Selective Intercalation of Charge Neutral Intercalators into GG and CG Steps: Implication of HOMO-LUMO Interaction for Sequence-Selective Drug Intercalation into DNA

Kazuhiko Nakatani,* Takahiro Matsuno, Kaoru Adachi, Shinya Hagihara, and Isao Saito*

Contribution from the Department of Synthetic Chemistry and Biological Chemistry, Faculty of Engineering, Kyoto University, CREST, Japan Science and Technology Corporation (JST), Kyoto 606-8501, Japan

Received November 14, 2000

Abstract: We have synthesized naphthopyranone epoxide **4** from D-isoascorbic acid together with its three diastereoisomers. DNA alkylation of ODNs containing 5′XGT3′ and 5′TGY3′ by **4** (11*R*, 13*R*), where X and Y are any nucleotide bases, occurred at all G residues except at G of the 5′TGC3′ sequence. In contrast, the three other diastereoisomers of **4** showed only weak G alkylation activity. Differential ^{1}H NMR NOE of the **4**-G adduct confirmed the G-N7 alkylation at the epoxide carbon of **4** with concomitant S_N2 ring opening of the epoxide. Quantitative HPLC analysis of G alkylation efficiency for **4** showed the order of G alkylation susceptibility as TGGT ≈ CGT ≫ TGA > AGT > TGT ≫ TGC. The order was fully consistent with those reported for aflatoxin B_1 oxide and kapurimycin A_3, suggesting that the sequence selectivity observed for these DNA alkylating agents is not structure dependent but most likely due to the intrinsic property of DNA sequences. We found that the order of G alkylation susceptibility obtained for **4** completely matched the calculated HOMO energy level of G-containing sequences. These results underscore that **4** is a unique molecular probe for ranking the HOMO level of G-containing sequences by well-known G alkylation chemistry and suggests that the intercalation of charge neutral intercalators is a HOMO-controlled process.

Introduction

Intercalation into stacked DNA base pairs is one of the most fundamental modes of interaction between planar aromatic molecules and duplex DNAs.[1−4] Numerous structures of DNA−intercalator complexes have been determined by X-ray and NMR analyses.[5,6] Intercalation of these molecules resulted in a significant change of the DNA conformation being accompanied by unwinding and elongation of the duplex at the intercalated base steps.[2,3,7] Intercalated molecules are stacked by flanking 3′ and 5′ base pairs. In DNA−intercalator complexes, a long axis of fused aromatic rings of intercalators may be parallel to the hydrogen bonds of neighboring base pairs, whereas an orthogonal orientation of intercalators is also known as a threading type intercalation.[8,9]

Certain types of intercalators are known to eventually produce covalent DNA complexes. Aflatoxin B_1 oxide (**1**),[10] one of the most potent environmental mutagens, and kapurimycin A_3 (**2**),[11,12] a member of pluramycin antibiotics,[13,14] are the representatives of such DNA-alkylating intercalators (Figure 1). Both **1**[15−20] and **2**[21−23] alkylate N7 of guanine (G) by an

(1) Lerman, S. *J. Mol. Biol.* **1961**, *3*, 13−80.
(2) Saenger, W. *Principles of Nucleic Acids Structure*; Springer-Verlag: New York, 1983; Chapter 16.
(3) Wilsonm, W. D.; Jones, R. L. In *Intercalation Chemistry*; Whittingham, M. S., Jacobson, A. J., Eds.; Academic Press: New York, 1982; Chapter 14.
(4) Wilson, W. D. In *Nucleic Acids in Chemistry and Biology*; Blackburn, M., Gait, M., Eds.; IRL Press Ltd.: Oxford, 1990; pp 295−336.
(5) Wilson, W. D.; Li, Y.; Veal, J. M. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Ltd.: Greenwich, 1992; Vol. 1, pp 89−166.
(6) Wang, A. H.-J. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Ltd.: Greenwich, 1996; Vol. 2, pp 59−101.
(7) Waring, M. J. *Annu. Rev. Biochem.* **1981**, *50*, 159−192.
(8) Daunomycin: Wan, A. H.-J.; Ughetto, G.; Quigley, G. J.; Rich, A. *Biochemistry* **1987**, *26*, 1152−1163.
(9) Nogalamycin: (a) Liaw, Y. C.; Gao, Y. G.; Robinson, H.; van der Marel, G. A.; van Boom, J. H.; Wan, A. H.-J. *Biochemistry* **1989**, *28*, 9913−9918. (b) Gao, Y. G.; Liaw, Y. C.; Robinson, H.; Wan, A. H.-J. *Biochemistry* **1990**, *29*, 10307−10316.
(10) Busby, W. F., Jr.; Wogan, G. N. In *Chemical Carcinogens*, 2nd ed.; Searle, C. E., Ed.; American Chemical Society: Washington, DC, 1984; pp 945−1136.
(11) Hara, M.; Mokudai, T.; Kobayashi, E.; Gomi, K.; Nakano, H. *J. Antibiot.* **1990**, *43*, 1513−1518.
(12) Yoshida, M.; Hara, M.; Saitoh, Y.; Sano, H. *J. Antibiot.* **1990**, *43*, 1519−1523.
(13) Séquin, U. *Fortschr. Chem. Org. Naturst.* **1986**, *50*, 57−122.
(14) Hansen, M. R.; Hurley, L. H. *Acc. Chem. Res.* **1996**, *29*, 249−258.
(15) Essigmann, J. M.; Green, C. L.; Croy, R. G.; Fowler, K. W.; Büchi, G. H.; Wogan, G. N. *Cold Spring Harbor Symp. Quant. Biol.* **1983**, *47*, 327−337.
(16) Essigmann, J. M.; Croy, R. G.; Nadzan, A. M.; Busby, W. F., Jr.; Reinhold, V. N.; Büchi, G.; Wogan, G. N. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 1870−1874.
(17) Lin, J. K.; Miller, J. A.; Miller, E. C. *Cancer Res.* **1977**, *37*, 4430−4438.
(18) (a) Muench, K. F.; Misra, R. P.; Humayun, M. Z. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6−10. (b) Misra, R. P.; Muench, K. F.; Humayun, M. Z. *Biochemistry* **1983**, *22*, 3351−3359.
(19) (a) Benasutti, M.; Ejadi, S.; Whitlow, M. D.; Loechler, E. L. *Biochemistry* **1988**, *27*, 472−481. (b) Loechler, E. L.; Teeter, M. M.; Whitlow, M. D. *J. Biomol. Struct. Dyn.* **1988**, *5*, 1237−1257.
(20) Gopalakrishnan, S.; Harris, T. M.; Stone, M. P. *Biochemistry* **1990**, *29*, 10438−10448.
(21) Chan, K. L.; Sugiyama, H.; Saito, I. *Tetrahedron Lett.* **1991**, *52*, 7719−7722.
(22) Hara, M.; Yoshida, M.; Nakano, H. *Biochemistry* **1990**, *29*, 10449−10455.
(23) Nakatani, K.; Okamoto, A.; Saito, I. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2794−2797.

10.1021/ja003956i CCC: $20.00
Published on Web 05/25/2001

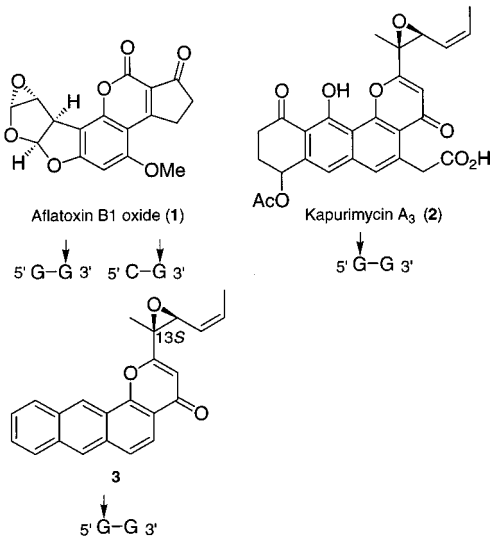


Figure 1. Structures of aflatoxin B_1 oxide (**1**), kapurimycin A_3 (**2**), and an aglycon model **3** of pluramycin antibiotics. Sequence selectivity for G alkylation of each molecule is shown below the structure. G's shown with an arrow are the alkylated G's.

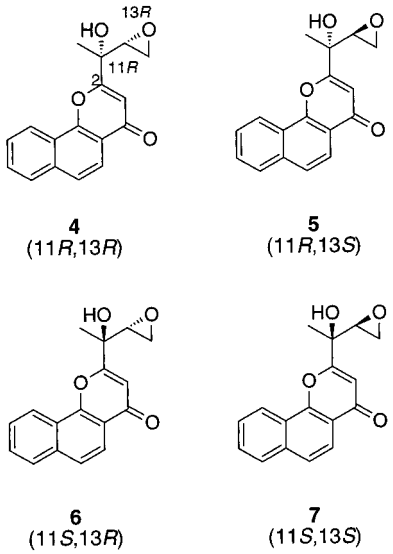


Figure 2. Naphthopyranone epoxides designed for alkylation of both 3′ and 5′ G's at the intercalation site.

electrophilic attack of the epoxide upon intercalation into DNA, forming a covalent bond between an epoxide carbon and G-N7 with concomitant S_N2 ring opening of the epoxide. Despite the lack of obvious DNA binding functionality, both **1** and **2** showed distinct sequence selectivity for G alkylation. Aflatoxin B_1 oxide (**1**) alkylates 3′ side G at the intercalation site with sequence preference of G*G* > C*G* > A*G* ≈ T*G*,[19] whereas 5′ side G was alkylated by **2** with the order of *G*G > *G*A > *G*T > *G*C (G's shown in italic are the site of alkylation).[23] We have recently reported that a synthetic aglycon model **3** of pluramycin antibiotics also selectively alkylates 5′ side G of the GG sequence.[24,25] It has been pointed out that the molecular basis for the G alkylation selectivity is ascribable to the different susceptibility of these base steps toward drug intercalation.[18,24] Accordingly, these three molecules preferentially intercalate into the GG step. On the basis of high level ab initio calculations, we have rationalized that the high preference for the GG step in the intercalation of **3** is due to the strong stabilization of the intercalated complex at the GG step by stacking of the intercalator with both 3′ and 5′ side G's.[24] However, the molecular basis for the difference between the high susceptibility of the CG sequence for **1** and the low reactivity of the GC sequence for **2** and **3** remain to be clarified.

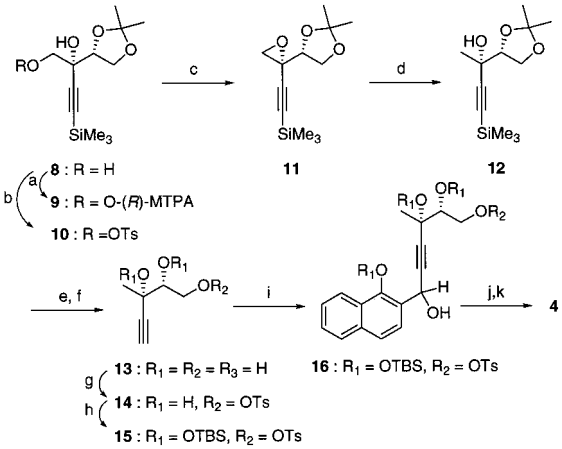
When **1** intercalated into the GC step, 5′ side G was not alkylated due to its intrinsic 3′ side G selectivity for the alkylation. Likewise, **2** cannot alkylate the 3′ side G of the **2**-CG complex. Therefore, it is not feasible to determine the efficiency of the intercalation of **1** into the GC step and **2** into the CG step by G alkylation experiments. This discrepancy of intrinsic alkylation selectivity between two drugs makes it difficult to directly compare their sequence selectivity for the intercalation. In the present work we have synthesized novel DNA alkylating intercalator **4** that alkylates both 3′ and 5′ side G's at the intercalation site to compare the susceptibility of GC and CG steps toward drug intercalation. Quantitative HPLC analysis of G alkylation at 5′XGY3′ by **4**, where X and Y are any nucleotide base, established the order of the G alkylation susceptibility of G-containing sequences. We here describe that the G selectivity observed for **4** is fully consistent with those previously observed for **1**, **2**, and **3**, suggesting that the G alkylation selectivity of charge neutral DNA intercalators is not structure dependent. Furthermore, the order of G-alkylation susceptibility is fully consistent with the calculated HOMO level of G-containing sequences. These observations indicate that interactions between the HOMO of DNAs and the LUMO of the intercalators may significantly contribute for the sequence selectivity of the drug intercalation.

Results

Molecular Design and Synthesis of a New DNA-Alkylating Intercalator. To alkylate both side G's at the intercalation site, we have designed naphthopyranone derivative **4** that has an electrophilic epoxide one carbon away from the pyranone ring as compared to **3** (Figure 2). Three diastereoisomers **5**, **6**, and **7** were also synthesized to determine the effect of the absolute configuration at the two chiral centers on the G alkylation efficiency. Synthesis of **4** starts from acetylenic alcohol **8**[26] possessing the correct absolute configuration of two chiral centers (see Scheme 1 and Supporting Information for the experimental procedure and spectral data). Alcohol **8** is readily obtained from D-isoascorbic acid. The enantiomeric purity of **8** was confirmed to be more than 98% by comparing ^{1}H NMR spectra of (*R*)-MTPA ester **9** and (*R*)-MTPA esters obtained from three diastereoisomers of **8**. The primary hydroxyl group of **8** was selectively removed by the following three-step sequence. Tosylation of **8** and a subsequent base treatment produced epoxide **11** which was reductively opened by L-Selectride to give **12**. Desilylation of **12** and a subsequent hydrolysis of acetonide produced triol **13** that was transformed into tosylate **14** without isolation. The remaining hydroxyl groups of **14** were protected as *tert*-butyldimethylsilyl ether to give **15**. Lithiation of **15** followed by addition to 1-*tert*-buthyldimethylsilyloxy-2-naphthoaldehyde gave adduct **16**. Oxidation of **16** with manganese dioxide followed by treatment of the resulting ketone with potassium fluoride in the presence of 18-crown-6 in dry DMF furnished both pyranone and epoxide formation to yield **4**.[27,28] Diastereoisomer **6** was synthesized

(24) Nakatani, K.; Okamoto, A.; Matsuno, T.; Saito, I. *J. Am. Chem. Soc.* **1998**, *120*, 11219−11225.
(25) Nakatani, K.; Okamoto, A.; Saito, I. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 3378−3381.
(26) Nakatani, K.; Arai, K.; Hirayama, N.; Matsuda, F.; Terashima, S. *Tetrahedron* **1992**, *48*, 633−650.
(27) Nakatani, K.; Okamoto, A.; Yamanuki, M.; Saito, I. *J. Org. Chem.* **1994**, *59*, 4360−4361.

Scheme 1[a]



[a] Reagents and conditions: (a) (*R*)-MTPA acid, DCC, DMAP, 63%; (b) TsCl, Py, DMAP, 98%; (c) LiHMDS, 79%; (d) LiB[CH(CH_3)-CH_2CH_3]$_3$H, 96%; (e) TBAF; (f) 1 N HCl; (g) TsCl, Py, 59% for 3 steps; (h) TBDMSOTf, 2,6-lutidine, 74%; (i) LiHMDS, $CeCl_3$, 1-*tert*-butyldimethylsilyloxy-2-naphthoaldehyde, 63%; (j) MnO_2; (k) KF, 18-crown-6, DMF, 40% for 2 steps.

from an epimer of **8** at the carbon attached to an acetylenic group. The two other diastereoisomers **5** and **7** possessing 13*S* configuration were synthesized from L-ascorbic acid (Supporting Information).

G Alkylation of the 5′XGY3′ Sequence by 4, 5, 6, and 7. G alkylations of DNA by **4**−**7** were examined with duplexes of 29-mer oligodeoxynucleotides d(TTATTGTTCGTTAGTTGGTATATTTAATA) (**ODN1**) and d(TTATGTTTGCTTGATTGGTATATTTAATA) (**ODN2**) with their complementary strands **ODN1c** and **ODN2c**. **ODN1** and **ODN2** have 5′XGT3′ and 5′TGY3′ sequences (underlined), respectively. Alkylated G sites were detected by PAGE analysis after strand cleavage induced by heating with piperidine. Efficient G alkylation was observed only for **4** (Figure 3). Three other epoxides **5**−**7** showed very weak G alkylation reactivity under the identical incubation conditions. The G alkylation efficiency of **ODN1** and **ODN2** by **4** was highly dependent on the flanking bases to the G being alkylated. The G's in TGGT, CGT, AGT, and TGA sequences were efficiently alkylated, whereas only weak alkylation was observed for the TGC sequence. Alkylation of the GG sequence by **4** was observed at both G sites.

Identification of G-Alkylation Products. To determine the chemistry of G alkylation, the reaction of **4** with 8-mer self-complementary duplex d(ATA CGT AT)$_2$ (**ACGT**/**ACGT**) possessing the 5′ACGT3′ sequence was monitored by HPLC. All oligomers used for HPLC analysis of G alkylation were specified by the sequence of the interior four bases for identification (Table 1). Incubation of **ACGT** with **4** at 0 °C produced two products **17** and **18** (Figure 4a). MALDI-TOF MS of **17** shows the molecular weight of 2690.31 that is in good agreement with the sum of the molecular weight of **ACGT** and **4**. A mixture of **17** and **18** was isolated by HPLC and heated at 90 °C for 30 min (Figure 4b). While **18** was stable under these conditions, **17** decomposed completely to yield two products **19** and **20**. Molecular weights of **18** and **19** separated by HPLC were determined by MALDI-TOF MS as 2689.48 and 2274.04, respectively. High-resolution FAB-MS of **20** showed the molecular formula of $C_{22}H_{19}N_5O_5$ (MW = 434.1458). The presence of the 1,2-diol functionality in **20** was suggested

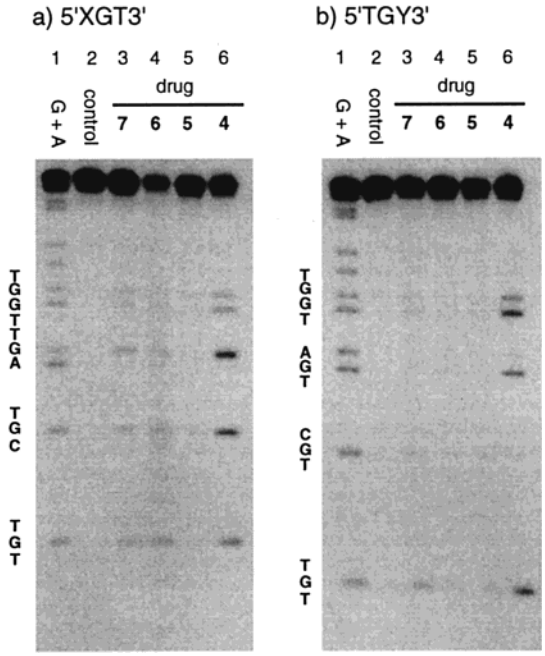


Figure 3. Sequence-selective G alkylation by naphthopyranone epoxides **4**−**7**. G alkylation reactions by **4**−**7** were examined with 29-mer duplexes of (a) d(TTA TTG TTC GTT AGT TGG TAT ATT TAA TA) (**ODN1**) and (b) d(TTA TGT TTG CTT GAT TGG TAT ATT TAA TA) (**ODN2**). ^{32}P-5′-end labeled **ODN1** and **ODN2** were hybridized with their complementary strands (1 μM) in sodium cacodylate buffer (10 mM, pH 7.0). The duplexes were incubated with drugs (100 μM) at 37 °C for 30 min. ODNs were recovered by ethanol precipitation and heated with piperidine (10% v/v) at 90 °C for 20 min. Recovered samples by ethanol precipitation were analyzed by electrophoresis on 15% polyacrylamide gel containing 7 M urea. Lane 1, Maxam−Gilbert G+A reaction; lane 2, control without drug; lane 3, **7**; lane 4, **6**; lane 5, **5**; lane 4, **4**. Partial sequences are shown in the left of the gel.

Table 1. ODNs Used for HPLC Analysis of G Alkylation with **4**

ACGT[a]	d(ATA CGT AT)
TAGT	d(ATT AGT AT)
ACTA	d(ATA CTA AT)
ATGT	d(ATA TGT AT)
ACAT	d(ATA CAT AT)
TGCA[a]	d(TAT GCA TA)
TGAT	d(ATA TGA TT)
ATCA	d(AAT CAT AT)
TGGT	d(TAT GGT ATA T)
ACCA	d(ATA TAC CAT A)

[a] **ACGT** and **TGCA** are self-complementary oligomers.

by rapid disappearance of **20** on HPLC analysis when treated with sodium periodate. On the basis of these results, **17** was assigned as the G-N7 adduct, which undergoes thermally induced depurination to produce adduct **20** and oligomer d(ATA CϕT AT) **19** containing an abasic site ϕ (Figure 5).[21] Since thermally stable DNA adduct **18** has the same molecular weight as **17**, it is presumably the adduct of the G 2-amino group. No further structure determination was attempted due to the limited amount of **18** available.

Large-scale preparation of **20** by incubation of **4** with herring sperm DNA enables us to assign the structure by ^{1}H NMR with differential NOE spectra (Figure S1 in Supporting Information). In the ^{1}H NMR spectra of **20** in DMSO-d_6, a distinct singlet assignable as the C8-H of guanine was observed at 7.74 ppm. Differential NOE spectra showed a strong NOE between the singlet of C8-H and a signal at 4.10 ppm that was assigned as one of two methylene hydrogens.

(28) Nakatani, K.; Okamoto, A.; Saito, I. *Tetrahedron* **1996**, *52*, 9427−9436.

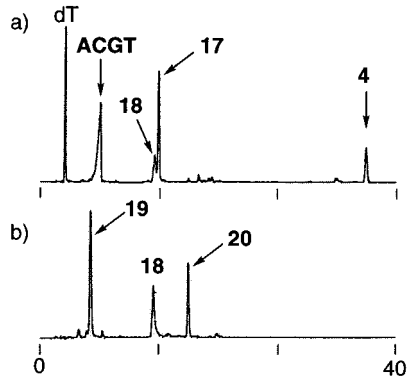


Figure 4. HPLC profiles for the reaction of self-complementary duplex **ACGT**/**ACGT** with **4. ACGT** (125 μM) was incubated with **4** (125 μM) in a 10% acetonitrile of sodium cacodylate buffer (50 mM, pH 7.0) at 0 °C. dT (500 μM) was added as an internal standard. (a) A HPLC profile obtained after incubation for 48 h. (b) A HPLC profile obtained by heating the mixture of **17** and **18** at 90 °C for 30 min.

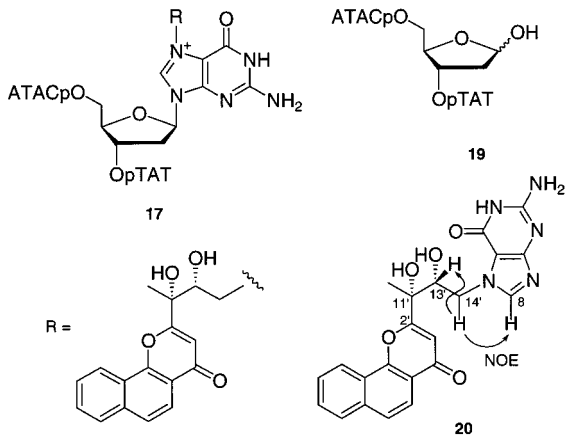


Figure 5. Structures of products obtained by the reaction of **ACGT** with **4**.

Quantitative Analysis for G-Alkylation Efficiency. We next examined the sequence dependency of the G alkylation using duplexes of short oligomers. Oligonucleotide 8-mers of **ACGT**, **TAGT**, **ATGT**, **TGCA**, and **TGAT** contain only one G residue in 5′XGT3′ or 5′TGY3′ sequence, whereas 10-mer **TGGT** contains 5′TGGT3′ sequence (Table 1). These duplexes were incubated with **4** at 0 °C and an aliquot of the mixture was analyzed by reverse-phase HPLC. Amounts of a starting oligomer and **4** were determined using dT as an internal standard. Adducts **4**-ODN were isolated by HPLC and digested to a nucleoside mixture with snake venom phosphodiesterase, P1 nuclease, and calf intestine alkaline phosphatase. The quantity of **4**-ODN adduct was determined from the amount of dT that is produced from **4**-ODN adduct by enzymatic digestion.

Initial rates of G alkylation of oligomer duplexes **ACGT**/**ACGT**, **TAGT**/**ACTA**, **ATGT**/**ACAT**, **TGCA**/**TGCA**, and **TGAT**/**ATCA** were determined from plots of the amount of **4-ODN** adduct versus the incubation period (Figure 6). A linear approximation for the plot with good correlation coefficient (R^2 = 0.97) was obtained at the beginning of G alkylation of ACGT, AGT, TGT, and TGA sequences. However, the **4**-**TGCA** adduct was not detected by HPLC analysis under the incubation conditions. This is quite consistent with the formation of an only weak cleavage band at the TGC site in PAGE analysis (cf. Figure 3b). Relative rates of G alkylation by **4** were 1.0 for CGT (standard), 0.42 for TGA, 0.34 for AGT, and 0.32 for TGT.

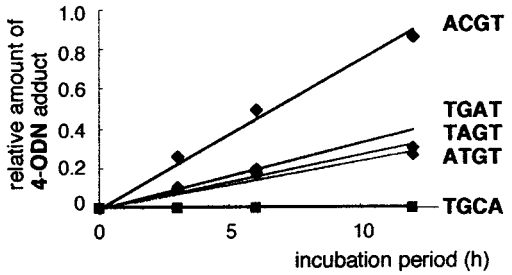


Figure 6. Relative rates for the initial G alkylation of G-containing oligomers in the presence of **4**. A linear approximation was obtained for all oligomers with good correlation coefficient (R^2 > 0.97 for all plots). Slopes for **ACGT**, **TGAT**, **TAGT**, and **TGCA** are 0.038, 0.016, 0.013, and 0.012, respectively.

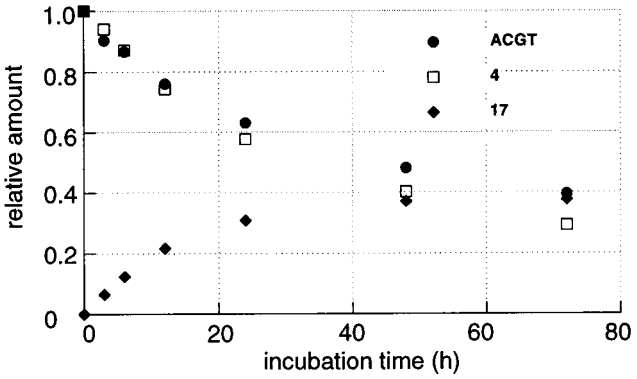


Figure 7. Time-course for the formation of **17**.

Because **4** is able to alkylate both side G's at the site of intercalation, two precovalent complexes of **ACGT**/**ACGT** with **4** with intercalation into the 5′CG3′ or 5′GT3′ step are conceivable as the precursor of **4**-**ACGT** adduct **17**. According to the nearest neighbor exclusion principle, intercalation of **4** into the CG step in the palindromic 5′ACGT3′/5′ACGT3′ sequence would result in an alkylation of G in only one strand, whereas G's in both strands would be susceptible to alkylation when **4** intercalates into the GT step.[20] Time-course experiments for the reaction of **ACGT** with **4** showed that the relative amount of **17** reaches a plateau with the ratio of ca. 0.4 (Figure 7), indicating that the alkylation is stopped at 40% conversion of **ACGT**.

HPLC analysis of the reaction of duplex **TGGT**/**ACCA** with **4** showed an efficient G alkylation of **TGGT**. However, the rate of G alkylation at the TGGT sequence could not be determined by the same procedure as for **ACGT** due to the formation of two **4**-**TGGT** N7 adducts at both 3′ and 5′ side G's. Accordingly, the rate of the G alkylation at the GG site relative to that at the CG sequence was estimated by comparing the rate of disappearance of starting oligomers **TGGT** and **ACGT** (Figure 8). As clearly seen from the figure, **TGGT** was consumed more rapidly than **ACGT**. While **TGGT** is consumed by the alkylation with **4** at G-N7 and presumably at G-2-NH_2, it is likely that the rate for G-N7 alkylation at the TGGT sequence is at least comparable to that at the CGT sequence.

Discussions

Effects of Chiral Centers on G-Alkylation Efficiency. The marked difference of the G alkylation reactivity among four diastereoisomers **4**−**7** indicated that the efficient G alkylation by **4** proceeds through a distinct precovalent complex. Molecular modeling studies of the intercalated complex of **4** with the GG step showed that (i) intercalation of the naphthalene ring parallel

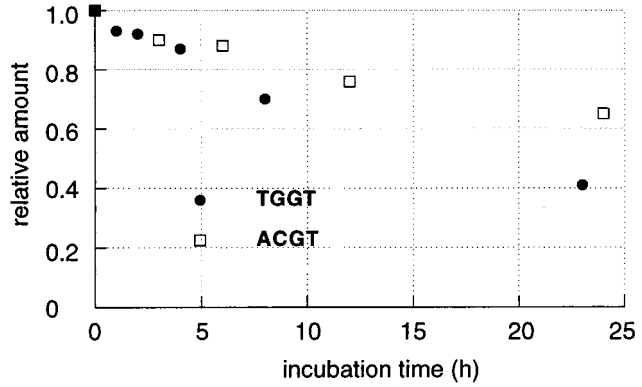


Figure 8. Time-course for the disappearance of **ACGT** and **TGGT** in the reaction with **4**.

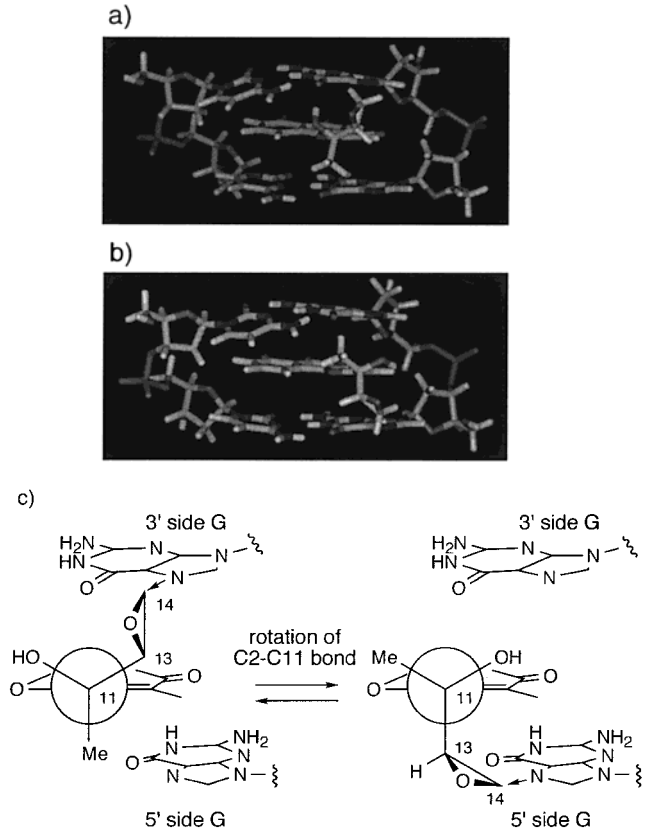


Figure 9. Proposed precovalent complexes of **4** at the GG step suitable for alkylation of (a) 3′ side and (b) 5′ side G's obtained by molecular modeling of the 8-mer duplex d(GAT GGT AC)/d(GTA CCA TC). Only the GG step and **4** were shown in the figure for clarity. (c) Partial illustration of two complexes with the Newman projection regarding a C2−C11 bond. Selected atoms were labeled for clarity. Dihedral angles of O1−C2−C11−C13 of **4** in the precovalent complex of parts a and b are −150 and 90°, respectively. Dihedral angles of C2−C11−C13−C14 of **4** for both complexes are about 90°.

to the neighboring GC base pairs seems to be essential for G alkylation, (ii) the optimal dihedral angle of C2−C11−C13−C14 for the efficient S_N2 ring opening of the epoxide would be 90° in the precovalent complex, and (iii) optimal dihedral angles of O1−C2−C11−C13 for 3′ and 5′ side G alkylation are −150 and 90°, respectively (Figure 9). The results of the conformational analysis of **4** were shown in Figure 10. The dihedral angle of C2−C11−C13−C14 for **4** is about 90° for all low-energy conformations (ΔE < 0.8 kcal/mol), whereas various values are feasible for the dihedral angle of O1−C2−C11−C13. In contrast, only one conformation that has the dihedral angles of 90° and −160° for O1−C2−C11−C13 and C2−C11−C13−C14, respectively, was found for **6** within ΔE of 1.6 kcal/mol. It has been shown that juxtaposition of N7 of G and the epoxide for in-line S_N2 attack is very important for efficient G alkylation.[24,25,29−31] Conformational analysis of **4** indicated that conformations suitable for 3′ and 5′ side G alkylation are only 0.3 and 0.5 kcal/mol above in energy the conformation at a global energy minimum. In contrast, conformations of **6** suitable for 3′ and 5′ side G alkylation are energetically less favorable by 2.2 and 1.6 kcal/mol, respectively (Figure 10). These energy differences required for conformational change of **4** and **6** reflect the efficiency for the G alkylation of these two alkylating drugs. Thus, the activation energy for G alkylation by **6** should be larger than that for **4** because an additional energy is required for the conformational change of **6**.

Molecular Basis of Sequence Selectivity for the G Alkylation. The order of G alkylation susceptibility obtained for **4** based on HPLC analysis shown in Figures 6 and 8 was compared to those reported for **1**,[19] **2**, and **3**[23] (Table 2). Despite the difference of drug structure and epoxide reactivity to G alkylation, the sequence selectivity obtained from 5′XGY3′ and **4** has a very similar tendency for the selectivity previously reported for the 5′XG3′ sequence with aflatoxin B_1 oxide **1**[19] and the 5′GY3′ sequence with kapurimycin A_3 **2** and its aglycon model **3**.[23] The GG sequence is the most preferred site for all three drugs. The CG sequence is also preferable for **1** and **4**, whereas the GC sequence is the least preferable site for **2**, **3**, and **4**.

In principle, both precovalent (intercalation) and covalent (G alkylation) binding steps are responsible for the G alkylation selectivity for DNA-alkylating intercalators. While it is very difficult to determine the rates of each step separately, we measured the relative overall rate of G alkylation by **4** at the given sequence as shown in Figures 6 and 8. To rationalize G alkylation selectivity for **1**, it has previously been proposed by two groups that the precovalent binding step is far more important for the overall G alkylation rate than the covalent binding step.[18,19] On the basis of high-level ab initio calculations and molecular dynamic simulations, we have also reached a similar conclusion, i.e., the efficiency of sequence-selective G alkylation by **3** is primarily controlled by the sequence-selective binding.[24] These conclusions are also applicable to the sequence-selective G alkylation by **4**. Thus, the results shown in Figure 6 indicate that **4** intercalates much more favorably into the CG step than the GT step. Weak preference for the TGA and AGT sequences over the TGT sequence suggests that the susceptibility of GA and AG steps for drug intercalation is slightly higher than that for TG and GT steps. It is not clear why the TGC sequence is inert toward G alkylation by **1** and **4**, because these drugs are able to intercalate into the TG step and alkylate 3′ side G.

Preferential intercalation into the GG and CG steps is not dependent on the individual drug structures. This suggests an important role of the intrinsic property of DNA sequences in the sequence selectivity. We previously reported that a precovalent complex formed between the GG step and **3** is most favorable among the complexes with other G-containing steps due to the favorable enthalpy change accompanied by drug intercalation.[24] We have calculated the enthalpy changes for the

(29) Sun, D.; Hansen, M.; Hurley, L. H. *J. Am. Chem. Soc.* **1995**, *117*, 2430−2440.
(30) Hansen, M.; Yun, S.; Hurley, L. H. *Chem. Biol.* **1995**, *2*, 229−240.
(31) Hansen, M.; Hurley, L. H. *J. Am. Chem. Soc.* **1995**, *117*, 2421−2429.

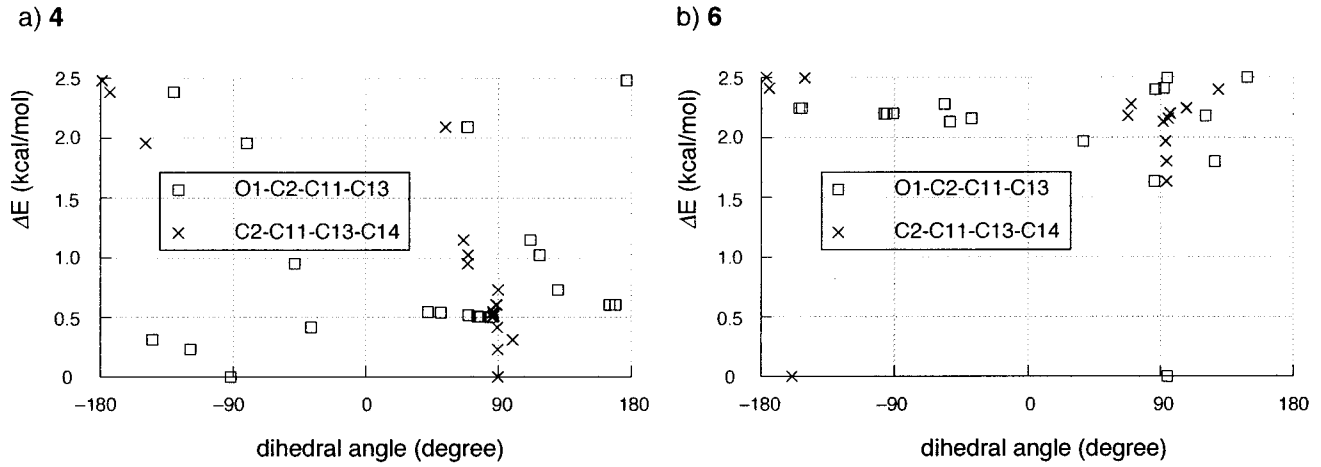


Figure 10. Conformational analysis of (a) **4** and (b) **6** regarding C2−C11 and C11−C14 bonds. Dihedral angles of O1−C2−C11−C13 and C2−C11−C13−C14 for each conformation are shown. The conformation at the lowest energy for **4** has −90 and 90° for dihedral angles of O1−C2−C11−C13 and C2−C11−C13−C14, respectively.

Table 2. Preferential Sequences for G Alkylation

drug	preferential sequence[a]
1	GGT > CGT ≫ AGT ≈ TGT
2, **3**	TGG > TGA > TGT > TGC
4	TGGT ≈ CGT ≫ TGA > AGT > TGT ≫ TGC

[a] Alkylated G's are underlined.

formation of a precovalent complex of **4** at the GC and CG steps at the B3LYP/6-31G(d) level and found that the energy difference for the formation of two complexes is only 0.1 kcal/mol.[32] These calculation results do not account for the enormous difference in the overall G-alkylation rate between these two G-containing sequences, GC and CG.

The preference of the runs of guanines for G alkylation by positively charged alkylating agents[33] such as nitrogen mustards and chloroethyltriazines has been rationalized in terms of the molecular electrostatic potential (MEP).[34] Negative MEP increases at the runs of guanines with an increase in the number of stacked G's.[34] The order of negative MEP is GG > GA > AG ≈ TG > GT > CG > GC. Therefore, positively charged drugs tend to bind more strongly to the more electronically negative G in the runs of G's.[33] However, the G-alkylation selectivity observed for **1**, **2**, **3**, and **4** could not be rationalized by MEP. Because these drugs are not positively charged, the order of each base step ranked by the magnitude of negative MEP does not fit with the order of susceptibility experimentally observed for the G alkylation, especially with regard to the marked difference in the reactivity of CG and GC steps. It has been proposed that selective alkylation at runs of G with methylating and ethylating agents via S_N2 would be rationalized by low barrier height for the transition state,[35,36] due to the low ionization potential of stacked G.[37−40] The base ionization potential is regarded as a marker for the easiness of the electronic reorganization that is necessary to accommodate significant charge transfer in the transition state.[41] Thus, the reactivity of the base for electrophiles increases as the ionization potential decreases. However, these arguments are not applicable to the G alkylation by **1**, **2**, **3**, and **4**, because these drugs intercalate into stacked base pair to form precovalent complexes prior to the covalent bond formation. Therefore, in the transition state of GG alkylation by **4**, two G's are no longer stacked. Intercalation into the GG step would collapse the stacking of the GG doublet to result in an increase in the oxidation potential.

The binding of metal ions to DNA is also known to be sequence selective.[42−45] The order of the preference for Co(II) binding determined by line broadening of G-H8 by ^{1}H NMR is GG > GA > GT ≫ GC. Mn(II) binding to the G of the GC sequence has not been detected.[42,43] The sequence selectivity has been partially rationalized by negative MEP.[44] Recently, we reported selective oxidation of DNA with Co(II) ion and benzoyl peroxide at various G-containing sequences including runs of G's.[37] The oxidation selectivity is well correlated with the distribution of HOMO of DNA,[38,39] but not with MEP, implying that the oxidation is initiated by HOMO-controlled binding of Co(II) ion to G-N7. On the basis of ab initio calculation and Co(II)-mediated G oxidation, we demonstrated the experimental HOMO mapping of duplex DNA.[37] The HOMO mapping method visualizes the susceptibility of G-containing sequences to HOMO-LUMO interaction with DNA-binding molecules.

(32) Calculations were carried out according to the method we previously described with the Gaussian 94 program:[29] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, Revision D.4.; Gaussian, Inc.: Pittsburgh, PA, 1995.

(33) Hartley, J. A. In *Molecular Basis of Specificity in Nucleic Acid−Drug Interactions*; Pullman, B., Jortner, J., Eds.; Kluwer Academic Publishers: Boston, 1990; pp 513−530.

(34) Pullman, A.; Pullman, B. *Q. Rev. Biophys.* **1981**, *14*, 289−380.

(35) (a) Kim, N. S.; Zhu, Q.; LeBreton, P. R. *J. Am. Chem. Soc.* **1999**, *121*, 11516−11530. (b) Zhu, Q.; LeBreton, P. R. *J. Am. Chem. Soc.* **2000**, *122*, 12824−12834.

(36) Kim, H. S.; LeBreton, P. R. *J. Am. Chem. Soc.* **1996**, *118*, 3694−3707.

(37) Saito, I.; Nakamura, T.; Nakatani, K. *J. Am. Chem. Soc*. **2000**, *122*, 3001−3006.

(38) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063−7068.

(39) Satio, I.; Nakamura T.; Nakatani, K.; Yoshioka, Y.; Yamaguchi, K.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 12686−12687.

(40) Prat, F.; Houk, K. N.; Foote, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 845−846.

(41) Ford, G. P.; Scribner, J. D. *Chem. Res. Toxicol.* **1990**, *3*, 219−230.

(42) Froystein, N. A.; Davis, J. T.; Reid, B. R.; Sletten, E. *Acta Chem. Scand.* **1993**, *47*, 649−657.

(43) Steinkipf, S.; Garoufis, A.; Nerdal, W.; Sletten, E. *Acta Chem. Scand.* **1995**, *49*, 495−502.

(44) Moldrheim, E.; Andersen, B.; Froystein, N. A.; Sletten, E. *Inorg. Chim. Acta* **1998**, *273*, 41−46

(45) Montrel, M.; Chuprina, V. P.; Poltev, V. I.; Nerdal, W.; Sletten, E. *J. Biomol. Struct. Dyn.* **1998**, *16*, 631−637.

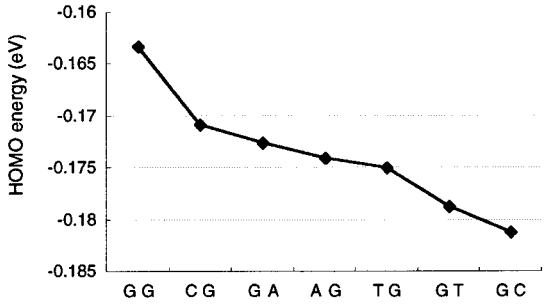


Figure 11. Calculated HOMO energy levels of two consecutive base pairs containing G in B-form structure at the B3LYP/6-31G(d) level. The HOMO energy level decreased in the order GG, CG, GA, AG, TG, GT, and GC.

We have then calculated the HOMO energy level of all sets of two base pairs containing G in the B-form structure at the B3LYP/6-31G(d) level (Figure 11).[32] The order of the HOMO level of G-containing sequences is GG > CG > GA > AG > TG > GT > GC, which is completely consistent with the order of the susceptibility experimentally observed for DNA alkylation. These results showed that **4** could be used as a unique molecular probe for raking the HOMO energy level of the G-containing DNA sequence by the well-known chemistry of DNA alkylation. Furthermore, the consistency of the HOMO energy level with the G alkylation susceptibility underscores the possibility that the intercalation of charge neutral intercalators is a HOMO-controlled process. These results further confirmed the importance of HOMO-LUMO interaction in DNA−drug interactions.

Conclusion

The molecular basis for the sequence selectivity observed for G alkylation by aflatoxin B_1 oxide (**1**) and kapurimycin A_3 (**2**) has been a subject of intense study, because these molecules lack the apparent DNA binding functionality to rationalize G-alkylation selectivity. By using synthetically designed G-alkylating drug **4**, we have shown that (i) intercalative DNA alkylating drugs have very similar sequence selectivity for G alkylation and (ii) this selectivity is derived from the intrinsic property of the G-containing DNA sequence. On the basis of these observations and the calculated HOMO energy levels of G-containing sequences, we propose that one of the most important driving forces for the intercalation of charge neutral intercalators is the interaction between the HOMO of DNA and the LUMO of intercalators. DNA-binding molecules that possess high sequence selectivity and covalently binding capability are important targets as therapeutic agents.[46−50] Our studies described here shed light on the mechanism of the sequence-selective intercalation of naturally occurring and synthetic drugs.

Experimental Section

^{1}H NMR spectra were measured with JEOL JNM α-400 (400 MHz) or Varian Mercury 400 (400 MHz) spectrometers. The chemical shifts are expressed in ppm downfield from tetramethylsilane, using residual chloroform ($\delta = 7.24$) as an internal standard. ^{13}C NMR spectra were measured with JEOL JNM α-400 (100 MHz) and Varian Mercury 400 (100 MHz) spectrometers. IR spectra were recorded on a JASCO FT-IR-5M spectrophotometer. Mass spectra were recorded on a JEOL JMS SX-102A and a JEOL JMS HX-110 for EI and FAB measurements, and a Perseptive Elite spectrometer for MALDI-TOF measurement. Trihydroxyacetophenone and ammonium citrate were used for the matrix. Wakogel C-200 was used for silica gel chromatography. Precoated TLC plates of Merck silica gel 60 F_{254} was used for monitoring the reactions and for preparative TLC. All reagents and solvents were used as received. HPLC analysis was performed on a CHEMCOBOND 5-ODS-H column (4.6 × 150 mm) with a Gilson Chromatography Model 305 using an UV detector Model 118 at 254 nm. Triethylammonium acetate (TEAA, 0.1 M) and acetonitrile were used for eluent for HPLC analysis. All enzymes and DNA oligomers used in these studies were from commercial sources. [γ-^{32}P]ATP (6000 Ci/mmol) was obtained from Amersham.

The synthetic procedure and spectral data of **4**−**7** ae available in the Supporting Information.

Alkylation of ^{32}P-5′-End-Labeled ODN1 and ODN2 with 4, 5, 6, and 7. Single-stranded 29-mer DNA oligomers 5′-d(TTA TTG TTC GTT AGT TGG TAT ATT TAA TA)-3′ (**ODN1**) and 5′-d(TTA TGT TTG CTT GAT TGG TAT ATT TAA TA)-3′ (**ODN2**) (500 pmol) were separately 5′-end labeled by phosphorylation with [γ-^{32}P]ATP (4 μL, 370 MBq/μL) and T4 polynucleotide kinase (4 μL, 10 units/μL) using standard procedures. The 5′-end-labeled ODNs were recovered by ethanol precipitation and further purified by electrophoresis on 15% denatured polyacrylamide gel and isolated by the crush and soak method. DNA alkylation experiments were carried out with a solution of 100 μL total volume. Isolated **ODN1** and **ODN2** were incubated with complementary strand **ODN1c** d(TAT TAA ATA TAC CAA CTA ACG AAC AAT AA) and **ODN2c** d(TAT TAA ATA TAC CAA TCA AGC AAA CAT AA) (1 μM), respectively, in 20 mM Tris-HCl buffer (pH 7.6) at 90 °C for 5 min, and cooled slowly to ambient temperature for duplex formation. A solution of **4**, **5**, **6**, and **7** (10 μL, 1 mM in CH_3CN, a final concentration of drug of 100 μM) was added to the DNA solution and the mixture was incubated at 37 °C for 30 min. The reaction was quenched by adding calf thymus DNA, and ODNs were recovered by ethanol precipitation, dissolved in 50 μL of 10% (v/v) piperidine, and heated at 90 °C for 20 min. The resulting mixture was concentrated in vacuo and resuspended in 10 μL of 80% formamide loading buffer (80% formamide, 1 mM EDTA, 0.1% xylene cyanole, and 0.1% bromophenol blue). The samples (1 μL) were loaded onto 15% polyacrylamide and 7 M urea sequence gel and electrophoresed at 1900 V for ca. 2 h. The gel was exposed to X-ray film with intensifying sheet at −70 °C and analyzed by densitometer.

Determination of the Intercalation Site of 4 at the 5′-ACGT-3′ Sequence. A solution (100 μL) of 125 μM of d(ATA CGT AT) **ACGT** and 500 μM of dT (internal standard) in 10% (v/v) CH_3CN/50 mM Na cacodylate buffer (pH 7.0) was incubated at 90 °C for 5 min and cooled slowly to 0 °C for duplex formation. **4** (a final concentration of 125 μM) was added and the mixture was incubated at 0 °C. An aliquot (10 μL) of the reaction mixture was analyzed by HPLC after incubation for 0, 3, 6, 12, 24, 48, and 72 h. Elution was performed with 0.1 M TEAA buffer, 10−40% acetonitrile linear gradient, for 0−40 min at a flow rate of 1.0 mL/min. Detection was carried out at 254 nm. Formation of **4**-**ACGT** adduct at G-N7 **17** was confirmed by MALDI-TOF MS (calcd 2691.92, found 2690.31). The mixture of **4**-**ACGT** at both N7 and amino groups was isolated by HPLC, lyophilized, and resolved in water. The mixture was heated to induce depurination and analyzed by HPLC. Formations of **4**-ACGT adduct **18** at an amino group and ODN containing abasic site **19** were confirmed by MALDI-TOF MS. **18**: calcd 2690.91, found 2689.48. **19**: calcd 2275.51, found 2274.04.

Synthesis and Purification of 4-G Adduct 20. To a solution of Herring Sperm DNA (20 mg) in 10% (v/v) CH_3CN/50 mM Na cacodylate buffer (20 mL, pH 7.0) was added **4** (2.8 mg, 0.01 mmol) and the mixture was incubated at 37 °C for 48 h. The resulting mixture was heated at 90 °C for 30 min. DNAs were precipitated with ethanol, and the supernatant solution was separated and concentrated in vacuo. Crude products were purified by HPLC to give **4**-G adduct **20** (2.0 mg, 46%) as a white solid: ^{1}H NMR (400 MHz, d_6-DMSO) δ 8.46 (d,

(46) Dervan, P. B.; Bulri, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688−693.
(47) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559−561.
(48) Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856−4864.
(49) Zhi-Fu, T.; Fujiwara, T.; Saito, I.; Sugiyama, H. *J. Am. Chem. Soc.* **1999**, *121*, 4961−4967.
(50) Okamoto, A.; Nakamura, T.; Yoshida, K.; Nakatani, K.; Saito, I. *Org. Lett.* **2000**, *21*, 3249−3251.

1 H, $J = 7.9$ Hz), 8.08 (d, 1 H, $J = 7.5$ Hz), 7.95 (d, 1 H, $J = 8.6$ Hz), 7.89 (d, 1 H, $J = 8.8$ Hz), 7.81−7.73 (2 H), 7.74 (s, 1 H), 6.59 (s, 1 H), 5.97 (s, 1 H), 5.89 (br, 1 H), 5.44 (d, 1 H, $J = 7.3$ Hz), 4.45 (dd, 1 H, $J = 2.4$, 13.7 Hz), 4.28 (m, 1 H), 4.10 (dd, 1 H, $J = 9.7$, 13.7 Hz), 1.65 (s, 3 H); MS (FAB) (NBA) *m*/*z* 434 [$(M + H)^+$]; HRMS calcd for $C_{22}H_{20}N_5O_5$ [$(M + H)^+$] 434.1463, found 434.1458.

HPLC Analysis of Sequence Selectivity of G Alkylation by 4. Duplexes (62.5 μM) and an internal standard of dT (500 μM) in Na cacodylate buffer (50 mM, pH 7.0) containing acetonitrile (10% v/v) were incubated at 90 °C for 5 min and cooled slowly to 0 °C for duplex formation. **4** (125 μM) was added to the DNA solution and the mixture was incubated at 0 °C. An aliquot of the reaction mixture was analyzed by HPLC after incubation for 0, 3, 6, 12, 24, 48, and 72 h. Elution was performed with 0.1 M TEAA Buffer, 10−40% acetonitrile linear gradient, for 0−40 min at a flow rate of 1.0 mL/min. Detection was carried out at 254 nm. **4**-**ODN** adducts were isolated by HPLC and confirmed by MALDI-TOF MS. **4**-**TAGT**: calcd 2706.93, found 2705.45. **4**-**ATGT**: calcd 2706.93, found 2705.09. **4**-**TGAT**: calcd 2706.93, found 2706.02. **4**-**TGGT**: calcd 3340.33, found 3339.72. Separately, **4**-ODN adducts were isolated by HPLC, lyophilized, resolved in water, and digested to nucleosides by sneak venom phosphodiesterase, P1 nuclease, and alkaline phosphatase at 37 °C for 2 h. The quantity of **4**-ODN adducts was determined by comparing the amount of the resulting dT with that of authentic sample.

Procedure for Molecular Modeling Shown in Figure 9. Proposed precovalent complexes of **4** at the GG step suitable for 3′ and 5′ side G's were obtained by molecular modeling of 8-mer duplex d(GAT GGT AC)/d(GTA CCA TC) with **4**. Manual docking of **4** to the duplex at the GG step from the major groove side produced an initial structure of **4**-DNA complex for molecular mechanics calculations. The epoxide side chain was set to point toward the 3′ side G of the GG step by rotating the C2−C11 bond. Energy minimization of the complex was carried out by using Amber* force field with GB/SA solvation treatment of water (distant-dependent dielectric electrostatic, cut off for van der Waals, electrostatic, and hydrogen bonding were 7, 12, and 4 Å, respectively). The epoxide side chain was then set to point the 5′ side G by rotating the C2−C11 bond to obtain the energy minimized structure suitable for alkylation of the 5′ side G. Two energy minimized structures were virtually indistinguishable from each other beside the conformation of the epoxide side chain. Only the GG step and **4** were shown in Figure **4** for clarity.

HOMO Calculations. HOMO energies of stacked *N*-methylated dinucleotide base pairs of GG, CG, GA, AG, TG, GT, and GC were calculated as follows: Geometries of stacked *N*-methylated dinucleotide bases at N1 (pyrimidine base) and N9 (purine base) were constructed by using the biopolymer module in Insight II (98.0) program with helical parameters of 3.38 Å for pitch, 36° for twist, and 1° for tilt. All the sugar backbones were removed except for the deoxyribose C1′ carbon and C1′ H and two H atoms were then attached to the C1′ methine to complete the stacked *N*-methylated dinucleotide base pairs. With these geometry, HOMO energies of stacked base pairs were calculated at the B3LYP/6-31G(d) level by using the Gaussian 94 program.

Conformational Analysis for 4 and 6 Regarding C2−C11 and C11−C13 Bonds. Forty-nine conformations generated by rotating both C2−C11 and C11−C13 bonds with an increment of 72° were minimized by energy with the PM3 method of semiempirical molecular orbital calculations. Conformers within the energy difference (ΔE) of 2.5 kcal/mol from the conformation at the global energy minimum were plotted in Figure 10.

Supporting Information Available: Synthetic procedure of **4**, **5**, **6**, and **7** and ^{1}H NMR spectra of **20** (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA003956I