The Purine 2-Amino Group as the Critical Recognition Element for Sequence-Specific Alkylation and Cross-Linking of DNA by Mitomycin C

Maria Tomasz,*,† Arunangshu Das,† Kit S. Tang,‡ Marjin G. J. Ford,‡ Andrew Minnock,‡ Steven M. Musser,§ and Michael J. Waring‡

Contribution from the Department of Chemistry, Hunter College, City University of New York, New York, New York 10021, U.S. Food and Drug Administration, Washington, D.C. 20204, and Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK

Received July 8, 1998

Abstract: Mitomycin C (MC) is a G·C-specific antitumor antibiotic that alkylates and cross-links DNA at 2-amino groups of guanine residues. Both reactions are known to be enhanced at guanines in the CpG sequence by independent mechanisms. The mechanisms were probed by substituting 2,6-diaminopurine (D) into oligonucleotides and into a 162-bp *tyr*T DNA restriction fragment and determining their alkylation and cross-linking by MC. Covalent D-MC adducts were isolated and structurally characterized. The results indicated that 2,6-diaminopurine functioned as a substrate analogue of guanine and displayed enhanced reactivity toward MC in all systems. The observed TpD sequence selectivity of the modifications by MC was analogous to the CpG sequence selectivity of guanine modifications. Selective monoalkylation and cross-linking was observed also at the TpG·CpD sequence, indicating that two purine 2-amino groups are necessary and sufficient for the selectivity regardless of whether they are supplied by G or by D. These findings reinforce the previously proposed mechanism in which the selectivity of monoalkylation by MC is attributed to a specific H-bond between the drug and the 2-amino group of a guanine. The specific sequence required for D−D and D−G cross-link formation was established as TpD·TpD and TpG·CpD, respectively, determined by the same minor groove structural factors as in the CpG·CpG cross-linking process. MC cross-linking was also probed in a 162-bp *tyr*T DNA fragment in which Gs and As were replaced by inosine and 2,6-diaminopurine, respectively, using PCR. Cross-linked sites were quantitated and mapped on the basis of an empirical correlation between the electrophoretic mobility of cross-linked DNA and the position of the cross-links relative to the center of the sequence. In the natural DNA sequence hotspots for formation of MC-cross-links were identified. The cross-links were shown to be translocated from CpG·CpG in the natural DNA to TpD·TpD or TpG·CpD in the substituted DNAs, demonstrating the dominant role of the purine 2-amino group for cross-link site selection by MC.

Introduction

A variety of DNA-interactive natural products display specificity for G·C at their binding sites in the minor groove of the double helix. The mechanism of this type of recognition usually involves the guanine 2-amino group, as can be demonstrated by the abolition of binding when guanine residues are replaced by inosine.[2] A different, positive test for the role of the 2-amino group of guanine was invented recently by Waring and Bailly[3] featuring the use of DNA in which adenines were substituted by 2,6-diaminopurine (DAP). DAP, like adenine, forms a specific base pair with thymine. The DAP-substituted DNA remains in the B-conformation[4] and has increased duplex stability.[5] The minor groove is widened at DAP·T basepairs, as compared to A·T basepairs. In this respect DAP·T mimics the G·C basepair. However, the major groove remains unaltered by the A→DAP substitution.[6] Using such DNA as substrate led to a change from G·C to DAP·T at the preferred binding site of numerous G·C-specific drugs. Echinomycin, a CpG-specific ligand, and actinomycin, specific to GpC, presented two dramatic examples of such relocation of G·C-specific binding to DAP·T sites, proving that the main determinant for the G·C specificity in these cases is simply the purine 2-amino group in the minor groove.[2b,3b] The test was

† City University of New York.
‡ University of Cambridge.
§ U.S. Food and Drug Administration.

(1) Abbreviations: AP, alkaline phosphatase; CD, circular dichroism; CE, cyanoethyl; DAP, 2,6-diaminopurine; D, 2,6-diaminopurine residue in a nucleic acid base sequence; dDAP, 2,6-diaminopurine-2′-deoxyriboside; E, molar extinction coefficient; dG, guanine-2′-deoxyriboside; D-M-G, DAP-MC-guanine bisadduct incorporated in an oligonucleotide sequence; dN, 2′-deoxyribonucleoside; ESIMS, electrospray ionization mass spectroscopy; MC, mitomycin C; M-D, MC-DAP monoadduct incorporated in an oligonucleotide sequence; M-G, MC-guanine monoadduct incorporated in an oligonucleotide sequence; SVD, snake venom phosphodiesterase; TEA, triethylammonium acetate.

(2) Marchand, C.; Bailly, C.; McLean, M. J.; Moroney, S. E.; Waring, M. J. *Nucleic Acids Res.* **1992**, *20*, 5601−5606. Iennewein, S.; Waring, M. J. *Nucleic Acids Res.* **1997**, *25*, 1502−1510.

(3) Waring, M. J.; Bailly, C. *Gene* **1994**, *149*, 69−79. Bailly, C.; Marchand, C.; Waring, M. J. *J. Am. Chem. Soc.* **1993**, *115*, 3784−3785. Bailly, C.; Payet, D.; Travers, A. A.; Waring, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13623−13628.

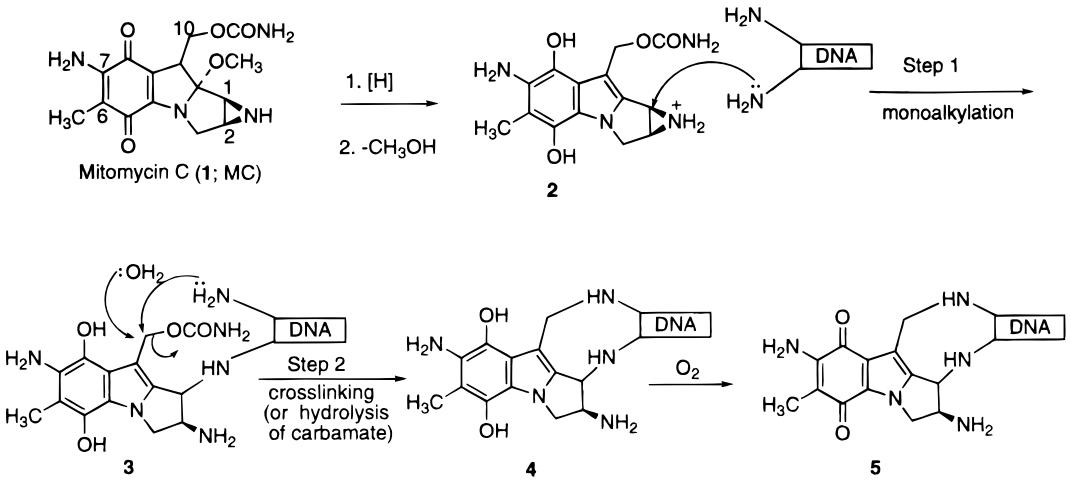
(4) Chazin, W. J.; Rance, M.; Chollet, A.; Leupin, W. *Nucleic Acids Res.* **1991**, *19*, 5507−5513.

(5) Hoheisel, J. D.; Lehrach, H. *FEBS Lett.* **1990**, *274*, 103−106.

(6) Bailly, C.; Mollegaard, N. E.; Nielsen, P. E.; Waring, M. J. *EMBO J.* **1995**, *14*, 2121−2131.

10.1021/ja9824019 CCC: $15.00
Published on Web 10/28/1998

Scheme 1. Mechanism of Reductive Activation of MC and Cross-Linking of DNA



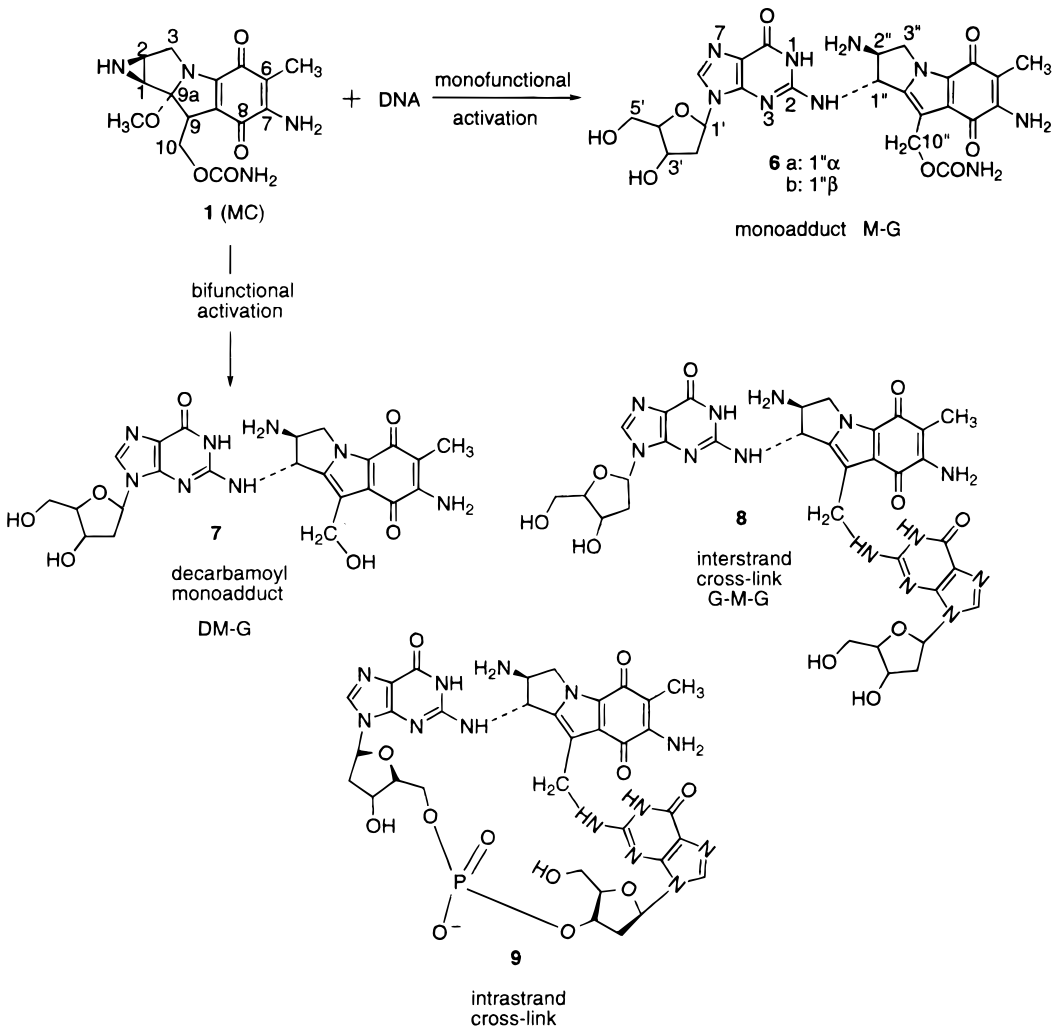
also used to elucidate unknown mechanisms of specific DNA sequence recognition by drugs.[7]

Mitomycin C (MC; **1**), a G·C-specific antitumor antibiotic,[8] alkylates DNA bifunctionally, and this results in covalent cross-links between the complementary DNA strands.[9] Both of its covalent alkylating functions, namely the C-1 aziridine and the C-10 carbamate, target exclusively the 2-amino group of guanine.[10,11] The bifunctional alkylating activity of MC is triggered by reduction of its quinone system as illustrated in Scheme 1. Once MC is activated, the C-1 aziridine function reacts first by alkylating the 2-amino group of a guanine residue in one DNA strand. The monoalkylated intermediate, adduct **6a**, can be isolated provided that the reaction is arrested at the stage of monofunctional activation of MC (Scheme 2). In the second step the C-10 carbamate function completes the bifunctional alkylation process by reacting with the 2-amino group of a second, suitably positioned guanine; this step gives rise to cross-linked DNA.[12] The interstrand and intrastrand cross-link adducts **8** and **9** have been isolated together with monoadduct **7**.[11,13,14] The latter is a product of hydrolysis at the C-10 carbamate (Step 2 in Scheme 1).

The 2-amino group of guanine is intrinsically the most reactive site toward activated MC (**2**) among the nucleophilic centers presented by the four bases in DNA, even when the reactivities are compared among simple nucleotides. However, the reactivity of this group with MC is strongly modulated by the DNA sequence at both covalent steps. Thus, the monoalkylation of guanine (Step 1) is greatly enhanced at the CpG sequence as compared to GpG, ApG, and TpC;[15,16] furthermore, the second step (cross-link formation) has an absolute specificity for the same CpG·CpG duplex site.[17−19] The origin of these DNA-sequence specificities of the alkylating action of MC was proposed to be as follows.[16,20] MC initially recognizes its preferred alkylation sites by identifying *two* guanine 2-amino groups simultaneously, one in each strand, as depicted in Figure 1a. One amino group holds MC by a specific H-bond to the 10-oxygen of the carbamate group while the other attacks the antibiotic covalently at its primary active site (the 1-position). The H-bond thereby facilitates the covalent reaction through its binding energy. Consequently, alkylation of guanine is faster at CpG than at other NpGs where such H-bonding is lacking. This hypothesis is experimentally supported by observations of alkylation of inosine-substituted oligonucleotides[16] and by the solution structures of MC monoadduct-DNA complexes determined by NMR.[21] Computer modeling is also consistent with this mechanism.[22] The observed selectivity for CpG is 5-fold[20] or 5−20-fold[16] over other NpG sequences. A much more stringent CpG specificity is involved in the second, cross-linking step, however, and it is independent of the above H-bond mechanism. Rather, the specific orientation of the precursor monoadduct in the minor groove is the decisive structural factor. This places the 10-carbamate pointing upstream *at all four NpG sequences*. At the CpG·CpG sequence this orientation results in a favorable juxtaposition of the second active site (C-10) of the MC monoadduct and the 2-amino group of the guanine in the opposite strand, leading to cross-link formation (Figure 1b). At ApG·CpT and TpG·CpA sequences only water will react with the reactive site of the MC monoadduct while at GpG·CpC steps intrastrand cross-linking is observed.[19] The NMR-derived solution structures of monoalkylated and cross-linked oligonucleotides strongly support this mechanism.[21,23] It has been speculated that the specific H-bonding of MC at CpG steps (Figure 1a) has evolved as a mechanism for pre-selecting cross-linkable guanines in DNA: considering that cross-links are more

(7) Bailly, C.; Waring, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 7311−7316.
(8) Hata, T.; Sano, Y.; Sugawara, R.; Matsume, A.; Kanamori, K.; Shima, T.; Hoshi, T. *J. Antibiot. Ser.* **1956**, *A9*, 141−46.
(9) Szybalski, W.; Iyer, V. N. *Fed. Proc.* **1964**, *23*, 946−957.
(10) Tomasz, M.; Chowdary, D.; Lipman, R.; Shimotakahara, S.; Veiro, P.; Walker, V.; Verdine, G. L. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6702−6706.
(11) Tomasz, M.; Lipman, R.; Chowdary, D.; Pawlak, J.; Verdine, G. L.; Nakanishi, K. *Science* **1987**, *235*, 1204−1208.
(12) Tomasz, M.; Chawla, A. K.; Lipman, R. *Biochemistry* **1988**, *27*, 3182−3187.
(13) Bizanek, R.; McGuinness, B. F.; Nakanishi, K.; Tomasz, M. *Biochemistry* **1992**, *31*, 4−3091.
(14) Tomasz, M.; Lipman, R.; McGuinness, B. F.; Nakanishi, K. *J. Am. Chem. Soc.* **1988**, *110*, 5892−5896.
(15) Li, V.; Kohn, H. *J. Am. Chem. Soc.* **1991**, *113*, 275−283.
(16) Kumar, S.; Lipman, R.; Tomasz, M. *Biochemistry* **1992**, *31*, 1399−1407.
(17) Weidner, M. F.; Millard, J. T.; Hopkins, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 9270−9272.
(18) Teng, S. P.; Woodson, S. A.; Crothers, D. M. *Biochemistry* **1989**, *28*, 3901−3907.
(19) Borowy-Borowski, H.; Lipman, R.; Chowdary, D.; Tomasz, M. *Biochemistry* **1990**, *29*, 2992−2999.
(20) Kohn, H.; Li, V.-S.; Tang, M-s. *J. Am. Chem. Soc.* **1992**, *114*, 5501−5509.
(21) Sastry, M.; Fiala, R.; Lipman, R.; Tomasz, M.; Patel, D. J. *J. Mol. Biol.* **1995**, *247*, 338−359.
(22) Remers, W. A.; Rao, S. N.; Wunz, T. P.; Kollman, P. *J. Med. Chem.* **1988**, *31*, 1612−1620.
(23) Norman, D.; Live, D.; Sastry, M.; Lipman, R.; Hingerty, B. E.; Tomasz, M.; Broyde, S.; Patel, D. J. *Biochemistry* **1990**, *29*, 2861−2876.

Scheme 2. Structures of Adducts of MC and Deoxyguanosine Formed in DNA under Monofunctional and Bifunctional Activation



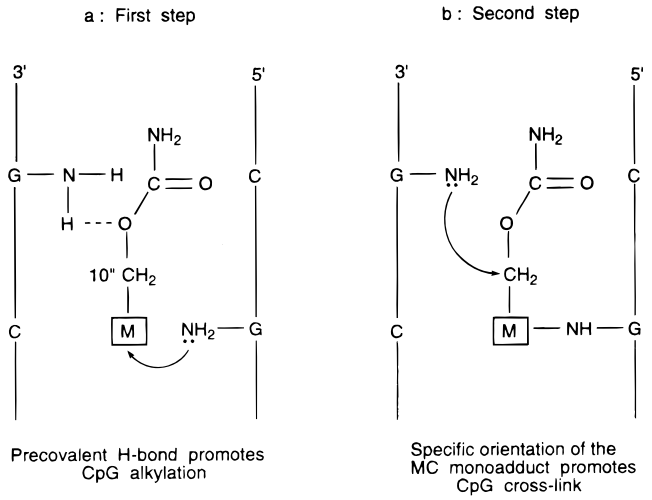


Figure 1. Distinct mechanisms of the CpG selectivities of (a) the first and (b) second DNA alkylation steps by MC.

lethal than monoadducts this mechanism maximizes the potency of MC by directing the available MC to sites where it does the most damage.[16]

According to this mechanism, recognition of DNA by activated MC relies on the antibiotic interacting with 2-amino groups of guanine at each of the three steps of the cross-linking trajectory, namely the pre-covalent binding via the specific H-bond and the two consecutive covalent reactions, shown in Scheme 1. It occurred to us that this unique mechanism could be put to the test by using the system of Waring and Bailly, i.e., artificial DNA, which lacks guanine but contains 2,6-diaminopurine (DAP)[3] as substrate for MC alkylation. Accordingly, our objective was to determine whether DAP-substituted DNA would be monoalkylated and cross-linked by MC and would follow the pattern of specificities expected from the above mechanism. In this endeavor, we utilized a 162-bp *tyr*T DNA restriction fragment and its DAP- and inosine-substituted variants, prepared by PCR, to probe the substrate specificities of cross-linking by MC. Since a study of *covalent* recognition necessarily involves characterization of the covalent adducts between drug and DNA, experiments were also conducted with a series of synthetic oligonucleotides as substrates. These substrates served as convenient sources for the isolation and characterization of MC-DAP adducts. The observed alkylation specificities of MC in the *tyrT* DNA and oligonucleotide systems together confirm the proposed MC-DNA recognition mechanism and extend its validity to natural DNA. The covalent reactions of MC with DAP and their sequence specificities are analogous to those with guanines. MC also forms mixed cross-links between guanine and DAP. However, DAP is significantly more reactive than guanine toward MC.

Chart 1. Structures of Adducts of MC and DAP Deoxyribonucleoside

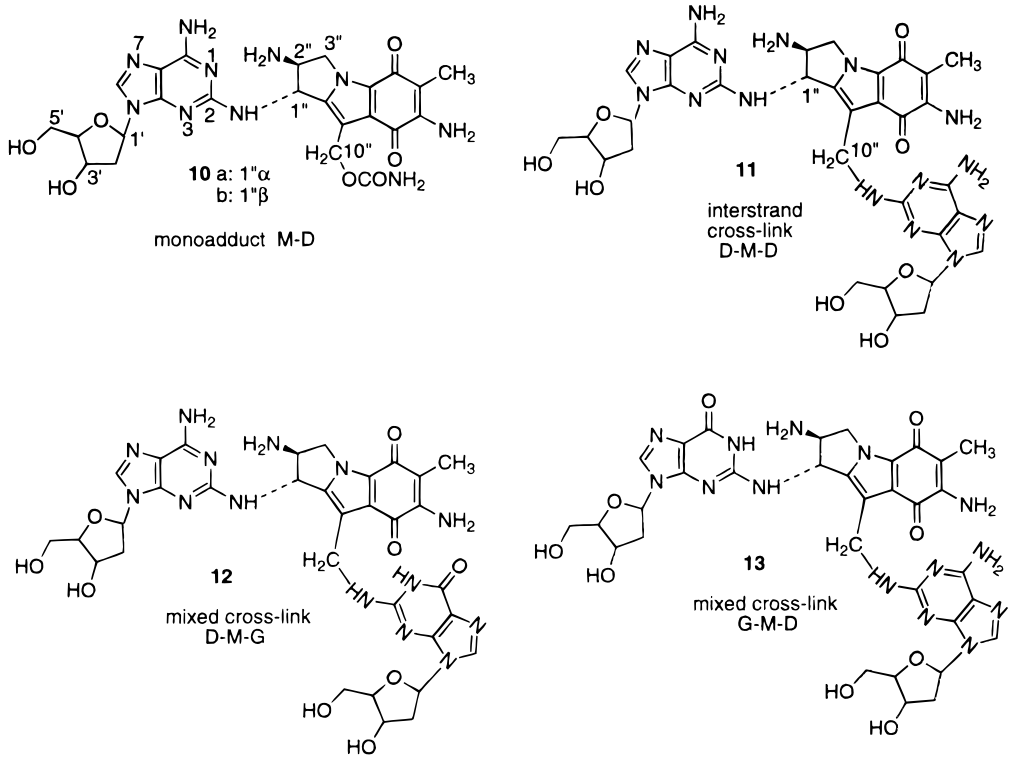


Chart 2. Structures of Oligonucleotides[a]

14 a:	5'-A C T A T A T T C A	**19 a:**	5'-A C T A C D T T C A
b:	3'-T G A T A T A A G T	**b:**	3'-T G A T G T A A G T
15 a:	5'-A C T A T D T T C A	**20 a:**	5'-A T A T A C G T A T A T
b:	3'-T G A T A T A A G T	**b:**	3'-T A T A T G C A T A T A
16 a:	5'-A C T A T D T T C A	**21 a:**	5'-A T D T A C G T A T D T
b:	3'-T G A T D T A A G T	**b:**	3'-T D T A T G C A T D T A
17 a:	5'-A C T A D T T T C A	**22:**	5'-A C T A T D T T C A
b:	3'-T G A T T D A A G T		
18 a:	5'-A T T A C G T T T A	**23:**	5'-A T T A C G T T T A
b:	3'-T A A T G C A A A T		

[a] The rectangle-enclosed sequences indicate the critical variants in the oligonucleotide sequence.

Results

Formation and Structural Characterization of Covalent MC-DAP Monoadducts and Bisadducts in Oligonucleotides. Incubation of activated MC with the DAP-containing oligonucleotides listed in Chart 2 resulted in monofunctionally alkylated oligonucleotides which could be isolated by HPLC of the reaction mixtures. For example, the complementary oligonucleotide substrates **16a** and **16b** as a duplex were converted almost completely to mitomycin-substituted forms **16a-M** and **16b-M** (Figure 2a). A covalent MC-DAP monoadduct **10a** was isolated by HPLC after digestion of the modified oligonucleotides or the unfractionated reaction mixture to nucleosides (Figure 3a). Its structural assignment is based on the following: The UV spectrum of **10a** is a 1:1 composite of that of standard dDAP and the standard mitosene chromophore,

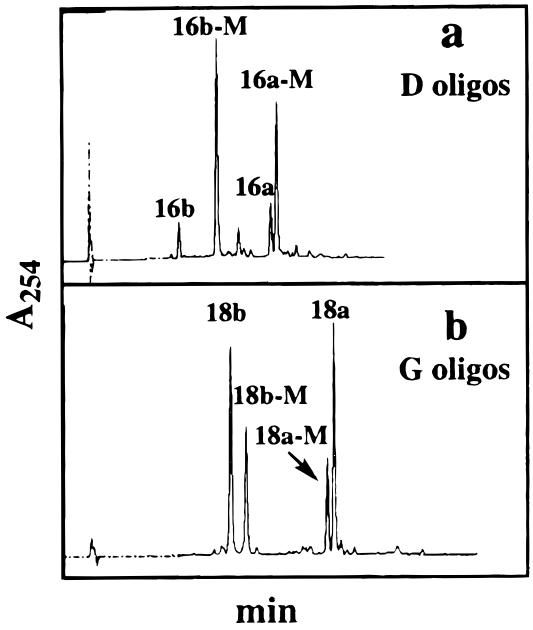


Figure 2. HPLC of oligonucleotides monoalkylated by MC: (a) monoalkylation reaction mixture of duplex **16** and (b) monoalkylation reaction mixture of duplex **18**. Peaks marked **16a**-M and **16b**-M (a) and **18a**-M and **18b**-M (b) correspond to the alkylated oligonucleotide strands. HPLC conditions: C-4 column (Beckman Ultrasphere-C-4; 4.5 × 250 mm); 6% to 11% acetonitrile in 0.1 M TEA, pH 7.0 buffer in 60 min; 1.0 mL/min flowrate.

as in 2,7-diaminomitosene.[25] Electrospray ionization mass spectrometry (ESIMS) verified its composition (calcd MH^+ 569; found 569). This indicated that in the adduct the C-10″ carbamate is intact and therefore the linkage site to the nucleoside must be the C-1″ position. Circular dichroism

(24) Guo, M.; Hildbrand, S.; Leumann, C. J.; McLaughlin, L. W.; Waring, M. J. *Nucleic Acids Res.* **1998**, *26*, 1863−1869.

(25) Tomasz, M.; Lipman, R. *Biochemistry* **1981**, *20*, 5056−5061.

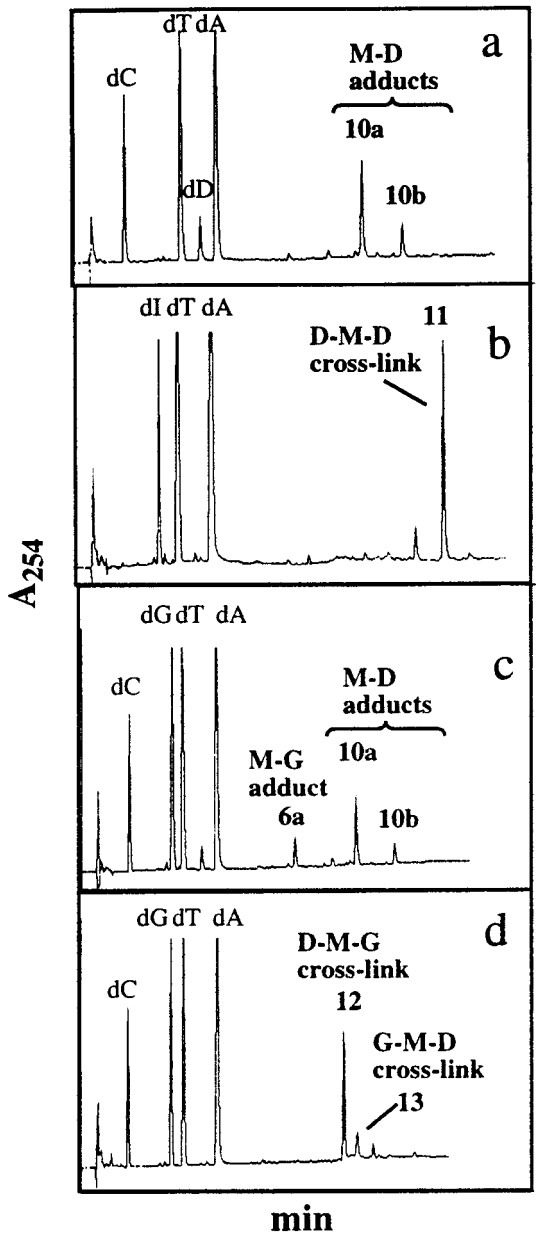
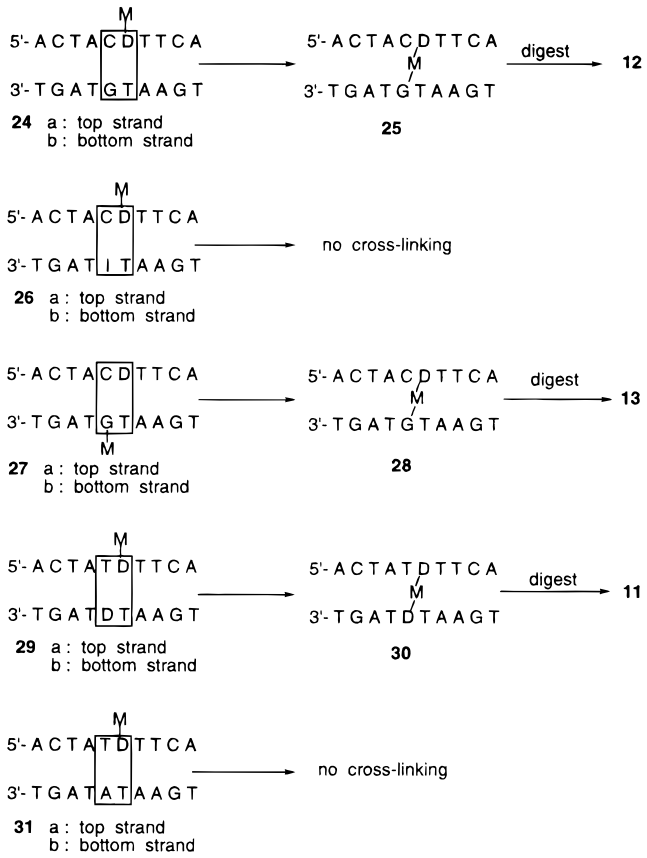


Figure 3. HPLC of SVD/AP digests of oligonucleotides alkylated by MC. (a) Isolation of the dD monoadducts formed under monofunctional alkylating conditions of d-(ATATATDTATAT)$_2$ by MC. (b) Isolation of bisadduct D-M-G (**11**) formed under bifunctional alkylating conditions in the same system. (c) G-adduct and D-adduct formed under monofunctional alkylation of duplex **19** by MC. (d) "Mixed" bisadducts D-M-G (**12**) and G-M-D (**13**) formed under bifunctional alkylating conditions of duplex **19** by MC. HPLC conditions: C-18 column (Beckman Ultrasphere-ODS; 4.5 c 250 mm); 1.5% to 18% acetonitrile in 0.03 M K-phosphate, pH 5.4 buffer in 75 min; 1.0 mL/min flowrate.

spectra matched closely that of **6a** in the entire UV region (data not shown). It is known that the C-1″ substituent stereochemistry of mitosenes can be determined by the CD spectrum regardless of the nature of the C-1″ substituent.[26] The observed match with authentic **6a** strongly suggested the C-1″α-configuration of the mitosene in **10a**.[10] The accompanying minor adduct **10b** (Figure 3a) displayed near-mirror image CD, closely matching that of **6b** (data not shown), and therefore **10b** is most likely the 1″*β*-stereoisomer of **10a**, by analogy with the MC-deoxyguanosine stereoisomer pair **6a** and **6b**.[27]

Scheme 3. Conversion of MC Monoadducts to Cross-Links by $Na_2S_2O_4$ in Oligonucleotides



The bisadduct **11** was formed in the reaction of bifunctionally activated MC with oligonucleotide **16**. The reaction mixture was separated into cross-linked and non-cross-linked oligonucleotide fractions (53% and 47%, respectively) by Sephadex G-50 chromatography.[19] The cross-linked oligonucleotide fraction was digested to nucleosides, and from the digest bisadduct **11** was isolated by HPLC (Figure 3b). Its UV spectrum was a composite of the dDAP and 2,7-diaminomitosene chromophores, similarly to the case of the monoadduct **10a** (data not shown). ESIMS confirmed that **11** is a MC-bis-dDAP adduct (calcd MH$^+$ 774; found 774). Bisadduct **11** was also formed from the **10a**-monoadduct oligonucleotide complex **29** by reactivation of the MC residue under anaerobic $Na_2S_2O_4$ treatment (Scheme 3). This conversion defined the direct structural relationship between **10a** and **11** and confirmed the structure of **11**.

The two "mixed bisadducts" **12** and **13** represent two isomeric cross-links formed between the indicated guanine and DAP residues in oligonucleotide **19**, under bifunctional MC activating conditions. They were isolated by the same procedures employed for bisadduct **11** in 47% overall yield, in 5:1 proportion (Figure 3d). ESIMS verified the molecular weights (calcd MH$^+$ 775; found 775, for both). The specific assignments of the isomeric structures are based on the results of controlled two-step cross-linking experiments as shown in Scheme 3 and described in detail in the Experimental Section.

Evidence that in DAP Adducts MC Is Linked to the 2-Amino Group of DAP. As described in the preceding section, the M-D monoadduct in the duplex **24** was readily converted to the D-M-G cross-link in **25**. However, in duplex **26** in which inosine replaces guanine the same monoadduct

(26) Tomasz, M.; Jung, M.; Verdine, G. L.; Nakanishi, K. *J. Am. Chem. Soc.* **1984**, *106*, 7367−7370.

(27) Gargiulo, D.; Musser, S. S.; Yang, L.; Fukuyama, T.; Tomasz, M. *J. Am. Chem. Soc.* **1995**, *117*, 9388−9398.

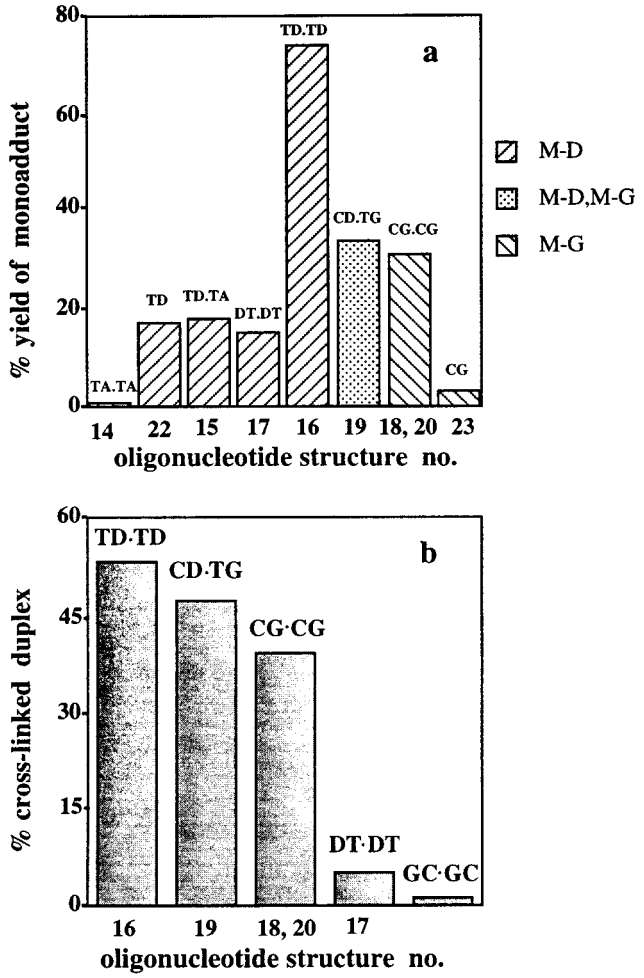


Figure 4. Yields of monoalkylation and bifunctional alkylation (cross-linking) of DAP or guanine by MC in oligonucleotides: (a) monoadduct M-D (**10a**) or M-G (**6a**) and (b) cross-linked duplex oligonucleotides. The dinucleotide sequences above the bars correspond to the alkylated or cross-linked sites of the oligonucleotides. Average deviation from the mean value of three experiments never exceeded ±15% in the cases when it was determined.

could not be converted to a cross-link (Scheme 3). This suggested that the D-M-G cross-link must involve the guanine 2-amino group that is located in the minor groove. If so, it follows that in addition to the D-M-G cross-link its precursor *the M-D monoadduct **10a** must also have been located in the minor groove.* Although it was not possible to verify this by NMR analysis due to scarcity of material, we could prove conclusively by a group of observations that the D-M-G cross-link **12** was indeed in the minor groove involving the two purine 2-amino groups as linkage points. Similarly, the DAP monoadduct (**10a**) and the DAP-to-DAP cross-link D-M-D (**11**) were proven to have their MC residues linked to DAP 2-amino groups, as described in the Experimental Section.

Sequence Specificity of Monoalkylation and Cross-Linking of DAP by MC in Oligonucleotides (Figure 4). (a) Monoalkylation: The yield of monoadducts formed under conditions of monofunctional activation of MC is presented in Figure 4a. The alkylated dinucleotide sequences within the oligonucleotides are indicated above the bars. It is apparent that the TD·TD target present in **16** gives the highest yield (74%) of **10a** in the series, indicating a specific enhancement of the alkylation at this sequence relative to single-stranded TD present in **22** (17%) as well as to the other D-containing sequences tested (**15** and **17**; 18 and 16%, respectively). The CD·TG sequence is also a favored site for alkylation of DAP: **10a** was formed in 34% yield, in contrast to the low yields already noted with **22**, **15**, and **17**. The enhanced alkylation of guanine in the CG·CG sequence, yielding **6a**,[16] is analogous to these findings with DAP.

(b) Cross-linking: The yield of cross-linked oligonucleotides formed under bifunctional activating conditions is seen in Figure 4b. The best cross-linking target was again the dinucleotide duplex sequence TD·TD. The dinucleotide in the "reversed" sequence DT·DT showed only ≤5% cross-linking. This preference for TD·TD is analogous to the CG·CG specificity of cross-linking of regular oligonucleotides[19] (Figure 4b). The "mixed" sequence CD·TG was cross-linked at intermediate efficiency between that of TD·TD and CG·CG.

Enhanced Reactivity of the 2-Amino Group of DAP Compared to That of Guanine (Figure 4). The oligonucleotide **19** (CD·TG) yielded two monoadducts upon monofunctional alkylation by MC: M-D monoadduct **10a** (along with a trace of stereoisomer **10b**) and M-G monoadduct **6a** in 34% and 5% yields, respectively. This is seen very clearly in Figure 3c. The same oligonucleotide **19** yielded upon bifunctional activation two bisadducts, **12** and **13**, in a 5:1 molar ratio (Figure 3d). Both sets of data indicate that the rate-determining first alkylation step by MC occurred faster at DAP than at guanine (by factors of 7 and 5, respectively).

In another set of experiments the TD·TD-containing duplex oligonucleotide **16** and the CG·CG-containing duplex oligonucleotide **18** were mixed in 1:1 molar proportion and the mixture was alkylated monofunctionally and bifunctionally by MC, in two separate reactions. Each reaction mixture was digested by snake venom diesterase (SVD) and alkaline phosphatase (AP) and the molar ratios of adducts were determined by HPLC, giving a **10a**:**6a** ratio of 2.6 and a **11**:**8** ratio of 2.0, respectively. Both of these results provide further evidence that DAP reacts more efficiently than guanine with MC.

The relatively high level (15−18%) of nonsequence-specific alkylation of DAP in oligonucleotides lacking the TD·TD target site (**15**, **17**, **22**) stands in contrast to the low yield (3%) of nonsequence-specific alkylation of guanine by MC in oligonucleotide **23**, which lacks the analogous CG·CG target site[16] (Figure 4a). This again indicates the higher reactivity of the DAP 2-amino group as nucleophile.

The same phenomenon was observed in the *tyr*T DNA series, where I/DAP·I/DAP DNA (Table 1) was cross-linked approximately 3-times more efficiently than G·G DNA, even though the latter contained more cross-linkable sites (13 CG·CG steps) than the I/DAP·I/DAP DNA (8 TD·TD sites) (see below.)

Replacement of Adenines with DAP in DNA Creates Additional Targets for Alkylation by MC. This is demonstrated explicitly by a comparison of the alkylation of oligonucleotide **20** with that of **21** (Chart 2). Oligonucleotide **20** has only one (central) CG·CG target, while in oligonucleotide **21** two additional DT·DT targets were created by replacing two adenines by DAP in each strand. The yields of both mono- and bis-alkylation products of **21** were higher than those from **20** (total monoadduct: 42% vs 30%; cross-linking: 76% vs 40%). These results are paralleled by the dramatic increase in cross-linking efficiency of the *tyr*T DNA in which A was replaced by DAP (Figure 5; see below).

Cross-Linking of *tyr*T DNA by MC. Effects of Guanosine→Inosine and Adenine→DAP Substitutions on Cross-Linking Patterns. Radioactively labeled PCR-developed DNAs containing natural or unnatural bases (Table 1) were subjected

Table 1. Sequences of *tyr*T DNA and Its Base-Analog-Substituted Variants

Descriptor	*tyr*T DNA variants	Sequences
Natural DNA (G•G)	Natural	75′ 60′ 55′ 42′ 5′- AATTCCGGTT ACCTTTAATC CGTTACGGAT GAAAATTACG CAACCAGTTC ATTTTTCTCA 3′- TTAAGGCCAA TGGAAATTAG GCAATGCCTA CTTTTAATGC GTTGGTCAAG TAAAAAGAGT 19′ 4′ 1′ 1 18 23 30 ACGTAACACT TTACAGCGGC GCGTCATTTG ATATGAAGCG CCCCGCTTCC CGATAAGGGA TGCATTGTGA AATGTCGCCG CGCAGTAAAC TATACTTCGC GGGGCGAAGG GCTATTCCCT 64 79 GCAGGCCAGT AAAAAGCATT ACCCCGTGGT GGGGGTTCCC GA -3′ CGTCCGGTCA TTTTTCGTAA TGGGGCACCA CCCCCAAGGG CT -5′
I DNA (I•I)	I-substituted	75′ 60′ 55′ 42′ 5′- AATTCCGGTT ACCTTTAATC CITTACIIAT IAAAATTACI CAACCAITTC ATTTTTCTCA 3′- TTAAIICCAA TIIAAATTAI ICAATICCTA CTTTTAATIC ITTIITCAAI TAAAAAIAIT 19′ 4′ 1′ 1 18 23 30 ACITAACACT TTACAICIIC ICITCATTTI ATATIAAICI CCCCICTTCC CIATAAIIIA TICATTITIA AATITCICCI CICAITAAAC TATACTTCIC IIIICIAAII ICTATTCCCT 64 79 ICAIICCAIT AAAAAICATT ACCCCITIIT IIIIITTCCC IA -3′ CITCCIITCA TTTTTCITAA TIGGGCACCA CCCCCAAGGG CT -5′
DAP DNA (D•D)	DAP-substituted	57′ 42′ 75′ 60′ 55′ 51′ 44′ 40′ 36′ 31′ 22′ 5′- AATTCCGGTT ACCTTTAATC CGTTDCGGDT GDDDDTTDCG CDDCCDGTTC DTTTTTCTCD 3′- TTDDGGCCDD TGGDDDTTDG GCDDTGCCTD CTTTTDDTGC GTTGGTCDDG TDDDDDGDGT 17′ 7′ 13 19′ 14′ 9′ 4′ 1′ 1 4 8 11 18 23 30 33 DCGTDDCDCT TTDCDGCGGC GCGTCDTTTG DTDTGDDGCG CCCCGCTTCC CGDTDDGGGD TGCDTTGTGD DDTGTCGCCG CGCDGTDDDC TDTDCTTCGC GGGGCGDDGG GCTDTTCCCT 41 46 49 56 59 64 79 GCDGGCCDGT DDDDDGCDTT DCCCCGTGGT GGGGGTTCCC GD -3′ CGTCCGGTCD TTTTTCGTDD TGGGGCACCA CCCCCAAGGG CT -5′
I/DAP DNA (I/DAP•I/DAP)	I/DAP-substituted	57′ 44′ 5′- AATTCCGGTT ACCTTTAATC CITTDCIIDT IDDDDTTDCI CDDCCDITTC DTTTTTCTCD 3′- TTDDIICCDD TIIDDDTTDI ICDDTICCTD CTTTTDDTIC ITTIITCDDI TDDDDDIDIT 17′ 9′ 11 33 DCITDDCDCT TTDCDICIIC ICITCDTTTI DTDTIDDICI CCCCICTTCC CIDTDDIIID TICDTTITID DDTITCICCI CICDITDDDC TDTDCTTCIC IIIICIDDII ICTDTTCCCT 48 58 ICDIICCDIT DDDDDICDTT DCCCCITIIT IIIIITTCCC ID -3′ CITCCIITCD TTTTTCITDD TIGGGCACCA CCCCCAAGGG CT -5′
Asymmetrical natural DNA (G•G)	Natural, except I-substituted in 5′-terminal regions	78′ 63′ 58′ 45′ 5′- AATTCCGGTT ACCTTTAATC CGTTACGGAT GAAAATTACG CAACCAGTTC ATTTTTCTCA 3′- TTAAGGCCAA TGGAAATTAG GCAATGCCTA CTTTTAATGC GTTGGTCAAG TAAAAAGAGT 22′ 7′ 4′ 2′ 14 19 26 ACGTAACACT TTACAGCGGC GCGTCATTTG ATATGAAGCG CCCCGCTTCC CGATAAGGGA TGCATTGTGA AATGTCGCCG GCCAGTAAAC TATACTTCGC GGGGCGAAGG GCTATTCCCT GCAGGCCAGT AAAAAGCATT ACCCCGTGGT GGGGGTTCCC GAGCCCCGG -3′ CGTCCGGTCA TTTTTCGTAA TIIIICACCA CCCCCAAIII CT -5′
Asymmetrical I/DAP DNA (G•I/DAP)	I/DAP-substituted in one strand	54′ 5′- AATTCCGGTT ACCTTTAATC CGTTACGGAT GAAAATTACG CAACCAGTTC ATTTTTCTCA 3′- TTDDIICCDD TIIDDDTTDI ICDDTICCTD CTTTTDDTIC ITTIITCDDI TDDDDDIDIT 4 9 ACGTAACACT TTACAGCGGC GCGTCATTTG ATATGAAGCG CCCCGCTTCC CGATAAGGGA TICDTTITID DDTITCICCI CICDITDDDC TDTDCTTCIC IIIICIDDII ICTDTTCCCT GCAGGCCAGT AAAAAGCATT ACCCCGTGGT GGGGGTTCCC GAGCCCCGG -3′ CITCCIITCD TTTTTCITDD TIIIICACCA CCCCCAAIII CT -5′

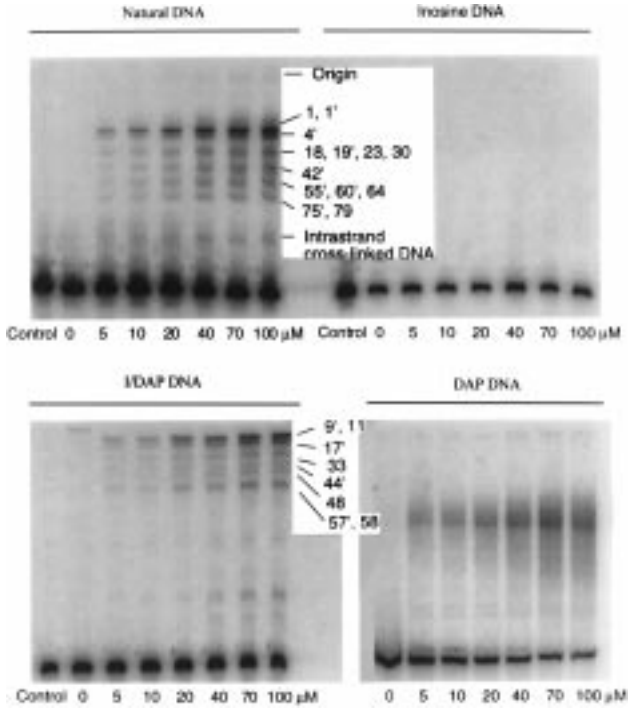


Figure 5. Cross-linking of *tyr*T DNA, inosine-substituted *tyr*T DNA, inosine/DAP-substituted *tyr*T DNA, and DAP-substituted *tyr*T DNA assayed by denaturing PAGE at various concentrations of MC. The control lanes are free from MC or $Na_2S_2O_4$. The remaining lanes show DNA treated with 0−100 μM MC in the presence of 2.8 mM $Na_2S_2O_4$. DNA molecules cross-linked at different positions are retarded to different extents, and centrally cross-linked DNA is the slowest. The numbers at the side of the bands indicate the assigned distance of the cross-linked site from the center of the DNA molecule.

to cross-linking by reductively activated MC and then separated by denaturing PAGE. Following denaturation by strand separation buffer and heating at 94 °C, un-cross-linked DNA migrates as a 162-base single stranded molecule which runs much faster than cross-linked DNA containing 324 bases. A cross-link between two complementary strands of DNA imparts considerable resistance to migration in denaturing gels, so that the proportion of cross-linked DNA can easily be determined by measuring the intensity of retarded species as a percentage of the whole intensity in the lane. The following control and substituted DNAs were used (Table 1): natural *tyr*T DNA (G·G DNA), inosine-substituted DNA (I·I DNA), DAP-substituted DNA (DAP·DAP DNA), and inosine/DAP-substituted DNA (I/DAP·I/DAP DNA). A typical phosphorimage is shown in Figure 5, which illustrates that various cross-linked natural DNA species can be detected at varying concentrations of MC. Un-cross-linked DNA, i.e., the starting material, appears as a single band with the highest mobility. Cross-linked DNA is resolved into 6 major bands, reflecting the heterogeneity of cross-link locations. Of course, the possible formation of intrastrand cross-links at GpG sites is not excluded. DNA containing this lesion or a simple MC monoadduct appears to migrate as a single band on denaturing PAGE; this band, migrating slower than the DNA starting material, is assigned as indicated in Figure 5 and corresponds most likely to these species as shown previously to be the case with shorter DNAs.[30]

(28) The glycosidic bond of deoxyguanosine, deoxyinosine, and deoxyadenosine residues alkylated at the N-3 or N-7 positions hydrolyzes rapidly on heating.[29]

(29) Singer, B.; Grunberger, D. *Molecular Biology of Mutagens and Carcinogens*; Plenum Press: New York, 1983; pp 19 and 313.

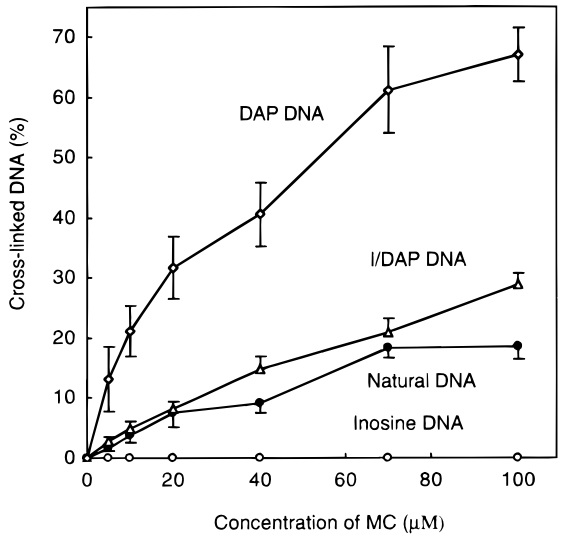


Figure 6. Plots of the overall extent of cross-linking as a function of MC concentration.

When the cross-linking experiment was repeated with inosine-substituted DNA no retarded species were formed. By contrast, with DAP-substituted DNA there was abundant evidence of extensive reaction leading to the formation of much retarded material (Figure 5). This is more or less as expected, and underpins the interpretation of the following cross-linking experiments on I/DAP DNA which provide a critical test of the hypothesis that relocating the purine 2-amino group in DNA should result in a new cross-linking pattern.

Indeed, I/DAP DNA can be efficiently cross-linked by MC under the usual experimental conditions and the reacted species are again resolved into 6 major bands. Interestingly, the cross-linked I/DAP DNA displays more retarded mobility in denaturing PAGE compared to natural DNA (Figure 5). It is known that the mobility of I/DAP double-stranded DNA on nondenaturing PAGE is slower than that of natural double-stranded DNA because the increased curvature of DNA induced by G→I substitution only partially compensates for the decreased curvature caused by A→DAP substitution,[31] but whether this might explain the more marked retardation caused by a MC cross-link in denatured I/DAP DNA we cannot tell.

Effects of the Substitutions on the Efficiency of Cross-Linking. Of the three DNA species which are susceptible to cross-linking by MC, natural DNA is the poorest substrate and DAP-DNA is by far and away the best. Figure 6 shows that at 100 μM MC, the average cross-linking of natural, I/DAP, and DAP DNAs amounts to 18, 29, and 66%, respectively. Clearly there are many potential cross-linking sites in DAP DNA yet in I/DAP DNA we expect only 8 cross-linkable (TpD) sites compared with 13 CpG sites in natural DNA. [The DpT sequence is unlikely to contribute to the cross-linking of DAP DNA since it showed only marginal cross-linkability in the oligonucleotide test system (Figure 4b).] Thus, the TpD sequence must be considerably more susceptible to cross-linking by MC than CpG.

Effects of the Substitutions on the Location of Cross-Links within the DNA Sequence. Given the specificity of MC for CpG sequences, 13 species of singly cross-linked natural DNA

(30) Basu, A. K.; Hanrahan, C. J.; Malia, S. A.; Kumar, S.; Bizanek, R.; Tomasz, M. *Biochemistry* **1993**, *32*, 4708−4718.

(31) Bailly, C.; Waring, M. J.; Travers, A. A. *J. Mol. Biol.* **1995**, 253, 1−7.

molecules would be expected because there are 13 such steps in *tyr*T DNA. It has been demonstrated that the position of the cross-link affects the mobility of DNA in denaturing gels, with centrally cross-linked molecules having the lowest mobility.[15,16] Mobility increases as the position of cross-linking moves toward either end of the DNA double helix. In addition, the rate of change of mobility as a function of the position of a cross-link is higher at the central region than at the termini of DNA.[15,16] The possible site(s) of cross-linking in each individual band can, therefore, be assigned by considering their distance from the center of the DNA. The numbers written above the sequences in Table 1 indicate the distance of each potential cross-linkage site from the center of the DNA molecule, with primed numbers indicating sites left of the center and unprimed numbers indicating sites right of the center. Because the experimental results reveal only 6 cross-linked bands it must be presumed that each band is likely to be a composite, consisting of DNA molecules each containing a cross-link about the same distance from one or the other of its ends so that primed and unprimed numbers run together, so to speak. On this basis the observed bands of retarded DNA have been assigned to particular groups of potential cross-linking sites.

The contribution of each resolved band to the total observed cross-linking at 100 μM MC is shown in Figure 7. The single-hit assumption is applicable because the overall cross-linkage is only about 30%. Groups 1 and 2 have the lowest mobility, and so must involve centrally located CpG sites. Together they contribute 62% of the total cross-linking, although only 3 out of the possible 13 species of cross-linked DNA appear to be involved. This probably reflects the strongly cross-linkable character of the sequence 5′-GCGCGT-3′ in the central region of the DNA. The yield of cross-linking per CpG site has been reported to be exceptionally high at regions containing successive CpG sites.[18] The varying extent of cross-linking at different positions in general is probably due to the variation of sequences flanking the CpG core.[32,33] 5′-PuCGPyr-3′ sequences are the most reactive, followed by 5′-PuCGPu-3′; 5′-PyrCGPu-3′ is the least reactive. Their relative reactivities are approximately 3:2:1.[32] Group 3 contributes 15% of the total cross-linking because it contains two readily cross-linkable 5′-PuCGPyr-3′ sites (18 and 19′). Group 4 consists of one 5′PuCGPyr-3′ site (42′) and contributes 11%. Group 5 contributes only 8% because it consists of three moderately cross-linkable 5′-PuCGPu-3′ sites (55′, 60′, and 64). Group 6 contributes 4% since it consists of two weakly cross-linkable 5′-PyrCGPu-3′ sites (75′ and 79).

In I/DAP DNA the 8 possible species of DNA singly cross-linked at TpD steps are similarly assigned to the 6 observed bands according to the distance of their cross-links from the center of the molecule (Figure 7). Again, the incompletely resolved group 1 and 2 band seems to originate from three strongly cross-linkable sites near the center of the DNA, whereas the other bands are mainly attributable to single, less favored sites.

With DAP DNA the large amount of retarded material points to the existence of a plethora of cross-linking sites such that discrete products cannot be resolved. Moreover, when as much as 70% of the DNA appears as cross-linked species after exposure to 100 μM MC the reaction conditions must depart completely from single-hit kinetics, rendering band assignment difficult if not impossible. Efforts to secure acceptable band resolution at lower levels of reaction were to no avail.

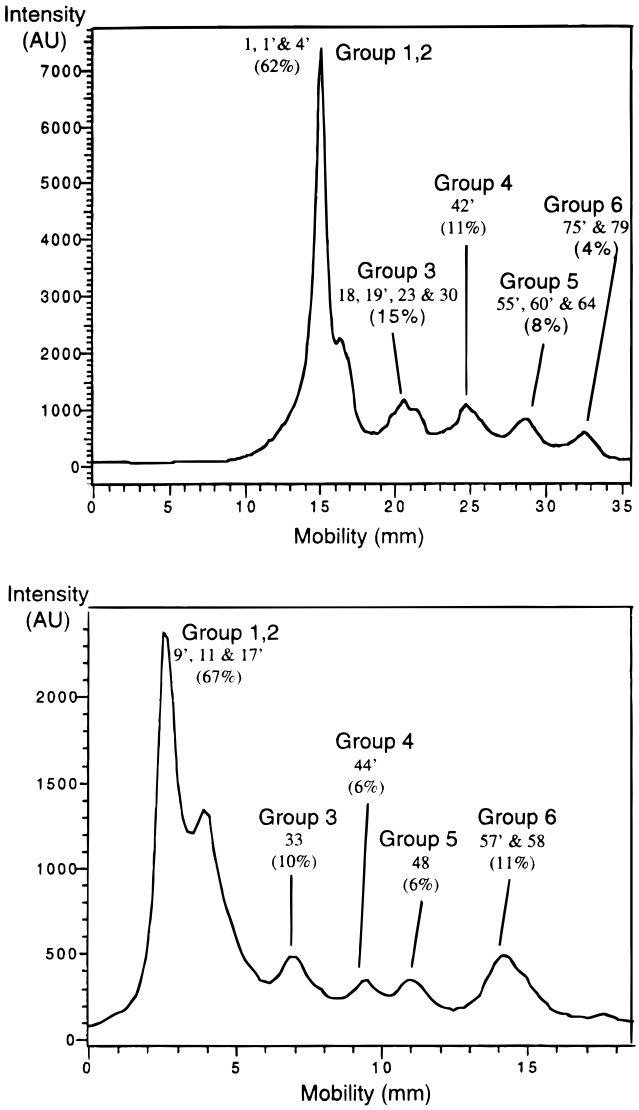


Figure 7. Densitometric plots showing peaks and assignment of positions of cross-linking for natural DNA (top) and I/DAP-substituted DNA (bottom), together with the relative proportion of the cross-linked DNA appearing in each peak. See Table 1 for the numbering scheme.

In natural DNA, the 5′-GCGC-3′ sequence is highly cross-linkable compared with sequences containing other 5′ and 3′ bases flanking the CpG core.[18] The equivalent 5′-DTDT-3′ sequence only appears at one position in I/DAP DNA, which is 11 bp from the center of the molecule and belongs to group 1. Indeed groups 1 and 2, representing 3 out of the total 8 species of singly cross-linked DNA, contribute 67% of the total cross-linking (Figure 7). We deduce that the highly cross-linkable 5′-GCGC-3′ sequence is functionally supplanted by the 5′-DTDT-3′ sequence following relocation of the purine 2-amino group.

In the case of CpG cross-linking, 5′-PyrCGPu-3′ and 5′-PuCGPu-3′ sequences are weakly cross-linkable. Likewise, in the TpD case (I/DAP DNA) groups 3, 4, 5, and 6 in Figure 7 consist of 5′-PyrTDPu-3′ or 5′-PuTDPu-3′ sequence(s). Each of these TpD sites contributes only 5−10% of the total cross-linking. This contrasts sharply with the 67% of groups 1 and 2 together. We conclude that the molecular basis of DNA recognition by MC is likely to be similar in both natural and I/DAP DNAs.

Mixed Cross-Link Formation between a Guanine and a DAP Residue in DAP-Substituted *tyr*T DNA. Formation of

(32) Borowy-Borowski, H.; Lipman, R.; Tomasz, M. *Biochemistry* **1990**, *29*, 2999−3004.

(33) Millard, J. T.; Weidner, M. F.; Kirchner, J. J.; Ribeiro, S.; Hopkins, P. B. *Nucleic Acids Res.* **1991**, *19*, 1885−1891.

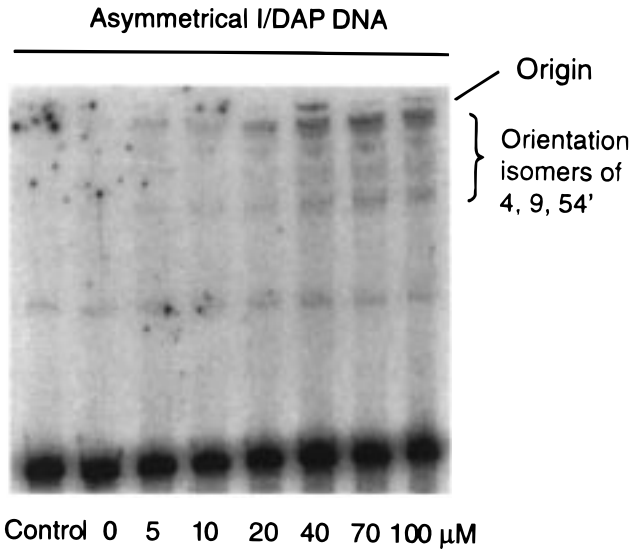


Figure 8. Cross-linking of "asymmetrical" *tyr*T DNA having one natural and one inosine- or I/DAP-substituted complementary strand (G·I/DAP DNA). Assay of the extent of cross-linking by increasing concentrations of MC with use of PAGE. Details as for Figure 5.

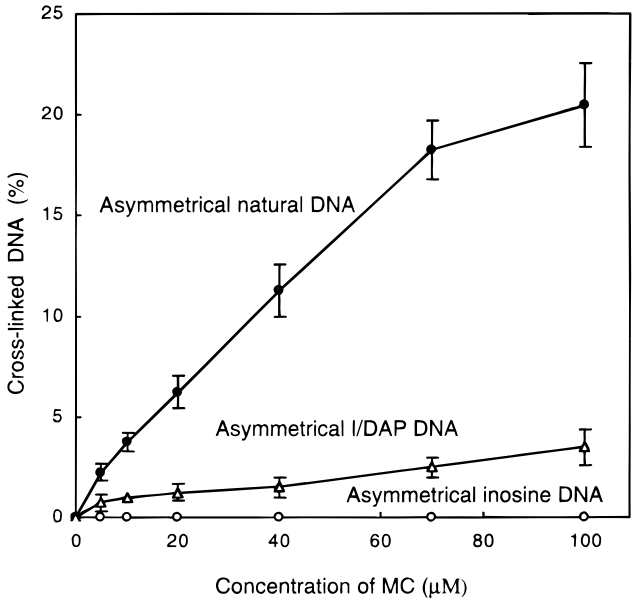


Figure 9. Plot of the extent of cross-linking of asymmetrical natural, asymmetrical I/DAP-substituted, and asymmetrical I-substituted *tyr*T DNA as a function of MC concentration.

mixed cross-links was demonstrated in the oligonucleotide series as described above (Figure 3d). To investigate whether the placement of two purine 2-amino groups in the minor groove of the double helix is both necessary and sufficient for cross-linking by MC, *asymmetrical DNAs* (Table 1) were synthesized and subjected to cross-linking. For these experiments *tyr*T DNA was modified by substituting inosine for guanosine, and DAP for adenine, selectively in only one of its complementary strands (G·I/DAP DNA). Two controls, natural *tyr*T DNA (G·G DNA) and *tyr*T DNA selectively substituted with inosine in one strand (G·I DNA), were used for comparison. Because different primers had to be used in the PCR to make these molecules, the numbering of potentially cross-linkable sites starting from the center of these fragments differs slightly from that which applies to fully substituted DNA (Table 1).

In the control, asymmetrical natural DNA (G·G) was readily cross-linked as seen by the detection of 5 well-resolved slow-moving bands in the gel. In contrast, asymmetrical inosine DNA (G·I) was not cross-linkable, showing that an I-MC-G cross-link cannot be formed, as predicted (data not shown).

Rather surprisingly, given that asymmetrical I/DAP DNA contains only three potentially cross-linkable TpG·CpD sites (Table 1), the gel showed at least four resolved bands (Figure 8). This can be explained by the formation of DAP-MC-G isomers, each having different mobilities in denaturing PAGE. As shown in the oligonucleotide experiments, two isomeric cross-links DAP-^{1}MC^{10}-G and G-^{1}MC^{10}-DAP can exist (Chart 1; Figure 3d), which may well migrate differently in denaturing gels.

In Figure 9 the percentage cross-linking of asymmetrically substituted DNA molecules is shown. The positive control is broadly similar to the natural DNA in Figure 6. However, less than 5% of asymmetrical I/DAP-substituted DNA is cross-linked at 100 μM MC. Evidently its three TpG·CpD sites are only weakly cross-linkable (two are 5′-PuCGPu-3′ and one is 5′-PyrCGPu-3′). Unfortunately, the relative reactivity of CpG·CpG and TpG·CpD sequences toward cross-linking cannot be directly inferred from Figures 5, 8, and 9.

(34) Warrren, A. J.; Hamilton, J. W. *Chem. Res. Toxicol.* **1996**, *9*, 1063−1071.

Discussion

The present work has examined in a new way the structural factors responsible for the sequence-dependent modulation of alkylation of guanines by MC. By employing DAP as a substrate analogue of guanine, incorporated into a natural 162-bp DNA restriction fragment, the system addresses the MC-DNA recognition problem differently from previous studies in two respects: (i) the putative guanine-specific recognition element is selectively relocated and (ii) a large natural DNA fragment with known sequences containing multiple MC cross-link sites is employed for the first time. In all previous studies only short ($\leq$23bp) synthetic oligonucleotides have been used.[34] The main reason for this was that, so far, no method has been devised for mapping MC-cross-linked sites in DNA directly. Only oligonucleotides which contain a unique MC cross-link site can be "footprinted" for the location of the cross-link.[33,34] We have now mapped multiple MC cross-link sites in *tyr*T DNA by an indirect method based on an empirical correlation between the relative electrophoretic mobility of individual, singly cross-linked DNA species and the position of their cross-link relative to the center of the DNA duplex.[35,36] Since this method requires a knowledge of the potential cross-linkable bases and their sequence context in DAP-substituted DNA they first had to be independently determined. Accordingly, the first goal of this work was to examine alkylation and cross-linking of synthetic DAP-substituted oligonucleotides containing a single potential monoalkylation or cross-link site. These studies established the following:

The isolation and structural characterization of the MC-DAP deoxynucleoside adducts **10**−**13** (Scheme 2) indicates that DAP does indeed function as a substrate analogue for the alkylation of DNA by MC. The adduct structures, including the stereochemistry of the attachment of drug to DAP, are analogous to the structures of MC-guanine adducts (**6**−**9**; Scheme 1), showing that the structural differences between the two purine substrates have no great significance for the specificity of the recognition

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

(36) Millard, J. T.; Luedtke, N. W.; Spencer, R. J. *Anti-Cancer Drug Design* **1996**, *11*, 485−492.

by MC of its alkylation site. Since considerable structural differences between a G·C and DAP·T basepair exist in the major groove but none in the minor groove, it may be concluded from this result that the sole determinant for alkylation site recognition is the 2-amino group of a purine in the minor groove. It is interesting, however, that the alkylation of DAP is faster[37] than that of guanine, as observed both in the oligonucleotide series and in inosine- and DAP-substituted *tyr*T DNA. The increased reactivity of DAP with MC can be attributed to a greater intrinsic nucleophilicity of the DAP 2-amino group compared to the 2-amino group of guanine. The fact that it is manifested even in *single-stranded* oligonucleotides, where precovalent binding does not play a role (Figure 4a), confirms this conclusion. It is notable, however, that in addition to increased nucleophilicity, the 2-amino group of DAP in DNA appears also to exhibit stronger H-bonding to various DNA-binding ligands than that of guanine.[3,7]

The observed sequence specificity of monoalkylation of DAP by MC (Figure 4a) reinforces the H-bond mechanism (Figure 1a) as follows. The yield of alkylation of each D residue in the TD·TD sequence was 4-fold enhanced over the yield of alkylation of D in the TD·TA sequence, indicating that a second 2-amino group, available at the TD·TD site, definitely participates, resulting in rate enhancement of the covalent step. Since at the TD·TA sequence no second 2-amino group is present, only the "basal" D-alkylation rate common to the single-stranded TD and DT·DT sites is observed. The alkylation rate of D at the CD·TG sequence was also enhanced. In this case the *H-bonding 2-amino group* was supplied by a *guanine* residue rather than a DAP residue. Interestingly, the yield of alkylation of DAP was only 34%, in contrast to 74% in the TD·TD sequence (Figure 4a). Since the 2-amino group of DAP in DNA generates stronger H-bonded binding sites for various drugs than does that of DNA-guanine,[3,7] this finding serves as independent evidence for the proposed role of an H-bond in the covalent reaction specificity (Figure 1a). Taken together, the new results provide compelling support for a mechanism in which H-bonding to a second purine 2-amino group in the minor groove assists MC to select guanine in the CpG sequence,[16,20] as shown in Figure 1a.

To determine the sequence specificity of cross-linking of DAP residues in oligonucleotides only one potential cross-link site was incorporated into each duplex. The results (Figure 4b) show unequivocally that DAP residues at the TD·TD sequence are cross-linkable while the DAPs in the reversed sequence, DT·DT, are not. This result is again completely analogous to the virtually absolute CG·CG specificity of the MC cross-links in natural oligonucleotides and it is fully consistent with the same mechanism, namely the specific orientation of the monoadduct toward a second, upstream purine 2-amino group (Figure 1b). The observed cross-linking of guanine to DAP at the CD·TG site reinforces this conclusion.

Cross-Linking of *tyr*T DNA by MC. The oligonucleotide study revealed that the three potential MC cross-link sites in DAP-containing DNA are CG·CG, TD·TD, and CD·TG. The 162-bp natural *tyr*T DNA fragment containing 13 CG·CG cross-link sites was successfully cross-linked and the distribution of the cross-links was mapped to the sequence of the DNA (Table 1, Figures 5 and 7). Since no direct chemical method exists for mapping MC cross-links in large DNA molecules containing multiple cross-linkable sites[33] an empirical method, based on the distance of the cross-link from the center of the DNA,[36] was employed. The central sequence 5′-GCGGCGCG containing 3 CG·CG sites is the "hotspot": 62% of cross-links are formed in this region even though these 3 CG·CG sites represent only 23% of a total of 13 CG·CG steps. It has been shown previously, using short oligonucleotides, that the nature of the 5′- and 3′-flanking basepairs at a CG·CG site strongly influences the cross-linking rate[32,33] and runs of several CGs in oligonucleotides have been reported to be particularly reactive.[18] This selectivity of MC is now observable in DNA containing multiple cross-link sites for the first time.

Removing the 2-amino groups from all guanines (I·I DNA) resulted in non-cross-linkable DNA. However, adding 2-amino groups to all adenines (I/DAP·I/DAP DNA) restored cross-linkability, due to the creation of 8 TD·TD sites. The corresponding map indicates 3 strong cross-link sites together comprising 67% of the total cross-links. It is notable that, on average, TD·TD sites were cross-linked 3 times more efficiently than CG·CG (Figure 6; calculated from the data at 100 μM MC). Similar relative reactivity ratios were observed with oligonucleotides, as discussed earlier. All in all, the 2-amino group of DAP appears to be more nucleophilic than that of guanine.

Cross-linking of "asymmetrical" DNA in which only one of the strands was substituted with DAP and inosine (G·I/DAP DNA) demonstrated dramatically that MC cross-link formation in DNA absolutely requires purine 2-amino groups *in both strands* and that "mixed" cross-links are readily formed between a guanine- and a DAP 2-amino group. The observed extent of cross-linking of this DNA was much lower than that of the natural control, but this is entirely consistent with the fact that only 3 cross-link sites (CG·TD) flanked by a nonideal arrangement of purines and pyrimidines were present, compared to 11 CG·CG sites in the control.

Experimental Section

Materials. Mitomycin C was obtained from Dr. D. M. Vyas, Bristol-Myers Squibb Co., Wallingford, CT. Protected 2,6-diaminopurine-2′-deoxyriboside-CE-phosphoramidite was purchased from Glen Research, Sterling, VA. 2,6-Diaminopurine-2′-deoxyribonucleoside (dDAP) was purchased from Sigma, St. Louis, MO.

Oligonucleotides were synthesized on an automated DNA synthesizer (model 380B, Applied Biosystems, Inc.), using the β-cyanoethyl phosphoramidite method. The crude products (1 μmol scale; "Trityl-off"), after deprotection by concentrated NH_4OH overnight at 55 °C, were purified with use of a Sephadex G-25 (fine) column (2.5 X 56 cm; 0.02 M NH_4HCO_3 eluant). The void volume fraction containing the oligonucleotide was lyophilized. HPLC on a C-4 reverse-phase column indicated ≥95% purity of the oligonucleotides. Oligonucleotides containing DAP were deprotected by a 1:1 ratio of CH_3NH_2 (40% solution in H_2O) and concentrated NH_4OH (30% in H_2O) for 16 h at 55 °C. A small amount (approximately 0.5 $O.D_{260}$ unit) of each oligo was digested by phosphodiesterase and alkaline phosphatase and was analyzed for nucleoside mole ratios by HPLC on a C-18 reverse-phase column followed by dividing the peak areas by the respective molar extinction coefficient. Satisfactory nucleoside composition (±5% average deviation from the calculated value) was established for all oligos used. Phosphodiesterase I (snake venom diesterase) and *E. coli* alkaline phosphatase (type III-R) were purchased from Worthington, Freehold, NJ.

The 162-bp *Escherichia coli tyr*T promoter-containing *tyr*T (A93) DNA used as a PCR template was isolated from pAT plasmid by digestion with restriction enzymes *Ava*I (Promega) and *Eco*RI (Boehringer) and purified by 6% nondenaturating PAGE (19:1 acrylamide:

(37) The activation of MC by $Na_2S_2O_4$ and its subsequent reactions with solvent, protonation, bisulfite ions, and DNA bases results in complex kinetics and rates which are too fast to measure. However, under the same conditions, the reactions of the activated MC with DNA follow pseudo-first-order kinetics (excess activated MC) and are terminated abruptly at the same time, due to the abrupt (exponential) decay of activated MC. Therefore, the observed *yields* are related to the first-order *rates* of DNA alkylations in the usual way.[16]

bisacrylamide). The DNA, located under UV light, was eluted out of the crushed gel slice with a unidirectional electroelutor (IBI Model UEA) at 120 V for 2 h. The DNA was further purified with a QIAquick PCR purification kit (QIAGEN) according to the suggested standard protocol. Primers were synthesized and quality-checked by mass spectrometry at the Protein and Nucleic Acid Chemistry Facility of Cambridge University Biochemistry Department. The Crick primers were 5′ end-labeled with [γ-^{32}P]ATP (DuPont/NEN) of 6000 Ci/mmol specific activity with T4 polynucleotide kinase (Promega). Unincorporated [γ-^{32}P] was separated from the primer by using a QIAquick nucleotide removal kit (QIAGEN) according to the manufacturer's protocol. Nucleoside triphosphates dATP, dGTP, dTTP, dCTP, and dITP were purchased from Pharmacia Biotech. dDTP was kindly synthesized by Dr. M. Guo as recently described.[24]

Methods. Quantities of oligonucleotides were measured by UV spectrometry by using the molar extinction coefficient 10 000 at 260 nm (average value for one mononucleotide unit; thus, 10 A_{260} units of an oligonucleotide corresponds to 1 μmol in mononucleotide). The quantity of MC was measured by using the molar extinction coefficient 21 840 at 360 nm. Quantities of nucleosides and nucleoside-MC adducts were measured by using the following molar extinction coefficients (*E*) at 254 nm: dA, 13 300; dG 13 000; dT, 6 600; dC, 6 300; dDAP, 8 600; adduct **6**, 24 000; adduct **8**, 30 000; adduct **10**, 28 000; adduct **11**, 16 600; adduct **12** and **13**, 18 300. Values of *E* for dDAP and various dDAP-MC adducts (**10**−**13**) were determined by digesting a purified oligonucleotide or purified adducted oligonucleotide with known composition followed by HPLC on a C-18 reverse-phase column as above. *E* was estimated from comparison of the corresponding HPLC peak area ratio to that of the dT peak, using the formula (dDAP) = area (dDAP) × mol ratio (dT/dDAP) × (dT)/area(dT).

HPLC: A Beckman instrument equipped with System Gold programmable solvent module 125 and diode array detector 168 controlled by System Gold Chromatography Software was used.

Spectroscopic Techniques. (a) UV spectra were determined by the LC-diode array method: A Beckman System Gold 165 UV−visible diode array scanning detector was used for recording the spectra of substances eluted from a Beckman Ultrasphere-ODS C-18 column (4.6 × 250 mm) with 0.03 M potassium phosphate, pH 5.4 containing 1.5−18% acetonitrile in a linear gradient. **Circular dichroism spectra:** A Jasco model J710 spectropolarimeter connected to a Digital Equipment Corp. PC was employed. **Mass spectroscopy (LC-ESIMS method):** HPLC employed a C-18 column (2 × 250 mm), linear gradient of 3−20% acetonitrile in 10 mM NH_4OCOH, pH 6.5 in 75 min at 0.2 mL/min flowrate. Mass spectrometry was performed on a Finnigan (San Jose, CA) Model TSQ-7000 triple quadrupole mass spectrometer equipped with a standard Finnigan electrospray ion source. UV (254 nm) absorbing compounds were detected in-line prior to entry into the mass spectrometer. Mass spectra were acquired in the positive ion mode at a rate of 2 scan/s over a mass range of 150−1200 daltons.

Monofunctional Alkylation of Oligonucleotides by MC: Substoichiometric Anaerobic $Na_2S_2O_4$ as Activating Agent.[16] Duplex oligodeoxynucleotides (self-complementary or complementary strands in a 1:1 molar strand ratio, 1 μmol mononucleotide unit/mL; usually 2 to 10 μmol total scale) were mixed with MC (4 μmol/mL) in 0.1 M sodium phosphate, pH 7.5 buffer at 40 °C, in an ice bath under stirring in air. Fresh anaerobically prepared $Na_2S_2O_4$ solution (40 mM in the same buffer) was added, to give 2 μmol/mL final concentration. After 1 h the mixture was chromatographed on a 2.5 × 56 cm Sephadex G-25 column with 0.02 M NH_4HCO_3 as eluant. The first UV absorbing peak fraction containing the alkylated and unalkylated oligonucleotide(s) was then collected and lyophilized. Alkylated oligonucleotides were isolated by HPLC (Figure 2).

Bifunctional Alkylation of Oligonucleotides: Excess Anaerobic $Na_2S_2O_4$ as Activating Agent. Duplex oligodeoxynucleotides as above (1 μmol mononucleotide unit/mL) were exposed to MC (5 μmol/mL) in 0.1 M Tris, pH 7.4 buffer at 4 °C with stirring. The solution was deaerated and kept anaerobic by purging with argon. Fresh, anaerobically prepared $Na_2S_2O_4$ solution (0.16 M in same buffer) was added to 7.5 μmol/mL final concentration in 5 portions at 10 min intervals. After 1 h the reaction mixture was chromatographed on a Sephadex G-50 (2.5 × 56 cm) column. The pure cross-linked oligonucleotide was isolated from the first-eluted fraction (Figure 3).

Conversion of MC-Monoalkylated Oligonucleotide to MC-Cross-Linked Duplex. Monoalkylated oligonucleotide was separated from the original oligonucleotide by HPLC by using a C-4 reverse phase column and 0.1 M TEA, pH 7.0 CH_3CN gradients. The HPLC fraction was desalted on a Sephadex G-25 column (2.5 × 56 cm, 0.02 M NH_4HCO_3 as eluant). A mixture of the monoadducted oligonucleotide (1 A_{260} unit) and the unmodified complementary strand (1.5 A_{260} unit; 1.5-fold molar excess) in 0.45 mL of 0.1 M Tris, pH 7.4 buffer was deaerated by argon at 4 °C. Excess fresh, anaerobic $Na_2S_2O_4$ (20 μL of a 0.05 M solution in the same buffer) was added at once. The solution was stirred under argon for 40 min at 4 °C and then opened to air. HPLC on a C-4 reverse phase column of a small aliquot of the reaction mixture before and after the reaction was used to monitor progress. The cross-linked fraction was purified on a Sephadex G-50 column (2.5 × 56 cm, 0.02 M NH_4HCO_3) at 40 °C to isolate the MC-cross-linked duplex.

Adduct Yield Determination. (i) Percent yield of monoalkylation adducts: The Sephadex G-25 fraction containing the oligonucleotide/alkylated oligonucleotide mixture-containing fraction was digested without further separation by SVD and AP by using a standard protocol (12). The digest was separated into its components by HPLC on a C-18 reverse phase column. The yield of nucleoside monoadduct was calculated from the relative mole ratio (peak area/extinction coefficient) with respect to that of dT (12). **(ii) Percent yield of cross-linked duplex:** This was determined by passing the oligonucleotide-MC reaction mixture through a Sephadex G-50 column heated at 40 °C, which separated the cross-linked duplex from the un-cross-linked, single stranded oligonucleotide components. The yield was determined by measuring total $O.D._{260}$ units of the cross-linked and un-cross-linked fractions.[19]

Heat-stability of the adducts was determined by heating them in solution at neutral pH for 30 min at 90 °C, and monitoring any change by HPLC.

Isomeric Structure Assignments of the Mixed Bisadducts 12 and 13. Oligonucleotide duplex **19** was *monoalkylated* by monofunctionally activated MC. The reaction mixture was separated by HPLC into four components: unreacted **19** (top strand), unreacted **19** (bottom strand), monoadducted oligo **24a**, and monoadducted oligo **27b**. The latter two were distinguished based on their digestion to give the MC-dDAP adduct **10a** and MC-dG adduct **6a**, respectively. In two separate reactions each monoadducted oligonucleotide was reannealed to its complementary strand and the resulting hybrid duplex was treated with $Na_2S_2O_4$, resulting in quantitative conversion from monoadduct to cross-link (**24**→**25** and **27**→**28**, respectively; Scheme 3). Digestion of the cross-linked duplexes **25** and **28** yielded the isomeric adducts **12** and **13**, respectively, identified by HPLC. Therefore, **12** is the C-1″-DAP-linked isomer and **13** the C-1″-guanine-linked isomer.

Proof That in DAP Adducts MC Is Linked to the 2-Amino Group of DAP. D-M-G adduct 12: The guanine-N3 and -N7 positions as conjectural linkage points are excluded by the observation that the adduct was heat-stable;[28] furthermore, the guanine-O^6 and -N1 positions are excluded by the lack of cross-linking of the inosine-substituted duplex **26**. This leaves G-N^2 as the only possible linkage of the mitosene to guanine. The linkage of the same mitosene to DAP in the D-M-G cross-link then must also be located in the minor groove: either *to DAP-N3 or to DAP-*N^2. Since DAP-N3 can be excluded by the above-mentioned heat-stability of the adduct,[28] this leaves the DAP-2-amino group as the only other possible linkage site to the mitosene in D-M-G adduct **12**.

From this conclusion the other DAP adduct structure assignments also fall into place: since the D-M-G cross-link adduct **12** is formed directly from the DAP monoadduct **10a** (Scheme 3; **24** → **25** → **12**), the latter (**10a**) is also a 2-amino-linked DAP adduct. Furthermore, since **10a** is also the precursor to the D-M-D cross-link **11** (Scheme 3; **29** → **30** → **11**), cross-link **11** is a minor groove 2-amino-linked adduct of DAP as well.

PCR for the Synthesis of Natural, Inosine, DAP, and I/DAP DNAs. PCR mixtures (100 μl each, 50 mM KCl, 10 mM Tris-HCl, 0.1% Triton) each containing 1 mM Watson 1 primer, 20 μl of the

labeled Crick 1 primer, 20 ng of *tyr*T template, and 0.25 mM of each appropriate dNTP were heated at 94 °C for 1 min. Two units of *Taq* Polymerase (Promega) were then added to each reaction. Finally 20 μL of mineral oil was added to prevent evaporation. Thirty amplification cycles were performed, each cycle consisting of the following segments: (a) for natural and DAP-containing DNAs, denaturation at 94 °C for 1 min, template-primer annealing at 37 °C for 2 min, and polymerization at 72 °C for 10 min and (b) for inosine and I/DAP DNAs, 84 °C for 1 min, 30 °C for 2 min, and 62 °C for 10 min. Following the last cycle the reaction mixtures were slowly cooled (5 min at 60 °C and 5 min at 37 °C) to facilitate complete annealing of complementary DNA strands. The PCR products were then purified and recovered by 6% nondenaturating PAGE, electroelution, QIAquick PCR purification, and lyophilization.

Synthesis of Asymmetric Natural, Inosine, and I/DAP DNAs. Two parental DNA molecules, P1 and P2, were synthesized for the preparation of asymmetrical DNAs. The sequence of P1 is the same as that of the PCR template except that it is 9 bp longer. It was prepared by the same method used for making the *tyr*T PCR template except that only *Eco*RI was used to digest the pAT plasmid; it was dissolved in 90 μL of water after purification by the QIAquick PCR purification kit. P2 was synthesized by the PCR method described above except that inosine-containing primers were used. For the preparation of asymmetric natural, inosine, and I/DAP DNAs, natural base-containing, inosine-containing, and inosine/DAP-containing DNAs were used as P2, respectively, dissolved in 90 μL of water. A 200 μL mixture (50 mM KCl, 10 mM Tris-HCl, 0.1% Triton) containing 90 μL of P1, 90 μL of P2, and 0.25 mM each dNTP was heated at 94 °C for 4 min and 37 °C for 1 min to allow the hybridization of P1 and P2. It was then heated to 72 °C for 1 min at which point 5 units of *Taq* polymerase were added to carry out 3′ extension. Finally it was cooled to 37 °C for 1 min. The DNAs in the mixture were resolved by nondenaturing PAGE at 400 V for 12 h. A 6% gel of 40 cm × 30 cm × 0.3 mm was used. The gel was subjected to autoradiography by exposing it to X-ray films (Kodak), which were developed and fixed by a film developer (Optimax, IGP). The gel was also stained with ethidium bromide solution. The radioactive band with lower mobility containing the asymmetrical DNA was cut out, electroeluted, purifed, and lyophilized.

Bifunctional Alkylation (Cross-Linking) of DNA: Excess $Na_2S_2O_4$ as Activating Agent. A stock solution of 1 mM MC was prepared in 30% methanol and stored at 20 °C in the dark. It was diluted to working concentrations with appropriate volumes of 30% methanol. The reaction mixture (25 μL) of TE buffer, pH 7.0, containing 10 μg of DNA and an appropriate amount of MC, was deaerated by bubbling with nitrogen for 2 min and then capped tightly. Water (500 μL) was deaerated for 5 min, which was followed by addition of sodium dithionite ($Na_2S_2O_4$) to make up a concentration of 20 mM. The solution was further deaerated for 5 min. The reaction mixture was supplemented with 4 μL of the 20 mM sodium dithionite solution, followed by deaeration for 30 s. It was capped tightly and kept in an ice bath in the dark for 1 h. The cross-linking reaction was stopped by adding 2 μg of calf thymus DNA (Sigma) in 2 μL of pH 7.4 TE buffer, 7.5 μg of tRNA (Sigma) in 25 μL of 0.3 M sodium acetate, and 130 μL of cold ethanol. It was kept on dry ice for 10 min and then microcentrifuged at 13 000 rpm for 10 min. The supernatant was discarded. The pellet was dried in air, then dissolved in 10 μL of strand separation buffer (30% dimethyl sulfoxide, 56% formamide, 1 mM EDTA, and 10% Sigma gel loading solution).

Gel Electrophoresis and Quantification by Storage Phosphorimaging. Cross-linked products were resolved by 8% denaturing PAGE (19:1 acrylamide:bisacrylamide, 50% w/v urea, 40 cm × 30 cm × 0.3 mm) run with 1 X TBE at 1000 W for 45 min. Such gels were prepared by mixing 20 mL of 40% acrylamide and bisacrylamide solution with 80 mL of 1 X TBE and caused to polymerize by adding 0.5 mL of 10% w/v ammonium persulfate and 50 μL of TEMED. Before loading the samples were denatured at 94 °C for 4 min followed immediately by chilling an ice bath. The gels were pre-run at 1000 W until the temperature reached 60 °C. After electrophoresis they were soaked in 10% acetic acid for 10 min, transferred to Whatman 3 MM paper, and dried on a gel-drier (Bio-Rad) at 80 °C for 45 min. They were then exposed to storage phosphor screens overnight at room temperature. A Molecular Dynamics 425E PhosphorImager was used to collect the data from the storage screens. Baseline corrected scans were analyzed by integrating all band densities between two selected boundaries using ImageQuant version 3.3 software. Each resolved cross-linked DNA-containing band with retarded mobility was assigned to one or more particular position(s) of cross-linking by comparison of its mobility relative to the others.

Acknowledgment. This research was supported by a grant (CA28681) from the National Cancer Institute and a Research Centers in Minority Institutions award (RR003037) from the Division of Research Resources, NIH (both to M.T.), and by grants from the Cancer Research Campaign, the Association for International Cancer Research, and the European Union (to M.J.W.). K.S.T. thanks the British Council for a scholarship.

JA9824019