DNA Interstrand Cross-Linking by Formaldehyde: Nucleotide Sequence Preference and Covalent Structure of the Predominant Cross-Link Formed in Synthetic Oligonucleotides

Huifang Huang and Paul B. Hopkins*

Contribution from the Department of Chemistry, University of Washington, Seattle, Washington 98195

*Received May 6, 1993**

Abstract: The nucleotide sequence specificity of the DNA interstrand cross-linking reaction of formaldehyde and the covalent structure of the predominant cross-link were established in synthetic DNA duplexes. A panel of synthetic DNAs was exposed to aqueous formaldehyde and analyzed by denaturing polyacrylamide gel electrophoresis (DPAGE). Those containing the sequence 5′-d(AT) afforded a relatively abundant interstrand cross-linked product. For one of these products, sequence random fragmentation using iron(II) EDTA/hydrogen peroxide/ascorbic acid and subsequent DPAGE analysis revealed the cross-link to bridge deoxyadenosine residues at the sequence 5′-d(AT). Enzymatic digestion of the sugar–phosphate backbone followed by HPLC resolution of the resulting residues afforded, in addition to the common deoxynucleotides, a substance identified by a combination of chemical and spectroscopic studies as bis(N^6-deoxyadenosyl)methane. The potential roles of monoadduct formation and progression of monoadducts to cross-links in the mechanistic origin of the selectivity for 5′-d(AT) are discussed. It is speculated that preferential monoadduct formation in AT-rich DNA accounts for formation of monoadducts of deoxyadenosine and that the combination of helical twist and propeller twist, both of which act to orient the amino groups at 5′-d(AT) for cross-linking, are responsible for selection of the sequence 5′-d(AT). It is noted that negative propeller hypertwisting is common to the best available structural model of several interstrand cross-linked DNAs and may therefore play a general role in determining which nucleotide sequences are susceptible to formation of interstrand cross-links.

Introduction

Formaldehyde is ubiquitous in the environment of this planet. The impressive contribution by industry of some 4 billion kilograms of formaldehyde annually is dwarfed by the estimated 400 billion kilograms of formaldehyde created annually by oxidation of methane in the troposphere.[1] Formaldehyde is a component of automobile exhaust and tobacco smoke. The impact of formaldehyde and higher aldehydes[2] on living organisms has been a topic of considerable interest. The acute toxicity of formaldehyde toward microorganisms is used to an advantage in its application as a disinfectant. In vitro tests have generally concluded that formaldehyde is a mutagen and causes chromosomal abberations and sister chromatid exchange. In vivo tests have generally been negative or inconclusive for these same effects.[1]

Formaldehyde has been shown to react with a variety of nucleophilic cellular constituents, forming, for example, methylol derivatives of the mercaptan group of glutathione and the amino groups of RNA and DNA.[3] The kinetics of the latter reaction have been studied in detail by McGhee and von Hippel[4] as a probe of conformational fluctuations in duplex DNA. In addition to methylol formation, formaldehyde acts as a cross-linking agent, having been shown to produce protein–DNA and DNA–DNA cross-links.[3] These lesions may result in the mutagenic[2a] and carcinogenic[2b] action of formaldehyde. Because the ability to form DNA–DNA interstrand cross-links is a property of several antibiotics, antitumor substances, and toxins, we have studied the DNA interstrand cross-linking reaction of formaldehyde with the goal of defining the sequences of nucleotides at which this linkage is formed and the covalent structure of the lesion.

The nucleic acid interstrand cross-linking reaction of formaldehyde is a stepwise process, resulting from a relatively rapid formation of the methylol derivative of the amino and imino functions followed by a slower conversion to cross-links.[3,5] Particularly relevant to the present study is the report of Chaw et al.[6] in which calf thymus DNA was incubated with formaldehyde for a prolonged period (40 days) followed by enzymatic digestion of the sugar–phosphate backbone and separation of the products by HPLC. In addition to returning the four component deoxynucleosides of DNA, this process afforded the five pairwise combinations of dA, dC, and dG (except dC/dC) linked through their former exocyclic amino group by a methylene bridge (e.g., **1**, **2**, and **3**). The structures of these substances were rigorously assigned by a combination of spectroscopic and chemical measures. While it seemed probable that *one or more* of these lesions might be responsible for an interstrand cross-link, we reasoned that these were not *all* equally likely to have arisen from interstrand cross-links. This reasoning was based upon past observations that in all well-characterized cases, the most abundant interstrand cross-links from a given agent bridge functional groups whose spacing in canonical B-DNA[7] is reasonably well-correlated with the spacing of the reactive centers in the cross-linking agent (e.g., mitomycin C[8] and other pyrrole-derived agents,[9] nitrogen mustard,[10] nitrous acid,[11] psoralen,[12] and CC-1065 analogs[13]). The

* Abstract published in *Advance ACS Abstracts,* September 1, 1993.

(1) WHO. Environmetnal Health Criteria for Formaldehyde. WHO Technical Report Series No. 89; WHO: Geneva, 1989.

(2) (a) Auerbach, C.; Moutschen-Dahmen, M.; Moutschen, J. *Mutat. Res.* **1977**, *39*, 317. (b) *Nature (London)* **1979**, *281*, 625.

(3) Feldman, M. Y. *Prog. Nucleic Acid Res. Mol. Biol.* **1973**, *13*, 1.

(4) (a) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1975**, *14*, 1281, 1297. (b) McGhee, J. D.; von Hippel, P. H. *Biochemistry* **1977**, *16*, 3267, 3276.

(5) Fraenkel-Conrat, H. *Biochim. Biophys. Acta* **1954**, *15*, 307.

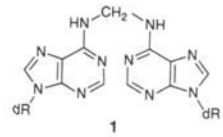
(6) Chaw, Y. F. M.; Crane, L. E.; Lange, P.; Shapiro, R. *Biochemistry* **1980**, *19*, 5525.

(7) Arnott, S.; Campbell-Smith, P.; Chandresekharan, P. In *CRC Handbook of Biochemistry*; CRC Press: Boca Raton, FL, 1976; Vol. 2, pp 411–422.

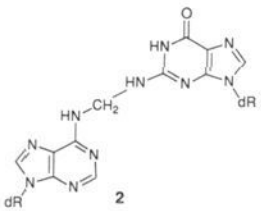
(8) (a) Chawla, A. K.; Lipman, R.; Tomasz, M. Covalent Crosslinks and Monofunctional Adducts of Mitomycin C in the Minor Groove of DNA: Effects on DNA Conformation and Dynamics. Structure and Expression; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Albany, NY, 1987; Vol. 2, DNA and its Drug Complexes. (b) Teng, S. P.; Woodson, S. A.; Crothers, D. M. *Biochemistry* **1989**, *28*, 3901. (c) Weidner, M. F.; Millard, J. T.; Hopkins, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 9270. (d) Borowy-Borowski, H.; Lipman, R.; Chowdary, D.; Tomasz, M. *Biochemistry* **1990**, *29*, 2992.

(9) Woo, J. W.; Sigurdsson, S. Th.; Hopkins, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 3407.

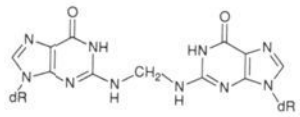
0002-7863/93/1515-9402$04.00/0



1



2



3

methylene bridge in formaldehyde cross-links would thus be expected to bridge amino groups in the same groove (the bridge being far too short to span from one groove to the other) and in adjacent base pairs (because no single base pair possesses two amino groups in a single groove and more separated base pairs greatly exceed the functional group spacing compatible with a single methylene bridge). This hypothesis leaves six duplex dinucleotides as candidates for preferential interstrand cross-linking, dA to dA at 5′-d(AT) or 5′-d(TA), dG to dG at 5′-d(CG) or 5′-d(GC), and dA to dC at 5′-d(AG) or 5′-d(GA). We report herein that in the short oligonucleotide duplexes we have studied, the most abundant interstrand cross-link is formed at the first of these six sequences, dA to dA at the sequence 5′-d(AT) and with the structure implicit in **1**.[14] Molecular mechanics modeling of this linkage indicated that in order to achieve the normal bond lengths and angles expected for the methylene linkage, some structural reorganization of B-DNA is required.[14] Most noteworthy in this regard is that the propeller twist at the linkage is predicted to be in the same direction (negative) as that found in B-DNA[15] but of greatly exaggerated magnitude. This is at least the fourth such example of such a correlation,[10,11,16] which may thus be general and of predictive value.

Results

Preliminary Experiments. It was necessary at the outset to determine what DNA sequences are susceptible to interstrand cross-linking by formaldehyde because this facilitates the preparation of structurally homogeneous samples critical for structural studies. A now standard approach to defining the nucleotide sequence preference of a cross-linking agent, an obvious prelude to the design of a DNA duplex cross-linkable at a single site, involves embedding a variety of potential "target" sequences in an otherwise inert flanking sequence.[8b,12] The indiscriminate reactivity of formaldehyde complicated the design of such an inert flanking sequence. Because only dT failed to appear in methylene-bridged digestion products,[6] only $d(A_n)$-$d(T_n)$, a DNA sequence known to have unusual structural properties, could be reasonably expected not to yield interstrand cross-links and thus serve as an inert flanking sequence. To avoid employing this structurally aberrant DNA sequence, we chose instead at the outset to survey some 10 DNA duplexes possessing a variety of nucleotide contents and sequences which were available in the laboratory, of which DNAs I and II in Table I are representative. These self-complementary DNAs were exhaustively 5′-phosphorylated (trace ^{32}P), incubated for 9 days with aqueous formaldehyde (25 mM, 25 °C, pH 6.0, 50 mM sodium phosphate buffer, 25 mM NaCl), and then analyzed for interstrand cross-linked products by denaturing polyacrylamide gel electrophoresis (DPAGE, Figure 1). Consistent with the notion that regions inert to interstrand cross-linking in a randomly chosen sequence would be rare, all of these DNAs afforded several (ca. three to six) products of low electrophoretic mobility, presumed to be interstrand cross-linked. (Structurally distinct interstrand cross-linked DNAs are frequently resolved by DPAGE.[17]) The distribution of product quantities was consistent with two

Table I. DNA Duplexes[a]

nucleotide sequence	descriptor
AATCGATAATATTATCG GCTATTATAATAGCTAA	I
AATATAATCCGGATTAT TATTAGGCCTAATATAA	II
$TACAACN_4GTTGT$ $TGTTGN_4CAACAT$	III–XIII
N_4 = ATAT	III
TATA	IV
AATT	V
TTAA	VI
CATG	VII
GATC	VIII
AGCT	IX
ACGT	X
GCGC	XI
CGCG	XII
CCGG	XIII
$TACCTN_8AGCT$ $TGGAN_8TCGAT$	XIV–XV
N_8 = TTAATTAA	XIV
AATTAATT	XV
$(AT)_{12}$	XVI

[a] The upper strand or indicated strand of each duplex is listed with 5′ → 3′ polarity. Unless otherwise noted in the text, all DNAs were 5′-radiolabeled by phosphorylation.

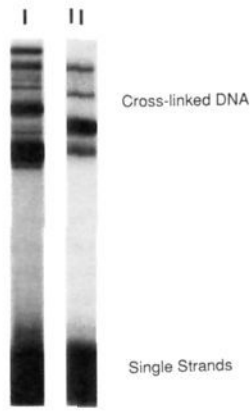


Figure 1. Autoradiogram of the DPAGE analysis of DNAs I and II (exhaustively 5′-phosphorylated) following exposure to formaldehyde.

(10) (a) Millard, J. T.; Raucher, S.; Hopkins, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2459. (b) Ojwang, J.; Grueneberg, D.; Loechler, E. L. *Proc. Am. Assoc. Cancer Res.* **1989**, *30*, 556.

(11) (a) Kirchner, J. J.; Hopkins, P. B. *J. Am. Chem. Soc.* **1991**, *113*, 4681. (b) Kirchner, J. J.; Sigurdsson, S. Th.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4021.

(12) Gamper, H.; Piette, J.; Hearst, J. E. *Photochem. Photobiol.* **1984**, *40*, 29.

(13) (a) Lee, C.-S.; Gibson, N. W. *Biochemistry* **1993**, *32*, 2592. (b) Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **1993**, *115*, in press.

(14) Huang, H.; Solomon, M. S.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 9240.

(15) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1988; p 253.

(16) (a) Lemaire, M. A.; Schwartz, A.; Rahmouni, A. R.; Leng, M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1982. (b) Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. Th.; Raucher, S. *Tetrahedron* **1991**, *47*, 2675.

(17) Millard, J. T.; Weidner, M. F.; Kirchner, J. J.; Ribeiro, S.; Hopkins, P. B. *Nucleic Acid Res.* **1991**, *19*, 1885.

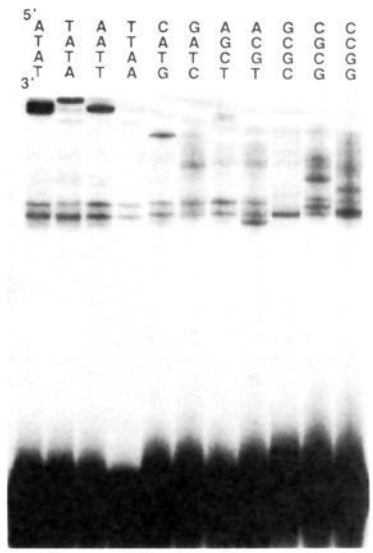


Figure 2. Autoradiogram of the DPAGE analysis of DNAs III–XIII (trace 5′-phosphorylated) containing the indicated central sequences following exposure to formaldehyde.

competing effects. The dominant effect was that of sequences nearer the strand termini being of greatest reactivity toward interstrand cross-linking. This was indicated by the observation that of the electrophoretically distinct cross-linked products, the more abundant tended to be of higher electrophoretic mobility. We have previously determined that such "high-mobility" products are cross-linked near the strand termini.[17,18] Superimposed upon this effect, however, was a variation in yield of interstrand cross-linked product quantities that did not correlate with electrophoretic mobility, no doubt resulting from just the sequence variation in cross-linking efficiency which we sought to study. The yield of individual cross-links (phosphorimagery) ranged from a few tenths of a percent to ca. 5% for cross-linking at the strand termini.

Having established that formaldehyde interstrand cross-linked synthetic oligonucleotides could not only be prepared but were formed with some selectivities and could be resolved by DPAGE, we pursued the goal of identifying an inert sequence in which to embed potential target sequences. Reasoning as described above that formaldehyde would link amino groups on opposite strands residing in adjacent base pairs, we devised the general, self-complementary DNA sequence 5′-d(TACAACN$_4$GTTGT). Inspection of the major and minor grooves of this sequence revealed no adjacent base pairs bearing amino groups on opposite strands in one groove. After preliminary experiments demonstrated that this approach was remarkably effective at both diminishing cross-linking at the strand termini (see Figure 2) and diminishing the number of interstrand cross-linked products, the panel of DNAs III–XIII (see Table I) was synthesized. This panel was treated with formaldehyde (25 mM, 25 mM NaCl, 50 mM sodium phosphate buffer, pH 6.0, 25 °C, 9 days) and analyzed by DPAGE (Figure 2). Immediately noteworthy was the formation in DNAs III (ATAT), IV (TATA), and V (AATT) of, respectively, two, one, and one major low-mobility cross-linked products. Because in this experiment the 5′-hydroxyl terminal strands were in large excess relative to the 5′-phosphoryl-bearing strands, the symmetry of these otherwise C_2 symmetric duplexes is broken in the detected (radiolabeled) duplexes. As such, DNAs III, IV, and V contain two, one, and one 5′-d(AT) sequences, respectively. The simplest explanation of these data is that cross-links are centered on 5′-d(AT).

The outcome of an independent experiment was fully consistent with the above conclusions. The duplex DNAs XIV and XV in singly phosphorylated form contain, respectively, one and two 5′-d(AT) sequences. Following exposure to formaldehyde, these duplexes afforded, correspondingly, one and two low-mobility interstrand cross-linked products (Figure 3). That the pair of products formed in duplex XV resulted from the symmetry-breaking monophosphate was confirmed by formaldehyde treatment of duplex XV in C_2 symmetric, doubly phosphorylated form, which now afforded a *single* low-mobility interstrand cross-linked product. All of these experiments are thus consistent with preferential formaldehyde cross-linking centered on the dinucleotide sequence 5′-d(AT).

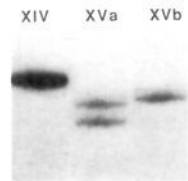


Figure 3. Autoradiogram of the DPAGE analysis of DNAs XIV and XV following exposure to formaldehyde. XIV and XVa are trace 5′-phosphorylated; XVb is exhaustively 5′-phosphorylated.

Table II. Yield of Interstrand Cross-Linked DNA

DNA	maj prodt (%)	total (%)	DNA	maj prodt (%)	total (%)
III	0.31, 0.38	1.44	IX	0.08	1.19
IV	0.18	0.93	X		1.45
V	0.35	0.91	XI		1.14
VI	0.04, 0.04	0.39	XII	0.19	1.76
VII	0.16	0.91	XIII	0.18	2.03
VIII		2.02			

One further point should be made concerning the formaldehyde cross-linking of DNAs III–XIII. It is important to distinguish the efficiency of cross-linking *at a specific site* from the efficiency of cross-linking of a *specific duplex*. The importance of this distinction is evident on inspection of Table II. Whereas on a per site basis the 5′-d(AT) dinucleotides in DNAs III and IV were most efficiently cross-linked, the most efficiently cross-linked duplexes were in fact several of the CG-containing DNAs (VIII, XII, and XIII) which provided more heterogeneous distributions of cross-links. The reasons for this have at present not been determined, but we speculate that this may be related to the competition in these experiments between denaturation[4] and cross-linking. If the CG-containing DNA duplexes by virtue of their greater stability persist longer in the duplex state (are more slowly denatured) even while bearing monoadducts, then these monoadducts at a variety of sites may have a greater chance of progressing to cross-links. Future studies will need to address these observations.

Nucleotide Connectivity of a Cross-Link Centered on 5′-d(AT). The nucleotide connectivity of one interstrand cross-linked DNA was determined by sequence random fragmentation with iron(II) EDTA/ascorbic acid/hydrogen peroxide. The least electrophoretically mobile band formed on formaldehyde exposure of DNA V (N_4 = AATT) was isolated from DPAGE. The site of cross-linking was established at nucleotide resolution by sequence random oxidative fragmentation followed by DPAGE analysis of the fragment mixture. Noteworthy in the resulting distribution of fragments (Figure 4) *shorter in length than a starting single strand* is the absence of fragments derived from cleavage to the nonradiolabeled, 3′-side of the deoxyadenosyl residue at the 5′-d(AT) sequence. This pattern is uniquely consistent with dA-to-dA connectivity at the 5′-d(AT) sequence.[8c] This result is fully consistent with the previously reported conclusions based upon the results of fragmentation of formaldehyde cross-linked DNA III.[14]

Optimization of the Yield of Cross-Linked DNA. Elucidation of the covalent structure of an interstrand cross-link is routinely

(18) Weidner, M. F.; Sigurdsson, S. Th.; Hopkins, P. B. *Biochemistry* **1990**, *29*, 9225.

Figure 4. Autoradiogram of the DPAGE analysis of DNA V (N_4 = AATT), trace 5′-phosphorylated: lane 1, DNA only; lane 2, G-lane[22]; lane 3, DNA, iron(II) EDTA, ascorbic acid, and hydrogen peroxide; lane 4, formaldehyde cross-linked DNA, iron(II) EDTA, ascorbic acid, and hydrogen peroxide.

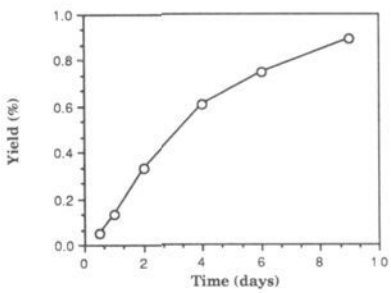


Figure 5. Total yield of the two major interstrand cross-linked products from formaldehyde-treated DNA III as a function of time.

accomplished by digestion of a cross-linked oligonucleotide whose connectivity of cross-linking has been previously established by the fragmentation protocol described in the previous section.[9,11b,19] Because the convenience and thoroughness of characterization is a function of the available quantity of lesion, we explored the impact of reactant concentration and reaction time on the yield of cross-linking.

The yield of the dA-to-dA interstrand cross-linked DNA III was determined (phosphorimagery of DPAGE) as a function of time (25 mM formaldehyde, 25 °C, 50 mM sodium phosphate buffer, pH 6.0, 25 mM NaCl) (Figure 5). The yield of the dA-to-dA cross-link rose nonlinearly and at a decreasing rate with time out to the longest time point measured, 9 days. This is consistent with a competition between cross-linking and denaturation of the DNA substrate. The yield of dA-to-dA interstrand cross-link was essentially invariant over the concentration range of 25–250 mM formaldehyde (data not shown), consistent with the switching of the DNA between a reactive and an unreactive state as the rate-limiting step.[4] From PAGE analysis of reactions conducted at 5, 15, and 25 °C, it was apparent that the yield of dA-to-dA cross-link increased with increasing temperature over this range, but the effect was not quantified. As a compromise between high yield and minimal time investment, we thus chose a 9-day exposure of DNA to 25 mM formaldehyde at 25 °C for preparative experiments.

Covalent Structure of the Cross-link. The covalent structure of the cross-link was determined by the combination of a chemical experiment and spectroscopic measurements on samples obtained

(19) Rink, S. M.; Solomon, M. S.; Taylor, M. J.; Rajur, S. B.; McLaughlin, L. W.; Hopkins, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2551.

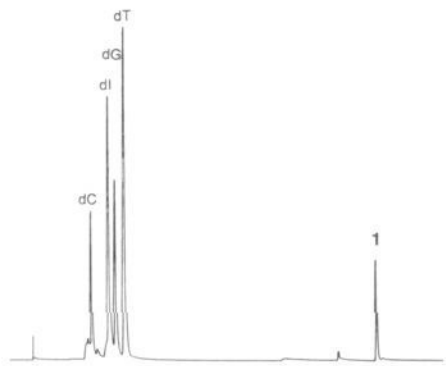


Figure 6. HPLC of the enzymatic digest of formaldehyde interstrand cross-linked DNA III (detection at 260 nm).

by enzymatic hydrolysis of the sugar–phosphate backbone of a formaldehyde-treated DNA. Two samples were important in this regard. The first of these was prepared from DNA III cross-linked dA-to-dA at 5′-d(AT).[14] This DNA, 20 OD, was exposed to formaldehyde under the above described conditions. The least electrophoretically mobile, interstrand cross-linked DNA was isolated by denaturing PAGE and exposed to a mixture of snake venom phosphodiesterase I, bovine spleen phosphodiesterase II, and calf intestinal alkaline phosphatase.[6] HPLC analysis (Figure 6) of the hydrolysate with detection at 260 nm revealed, in order of elution, dC, dI (presumably from deamination of dA by contaminating adenosine deaminase), dG, dT, and finally a strongly retained substance which we attribute to the nucleus of the cross-link. Quantitation of the released nucleosides using ϵ_{260} revealed a ratio of 1.9(dC):2.2(dG):6.0(dT):4.1(dI). Assuming that dI was derived from dA, this is one dA residue short of the calculated (and experimentally verified) starting single strand ratios of 2(dC):2(dG):6(dT):5(dA), consistent with loss of one dA residue per strand for formation of the cross-link. If one assumes as a first approximation that the molar extinction coefficient of the last-eluting substance is twice that of deoxyadenosine (15 400 M^{-1} cm^{-1} at 260 nm), it can be estimated that 0.77 mol of this lesion was detected per mole of hydrolyzed duplex DNA.

A second more abundant sample of what proved to be the same substance was obtained from DNA 5′-d$(AT)_{12}$ following formaldehyde treatment and enzymatic digestion without an intervening DPAGE separation of interstrand cross-links from residual single-stranded DNA. Thus, unlike the sample from DNA III, the sample derived from 5′-d$(AT)_{12}$ did not necessarily have its origin in an interstrand cross-link. However, that sample was considerably more conveniently available owing to the higher content of dA in that DNA, the deletion of the DPAGE step, and the simplification of the enzymatic hydrolysate to be resolved by HPLC. In the event, 11% (isolated yield, HPLC) of the deoxyadenosine residues in 5′-d$(AT)_{12}$ was converted to the strongly retained substance following treatment with 100 mM formaldehyde at pH 6.0 and 25 °C for 13 days. The elution times of the low-mobility RP HPLC peaks from DNA III and 5′-d$(AT)_{12}$ were identical to one another in two gradient solvent systems.

The covalent structure of this substance was determined essentially as previously described by Chaw et al.[6] The positive-ion electrospray-ionization mass spectrum of the strongly retained substance from either source showed a parent ion of m/e 515. Assuming that this is the M + H^+ ion derived from two deoxynucleoside residues absent a pair of hydrogens (–2 amu) and augmented by the methylene bridge (14 amu), only linkage of a pair of deoxyadenosine residues could afford this mass. The proton NMR spectrum of the strongly retained substance, obtained from 5′-d$(AT)_{12}$ and determined at 500 MHz in D_2O, required that the pair of dA residues be related by symmetry,

presumably a C_2 axis, as only eight resonances were observed. The coincidence of H^2 and H^8 in this spectrum was suggested by a singlet integrating for four hydrogens. Finally, then, the attachment point of the methylene linkage on both dA residues was established by reduction of the sample obtained from $d(AT)_{12}$ with sodium borohydride in 0.25 M KOH. HPLC analysis showed that almost all of the strongly retained substance was converted to two substances in a ratio of 4:1 identified by coininjection of authentic samples of deoxyadenosine and N^6-methyldeoxyadenosine. Taken together, these results are best accommodated by structure **1**.

Discussion

We report herein the first study of the DNA interstrand cross-linking reaction of formaldehyde in the well-defined setting of synthetic oligodeoxynucleotide duplexes. Using a panel of self-complementary DNA duplexes designed such that a variety of central "targets" for formaldehyde cross-linking were flanked by duplex termini effectively inert to interstrand cross-linking, it was shown that DNAs containing the sequence 5′-d(AT) in an AT-rich context provide dA-to-dA interstrand cross-linked DNA. DNAs containing other sequences also afforded comparable quantities of products whose DPAGE mobilities were consistent with an interstrand cross-link, but these products were electrophoretically heterogeneous, e.g., a "smear". Thus, while 5′-d(AT) appears to be an efficiently cross-linked *sequence* in some contexts, it also appears that some duplexes not containing this sequence are cross-linked in comparable efficiency. The attributes of the latter duplexes and the structure(s) of the resulting products have not been determined. DPAGE analysis of the product distribution resulting from sequence random cleavage of an interstrand cross-linked DNA duplex containing the sequence 5′-d(AT) revealed that the linkage bridged deoxyadenosine residues at this duplex sequence. Enzymatic digestion of the sugar–phosphate backbone of such a dA-to-dA, interstrand cross-linked DNA afforded a residue in which two deoxyadenosyl moieties were linked N^6-to-N^6 by a bridging methylene linkage as in **1**, one of several such nucleotide-derived substances previously isolated by Chaw et al.[6] from formaldehyde-treated nucleic acids.

Further studies will be necessary to resolve several remaining interesting aspects of the DNA–formaldehyde reaction. For example, Chaw et al.[6] have isolated from formaldehyde-treated calf thymus DNA all possible combinations of dA, dC, and dG linked through their amino groups by a methylene bridge (except dC-to-dC) in roughly similar quantities. In synthetic oligonucleotides, we find that the dA-to-dA bridge predominates as an interstrand cross-link. Do these other substances arise from denatured regions of calf thymus DNA, the abundance of which may increase during the longer exposure times in that study (40 days)? Alternatively, if these substances arose from intrastrand cross-links, they would not be detected in the present study because only interstrand cross-linked products were analyzed. These substances may figure prominently in accounting for the heterogeneous family of interstrand cross-linked products observed in this study in DNAs not containing a 5′-d(AT) sequence in an AT-rich context.

A second issue deserving further experimental study is that of the mechanistic origin of the accumulation of dA-to-dA cross-links at the 5′-d(AT) sequence. The kinetics and thermodynamics of monoadduct formation and progression of monoadducts to cross-links are of obvious relevance in this regard. It is possible that the selectivity for dA-to-dA cross-links observed in this study has its origin in the known preferential formation of monoadducts of dA in AT-rich regions of DNA, likely a consequence of the lower kinetic and thermodynamic stability of AT-rich regions of DNA affording a higher concentration of some non-ground-state conformation of DNA which reacts with formaldehyde.[4] Given this situation, it is then natural to ask why the sequence 5′-d(AT) is cross-linked in preference to, for example, 5′-d(TA), an alternative sequence which brings the reactive N^6 amino groups of two deoxyadenosine residues into close spatial proximity. Specificity of monoadduct formation alone cannot account for the selection of cross-linking at 5′-d(AT) in preference to 5′-d(TA), given the observation that the sequence 5′-d(TATA) shows 5′-d(AT) selectivity. Assuming a stepwise process, a monoadduct at the sequence 5′-d(TAT) must have been an intermediate. If the progression of monoadduct to cross-links was sequence random, then monoadducts at 5′-d(TAT) would yield dA-to-dA cross-links at both 5′-d(AT) and 5′-d(TA). Because this is not the case, some combination of a kinetic preference for closure to or a relatively greater thermodynamic stability of the cross-link at 5′-d(AT) must obtain.

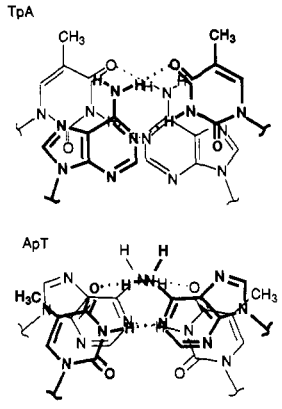


Figure 7. View down the helix axis of B-DNA at 5′-d(TA) (upper) and 5′-d(AT) (lower) steps.

Inspection of the local structure of 5′-d(AT) and 5′-d(TA) steps in B-DNA suggests two parameters which impact that spatial relationship of functional groups in the biopolymer and may contribute to the observed selectivity for cross-linking of 5′-d(AT). Firstly, the handedness and magnitude of the helical twist of B-DNA result in both quantitative and qualitative differences in the geometric relationship of the reactive N^6 amino groups of deoxyadenosine at these two steps. Inspection of a helical projection of these two sequences in B-DNA[7] (see Figure 7) reveals the nitrogen atoms of the reactive amino groups which become linked to be nearly superimposed in the sequence which is cross-linked, 5′-d(AT), but more isolated from one another at the 5′-d(TA) sequence. (Propeller twist, discussed below, further exaggerates this effect.) This relationship is not altered if one considers the best currently available, high-resolution solution structure of B-DNA from proton NMR studies[20] rather than that from the Arnott model based on fiber diffraction measurements.[7] Furthermore, the hydrogen atoms of these amino groups not involved in Watson–Crick base pairing and thus presumably replaced by the bridging methylene group are closer to one another in the 5′-d(AT) sequence. Based upon this observation, it is a straightforward exercise to construct kinetic (rate at which an iminium ion or O-protonated methylol derivative captures the neighboring amino group) or thermodynamic arguments favoring cross-linking at 5′-d(AT) in preference to that at 5′-d(TA) owing to minimal structural distortion of the biopolymer. A closely related argument has been made to account for the preferential dG-to-dG cross-linking observed at 5′-d(CG) in preference to that at 5′-d(GC) upon exposure of B-DNA to nitrous acid.[11]

The second factor is propeller twist. The negative propeller twist commonly observed in B-DNA[15] diminishes the spacing between amino groups at 5′-d(AT) but increases this spacing at 5′-d(TA) (see Figure 8). (Propeller twist is negative if, when viewed along the long axis of the base pair, the near base is rotated

(20) Cheng, J.-W.; Chou, S.-H.; Salazar, M.; Reid, B. R. *J. Mol. Biol.* **1992**, *228*, 118.

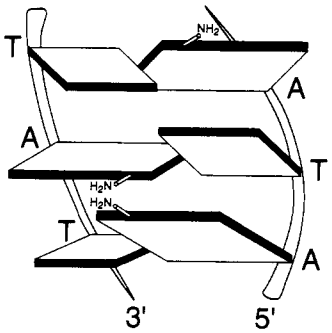


Figure 8. Schematic depiction of a view into the major groove of B-DNA emphasizing the impact of negative propeller twisting (exaggerated for emphasis) on the location of the exocyclic amino groups of deoxyadenosine residues on opposite strands at the sequences 5'-d(TA) and 5'-d(AT).

clockwise relative to the distant base.[21]) It could be that the propeller twist, acting in this case in concert with the helical twist, assists in bringing reactive groups in the ground state into close proximity favoring reaction. That this effect may be quite important is suggested by the observation that in every well-characterized case of an interstrand cross-link in which the bridged atoms of B-DNA must approach one another to satisfy the bond lengths and angles of the cross-linkage, the simplest structural model for the perturbation involves a hyperextended *negative* propeller twisting. These four cases span an impressive range of sites on DNA including one minor groove cross-link (nitrous acid[11]), three major groove cross-links (mechlorethamine,[10] formaldehyde,[14] and cisplatin[16]), four different preferred nucleotide sequences (dG-to-dG at 5'-d(CG), dG-to-dG at 5'-d(GNC), dA-to-dA at 5'-d(AT), and dG-to-dG at 5'-d(GC), respectively, and three different reactive sites within the deoxynucleosides (N^7 and N^2 of dG and N^6 of dA). A critical test of the hypothesis, a case in which helical twist and propellor twist fail to operate in concert, is currently unavailable.

Experimental Section

Material and Methods. Materials and their sources were as follows: DNA synthesis reagents, Applied Biosystems; [γ-^{32}P]ATP, New England Nuclear; Klenow fragment and T4 polynucleotide kinase, Amersham; phosphodiesterase I (crotalus adamanteus venom), Pharmacia; phosphodiesterase II (bovine spleen), Sigma; alkaline phosphatase (calf intestinal), Pharmacia; formaldehyde, J. T. Baker; and N^6-methyldeoxyadenosine, Sigma. Water was purified on a Millipore Milli-Q deionizer. Oligonucleotides were synthesized on an Applied Biosystems Model 380A or 392 synthesizer. All other chemicals were of commercial origin and used as received. Samples were concentrated in vacuo with a Savant speed vacuum concentrator. Except for purification of synthetic oligonucleotides, all PAGE was conducted on a Hoeffer thermojacketed Poker Face gel stand. Formamide dye solution was a 9:1 mixture (v/v) of deionized formamide and 100 mM Tris (pH 7.5), 1% xylene cyanole, and 1 mM EDTA. Gels were dried by using a Bio-Rad Model 583 gel drier onto Whatman 3M paper; autoradiograms were with Amersham Hyperfilm-MP. Phosphorimaging used a Molecular Dynamics 400A phosphorimager. All scanning operations and data display and analysis were performed using Molecular Dynamics' ImageQuant software operating on Intel 80386 or 80486 microprocessers. UV spectra were measured on a Hewlett-Packard 8452A spectrophotometer. Optical density measurements were made on a Perkin-Elmer Lambda 3A spectrophotometer. Analytical and preparative HPLC were performed on an Alltech, C18, 5-μm, 250- × 4.6-mm (analytical scale) or an Alltech Macrosphere 300, C18, 7-μm, 250- × 10-mm (preparative scale) column using SSI 200B pumps controlled by an SSI controller and a sequential SSI 500 UV/vis (output to both an HP 3390A electronic integrator and a Linear Model 255/MM recorder) detector. 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); solvent B, CH_3CN; isocratic 92% A for 7 min, 13-min linear gradient to 70% A, 10-min linear gradient to 60% A, and then 10-min linear gradient to initial conditions. Gradient B: the same program except solvent A was 0.1 mM ammonium acetate (pH 7.0). The ^{1}H NMR spectrum was determined on a Bruker AM 500 (500 MHz) spectrometer and is reported in parts per million downfield from external DSS (0.00 ppm). Electrospray-ionization mass spectra (ESMS) were measured on a Sciex Atmospheric Pressure Ionization (API) triple quadrapole mass spectrometer.

Preparation of Singly Phosphorylated, 5'-^{32}P-Radiolabeled DNA Duplexes. Oligonucleotides (1-μmol scale synthesis) were purified by denaturing PAGE (ca. 40 OD_{260} of crude DNA, 20% polyacrylamide, 19:1 acrylamide:bisacrylamide, 8 M urea, 1.5-mm thick, 14 × 16 cm, using a 5-toothed comb), run until the xylene cyanole dye had traveled 8–9 cm from the origin. DNA was visualized by UV shadowing. DNA was isolated from the gel by a crush and soak procedure: gel slices were crushed with a glass rod into fine particles and incubated at 37 °C for at least 4 h in 0.5 M aqueous NH_4OAc and 1 mM aqueous EDTA. The elutant 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) 3 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 0.2 OD_{260} of DNA (8 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), 10 μCi of [γ-^{32}P]ATP, and 10 units of T4 polynucleotide kinase at 37 °C for 30 min. Radiolabeling was stopped by heating the reactions at 68 °C for 10 min. In all cases, unincorporated radionucleotides were removed by ethanol precipitations.

Preparation of Doubly Phosphorylated, 5'-^{32}P-Radiolabeled DNA Duplexes. Oligonucleotides were radiolabeled as described above under the preparation of radiolabeled DNA duplexes except that after incubating the reaction at 37 °C for 30 min, 6 μL of 5 mM aqueous ATP and 1 μL of 10× kinase buffer were added. The reaction was incubated at 37 °C for an additional 20 min and then stopped, and ethanol precipitated as described above under the preparation of radiolabeled DNA duplexes.

Preparation of Formaldehyde Cross-Linked DNA Duplexes. The following conditions were used for analytical experiments. Preparative conditions refers to a 100-fold-scale increase with the exception of the quantity of the radiolabel which remained unchanged. Synthetic radiolabeled DNA, 0.2 OD_{260} (8 nmol of base pairs), in 50 μL of 50 mM aqueous Na_xPO_4 buffer (pH 6.0), 25 mM aqueous NaCl, and 25 mM aqueous formaldehyde was incubated at 25 °C for 9 days. The DNA solution from the cross-linking mixture was ethanol-precipitated, the supernatant was removed, and the pellet was dried. Pelleted DNA was dissolved in 10 μL of purified water, and an equal amount of loading buffer (90% deionized formamide, 10 mM Tris (pH 7.5), 0.1% xylene cyanole, and 0.1 mM EDTA) was added. The resulting solution was heated at 90 °C for 2 min, iced for 2–3 min, and analyzed by denaturing PAGE. Denaturing PAGE was conducted on a 25% gel (19:1 acrylamide:bisacrylamide, 8 M urea, 0.35 mm thick, 33 × 41 cm, using a 20-toothed comb) at 65–70 W and at ca. 55 °C until xylene cyanole had run 14–16 cm. If necessary, the gel was transferred onto filter paper covered with Saran Wrap and dried for 50 min followed by autoradiography or phosphorimagery. Autoradiography was used to visualize the single-stranded and cross-linked DNA (roughly half the mobility of the corresponding single strand).

Excision of Cross-Linked DNA Bands. Gels slices of an approximate dimension of 1.5 × 0.5 cm containing cross-linked DNA were excised from the above-described 25% denaturing polyacrylamide gels. DNA was eluted from the gel slice by using a crush and soak procedure. The DNA-containing supernatant was loaded onto a Sep-Pak and treated as described above under the preparation of radiolabeled DNA duplexes. Cross-linked DNA was recovered by concentration of the acetonitrile/water eluant.

DNA Cleavage Reactions. Iron(II) EDTA cleavage of DNA was generally performed as described by Tullius and Dombroski[23] with the following modifications. To a solution of DNA in 7 μL of 10 mM Tris (pH 7.5) and 10 mM NaCl was added 2 mL each of 10 mM aqueous ascorbic acid, 0.3% aqueous H_2O_2, and 1 mM aqueous $(NH_4)_2Fe(SO_4)_2$. After 1 min at 25 °C, the reaction was stopped by the addition of 2 μL of 0.1 M aqueous thiourea. These samples, along with comparison lanes containing uncut, cross-linked, and native (uncross-linked) DNA and a Maxam–Gilbert guanine-specific sequencing reaction[22] on native DNA,

(21) Dickerson, R. E. *J. Biomol. Struct. Dyn.* **1989**, *6*, 627.

(22) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499–560.
(23) Tullius, T. D.; Dombroski, B. A. *Science* **1985**, *230*, 679.

were dried and analyzed by denaturing PAGE as described above under Preparation of Formaldehyde Cross-Linked DNA Duplexes.

Yield Optimization of Cross-Linking. Synthetic, radiolabeled DNA III (0.2 OD; 8 nmol of base pairs) was dissolved in 25 μL of 100 mM aqueous Na_xPO_4 buffer (pH 6.0). The reactions were initiated by addition of an equal volume of 50 mM formaldehyde at times such that the reactions which had proceeded for 0.5, 1, 2, 4, 6, and 9 days could all be stopped simultaneously by ethanol precipitation. The supernatant was discarded, and the pellet was washed with cold (–20 °C) ethanol, which was in turn discarded. Ethanol precipitation involved addition of 10 μL of 3 M sodium acetate (pH 5.4) and 1.0 mL of –20 °C 90% aqueous ethanol, and the mixture was allowed to stand at –20 °C for 20 min. The cool suspension was centrifuged for 15 min (4 °C), and the supernatant was removed and discarded. The pellet was analyzed by denaturing PAGE as described above under Preparation of Formaldehyde Cross-Linked DNA Duplexes.

Enzymatic Hydrolysis, Quantitation of Nucleosides, and Isolation of Bis(N^6-deoxyadenosyl)methane (1) from DNA III. Cross-linked (0.1 OD_{260}) or synthetic DNA III (0.1 OD_{260}) in 30 μL of aqueous 10 mM $MgCl_2$ and 50 mM Tris (pH 8.9) was treated with 1 μL (1 unit) of alkaline phosphatase, 3 μL (2 units) of phosphodiesterase I, and 1 μL (0.2 unit) of phosphodiesterase II at 37 °C for 4 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, dG, dT, and dI (dA; because of presumably contamination of deaminase in phosphodiesterase II, all dA was transferred to dI) at 260 nm which were as follows: dC, 1.0; dG, 1.6; dT, 1.2; and dI, 1.1. Using these response factors, the uncross-linked DNA analyzed as 2.2-(dC):2.2(dG):6.0(dT):5.2(dA) (calcd 2.0:2.0:6.0:5.0); cross-linked DNA analyzed as 1.9(dC):2.2(dG):6.0(dT):4.1(dA) (calcd 2.0:2.0:6.0:4.0).

Compound **1** was isolated from the hydrolysate (30 μL) by HPLC using gradient B, the analytical column, and a 1 mL/min flow rate. The most strongly retained peak, with a retention time of 23 min, was collected and subjected to UV and electrospray MS analyses, as described below.

Isolation of Bis(N^6-deoxyadenosyl)methane (1) from 5′-d(AT)$_{12}$. Crude DNA d(AT)$_{12}$ (100 OD_{260}, 4 mmol of base pairs) in 20 mL of 50 mM K_xPO_4 buffer (pH 6.0) and 100 mM formaldehyde was incubated at 37 °C for 13 days. DNA was recovered by passing it through a Waters Sep-Pak C_{18} cartridge, as described above under the preparation of radiolabeled DNA duplexes. The desalted DNA was subjected to enzymatic hydrolysis as described above under the enzymatic hydrolysis and quantitation of nucleosides by HPLC except in 50× scale. HPLC separation of **1** was carried out on 50 μL of the above reaction mixture using gradient B, the preparative scale column, and a 2 mL/min flow rate. The peak with the 25.7-min retention time was collected and subjected to UV, electrospray MS, and ^{1}H NMR analyses.

Borohydride Reduction of Bis(N^6-deoxyadenosyl)methane (1). Compound 1, 0.5 OD_{260} (isolated from 5′-(AT)$_{12}$) was dissolved in 400 μL of 0.25 M KOH. Sodium borohydride (100 mg) was added, and the mixture was stirred at 60 °C for 30 h. The reaction was stopped by the addition of glacial acetic acid dropwise until bubbling ceased, and the mixture was analyzed by HPLC in analytical scale using gradient A. After 30 h, **1** had been converted to deoxyadenosine and N^6-methyldeoxyadenosine in a ratio of 4:1. The products were identified by coinjecting authentic, commercial samples with the reaction mixture.

Spectroscopic Analysis of Bis(N^6-deoxyadenosyl)methane (1). Samples of **1** from both DNA III and 5′-d(AT)$_{12}$ were prepared for electrospray MS by the addition of 5 μL of formic acid followed by 100 μL of 1:1 water/methanol. The electrospray was run with an inlet voltage of 240 V and a needle voltage of 5000 V; the sample was injected at a rate of 3 μL/min. The reported spectrum is an average of 10 scans. ESMS: 515 (1 + H^+), 399 (1 + $2H^+$ – deoxyribosyl$^+$), 264 (dA – H^+ + CH_2^{2+}), 252 (dA + H^+). The NMR sample (7 OD_{260}) of **1** (isolated from 5′-d(AT)$_{12}$) was repeatedly dried from 99.9% D_2O in a speed vacuum and then dissolved in 0.45 mL of 99.96% D_2O. ^{1}H NMR (500 MHz, D_2O, external DSS): δ 2.47 and 2.72 (4H, m, H-2′), 3.71 and 3.77 (4H, m, H-5′), 4.11 (2H, m, H-4′), 4.57 (2H, m, H-3′), 5.32 (2H, s, N^6–CH_2–N^6), 6.36 (2H, s, H-1′), 8.22 (4H, s, H-2 and H-8). UV spectrum: λ_{max} 278 (pH 0.8, aqueous HCl), 272 (pH 6.7, H_2O), 273 nm (pH 12.1, aqueous NaOH).

Acknowledgment. This work was supported by the NSF (CHE-9019129 and DIR-8220099) and NIH (AG00417). Phosphorimagery was performed by the Phosphorimager Facility, Markey Molecular Medicine Center, UW. P.B.H. is a Cope Scholar. We thank Stephen Alley for editorial assistance and Leiming Zhu for assistance with computation.