VOLUME 115, NUMBER 7
APRIL 7, 1993

JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

Covalent Structure of a Nitrogen Mustard-Induced DNA Interstrand Cross-Link: An N^7-to-N^7 Linkage of Deoxyguanosine Residues at the Duplex Sequence 5′-d(GNC)

Stacia M. Rink,[†] Marjorie S. Solomon,[†] Matthew J. Taylor,[†] Sharanabasava B. Rajur,[‡] Larry W. McLaughlin,[‡] and Paul B. Hopkins*,[†]

Contribution from the Departments of Chemistry, University of Washington, Seattle, Washington 98195, and Boston College, Chestnut Hill, Massachusetts 02167

Received September 18, 1992

Abstract: The covalent structure of the major product of the interstrand cross-linking reaction of mechlorethamine [bis(2-chloroethyl)methylamine, nitrogen mustard] with duplex DNA was determined using synthetic oligodeoxyribonucleotide duplexes. Analysis of the mobility of cross-linked products by denaturing polyacrylamide gel electrophoresis (DPAGE) revealed the reaction to provide intrahelical, interstrand cross-links. These cross-links were cleaved on exposure to aqueous piperidine at deoxyguanosine residues, suggesting N^7 of deoxyguanosine residues as the site of alkylation on DNA. The resulting fragment sizes were consistent with the predominant cross-links bridging deoxyguanosine residues at the sequence 5′-d(GNC). A DNA duplex in which one of the nucleophilic N^7 atoms of deoxyguanosine implicated in cross-link formation was replaced with a methine (CH) was cross-linked with greatly reduced efficiency by mechlorethamine. Preparative HPLC of hydrolyzed interstrand cross-linked DNA afforded a conjugate of two guanine residues with a single nitrogen mustard molecule, which was identified by spectroscopic and chromatographic comparison to an authentic sample. This study provides the first unequivocal demonstration of the covalent connectivity of mechlorethamine–interstrand cross-linked DNA.

Introduction

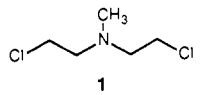
Mechlorethamine [bis(2-chloroethyl)methylamine] (**1**) was the first clinically useful antitumor substance and is among the simplest members of the nitrogen mustard family.[1] In an attempt to account for the general observation that bifunctional electrophiles are considerably more toxic than their monofunctional counterparts, the hypothesis was put forth over 40 years ago that "two groups are required to permit the molecule to react at two distinct points, lying either on a single surface or fibre or, more especially, on two contiguous fibres."[2] Nitrogen mustards were, in fact, subsequently demonstrated to create interstrand cross-links in duplex DNA.[3] It has been speculated that DNA interstrand cross-links might exert a lethal influence on cells by disrupting replication and transcription.[4] This hypothesis is supported by numerous reports that show a correlation of DNA cross-linking and cytotoxicity.[5] This paper addresses the question of the covalent structure of the mechlorethamine-induced DNA interstrand cross-link.

Geiduschek reported in 1961 that mechlorethamine-treated DNA possesses an anomalously high rate of renaturation, which he interpreted as evidence for the presence of interstrand cross-links that nucleate renaturation.[3] Brookes and Lawley reported in the following year that conjugate **2**, in which two guanine residues are bridged through their N^7 atoms by a single

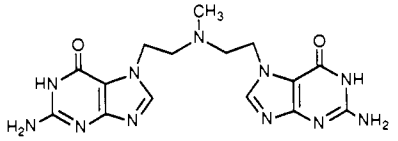
[†] University of Washington.
[‡] Boston College.

(1) (a) Gilman, A.; Philips, F. S. *Science* **1946**, *103*, 409. (b) Pratt, W. B.; Ruddon, R. W. *The Anticancer Drugs*; Oxford University Press: New York, 1979.
(2) Goldacre, R. J.; Loveless, A.; Ross, W. C. J. *Nature* **1949**, *163*, 667.
(3) Geiduschek, E. P. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 950.
(4) Kohn, K. W. *Methods Cancer Res.* **1979**, *16*, 291.

0002-7863/93/1515-2551$04.00/0



1



2

mechlorethamine residue, could be prepared by treatment of GMP with mechlorethamine.[6] Importantly, the same substance was isolated from mechlorethamine-treated yeast RNA.[7] Whether this substance arose from interstrand, intrastrand, or even interhelical cross-links was not directly demonstrated. On the basis of Watson and Crick's model of double helical DNA and the assumption that **2** was derived from an interstrand cross-link, it was proposed that the spacing of N^7 atoms in DNA at the nucleotide sequence 5′-d(GC) was well suited to interstrand cross-linking.[7] This conclusion has been confirmed by computer modeling of a closely related cross-linking agent.[8] This suggestion stood for some 30 years, until two laboratories independently demonstrated that in vitro it is the distal deoxyguanosine residues in the duplex nucleotide sequence 5′-d(GNC), N = A, C, G, or T, which are in fact preferentially cross-linked by mechlorethamine in small, synthetic DNA fragments.[9-11] This finding was unexpected, because the tether which links the N^7 atoms of the guaninyl substituents of **2** is several angstroms too short to accommodate the spacing of N^7 atoms at the duplex DNA sequence 5′-d(GNC). This clearly implies that distortion of the double helix must accompany formation of such a cross-link. This distortion might be important in the eventual understanding of the mechanistic origin of this sequence preference, the ground state conformation of the cross-linked DNA, and the recognition of this lesion by repair enzymes.

An alternative to the conformational distortion necessitated by incorporation of substructure **2** at 5′-d(GNC) is the possibility that **2** is actually *not* the covalent lesion responsible for cross-linking. This suggestion deserves consideration for the following reasons. Oligomerization of mechlorethamine on the time scale of minutes to hours is well-known[12] for aqueous solutions of mechlorethamine similar in concentration (100 mM) to those used in the in vitro cross-linking reactions which led to the sequence specificity assignment (2.5 to 250 mM).[9-11] Inspection of molecular models suggests that a *dimer* of mechlorethamine could bridge the N^7 atoms of deoxyguanosine residues at 5′-d(GNC) in B-DNA without appreciable distortion of the biopolymer. A chemically plausible sequence of events is thus illustrated in Figure 1: (a) alkylation of N^7 by an aziridinium ion derived from mechlorethamine to yield a relatively long lived monoadduct (**3**); (b) capture by monoadduct **3** of a second aziridinium ion from solution to yield **5** rather than cyclization with distortion to form **4**; and (c) reaction with the opposing strand to yield **6**, an interstrand cross-linked DNA free of significant distortion of the biopolymer. Isolation and characterization of the lesion responsible for cross-linking at 5′-d(GNC) is thus critical to resolving this ambiguity. We report herein that the intrahelical, interstrand cross-link now known to occur preferentially at 5′-d(GNC)[9-11] in duplex DNA is, as proposed by Brookes and Lawley,[7] comprised of the substructure **2**.

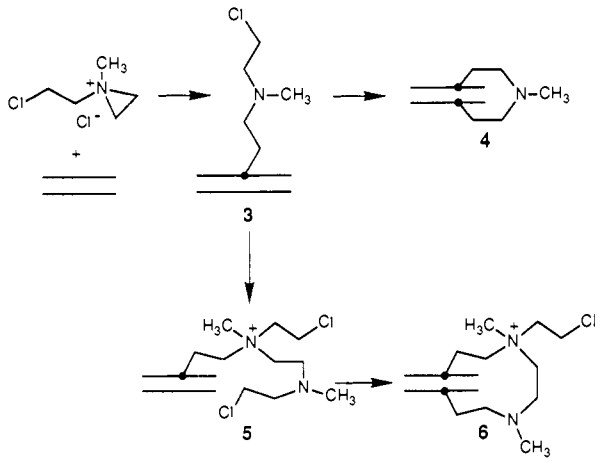


Figure 1. Hypothetical scheme by which a monoadduct of mechlorethamine with DNA (**3**) could progress directly to a conformationally distorted interstrand cross-linked DNA (**4**) or capture an aziridinium ion from solution and progress to a conformationally unperturbed interstrand cross-linked DNA (**6**).

Results and Discussion

Conditions for Interstrand Cross-Linking. Conditions which maximized the yield of structurally homogeneous, interstrand cross-linked DNA were sought by variation of the reaction temperature (0–37 °C), reaction time (1–9 h), and mechlorethamine concentration (0.25–1.0 mM). A 5–10% yield of interstrand cross-linked DNA was obtained under the optimized reaction conditions: DNA (ca. 0.1 mM in duplex) at pH 8.0 (40 mM sodium cacodylate) at 37 °C was exposed to 0.40 mM mechlorethamine for 3 h. DPAGE analysis revealed three major products in a 65:25:10 ratio (see Figure 4). The first two of these were tentatively assigned to be unchanged single strands and monoalkylated single strands, respectively, and were not studied further. The minor, least electrophoretically mobile component of the ternary mixture was assigned as interstrand cross-linked DNA.

Interstrand Cross-Links Are Intrahelical. To demonstrate that the interstrand cross-links studied herein resulted from an intrahelical reaction, that is, reaction within a single helix linking the two strands, rather than an interhelical reaction linking two helices to one another or a reaction joining non-associated single strands, we cross-linked a ca. 1:1 molar mixture of two DNA duplexes which differed in length, one 12 base pairs (GGGCCC-12) and the other 20 (GGGCCC-20). We anticipated that the two intrahelical cross-linked products (24 nucleotides and 40 nucleotides, respectively) would each show distinct DPAGE mobility from the product of interhelical reaction consisting of a 12-mer joined to a 20-mer (32 nucleotides) (see Figure 2). In

(5) (a) Chun, E. H. L.; Gonzales, L.; Lewis, F. S.; Jones, J.; Ruiman, R. J. *Cancer Res.* **1969**, *29*, 1184. (b) Zwelling, L. A.; Michaels, S.; Schwartz, H.; Dobson, P. P.; Kohn, K. W. *Cancer Res.* **1981**, *41*, 640. (c) Murname, J. P.; Byfield, J. E. *Chem.-Biol. Interact.* **1981**, *38*, 75. (d) Fleer, R.; Brendel, M. *Chem.-Biol. Interact.* **1981**, *37*, 123. (e) Ducore, J. M.; Erickson, L. C.; Zwelling, L. A.; Laurent, G.; Kohn, K. W. *Cancer Res.* **1982**, *42*, 897. (f) Hilton, J. *Biochem. Pharmacol.* **1984**, *33*, 1867. (g) Baker, J. M.; Parish, J. H.; Curtis, J. P. E. *Mutation Res.* **1984**, *132*, 171. (h) Garcia, S. T.; McQuillan, A.; Panasci, L. *Biochem. Pharmacol.* **1988**, *37*, 3189. (i) Geleziunas, R.; McQuillan, A.; Malapetsa, A.; Hutchinson, M.; Kopriva, D.; Wainberg, M. A.; Hiscott, J.; Bramson, J.; Panasci, L. *J. Natl. Cancer Inst.* **1991**, *83*, 557. (j) Ross, W. E.; Ewig, R. A. G.; Kohn, K. W. *Cancer Res.* **1978**, *38*, 1502. (k) Hansson, J.; Lewensohn, R.; Ringborg, U.; Nilsson, B. *Cancer Res.* **1987**, *47*, 2631.

(6) Brookes, P.; Lawley, P. D. *J. Chem. Soc.* **1961**, 3923.

(7) Brookes, P.; Lawley, P. D. *Biochem. J.* **1961**, *80*, 496.

(8) Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Colvin, O. M. *Anti-Cancer Drug Design* **1989**, *4*, 281.

(9) (a) Ojwang, J.; Grueneberg, D.; Loechler, E. L. *Proc. Am. Assoc. Cancer Res.* **1989**, *30*, 556. (b) Ojwang, J. O.; Grueneberg, D. A.; Loechler, E. L. *Cancer Res.* **1989**, *49*, 6529. (c) Grueneberg, D. A.; Ojwang, J. O.; Benasutti, M.; Hartman, S.; Loechler, E. L. *Cancer Res.* **1991**, *51*, 2268.

(10) Millard, J. T.; Raucher, S.; Hopkins, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2459.

(11) Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. Th.; Raucher, S. *Tetrahedron* **1991**, *47*, 2475.

(12) Golding, B. T.; Kebbell, M. *J. Chem. Soc., Perkin Trans. II* **1987**, 705.

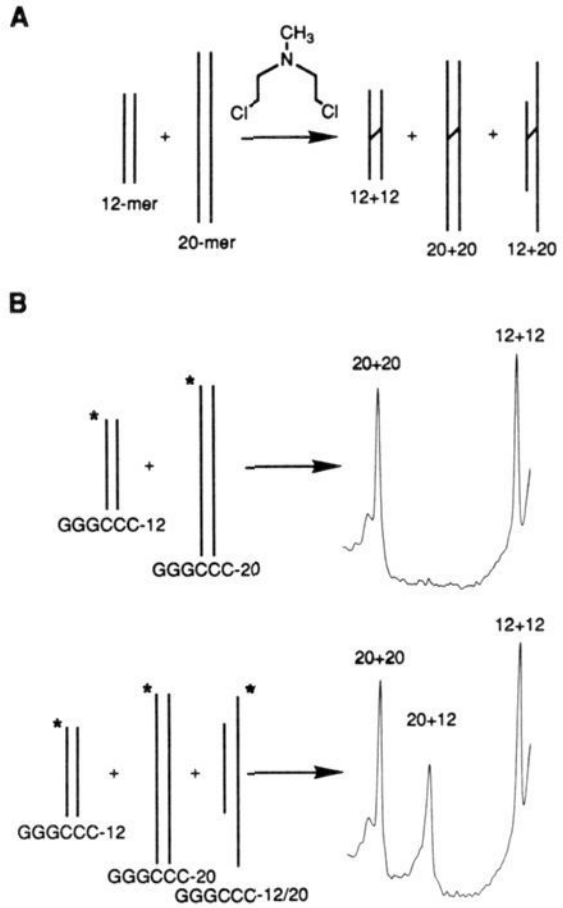


Figure 2. Mechlorethamine-induced cross-links are intrahelical. (A) If mechlorethamine cross-linking proceeds in an interhelical fashion, then a mixture of 12-mer duplex and 20-mer duplex would be expected to yield the heterodimer 12+20 in addition to the homodimers 12+12 and 20+20. (B) Densitometry of DPAGE analysis of binary (upper) and ternary (lower) mixtures of DNA duplexes (see text) shows a failure of the binary mixture to yield a 20+12 heterodimer, proving that interstrand cross-linking with mechlorethamine is an intrahelical reaction.

fact, DPAGE analysis of the cross-linking reaction of the binary mixture showed two interstrand cross-linked products of distinct mobility. That these were in fact the 12/12 and 20/20 combinations was confirmed by cross-linking a ternary mixture of GGGCCC-12, GGGCCC-20, and GGGCCC-12/20, the last consisting of a partially Watson-Crick paired 12-mer and 20-mer (see Table I). The ternary mixture afforded three products, two identical in electrophoretic mobility to those formed in the binary mixture (12/12 and 20/20) plus a third of intermediate mobility (12/20) (Figure 2). The absence of the latter in the binary mixture clearly demonstrated that the interstrand cross-linking reactions studied herein proceed intrahelically. This conclusion is fully consistent with the observation that non-self complementary single strands do not form interstrand cross-links in the absence of the complementary strand.[9a]

The Cross-Links Are N^7-to-N^7, dG-to-dG, at 5′-d(GNC). Mechlorethamine-cross-linked DNA GGGCCC-12, radiolabeled at either the 5′- or the 3′-end, afforded on exposure to 1 M aqueous piperidine (90 °C for 1 h) virtually complete cleavage to radiolabeled fragments corresponding in DPAGE mobility to the Maxam-Gilbert G-reaction products at G2 and G3 (Figure 3). This was fully consistent with N^7 as the site of alkylation on both strands,[10] in agreement with the results and conclusions of Loechler et al.[9] The ratio in which these products were formed provided information regarding the sequence-selectivity of cross-linking. In interpreting the results of experiments using a related duplex containing the same central sub-sequence 5′-d(GGCC), we have

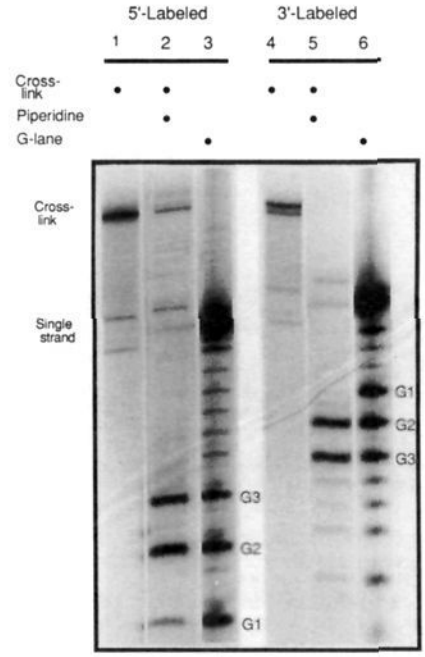


Figure 3. Piperidine fragmentation of mechlorethamine cross-linked DNA proceeds to yield a ca. 1:1 ratio (see text) of fragmentation products at G2 and G3. Lanes 1, 4: interstrand cross-linked DNA. Lanes 2, 5: piperidine treated interstrand cross-linked DNA. Lanes 3, 6: Maxam-Gilbert G-lane (dimethyl sulfate/piperdine).

noted that the formation of these fragments in a ca. 1:1 ratio is most simply explained by a preference for cross-linking at 5′-d(GNC).[10] An alternative interpretation, that this ratio might derive from a 1:1 mixture of cross-links at 5′-d(GC) (G3-to-G3) and 5′-d(GNNC) (G2-to-G2), is not only more complex but also inconsistent with results obtained using DNA GDGCCC-12 (see below), in which D is an N^7-deazadeoxyguanosine residue.

While 5′-d(GC) and 5′-d(GNNC) were not candidates for *predominant* cross-link sites, it should be noted that the data were consistent with the presence of cross-links at sites other than 5′-d(GNC). The ratio of yields of cleavage fragments with DNA GGGCCC-12 was found (phosphorimager) to deviate significantly from 1.0:1.0, being 1.25:1.00 (3′-labeled) and 1.41:1.00 (5′-labeled), G2:G3 (Figure 3). The excess cleavage product at G2 was consistent with derivation from a 5′-d(GNNC) (G2-to-G2) cross-link; given this interpretation, the observed cleavage fragment ratio was then consistent with a ca. 85:15 ratio of cross-linking at 5′-d(GNC) and 5′-d(GNNC), respectively. In another DNA duplex, we have observed a fragment ratio consistent with a 75:25 ratio of the same cross-link isomers.[11] The appearance of a trace of cleavage fragment corresponding to alkylation at G1 in the 5′-labeled DNA was consistent with a GNNC cross-link (G1-to-G3). However, were this the case, this band should be present in the same proportion in the 3′-labeled experiment, where it is, in fact, much reduced in intensity. An alternative and more likely explanation derives from the possibility of some monoadduct at G1. Because monoadducts are only revealed by a radiolabeled fragment when the monoadduct is between the radiolabel and the cross-link, such a lesion would appear only in the 5′-labeled sample, as is observed. Explanation of this discrepancy in quantitation and a more definitive demonstration of dG-to-dG cross-linking at 5′-d(GNNC) must await further experimentation.

A second set of experiments independently suggested an N^7-to-N^7 linkage and, as mentioned above, discounted the alternative interpretation of the fragmentation experiment that the cross-linked sample was a 1:1 mixture of linkage 5′-d(GNNC) (G2-to-G2) and 5′-d(GC) (G3-to-G3). The DNA GDGCCC-12

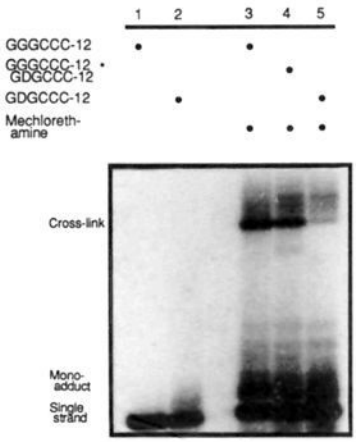


Figure 4. Demonstration that N^7 of G2 is necessary for cross-linking at 5′-d(GNC). GGGCCC-12, GDGCCC-12·GGGCCC-12, and GDGCCC-12 (Table I) contain respectively 2, 1, and 0 5′-d(GNC) sites; the decreasing number of sites parallels the decreasing yield of interstrand cross-linked product.

contains a 7-carba-7-deazadeoxyguanosine residue[13] (D) at what was originally G2; this substitution removes one of the two N^7 atoms potentially required for cross-linking at 5′-d(GNC). The availability of this DNA relied upon the availability of the appropriate phosphoramidite which was prepared by chemical synthesis in the Boston College laboratories. We compared the yield of interstrand cross-linked product from the 5′-end labeled homodimer GGGCCC-12, the heterodimer GDGCCC-12·GGGCCC-12, and the homodimer GDGCCC-12, which contain respectively 2, 1, and 0 N^7-to-N^7, G2-to-G3 cross-linkable sites. Fully consistent with this model, the yield of the major interstrand cross-linked product (7.9%) from the heterodimer (prepared by admixture of radiolabeled GDGCCC-12 with a large excess of GGGCCC-12) was roughly half that of the fully functional homodimer (11.7%, Figure 4). The N^7-depleted homodimer afforded a 1% yield of a product with comparable electrophoretic mobility, indicating the absence of even a single site which could be relatively efficiently cross-linked (Figure 4). This not only implicates N^7 as the site of alkylation at G2 but also eliminates from serious consideration the possibility that a G3-to-G3 cross-link is a major component of the interstrand cross-linked products, the latter being fully consistent with the work of others.[9] Native PAGE clearly demonstrated that GDGCCC-12 existed as a duplex (equivalent in electrophoretic mobility to GGGCCC-12 and distinct from a single strand 12 residues in length), discounting the trivial possibility that the inefficiency of cross-linking of GDGCCC-12 stemmed from instability of the duplex form.

Isolation and Characterization of the Cross-Link Lesion. The above studies provided substantial evidence that the mechlorethamine-induced DNA interstrand cross-link bridges N^7 atoms of two deoxyguanosine residues on opposite strands at the sequence 5′-d(GNC) but yielded no information concerning the structure of the mechlorethamine-derived component of the lesion. As noted in the introductory material, it has long been assumed that a single mechlorethamine molecule forms this bridge.[7] However, given constraints of bond angles and lengths, the span of a single mechlorethamine is inadequate to bridge N^7 atoms of two dG residues at this duplex sequence.[9-11] Distorted structures of this duplex sequence in either the original duplex or the cross-linked product would resolve this apparent paradox. Alternatively, as described above and illustrated in Figure 1, a second molecule of mechlorethamine might become involved in cross-linking.

The covalent structure of the cross-link was unequivocally established by direct isolation and characterization of the lesion

(13) Prepared by a modification of the literature procedure (Seela, F.; Westermann, B.; Bindig, U. *J. Chem. Soc., Perkin Trans. 1* **1988**, 697) to be reported elsewhere.

Table I. DNAs Used in This Study

Descriptor	DNA[a]
GGGCCC-12	123 5′HO-TATGGGCCCATA-OH3′ 3′HO-ATACCCGGGTAT-OH5′ 321
GGGCCC-20	5′HO-ATAATTAAGGGCCCTTAATT-OH3′ 3′HO-TTAATTCCCGGGAATTAATA-OH5′
GGGCCC-12/20	5′HO-TAAGGGCCCAAT-OH3′ 3′HO-TAATATTCCCGGGTTAATTA-OH5′
GGGCCC-12· GDGCCC-12	123 5′HO-TATGGGCCCATA-OH3′ 3′HO-ATACCCGDGTAT-OH5′ 321
GDGCCC-12	123 5′HO-TATGDGCCCATA-OH3′ 3′HO-ATACCCGDGTAT-OH5′ 321

[a] D refers to a 7-carba-7-deaza-2′-deoxyguanosine residue.

from a sample of interstrand cross-linked DNA. DNA GGGCCC-12 (see Table I) was treated with mechlorethamine and the major, interstrand cross-linked product was isolated by DPAGE. The studies described above ensured that this substance was intrahelically, interstrand cross-linked through deoxyguanosine residues G2 and G3. A portion of the interstrand cross-linked product was heated to 90 °C for 0.5 h (pH 7.0, 5 mM phosphate buffer) to achieve depurination. Subsequent digestion of the resulting solution with snake venom phosphodiesterase and calf intestinal alkaline phosphatase followed by HPLC analysis revealed a dA:dC:dT ratio (5.9:6.0:5.6) that was insignificantly changed from that of the untreated DNA (experimentally determined to be 6.0:6.0:6.3, calculated 6:6:6). As expected, the content of dG was diminished relative to the uncross-linked duplex but to an extent greater than predicted (2.6 found vs 4 calculated to remain of the 6 present in each duplex before cross-linking) given the involvement of two dG residues per cross-link. We speculated that the abasic site revealed by depurination interferes with the enzymatic hydrolysis step, but this point was not further investigated. In a separate experiment, excission of the cross-link was achieved by treatment of the cross-linked DNA with 88% aqueous formic acid at 150 °C for 4 h. The resulting hydrolysate was analyzed by RP HPLC and compared with an authentic sample of **2** prepared by a variant of the literature procedure involving direct treatment of GMP with mechlorethamine.[6] The chemically synthesized product, hereafter referred to as "authentic **2**", afforded proton and carbon NMR spectra, mass spectra (including high-resolution mass spectrum), and combustion analysis consistent with the depicted structure, as well as pH-dependent UV spectra consistent with those described by Brookes and Lawley.[6] HPLC analyses (in two different solvent systems) of the formic acid-hydrolyzed, interstrand cross-linked DNA clearly revealed the presence of a material with an elution time coincident with authentic **2**. Attempts to quantify **2** released from DNA were complicated by its low water solubility and strong affinity for RP HPLC columns. Ultimately, a procedure in which a formamide solution of the concentrated hydrolysate was analyzed on an "aged" C18 HPLC column returned 0.90 mol of **2** per mol of cross-linked duplex when analyzed by UV detection at either 260 nm, where the molar extinction coefficient of **2** (ϵ 6 600 M^{-1} cm^{-1}, pH 7) is roughly half that of dG, or 284 nm,

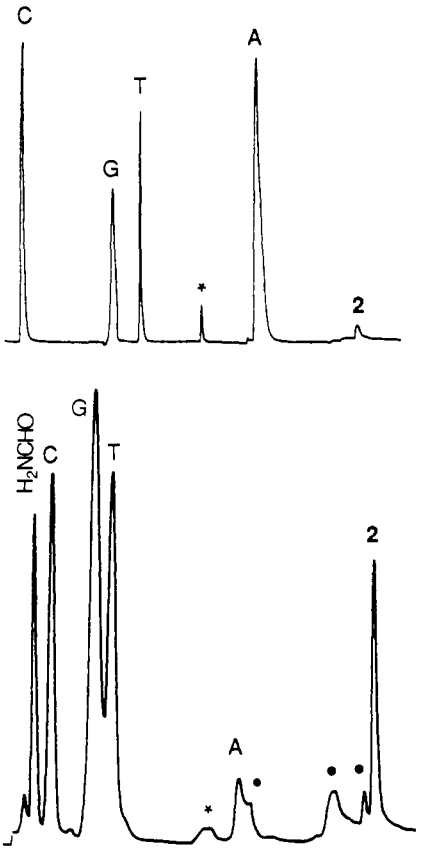


Figure 5. HPLC analysis of formic acid hydrolysate of interstrand cross-linked GGGCCC-12 (Table I) with (top) a high-efficiency RP-HPLC column with detection at 260 nm (see text) and (bottom) a low efficiency column with detection at 284 nm. Unidentified peaks unique to hydrolysate of cross-linked DNA (•) and common to hydrolysates of uncross-linked and cross-linked DNA (*) are indicated.

where the coefficient of **2** (ϵ 12 500 M^{-1} cm^{-1}, pH 7.0) is roughly twice that of dG (Figure 5).[14]

Further evidence in favor of the structural identity of lesion **2** isolated from interstrand cross-linked DNA and the synthetic sample of **2** was sought by spectroscopic comparison of the two samples. The available quantity of **2** from cross-linked DNA (submicromole) was near the practical analytical limits of standard UV, MS, and NMR measurements, and in all cases it showed lower signal-to-noise values and higher levels of impurities than the synthetic sample. Nevertheless, these measurements were fully consistent with the identity of the major components of the two samples. Lesion **2** (ca. 0.1 O.D.) isolated from hydrolyzed, interstrand cross-linked DNA gave strongly pH-dependent UV spectra with λ_{max} and λ_{min} values at acidic, neutral, and basic pH's within 2 nm of those of the authentic (synthetic) sample. The positive ion electrospray mass spectrum of **2** isolated from hydrolyzed DNA possessed the parent ion *m/e* 408 ($M + Na^+$), assuring that a single mechlorethamine-derived fragment bridged the guaninyl termini. MS/MS of the *m/e* 408 ion from both authentic and interstrand cross-link-derived samples afforded the common fragment ions *m/e* 365 ($M + Na^+$ – HNCO), 235 ($M + H^+$ – guanine) and 178 further assuring the identity of the two samples. The 300-MHz FT ^1H NMR spectra of **2** from interstrand cross-linked DNA (ca. 20 000 scans) and authentic **2** had in common all four of the non-equivalent proton resonances consistent with the assigned structure. Taken together, these data made a strong case in favor of the lesion **2** as the nucleus of the covalent interstrand cross-link caused by mechlorethamine at the duplex DNA sequence 5′-d(GNC).

(14) Because it was central to defining the stoichiometry of the hydrolysate, we independently determined the extinction coefficient of **2** at 260 nm by quantifying an aliquot of authentic **2** by NMR with acetonitrile as an internal standard. Our values differ insignificantly from those reported by Brookes and Lawley.[6]

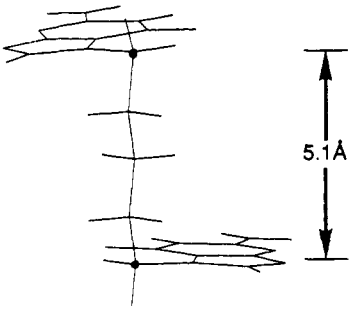


Figure 6. Guaninyl moieties constrained to reside in parallel planes and bridged through their N^7 atoms by an unstrained pentylene tether can be separated by at most 5.1 Å, substantially less than the 6.8 Å separation found at 5′-d(GNC) in B-DNA.

Impact of the Mechlorethamine Cross-Link on Duplex DNA Structure. These studies clearly demonstrate that a single mechlorethamine molecule creates an intrahelical, N^7-to-N^7, dG-to-dG interstrand cross-link in duplex DNA preferentially at the nucleotide sequence 5′-d(GNC). As noted in the introductory material, the result *necessitates* that the cross-linked DNA possesses some structure other than that of canonical B-DNA. This conclusion can be reached without recourse to advanced computer modeling or even a space-filling model of B-DNA, requiring only the knowledge that the planes occupied by adjacent base pairs in B-DNA are separated by roughly 3.4 Å.[15] Thus, the planes containing opposing dG residues at the nucleotide sequence 5′-d(GNC) reside in parallel planes 6.8 Å apart. Examination of a scale model of substance **2**, with the optimal bond lengths and angles, clearly reveals that when the two guaninyl substituents are constrained to reside in parallel planes, then the maximal distance between these planes is 5.1 Å, some 2 Å short of the required 6.8 Å (Figure 6). However, the actual spacing of N^7 atoms in B-DNA (8.9 Å) at the 5′-d(GNC) sequence is greater than 6.8 Å because these two N^7 atoms do not reside on a single line perpendicular to the parallel planes.

What then is expected for the three-dimensional structure of mechlorethamine-cross-linked duplex DNA? There are at present no experimental data bearing on this issue. This question may well have relevance to the ability of repair enzymes to recognize this lesion or to the lesion's ability to adventiciously interact with other DNA binding proteins, as has recently been found for the cisplatin-derived intrastrand cross-link.[16] At least two extreme models for the structure of the mechlorethamine-induced cross-link can be imagined. In both, we speculate that the Watson–Crick hydrogen bonding pattern is maintained, while the reduced spacing of the linked N^7 atoms might be achieved by hyper-propeller twisting of the two relevant deoxyguanosine residues. If the bases external to the cross-link maintain maximal stacking on these twisted residues, the result might be a kink in the helix axis resulting in a bend toward the major groove. Alternatively, it might be that this propeller twisting is dissipated over several adjacent base pairs, without global distortion of the helix axis. If technical problems involving the thermal instability of mechlorethamine-cross-linked DNA can be overcome, it is not unreasonable to imagine that the structures of these cross-links in small DNA fragments will ultimately be revealed by high-resolution X-ray or NMR measurements.

(15) Arnott, S.; Campbell-Smith, P.; Chandresekharan, P. In *CRC Handbook of Biochemistry and Molecular Biology Nucleic Acids*; Fasman, G., Ed.; CRC Press, Inc.: Boca Raton, FL, 1986; Vol. 2, pp 411–422.

(16) (a) Toney, H. H.; Donahue, B. A.; Kellett, P. J.; Bruhn, S. L.; Essignmann, J. M.; Lippard, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8332. (b) Donahue, B. A.; Augot, M.; Bellon, S. F.; Treiber, D. K.; Toney, J. H.; Lippard, S. J.; Essigmann, J. M. *Biochemistry* **1990**, *29*, 5872.

Conclusion

The dG-to-dG interstrand cross-link caused by mechlorethamine is thus shown to occur preferentially at the nucleotide sequence 5′-d(GNC), to link opposite strands of a single duplex (intrahelical), and to possess the sub-structure **2** in which the residue of a single mechlorethamine links the N^7 atoms of deoxyguanosine residues. This structure cannot be accommodated without significant conformational distortion of B-DNA, because the linking tether is insufficient to span the appropriate distance in B-DNA. How duplex DNA accommodates this lesion through structural reorganization is at this time unknown.

A significant and interesting question posed by these studies is the underlying basis for the 5′-d(GNC) specificity of mechlorethamine. This question must remain unanswered until further mechanistically relevant details can be obtained. It is proper to comment that the failure of the 30-year-old prediction of specificity for 5′-d(GC)[17] is, in hindsight, not surprising. This prediction was based on the fact that N^7 atoms at this sequence have an internuclear spacing which is nearly identical with that of two atoms bridged by a fully extended pentylene tether. Missing from that analysis was the impact of the angular constraints on bonding imposed by the relative orientations in space of the dG residues at this sequence. When this is taken into account, it is obvious that *no* nucleotide sequence in DNA can be interstrand cross-linked N^7-to-N^7 by a pentylene tether without significant conformational reorganization of DNA. To progress from this stage to that of understanding why 5′-d(GNC) is in reality the preferred sequence will require that we not only learn further mechanistic details of the mechlorethamine cross-linking reaction but also develop a considerably more advanced quantitative understanding of the structure and energetics of non-ground-state DNA conformations.

Experimental Section

Materials and Methods. The materials and methods were identical with those of Kirchner et al.[18] except for the following. Source of materials: Klenow fragment of DNA polymerase I, Boehringer Mannheim; mechlorethamine hydrochloride, Aldrich. O.D. refers to O.D. units at 260 nm. ϵ_{260} for DNAs were estimated to be 10 000 M^{-1} cm^{-1}/residue.[19] Aqueous solutions of mechlorethamine in pH 8.0 sodium cacodylate are unstable and were freshly prepared for all experiments, standing no more than 60 s before admixture with a DNA-containing solution. Autoradiography was performed on Amersham Hyperfilm-MP. Densitometry was performed on a Hoeffer GS-300 densitometer interfaced to an IBM PC. Data were smoothed and plotted using the program Spectra Calc. Unless otherwise specified, solutions were aqueous. Analytical and preparative HPLC were performed on an Alltech, 5 μm, C18, 250 mm × 4.6 mm or an Alltech Macrosphere 300, C18, 7 μm, 250 mm × 10 mm column using a Beckman 110B dual pump system with a Beckman 421A controller and sequential Waters Lambda-Max Model 481LC UV/vis (output to both an HP 3390A electronic integrator and a Linear Model 156 recorder) and Ranin Dynamax UV-1 variable wavelength UV/vis (output to both an HP 3390A electronic integrator and a Linear Model 255/MM recorder) absorbance detectors. Solvent gradients were run at 1 mL/min (analytical) and 2 mL/min (preparative) as follows. Gradient A, solvent A: 100 mM ammonium acetate (pH 7.0). Gradient A, solvent B: 80% CH_3CN; isocratic 90% A for 8 min, 12 min linear gradient to 62.5% A, 5 min linear gradient to 10% A, then a 5 min linear gradient to initial conditions. Gradient B, solvent A: 10 mM ammonium formate (pH 7.0). Gradient B, solvent B: 50% MeOH; isocratic 98.2% A for 10 min, a 10 min linear gradient to 8.2% A, isocratic for 7 min, then 5 min linear gradient to initial conditions. Gradient C, solvent A: 1.0 mM ammonium formate (pH 7.0). Gradient C, solvent B: 80% CH_3CN; isocratic 90% A for 8 min, 12 min linear gradient to 62.5% A, 5 min linear gradient to 10% A, then a 5 min linear gradient to initial conditions. Proton nuclear magnetic resonance spectra (^1H NMR) were determined on Bruker AC200 (200 MHz), AF300 (300 MHz), or AM500 (500 MHz) spectrometers and, unless otherwise noted, are reported in parts per million downfield from external DSS (0.00 ppm). ^{13}C NMR spectra were determined on a Bruker AM500 (125 MHz) spectrometer and are reported in parts per million downfield from external DSS (−1.6 ppm). Ultraviolet (UV) spectra were measured on a Hewlett-Packard Model 8450A or 8452A UV/vis spectrophotometer and are reported as wavelength in nanometers. High-resolution (HRMS) and low-resolution mass spectra (LRMS) were measured on a VG 70SEQ double-focusing mass spectrometer; electrospray ionization mass spectra (ESMS) were measured on a Sciex Atmospheric Pressure Ionization (API) triple quadrapole mass spectrometer. Except where otherwise noted, selected ions, diagnostic for the substance of interest, are reported.

Preparation of Radiolabeled Duplex DNA. Synthesis and purification of DNA were as described by Kirchner et al.,[18] except that oligonucleotides were synthesized on both Applied Biosystems Models 380A and 392 automated synthesizers. 5′-End labeling reactions were performed on 0.01 O.D. of DNA using 10 μCi of [γ-^{32}P] ATP 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] and 10 units of T4 polynucleotide kinase at 37 °C for 30 min. Radiolabeling was stopped by heating reactions to 90 °C for 1–2 min followed by addition of 2 μL of 3 M sodium acetate (pH 5.2) and 1 mL of ethanol at −20 °C. After 0.25 h at −20 °C, the resulting precipitate was pelleted by centrifugation, the supernatant was discarded, and the remaining solid was resuspended in 1 mL of 85% ethanol at −20 °C. After 0.25 h at −20 °C, the resulting precipitate was pelleted by centrifugation, the supernatant was discarded, and the remaining precipitate of radiolabeled DNA was used in the following experiments. 3′-Radiolabeled GGGCCC-12 was prepared from the corresponding 11-mer (0.001 O.D.) (truncated by a single residue at the 3′-terminus relative to the 12-mer in Table I) using the Klenow fragment of DNA polymerase I/[α-^{32}P]-dATP (3000 Ci/mmol) under standard conditions.[19] To form duplex DNA, radiolabeled oligomers were added to the appropriate quantity (see below) of complementary, unlabeled DNA in 40 mM sodium cacodylate, heated to 90 °C for 1–2 min, and then cooled to 37 °C prior to cross-linking with nitrogen mustard. Analytical scale cross-linking experiments utilized 0.3 O.D. of unlabeled DNA and ca. one-half of the product of a radiolabeling reaction (ca. 0.0005 OD); preparative scale experiments used 30 O.D.s of unlabeled GGGCCC-12 (Table I) and one-half of the product of a 3′-radiolabeling of 11-mer. The interhelical/intrahelical cross-linking experiments used 5′-radiolabeled GGGCCC-12, GGGCCC-20, and the 20-mer strand of the GGGCCC-12/20 duplex (ca. 150 000 cpm of each as determined by Geiger counter, ca. one-quarter of the products of a labeling reaction, ca. 0.0025 O.D.) each admixed with 0.3 O.D. of the appropriate unlabeled strand. Experiments probing the site of cross-link formation by nitrogen mustard utilized 5′-end-labeled DNAs. 5′-Labeled GGGCCC-12 was prepared as described above. 5′-Labeled GDGCCC-12 (ca. 0.005 O.D., see Table I) was admixed with 0.3 O.D. of the appropriate unlabeled, complementary strand to form the homoduplex GDGCCC-12 and the heteroduplex GGGCCC-12·GDGCCC-12.

Yield Optimization of Cross-Linking. To study the impact of temperature on yield of cross-linking, 5′-labeled GGGCCC-12 (0.3 O.D., ca. 1 nmol duplex) in 5 μL of 40 mM sodium cacodylate (pH 8.0) equilibrated at 0, 24, or 37 °C was treated with 5 μL (2.5 nmol) of 0.5 mM mechlorethamine hydrochloride in 40 mM sodium cacodylate. After being mixed briefly, the reactions were held for 3 h at the specified temperatures and were stopped by addition of 10 μL of 100 mM HOAc. Stopped reactions were purified and analyzed as described below. The relative yields of cross-linked DNA were 37 °C > 24 °C > 0 °C, with a yield of ca. 5–10% at 37 °C.

To study the effect of nitrogen mustard concentration and time on the yield of cross-linked DNA, reaction mixtures were prepared as described above in the temperature experiments and were initiated by addition of nitrogen mustard solutions such that the total final nitrogen mustard concentrations were 0.25, 0.40, 0.60, and 1.00 mM in a total volume of 10 μL. Reactions were incubated at 37 °C, stopped at 1, 3, and 9 h time points, purified, and analyzed as described below. In general the relative yields of cross-linked DNA were 9 h > 3 h > 1 h at a given mechlorethamine

(17) What Brookes and Lawley (BL) predicted in their pioneering work is actually arguable. The text concludes that cross-linking should occur preferentially at "deoxyguanylyl-(3′→5′)-deoxycytidylyl", presumably what we would today call 5′-d(CG), but this is completely inconsistent with the spirit of the argument made therein. Specifically, in describing the attributes of the predicted sequence, BL clearly describe the relationship of atoms at 5′-d(GC). Given the logical assumption that Figure 3 in BL depicts the major groove of duplex DNA, then the cross-link bridges dG-to-dG at 5′-d(GC).

(18) Kirchner, J. J.; Sigurdsson, S. Th.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4021.

(19) Maniatis, T.; Fritsch, E. F.; Sambrook, J. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

concentration. Furthermore, the yields were 1.00 mM > 0.60 mM > 0.40 mM > 0.25 mM for a given reaction time.

Yields of interstrand cross-linking were determined by DPAGE as described by Kirchner et al.[18] with the following exceptions. Cross-linked samples, after passage through a Sep-Pak, were dried in vacuo at 40 °C, were not heat denatured prior to DPAGE analysis, and were electrophoresed at reduced temperatures (20–30 °C) to minimize the thermally promoted decomposition of the cross-linked DNA. Data were quantified by one-dimensional densitometric analysis of autoradiograms or phosphorimagery.[18]

General Conditions for Preparation of Interstrand Cross-Linked DNA Duplexes. The following conditions were based on the outcome of the optimization experiments. Cross-linking experiments were performed on both analytical and preparative scales. The following conditions were used for analytical experiments. Preparative conditions refers to a 100-fold scale increase with the exception of the quantity of radiolabel which remained unchanged. A microfuge tube containing DNA (ca. 0.3 O.D., 1 nmol duplex) in 6 μL of 40 mM sodium cacodylate (pH 8.0) at 37 °C was treated with 4 μL of 1.0 mM mechlorethamine hydrochloride in 40 mM sodium cacodylate. After the samples were briefly mixed, and microfuged, they were maintained at 37 °C for 3 h, stopped with 10 μL of 100 mM HOAc, Sep-Pak purified, and DPAGE analyzed as described above. Cross-linked material to be used in subsequent fragmentation experiments was isolated from the gel as described by Kirchner et al.[18] with the exception that the gel slices were soaked in elution buffer at 25 °C for 12 h prior to Sep-Pak purification. Purification of cross-linked DNA from analytical scale reactions was achieved in a single lane of an analytical gel; purification of preparative-scale reactions used 3.0–6.0 O.D. of DNA per lane of an analytical gel.

Distinction of Interhelical from Intrahelical Cross-Linking. Two microfuge tubes, one containing a 1:1:1 molar ratio of 5′-radiolabeled GGGCCC-12, -20, -12/20 (0.075, 0.125, and 0.1 O.D., respectively, ca. 0.3 nmol of each duplex) and the second containing only the first two of these in a 1:1 molar ratio, were prepared as described above and cross-linked on an analytical scale using the general conditions. Reaction quenching, Sep-Pak purification, and DPAGE analysis were performed as described above under Yield Optimization. The results are shown in Figure 2.

Piperidine Cleavage of 5′- or 3′-Labeled Cross-Linked GGGCCC-12. Cross-linked DNA in 50 μL of 1 M piperidine in a sealed microfuge tube was heated to 90 °C for 1.0 h. The mixture was dried in vacuo, lyophilized twice from 25 μL of H_2O, dissolved in loading buffer, and analyzed by 25% DPAGE at 40 °C. The gel was dried and quantified by phosphorimagery as described by Kirchner et al.[18] The results are shown in Figure 3.

Cross-Linking of N^7-Deazadeoxyguanosine-Containing DNA Duplexes. The radiolabeled duplexes GDGCCC·GGGCCC and GDGCCC·GDGCCC, prepared as described above, were cross-linked on an analytical scale under the general reaction conditions and, after Sep-Pak purification, analyzed by DPAGE as described above. The results are shown in Figure 4.

Enzymatic Hydrolysis and Quantification of Nucleosides by HPLC. Two microfuge tubes, each containing 0.2 O.D. of interstrand cross-linked or non-cross-linked DNA, were heated at 90 °C for 30 min in 5 mM potassium phosphate buffer (pH 7.0) and then enzymatically hydrolyzed using a total reaction volume of 60 μL while maintaining reagent and enzyme concentrations as described by Kirchner et al.[18] HPLC analysis was performed on 30 μL of reaction mixture using gradient A. Quantitation was based on the peak areas ratios obtained using the method of Kirchner et al.[18] and were as follows: dC, 1.00:dG, 1.67:dT, 1.15:dA, 1.77. The uncross-linked duplex DNA analyzed as 6.0 (dC):6.3 (dG):6.3 (dT):6.0 (dA) (calcd: 6:6:6:6); cross-linked DNA analyzed as 6.0 (dC):2.6 (dG):5.6 (dT):5.9 (dA) (calcd: 6:4:6:6). Representative traces are shown in Figure 5.

Authentic Bis[2-(N^7-guaninyl)ethyl]methylamine (2). Guanosine 5′-monophosphate, disodium salt trihydrate (50 n.g, 0.11 mmol) in 250 mL of water at 37 °C was treated with a solution of mechlorethamine hydrochloride (50 mg, 0.26 mmol) in 250 mL of water. The gelatinous reaction mixture stood at 37 °C for 6 h, then was heated to 100 °C for 2 h, and finally was cooled to 25 °C. After 20 h at 25 °C, the supernatant was decanted and the residual precipitate was washed with water and dried in vacuo to yield **2** as a white solid (4 mg, protonation state and degree of hydration not determined). ^{1}H NMR (300 MHz, 1.2 N DCl/D_2O): δ 3.30 (3H, s, CH_3), 4.08 (4H, m, CH_2), 5.07 (4H, m, CH_2), 8.82 (2H, s, CH). ^{13}C NMR (125 MHz, 1.2 N DCl/D_2O): δ 42.0 and 43.2 (NCH_2CH_2), 55.7 ($-CH_3$), 108.9 ($C^{5'}$), 142.2 ($C^{8'}$), 148.9 ($C^{4'}$), 153.9 ($C^{2'}$), 154.7 ($C^{6'}$). LRMS (FABS, thioglycerol): *m/e* 408 (M + Na^+), 386 (M + H^+). LRMS/MS of *m/e* 408 [electrospray from CH_3CO_2H:MeOH:H_2O (1:1:2 v/v/v)]: *m/e* 408 (M + Na^+), 386 (M + H^+), 365 (M − HNCO + Na^+), 235 (M + H^+ − guanine), 178. HRMS: *m/e* 386.1770 (calcd 386.1801). UV (lit.[6] values in parentheses): λ_{max} 280 (281) nm, λ_{min} 258 (257) nm (0.1 N NaOH); λ_{max} 285 (284) nm, λ_{min} 262 (261) nm (H_2O); λ_{max} 252 (252) nm (1.6 N HCl). HPLC: 19.8 min (gradient A), 27.0 min (gradient B). A sample for combustion analysis, obtained from preparative HPLC using solvent system C, was dried at 80 °C in vacuo for 74 h and then brought to ambient pressure and temperature under argon. Anal. Calcd for $C_{15}H_{19}N_{11}O_2 \cdot H_2O \cdot 0.66SiO_2$. Calcd: 40.65, C; 4.75, H; 34.70, N. Found: 40.2, C; 4.6, H; 34.9, N.

Bis[2-(N^7-guaninyl)ethyl]methylamine from Interstrand Cross-Linked DNA. Cross-linked GGGCCC-12 (ca. 4 O.D.) was taken up in 300–400 μL of concentrated formic acid and hydrolyzed in a sealed, Teflon-lined conical vial for 4 h at 150 °C. After drying in vacuo, the sample was concentrated twice from 200 μL of H_2O and dissolved in 50 μL of H_2-NCOH to provide multiple 10–25-μL portions for HPLC analysis. The substance **2** was isolated by analytical HPLC using gradient B. The most strongly retained material, retention time 27 min, was collected and concentrated in vacuo. After being concentrated three times from 1 mL of water, the sample in CH_3COOH/MeOH/H_2O (1:1:1, v/v/v), final concentration of 170 pmol/μL based on ϵ_{284} 12 500 M^{-1} cm^{-1} (H_2O),[6,14] gave the following results. MS (electrospray, 250 V inlet voltage, 5000 V needle voltage, 3 μL/min): *m/e* 452 (M + 3Na^+ − 2H^+), 408 (M + Na^+), 365 (M − HNCO + Na^+). LRMS/MS: *m/e* 408 (M + Na^+), 365 (M − HNCO + Na^+), 235 (M + H^+ − guanine), 178. UV [values from synthetic **2** (vide supra) in parentheses]: λ_{max} 280 (280) nm, λ_{min} 258 (258) nm (0.1 N NaOH); λ_{max} 285 (285), λ_{min} 262 (262) nm (H_2O); λ_{max} 254 (252) nm (1.6 N HCl). The combined hydrolysates from five 4.0 O.D. samples of cross-linked DNA were HPLC purified to afford a sample for ^{1}H NMR (300 MHz, 1.2 N DCl/D_2O): δ 3.31 (3H, s, CH_3), 4.09 (4H, m, CH_2), 5.08 (4H, m, CH_2), 8.83 (2H, s, CH). This sample contained unidentified impurities with the following resonances: δ 2.23 (s), 2.10 (s), 1.84 (s), 0.31 (s). Authentic **2** (HPLC purified, from guanosine monophosphate and nitrogen mustard) and **2** from cross-linked GGGCCC-12 (HPLC purified) admixed in a 1:1 ratio (UV) were shown to coelute in both gradients A and B.

Acknowledgment. This work was supported by the NIH (AG00417 and GM45804) and NSF (DIR821099). We thank Professor F. Seeman for thought-provoking correspondence. Phosphorimaging was performed by the PhophorImager Facility, Markey Molecular Center, UW. P.B.H. is a Sloan Fellow and Cope Scholar. S.R. is an NIH predoctoral fellow (GM08437).

Supplementary Material Available: Spectroscopic data on authentic **2** and **2** from interstrand cross-linked GGGCCC-12 (5 pages). Ordering information is given on any current masthead page.