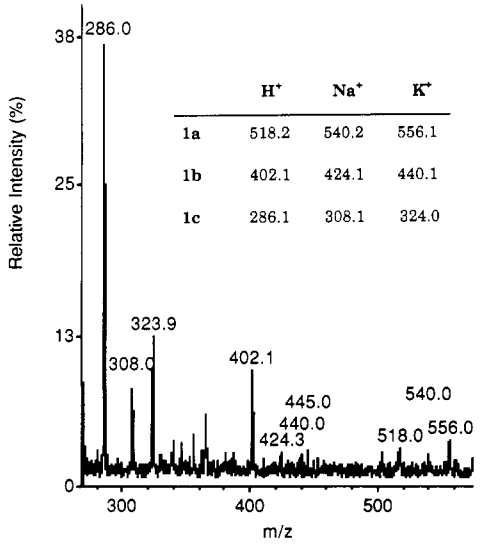


Figure 4. Ultraviolet spectra of **1a** isolated from nitrous acid cross-linked calf thymus DNA (--) and $(CG)_6$ (—) at pH 2.5 (left), 7.0 (center), and 13.0 (right). Data for cross-linked calf thymus DNA are from ref 11.



	H^+	Na^+	K^+
1a	518.2	540.2	556.1
1b	402.1	424.1	440.1
1c	286.1	308.1	324.0

Figure 5. Electrospray ionization mass spectrum of **1a** obtained from enzymatic hydrolysis of interstrand cross-linked CCGG (Table I).

are formed in highest yield in DNAs containing at least two adjacent dC·dG pairs. The nucleotide sequence 5′-GC was cross-linked somewhat less efficiently than 5′-CG. For the latter sequence, it was conclusively demonstrated that deoxyguanosine residues on opposite strands were linked. These *nucleotide* resolution studies left unresolved the question of the actual covalent connectivity. The present study remedies that situation. We demonstrate herein that the dG-to-dG cross-link formed in synthetic DNAs at the nucleotide sequence 5′-CG has as its core structure the same substance isolated from nitrous acid-treated calf thymus DNA and assigned structure **1a** by Shapiro et al.[9] The data reported herein, particularly the 500-MHz ^1H NMR spectrum of **1a**, lend further support to this structural assignment.

Establishment of the structure of the dG-to-dG interstrand cross-link as is implied by substance **1a** sets the stage for addressing several remaining questions concerning nitrous acid's reactions with DNA. One issue is that of the mechanistic origin of the preference for cross-linking at the duplex sequence 5′-CG over 5′-GC. We have previously noted that this might result from the minimal structural reorganization required in the cross-link forming transition state which links N^2 of dG on one strand to C2 of a diazotized dG on the opposite strand.[13-15] Equally well accounting for this sequence specificity would be preferential diazotization at 5′-CG or preferential hydrolysis of diazotized dG at 5′-GC.[13] Having shown that the exocyclic amino groups of dG are in fact involved in cross-linking, relevant experiments can now be pursued. A second issue of interest is the conformation of duplex DNA containing this cross-link. Molecular mechanics studies indicate that dG residues at the sequence 5′-CG can share a common N^2 atom, as in **1a**, with relatively minor structural distortion.[14] It may be that rational chemical synthesis, rather than direct treatment of DNA with nitrous acid, will be the preferred method for obtaining quantities of cross-linked DNA sufficient for conducting appropriate conformational studies.[23] Knowledge of the structure of this cross-link clearly benefits efforts to secure samples by this means.

It is appropriate to comment, in closing, on the failure of our studies in synthetic oligodeoxynucleotides to reveal any substantial role for dG-to-dA interstrand cross-links, as might have been inferred from the isolation of **2** in quantities similar to **1a** from nitrous acid-treated calf thymus DNA.[9] Several explanations are worthy of consideration. First, we have studied exclusively the population of *interstrand* cross-linked DNA molecules following nitrous acid treatment: If dG-to-dA cross-links are exclusively (or nearly so) *intrastrand*, they would not have been detected. Second, it should be noted that we have not exhaustively investigated the possible sequences for dG-to-dA interstrand cross-linking, although we have studied DNAs containing the duplex sequences 5′-TG and 5′-GT in which dA and dG residues are close to another in space on opposite strands.[13] Interstrand cross-linking at those two sequences similar in efficiency to that of dG-to-dG at 5′-CG is ruled out by those experiments. Perhaps other, as yet unexplored sequences are cross-linked dG-to-dA. Third, the linkage found in **2** indicates the participation of exocyclic amino groups of dA and dG which reside in opposite grooves of A and B DNA. This suggests the possibility that denatured regions of commercial calf thymus DNA may be critical for the formation of **2** because significant reorganization of A or B DNA would necessarily accompany linking of dG-to-dA as in **2**. This runs opposite current evidence that efficient interstrand cross-linking of duplex DNA strands can be accounted for with only limited structural reorganization. Regions in some conformation other than A or B (e.g., denatured) would thus be more likely to afford dG-to-dA interstrand cross-links.

Experimental Section

Materials and Methods. Materials and their sources were as follows: DNA synthesis reagents, Applied Biosystems; [γ-^{32}P] ATP, T4 polynucleotide kinase, and alkaline phosphatase (calf intestinal), Amersham; phosphodiesterase I (Crotalus adamanteus venom), Pharmacia; sodium nitrite, Aldrich. Water was purified on a Millipore Milli-Q deionizer. All other reagents were commercial and used as received. Samples were concentrated on a Savant Speed Vac concentrator. Loading buffer was 90% aqueous deionized formamide containing 10 mM Tris (pH 7.5), 0.1% xylene cyanol, and 0.1 mM sodium EDTA. TE buffer was 10 mM aqueous Tris (pH 7.5), 1.0 mM EDTA. For autoradiography, gels were dried (Bio-Rad Model 583) onto Whatman 3MM paper and autoradiographed on Kodak XAR-5 film. Cerenkov counting was performed on a Packard 2000 CA Tri Carb scintillation analyzer with a window setting 1–1000. Samples consisted of excised DPAGE bands in a 1.5-mL microfuge tube placed, in turn, in a 20-mL glass scintillation vial. Samples were counted for 5 min. Phosphorimaging used a Molecular Dynamics 400A PhosphorImager. All scanning operations, data display, and analysis were performed using Molecular Dynamics' ImageQuant software operating on Intel 80386 or 80286 microprocessors. UV spectra were measured on a Hewlett-Packard 8450A or 8452A spectrophotometer. The resolution of spectra from the latter was enhanced from 4 to 0.5 nm using the spline function of the software package 386-MATLAB (MathWorks, Inc.; Natick, MA). HPLC analytical and preparative separations were performed on an Alltech, 5 Å, C18, 250 mm × 4.6 mm or an Alltech Macrosphere 300, C18, 7μ, 250 mm × 10 mm column, using SSI 200B pumps controlled by an SSI controller, and sequential SSI 500 UV/vis (output to both an HP 3390A electronic integrator and a Linear Model 255/MM recorder) and Waters Lambda-Max Model 481LC (output to a Linear Model 156 recorder) detectors. Solvent gradients were run at 1 mL/min as follows: gradient A, solvent A, 100 mM triethylammonium acetate (pH 7.0); solvent B, acetonitrile; isocratic 95% A for 2 min, 13 min linear gradient to 70% A, 15 min linear gradient to 60% A, then a 10 min linear gradient to initial conditions; gradient B, solvent A, water; solvent B, 50% aqueous methanol; isocratic 99% A for 1 min, 10 min linear gradient to 49% A, isocratic for 25 min, then a 5 min linear gradient to initial conditions; gradient C, solvent A, 100 mM aqueous ammonium formate; solvent B, 50% 100 mM aqueous ammonium formate/methanol; isocratic 99% A for 10 min, a 10 min linear gradient to 95% A, isocratic for 5 min, then 15 min linear gradient to 30% A, isocratic for 10 min, 15 min linear gradient to initial conditions; gradient D, solvent A, water; solvent B, acetonitrile; isocratic 99% A for 2 min, 13 min linear gradient to 70% A, 15 min linear gradient to 60% A, 10 min linear gradient to initial conditions.

Preparation and Radiolabeling of DNA.[24] Oligodeoxynucleotides were synthesized (Applied Biosystems Model 380A) using the phosphoramidite method on a 1 μmol scale. Except for $(CG)_6$ (see below) deoxyoligonucleotides were purified by DPAGE (ca. 80 o.d. crude DNA, 20% polyacrylamide, 19:1 acrylamide/bisacrylamide, 8 M urea, 1.5 mm thick, 14 × 16 cm, using a five-toothed comb) run until the xylene cyanol dye had traveled 9 cm from the origin. DNA was visualized by UV shadowing and eluted from crushed gel slices by incubation at 37 °C in 0.5 M NH_4OAc/1 mM sodium EDTA for 16 h. The eluant was passed through a Waters Sep-Pak C_{18} cartridge which had previously been washed with 10 mL of CH_3CN followed by 10 mL of water. The Sep-Pak was then sequentially eluted with ca. (1) 10 mL of 10 mM aqueous NH_4OAc, (2) 10 mL of water, and (3) 3 mL of 25% aqueous CH_3CN. DNA was recovered by concentration of the acetonitrile/water eluant. DNA was radiolabeled at the 5′-terminus using 0.5 o.d. of DNA, (20 nmol of base pairs) in 20 μL of kinase buffer [50 mM Tris (pH 7.5), 10 mM $MgCl_2$, 0.1 mM spermidine, 5 mM DTT, 0.1 mM EDTA], 30 μCi [γ-^{32}P] ATP and 10 units of T4 polynucleotide kinase at 37 °C for 10 min. Radiolabeling was stopped by addition of 10 μL of 3 M aqueous sodium acetate (pH 5.2), followed by 1 mL of ethanol at −20 °C. After 0.5 h at −78 °C, the resulting precipitate was collected by centrifugation, redissolved in 100 μL of 0.3 M sodium acetate (pH 5.2), and precipitated by addition of 1 mL of ethanol at −20 °C. After 0.5 h at −78 °C, the precipitate was collected by centrifugation and used in the following experiments. Crude $(CG)_6$ (85 o.d.) was purified by dissolving in 150 μL of water and adding 1.3 mL of −20 °C ethanol. After 0.25 h at −78 °C, the resulting precipitate of DNA was pelleted by centrifugation at 4 °C for 0.25 h, and the supernatant was discarded.

UV Monitored Thermal Denaturation. Samples were prepared by dissolving 2.5 o.d. of DNA (ca. 100 nmol of base pairs; 5′- and 3′-termini as free hydroxyls) in 500 μL of 0.3 M sodium acetate buffer (pH 4.15, 63 mM Na^+). In some runs, the total Na^+ concentration was raised to 563 mM using sodium chloride. The samples were heated to 90 °C for 10 min and cooled to 25 °C over 3 h, followed by purging with helium for 5 min. UV monitored thermal denaturation was performed in a 0.1-cm path length cell in a locally assembled apparatus consisting of a Physitemp type IT-18 thermocouple, a Perkin-Elmer Lambda 3A UV/vis spectrophotometer, and a Techne water bath (Model TU-16D), all interfaced to an IBM PC. The temperature was increased at a rate of 0.5 °C/min. The T_m value reported in Table I is the temperature at the midpoint of the baseline corrected spectroscopic transition.

Yield Optimization of Cross-Linking. Effect of Nitrite Concentration. Synthetic, radiolabeled DNA (0.5 o.d.; 20 nmol bp) in 0.3 M aqueous sodium acetate buffer (pH 4.15) at 25 °C was treated with an appropriate volume (1–10 μL) of 5.0 M or 0.50 M aqueous sodium nitrite such that the total nitrite concentration ranged from 12.5 to 500 mM in a total volume of 100 μL. After 100 min at 25 °C, the DNA was precipitated by addition of 1.0 mL of −20 °C ethanol and cooled to −78 °C for 20 min. The cool suspension was centrifuged for 15 min (0 °C), and the supernatant was removed and discarded. The pellet was analyzed by DPAGE as described below.

Effect of Time. Reaction mixtures were prepared as described above in the nitrite concentration experiment and were initiated by addition of 10 μL of 5.0 M sodium nitrite at times such that reactions which had proceeded for 10, 50, 100, 500, 1000, and 5000 min could all be stopped simultaneously by precipitation of the DNA as described above and analyzed by DPAGE as described below.

Effect of Temperature. Reaction mixtures were prepared as described in the study of nitrite concentration except that, prior to addition of 10 μL of 5.0 M aqueous sodium nitrite, they were equilibrated for 10 min at 10, 24, 30, 35, 49, 55, or 65 °C. Following addition of sodium nitrite, the reactions were allowed to proceed for 100 min at the specified temperature, then worked up as described above, and analyzed by DPAGE as described below.

DPAGE Analysis of Cross-Link Yield. Pelleted DNA was dissolved in 10 μL of loading buffer [90% deionized formamide, 10 mM Tris (pH 7.5), 0.1% xylene cyanol, 0.1 mM EDTA]. The samples were denatured at 90 °C for 4 min and chilled on ice prior to 20% PAGE (19:1 acrylamide/bisacrylamide, 50% urea, 0.35 mm thick, 41 × 37 cm). Gels were prepared as follows: 19 g of acrylamide, 1 g of bisacrylamide, and 50 g of urea were dissolved in 10 mL of 10 × TBE[24] and 20 mL of water. The volume was brought to 100 mL with water, and 350 μL of 20% aqueous ammonium persulfate was added prior to filtering through Whatman no. 5 filter paper. To the solution was added 25 μL of TEMED[24] to induce polymerization, and the gel was poured. After 1 h, electrophoresis on a Hoeffer thermojacketed Poker Face gel stand was performed until the gel reached 65 °C. Samples were loaded using flat sequencing tips (Marsh Biomedical) on a Rainin P-200 Pipetman, and the gel was run at 75 W with a Bio-Rad Model 3000 XI power supply until the xylene cyanol dye traveled 12–14 cm (~4 h). The gel was transferred onto filter paper covered with Saran Wrap and dried for 1 h followed by autoradiography or phosphorimagery. Audioradiography was used to visualize the single-strand and cross-linked DNA. For Cerenkov counting, the cross-linked material (roughly half the mobility of the corresponding single strand) and the single strand DNA were cut from the gel and counted.

Enzymatic Hydrolysis and Quantitation of Nucleosides by HPLC. Cross-linked (0.1 o.d.) or synthetic CCGG (0.5 o.d.) in 30 μL of aqueous 10 mM $MgCl_2$, 50 mM Tris (pH 8.9) was treated with 1 μL (10 units) of alkaline phosphatase and 3 μL (2 units) of phosphodiesterase I at 37 °C for 8 h. HPLC analysis was carried out on 10 μL of the above reaction mixture using gradient A. Peaks were identified by comparison of retention times to those of authentic, commercial samples. Quantitation was based on the peak area ratios obtained from a standard, equimolar mixture prepared by weight of dC, dI, dG, dT, and dA at 260 nm which were as follows: dC,1:dI,1.04:dG,1.67:dT,1.15:dA,1.98. Using these response factors, the uncross-linked DNA analyzed as 2.0-(dC):2.1(dG):6.6(dT):7.2(dA) (calcd 2:2:6:7); cross-linked DNA analyzed as 2.0(dC):1.1(dG):6.5(dT):7.2(dA) (calcd 2:1:6:7).

Preparative DNA Cross-Linking. CCGG. 5′-^{32}P-Radiolabeled CCGG (5 o.d.; 200 nmol base pairs) was placed in each of 4 microfuge tubes. To each tube was added 900 μL of 0.3 M aqueous sodium acetate (pH 4.15), followed by treatment with 100 μL of 5.0 M aqueous sodium nitrite (total $[Na^+]$ = 562 mM) at 25 °C. After vortexing, the mixtures were allowed to stand 100 min at 25 °C. Each of the four mixtures was partitioned into two equal volumes (a total of eight tubes), and to each was added 1 mL of −20 °C ethanol, followed by cooling to −78 °C for 20 min and then centrifugation for 15 min (0 °C). The supernatant was removed and discarded. To four of the reaction tubes, 50 μL of water was added, and then each was transferred to one of the microfuge tubes that did not have water added, to yield a total of four reaction tubes. The samples were individually concentrated to dryness, dissolved in 10 μL of loading buffer and analyzed by DPAGE as described above with the exception that instead of transferring the gel to filter paper, one glass plate was removed from the gel, and the gel was covered with Saran Wrap followed by autoradiography to visualize the cross-linked DNA.

The film was then aligned with the gel and gel slices containing cross-linked DNA were excised. Gel slices were crushed in a microfuge tube, suspended in ca. 1 mL of TE buffer, and allowed to stand 12 h at 37 °C. The DNA-containing supernatant was removed by pipette. The DNA was loaded onto a Sep-Pak and treated as described above under **Preparation and Radiolabeling of DNA**. After concentration of the acetonitrile/water eluant to dryness, the residue was dissolved in 800 μL of water, and the UV spectrum was recorded.

$(CG)_6$. DNA (20 o.d., 800 nmol base pairs) was dissolved in 560 μL of water and treated sequentially with 80 μL of 3.0 M aqueous sodium acetate (pH 4.15) and 160 μL of 5.0 M aqueous sodium nitrite. The mixture was vortexed and allowed to stand at 25 °C for 43 h. The mixture was divided into eight 100-μL portions, from each of which the DNA was ethanol-precipitated as described above.

Isolation and Characterization of 1a from CCGG. Cross-linked CCGG (0.4 o.d.) in 30 μL of aqueous 10 mM $MgCl_2$, 50 mM Tris (pH 8.9) was treated with 1 μL (10 units) of alkaline phosphatase and 3 μL (2 units) of phosphodiesterase I at 37 °C for 8 h. The substance **1a** was isolated by preparative HPLC using gradient B. The most strongly retained peak, with a retention time of 33 min, was collected and concentrated in a speed-vac. The sample was prepared for electrospray MS by sequential addition of 1 μL of formic acid and 20 μL of 1:1 water/methanol. The electrospray was run with an inlet voltage of 250 V and a needle voltage of 5000 V; the sample was injected at 3 μL/min. The reported spectrum is an average of 20 scans.

Isolation of 1a from $(CG)_6$. UV and MS Sample. Crude-cross-linked $(CG)_6$ (seven eighths of the material described above) was dissolved in 75 μL of aqueous 10 mM $MgCl_2$, 50 mM Tris (pH 8.9) and treated with 2 μL (20 units) of alkaline phosphatase and 6 μL (4 units) of phosphodiesterase I at 37 °C for 8 h. The reaction was diluted with 27 μL of water to bring the total volume to 60 μL. The substance **1a** was isolated by preparative HPLC using gradient A. The most strongly retained substance, retention time 17.2 min, was collected and concentrated to dryness.

UV Spectra of 1a. The dried DNA was dissolved in 600 μL of water, and the pH was adjusted to 7.0 using 2.5% aqueous phosphoric acid. The UV spectrum was determined relative to a blank sample prepared identically from analyte-free HPLC eluant. The blank and the cross-linked samples were adjusted to pH 2.5 using ca. 10 μL of 2.5% aqueous phosphoric acid, and the UV spectrum was recorded. Blank and cross-linked samples were adjusted to pH 13.0 using concentrated ammonia and aqueous NaOH. The samples were finally transferred back to microfuge tubes and adjusted to pH 7.0 using concentrated aqueous phosphoric acid followed by concentration to dryness. Each spectrum of **1a** from $(CG)_6$ in Figure 4 was scaled vertically such that $A_{\lambda max} - A_{350}$ was identical to that reported for **1a** from calf thymus DNA.[11]

Mass Spectrum of 1a from $(CG)_6$. Following determination of the UV spectra, the sample was desalted by dissolution in 70 μL of water and preparative HPLC using gradient D. The peak with a retention time of 6.5 min was collected and concentrated to dryness. The sample was prepared for electrospray MS by addition of 5 μL of formic acid followed by 100 μL of 1:1 water/methanol. The electrospray mass spectrum was determined under the same conditions as described above, ESMS: *m/e* 584 (**1a** + 3Na − 2H)$^+$; 562 (**1a** + 2Na − H)$^+$; 556 (**1a** + K)$^+$; 540 (**1a** + Na)$^+$; 518 (**1a** + H)$^+$; 468 (**1b** + 3Na − 2H)$^+$; 446 (**1b** + 2Na − H)$^+$; 424 (**1b** + Na)$^+$; 402 (**1b** + H)$^+$; 352 (**1c** + 3Na − 2H)$^+$; 330 (**1c** + 2Na − H)$^+$; 324 (**1c** + K)$^+$; 308 (**1c** + Na)$^+$; 286 (**1c** + H)$^+$.

NMR Sample. DNA (320 o.d., 13 μmol base pairs) was dissolved in 700 μL of H_2O and divided equally among 3 microfuge tubes. To each tube was added sequentially 50 μL of 3.0 M aqueous sodium acetate (pH 4.15) and 100 μL of 5 M $NaNO_2$. After vortexing, the solutions were allowed to stand for 47.5 h at 25 °C. Each of the three solutions was partitioned into four equal volumes (a total of 12 tubes), from each of which the DNA was ethanol-precipitated as described above. To six of the tubes was added 100 μL of H_2O and this was in turn, transferred to one of the dry samples. The resulting six samples were individually concentrated to dryness.

To each of these six tubes was added 75 μL of aqueous 10 mM $MgCl_2$, 50 mM Tris (pH 8.9) followed by 2 μL (20 units) of alkaline phosphatase and 6 μL (4 units) of phosphodiesterase I. The samples were then incubated at 37 °C for 8 h. The substance **1a** was isolated by HPLC using gradient A except at a flow rate of 2 mL/min and with solvent A: 10 mM NH_4HCO_2 on the 250 mm × 10 mm column.

Substance **1a**, the most strongly retained substance, retention time ca. 23 min, was collected and concentrated to dryness to give 7 OD (0.4 μmol based on ϵ = 16 800 M^{-1} cm^{-1}; 6% yield). The sample was twice concentrated from 99.96 atom percent D_2O and then dissolved in 0.4 mL of D_2O (99.96%): ^1H NMR (D_2O, 500 MHz, DSS as an external standard) δ 2.60 (2 H, ddd, *J* = 4, 7, 14 Hz, 2′ or 2″), 2.87 (2 H, ddd, *J* = 7, 7, 14 Hz, 2′ or 2″), 3.74 (2 H, dd, *J* = 4, 12 Hz, 5′ or 5″), 3.79 (2 H, dd, *J* = 5, 12 Hz, 5′ or 5″), 4.12 (2 H, ddd, *J* = 4, 4, 5 Hz, 4′), 4.69 (2 H, ddd, *J* = 4, 4, 6 Hz, 3′), 6.35 (2 H, dd, *J* = 7, 7 Hz, 1′), 7.98 (2 H, s, 8).

HPLC Comparison of 1a from CCGG and $(CG)_6$. **1a** from CCGG and $(CG)_6$ were shown to coelute using gradients B and C.

Acknowledgment. This work was supported by the NIH (AG00417, GM35466) and NSF (DIR-822099 and CHE-9019129). Phosphorimaging was performed by the Phosphor-Imager Facility, Markey Molecular Medicine Center, UW. P.B.H. is a Sloan Fellow and Cope Scholar. We thank Professor R. Shapiro for providing relevant spectroscopic data and for his encouragement, E. Hustedt for writing the software used in measuring and analyzing the DNA melting curves, and K. Fincher, M. Gibson, S. McKenzie, S. Ribeiro, J. Roe, and L. Rose for technical assistance.

Supplementary Material Available: UV-monitored thermal denaturation profiles for deoxyoligonucleotides used in this study and the mass and ^1H NMR spectra of **1a** isolated from $(CG)_6$ (9 pages). Ordering information is given on any current masthead page.