Synthesis and Characterization of a Set of Four Dodecadeoxyribonucleoside Undecaphosphates Containing O^6-Methylguanine opposite Adenine, Cytosine, Guanine, and Thymine†

Barbara L. Gaffney, Luis A. Marky, and Roger A. Jones*

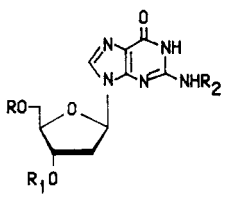
ABSTRACT: A set of four self-complementary dodecanucleoside undecaphosphates, d[CGNGAATTC(O⁶Me)GCG],[1] where N = A, C, G, or T, has been synthesized by a phosphoramidite procedure. A single large-scale preparation of the nonamer d[DMT-*G*p*A*p*A*pTpTp*C*p(O⁶Me)*G*p*C*p*G*] was divided into four portions for synthesis of the dodecamers. The synthesis, purification (high-performance liquid chromatography), and characterization of each of these molecules are described. Each sequence forms a stable duplex, with a T_m between 19 and 26 °C lower than the T_m of the "parent" molecule d(CGCGAATTCGCG). The lowest melting sequence is the N = T molecule; the overall order is N = C > A > G > T. Thus O⁶-methylation of guanine creates a region of localized instability in DNA regardless of the base opposite the lesion. This instability, which could disrupt some regulatory process or event, may be as significant as or more significant than is the mutation itself to the oncogenic process initiated by alkylating agents.

The formation of O^6-alkylguanine derivatives in DNA is thought to be the fundamental molecular event that initiates some types of carcinogenesis (Cairns, 1981; Lewis & Swenberg, 1980; Goth & Rajewsky, 1974). Compounds capable of causing this modification include not only chemical agents like diazoalkanes but also the nitrosamines present in tobacco products and the ubiquitous nitroso compounds generated in the digestive system (Singer & Grunberger, 1983). A single dose of *N*-methylnitrosourea (MeNU) has been shown to induce mammary carcinomas in rats (Sukumar et al., 1983). The molecular basis for these effects has been thought to reside in the promutagenic nature of O^6-alkylguanine. For example, each of the nine mammary carcinomas produced by MeNU was found to contain a transforming H-ras-1 gene, and in at least one case a single G → A transition was found to be responsible for activation of this oncogene. It had earlier been shown that human ras oncogenes are also produced by modification at the same position of the ras gene, either by a G → A transition (Santos et al., 1983) or by a G → T transversion (Tabin et al., 1982; Reddy et al., 1982; Taparowsky et al., 1982).

The G → A transition caused by MeNU could be the result of the mispairing of O^6-methylguanine with thymine, which would subsequently direct incorporation of adenine, overall a replacement of G with A. The preferential incorporation of thymidine (or uridine) opposite O^6-methylguanine in copolymers has been reported for *Escherichia coli* polymerase I (Abbott & Saffhill, 1979) and RNA polymerase (Gerchman & Ludlum, 1973) in in vitro replication studies. In the latter case a low level of adenosine incorporation was also noted. Other studies have focused on the incorporation of the O^6-methyldeoxyguanosine triphosphate on a template of calf thymus DNA (Hall & Saffhill, 1983) or pBR322 DNA (Toorchen & Topal, 1983). Preferential incorporation opposite thymine was again reported.

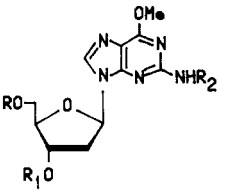
A different perspective on O⁶MeG[1] pairing was obtained in a study of O⁶MeG sites generated in *E. coli* using *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG). The *E. coli* mismatch correction system recognized all O⁶MeG·N base pairs as mismatches. In particular, the O⁶MeG·T base pair, presumably the predominant base pair present after replication, was treated as a mismatch by this system (Karran & Marinus, 1982).

In order to define the effects of guanine O⁶-alkylation on DNA structure and stability, we have begun a program of defined chemical synthesis and physical characterization of DNA fragments containing O^6-methyldeoxyguanosine [d(O⁶Me)G, **2a**]. To this end we first devised a practicable



1a: R=R1=R2=H

b: R=R1=R2=Ac



2a: R=R1=R2=H

b: R=R1=H; R2=Ac

synthesis of d(O⁶Me)G itself via O⁶-sulfonylation and displacement (Gaffney & Jones, 1982a). Subsequently, we reported the synthesis and characterization of two hexanucleoside pentaphosphates containing d(O⁶Me)G (Kuzmich et al., 1983). We now wish to report the synthesis and characterization of a set of four dodecanucleoside undecaphosphates, d[CGNGAATTC(O⁶Me)GCG], where N = A, C, G, or T. These are the longest synthetic sequences containing d-

† From the Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903. *Received September 4, 1984.* This work was supported by Biomedical Research Support Grant and grants from the American Cancer Society (CH-248) and the National Institutes of Health (GM 31483).

[1] Abbreviations: DMT, 4,4'-dimethoxytrityl; dA, deoxyadenosine; dC, deoxycytidine; dG, deoxyguanosine; dT, thymidine; d(O⁶Me)G, O^6-methyldeoxyguanosine; d*A*, *N*-benzoyldeoxyadenosine; d*C*, *N*-benzoyldeoxycytidine; d*G*, *N*-isobutyryldeoxyguanosine; *d(O⁶Me)G*, O^6-methyl-*N*-acetyldeoxyguanosine; p, internucleotide methyl phosphotriester; *p*, methyl diisopropylphosphoramidite; Ac, acetyl; DMAP, 4-(dimethylamino)pyridine; TPSCl, triisopropylbenzenesulfonyl chloride; DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene; TEAA, triethylammonium acetate; AA, ammonium acetate; CD, circular dichroism; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; TCA, trichloroacetic acid; THF, tetrahydrofuran.

0006-2960/84/0423-5686$01.50/0

(O^6Me)G that have been reported and the first prepared by the phosphoramidite method.

Experimental Procedures

Materials. Fractosil 500 silica was derivatized according to the literature (Chow et al., 1981) and loaded as reported previously (Gaffney et al., 1984). Deoxynucleosides were protected as reported elsewhere (Ti et al., 1982) or below. Chloro(*N*,*N*-diisopropylamino)methoxyphosphine was prepared and used according to the literature (McBride & Caruthers, 1983).

Methods. Pyridine, triethylamine, diisopropylethylamine, diisopropylamine, and acetonitrile were dried by distillation from calcium hydride. Purification of oligonucleotides was effected on a 7.8 mm × 30 cm μBondapak C_{18} column manufactured by Waters Associates. Analytical HPLC was performed with a Radial-Pak Bondapak C_{18} cartridge (10 μm unless otherwise indicated) in a Waters Z module. Two Waters M6000A pumps, a model 660 solvent programmer and a model 440 UV detector, were employed. The absorption spectra, melting curves, and circular dichroism spectra were obtained as described previously (Gaffney et al., 1984). The concentrations in single strands, C_T, of d[CGNGAATTC-(O^6Me)GCG] were determined with molar absorptivities, ϵ, at 260 nm and 80 °C of 1.17×10^5 for N = A, 1.09×10^5 for N = C, 1.17×10^5 for N = G, and 1.10×10^5 for N = T, as determined by enzymatic degradation.

2-N,3′-O,5′-O-Triacetyl-2′-deoxyguanosine (**1b**). To 1.3 g (5.0 mmol) of 2′-deoxyguanosine (**1a**) and 200 mL of dry pyridine were added 4.7 mL (50 mmol) of acetic anhydride, 61 mg (0.5 mmol) of DMAP, and 7.6 mL (55 mmol) of triethylamine. The mixture was stirred at 50 °C for 20 h, and then 50 mL of cold H_2O was added to the chilled, dark solution. Pyridine was removed by evaporation, an additional 100 mL of H_2O was added, and the solution was concentrated to induce crystallization. The product was collected by filtration to give 1.50 g (76%). The crystals softened at 190 °C and decomposed at 225 °C. Spectral data: ^{1}H NMR ($CDCl_3$/Me_2SO-d_6) δ 2.05 (s, 3, CH_3CO), 2.10 (s, 3, CH_3CO), 2.22 (s, 3, CH_3CO), 2.48–3.48 (m, 2, $H_{2',2''}$), 4.28 (m, 3, $H_{4'}$, $H_{5',5''}$), 5.38 (m, 1, $H_{3'}$), 6.28 ("t", 1, J_{app} = 7 Hz, $H_{1'}$), 8.08 (s, 1, H_8), 11.62 (br s, 1, NH), 12.05 (br s, 1, NH); UV_{max} (MeOH) 255 nm (ε 13 600); UV_{sh} 278 nm (ε 9500); UV_{min} 224 nm (ε 1500). Anal. Calcd for $C_{16}H_{19}N_5O_7$: C, 48.85; H, 4.87; N, 17.81. Found: C, 48.65; H, 4.96; N, 17.65.

2-N-Acetyl-6-O-methyl-2′-deoxyguanosine (**2b**). To 4.3 g (10.9 mmol) of **1b** dried 3 times by evaporation of toluene and dissolved in 40 mL of methylene chloride were added 6.6 g (22 mmol) of TPSCl, 6.1 mL (44 mmol) of triethylamine, and 66 mg (0.55 mmol) of DMAP. After 2 h, the methylene chloride was removed by evaporation and 15 mL of methanol was added. The solution was chilled in an ice bath and 1 mL (11 mmol) of cold trimethylamine was added, followed after 10 min by 8.1 mL (55 mmol) of DBU. After 1 h, 20 mL of H_2O and 2.9 g (55 mmol) of ammonium chloride were added, and the mixture was concentrated and dried by evaporation of pyridine. The mixture was suspended in methylene chloride, filtered, and purified twice by flash chromatography on silica gel. The product was crystallized from methanol/ethyl acetate to give 1.42 g (40%), mp 165–170 °C dec. Spectral data: ^{1}H NMR ($CDCl_3$/Me_2SO-d_6) δ 2.32 (s, 3, CH_3CO), 2.05–2.98 (m, 2, $H_{2',2''}$), 3.12–3.78 (m, 3, OH, $H_{5',5''}$), 3.98 (m, 1, $H_{4'}$), 4.10 (s, 3, CH_3), 4.52 (m, 1, $H_{3'}$), 5.22 (br s, 1, OH), 6.40 ("t", 1, J_{app} = 7 Hz, $H_{1'}$), 8.33 (s, 1, H_8), 10.0 (br s, 1, NH); UV_{max} (MeOH) 265 nm (ε 16 400); UV_{min} 234 nm (ε 5900). Anal. Calcd for $C_{13}H_{17}N_5O_5 \cdot {}^1/_4CH_3CO_2C_2H_5$: C, 48.69; H, 5.55; N, 20.28. Found: C, 48.91; H, 5.81; N, 20.42.

Oligonucleotide Synthesis. A 25-mL flask fitted with a frit and a stopcock was used. Detritylation of the silica-bound oligonucleotide was accomplished by one of two methods. When adenosine was the 5′-terminal nucleoside, two 30-min treatments at room temperature with 20 mL of 1 M $ZnBr_2$ in CH_2Cl_2/2-propanol (85/15) alternating with a wash of the same solvent were employed, followed by quenching with 0.5 M TEAA in dimethylformamide. In all other cases, four to six 30-s treatments with 10 mL of 2% TCA in CH_2Cl_2 at about −20 °C alternating with washes of CH_2Cl_2 at −20 °C were employed, followed by quenching with pyridine. The silica was washed with three portions of CH_2Cl_2 and three portions of CH_3CN. The flask was then capped with a septum, and the silica was washed with three more portions of dry CH_3CN and dried by passing N_2 through it until it was at constant weight.

A stock solution of tetrazole was prepared by drying it twice by evaporation of CH_3CN, adding excess CH_3CN, and evaporating to a concentration of 50 mg/mL. The appropriate phosphoramidite (10–20 equiv) was dried 3 times by evaporation of acetonitrile, the last time to a volume of about 1 mL. A portion of the tetrazole solution (3 equiv relative to the phosphoramidite) was added by syringe to the silica, followed by the phosphoramidite. A flow of N_2 through the flask was maintained during these transfers. The silica was shaken for 20 min, filtered, and washed with three portions of CH_3CN and one portion of lutidine/THF/H_2O (2/2/1).

The phosphite was oxidized by treatment with 10 mL of 0.2 M I_2 in lutidine/THF/H_2O (2/2/1) for 1 min, followed by rinsing with three portions of CH_3CN. Coupling efficiency could be determined at this point by a trityl assay as described above on 5–7 mg of the dried silica or could be calculated by assay of the combined filtrates after the next detritylation. Unreacted sites were capped by shaking for 5 min with pyridine/Ac_2O/*N*-methylimidazole (9/1/0.1), followed by washing with three portions of CH_3CN and three portions of CH_2Cl_2. After the last cycle, the oligomer was oxidized but not capped or detritylated.

Deprotection and Purification. The phosphate triesters were deprotected by a 45-min treatment with 15 mL of dioxane/triethylamine/thiophenol (2/1/1), followed by washing with CH_3CN. The silica was then shaken with 20 mL of concentrated NH_3 for 16 h, filtered, and washed with CH_3OH. The filtrate and washes were concentrated with addition of pyridine to prevent detritylation, dried thoroughly by evaporation of pyridine, and treated for 3 days with CH_3OH/pyridine/DBU (7/2/1). This mixture was then concentrated and applied to a Bio-Gel P2 column in 20% CH_3CH_2OH. The combined product fractions were concentrated in the presence of pyridine and purified on a semipreparative C_{18} μBondapak column with a gradient of 10–40% CH_3CN/0.1 M TEAA in 30 min at 2 mL/min. The combined product fractions were concentrated and detritylated with 80% AcOH for 20 min. The AcOH was removed by evaporation under reduced pressure, and the dimethoxytritanol was removed by extraction with ether. The product solution was concentrated and purified on the semipreparative column with a gradient of 12–30% CH_3CN/0.1 M TEAA in 45 min at 2 mL/min. If it was not completely pure, the product was again purified on the semipreparative column with a gradient of 10–30% CH_3CN/0.1 M AA in 45 min at 2 mL/min. The combined product fractions were lyophilized, desalted on a Bio-Gel P2 column, and again lyophilized. Samples of each pure oligonucleotide were treated with venom phosphodiesterase and bacterial alkaline phos-

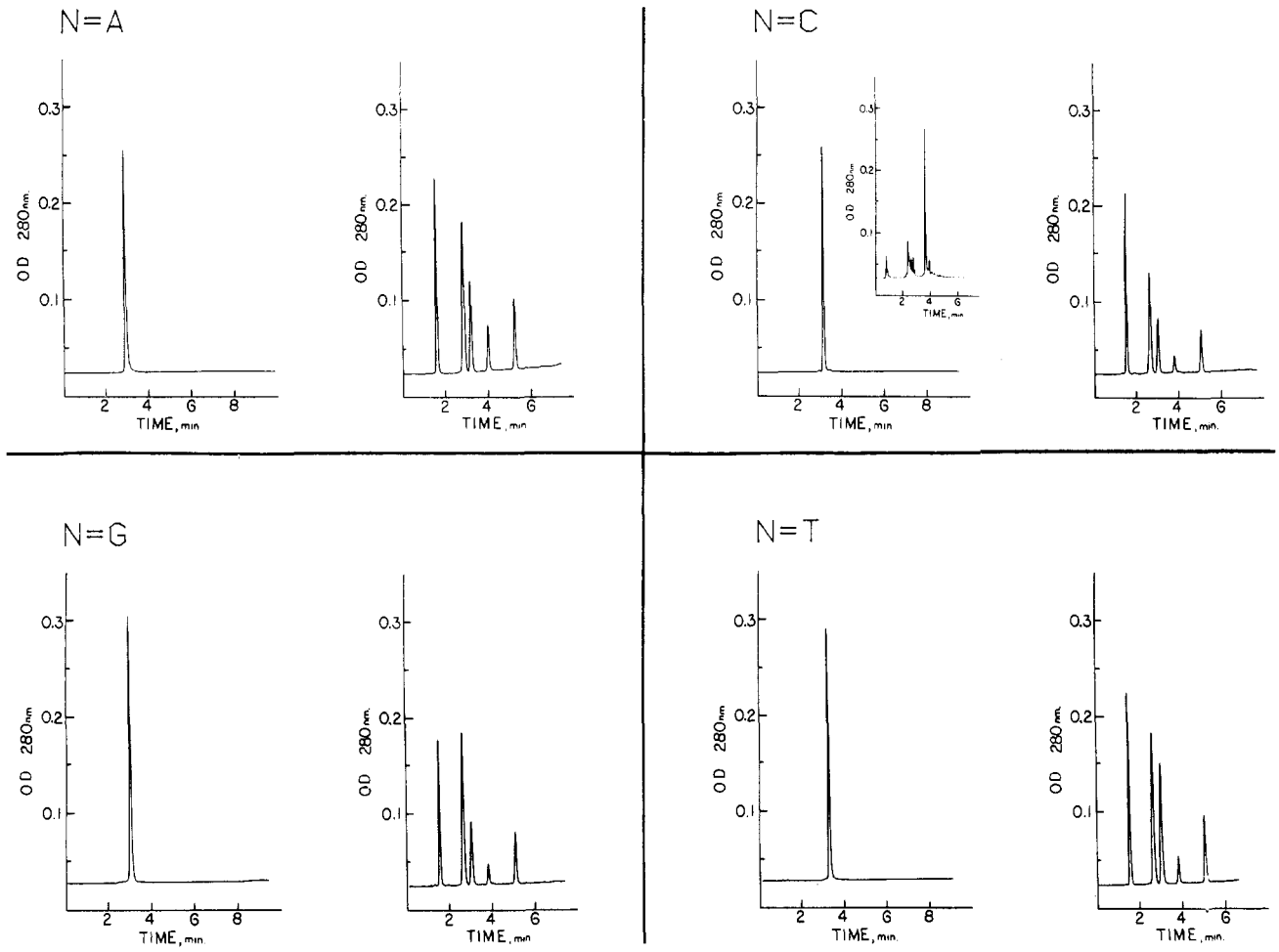


FIGURE 1: Reversed-phase HPLC profiles of d[CGNGAATTC(O^{6}Me)GCG], N = A, C, G, or T, as indicated. (Left) After purification, the gradient is 12–30% acetonitrile/0.1 M TEAA in 5 min at 4 mL/min; (right) after degradation with venom phosphodiesterase and alkaline phosphatase, the gradient is 6–20% acetonitrile/0.1 M TEAA in 5 min at 4 mL/min. The deoxynucleosides elute in the order dC, dG, dT, dA, and d(O^{6}Me)G. The inset in panel N = C is the partially deprotected sequence immediately after cleavage from the silica support; the gradient is 10–40% acetonitrile/0.1 M TEAA in 5 min, on a 5-μm Nova-Pak cartridge.

phatase. The expected ratios of monomers were obtained, as determined by analytical HPLC.

Results and Discussion

Synthesis. The synthesis of the four dodecamers was carried out from a single large-scale synthesis of the nonamer **4**. From

d(G)-silica
3
d[CGNGAATTC(O^{6}Me)GCG]
6

d[GpApApTpTpCp(O^{6}Me)GpCpG]-silica
4
d[CpGpNpGpApApTpTpCp(O^{6}Me)GpCpG]-silica
5

3 g of d(DMT-*G*)–silica [**3** (DMT); 59 μmol/g] was obtained the nonamer in 36% yield (by trityl assay) by successive additions of the appropriate nucleoside diisopropylphosphoramidites. Zinc bromide was used for detritylation after the two condensations with the deoxyadenosine phosphoramidite [d(DMT-*Ap*)]; in all other cases 2% trichloroacetic acid (TCA) was used. On this large scale we used a reaction time of 20 min for the coupling and 10–20 equiv of each phosphoramidite. It is likely that fewer equivalents of phosphoramidite would give the same results; certainly there was no clear correlation with yield within the 10–20 range. Although it has been suggested that capping is not necessary, we included a capping step because our yields (trityl assay) were often below 95%; the capping step also helped to dry the silica, because on this scale it is insufficient just to blow nitrogen through it.

The nonamer was then divided into four equal portions for synthesis of the dodecamers. The overall yields (from **3**) were 30, 32, 25, and 24%, respectively, for the N = A, C, G, and T sequences (**5**). The average yield per coupling step was only 91%, which is somewhat lower than the >97% efficiencies reported for small-scale syntheses, but is substantially better than our average yields with phosphate triester syntheses. Moreover, the coupling reactions are so much cleaner than is the case with the phosphate method that we have not used a guanine O^6 protecting group (Gaffney & Jones, 1982b) in this phosphoramidite approach. Cleaner coupling reactions also mean that a larger proportion of pure product can be obtained from the "crude" product upon preparative HPLC. The isolated yield of pure dodecamer **6** varied from 500 to 650 OD_{260} per gram of **3**. The difference in actual yield between the two methods is thus even larger than is apparent from the difference in coupling efficiency. This makes the phosphite method more economical than the phosphate, as well as much faster, even on this scale and at a 10–20-fold excess per coupling.

Because of the d(O^{6}MeG) present in these sequences, deprotection requires unusual care. We have found that thiophenoxide ion is capable of slowly demethylating d(O^{6}MeG); after 24 h a sample of d(O^{6}Me)G showed about 5% demethylation. Hence this reaction time should be kept to a minimum. Forty-five minutes was satisfactory. We had previously reported (Kuzmich et al., 1983) that ammonolysis

of the O^6-methyl group is also a potential problem, especially when heated. However, at room temperature d(O^6Me)G showed less than 2% ammonolysis (by HPLC) after 48 h in concentrated ammonia. Therefore, we did use ammonia at room temperature for 16 h to cleave the crude product from the silica support. A representative HPLC profile of the crude product obtained at this stage is shown in Figure 1 (inset, panel N = C). The non-trityl failure sequences are clustered at 2.4 min (9.6 mL), while the tritylated product appears at 3.7 min (14.8 mL). It is likely, however, that some of the amino protecting groups are still present, although the HPLC does not show this heterogeneity because of the overriding effect of the dimethoxytrityl group on retention.

We had found in our earlier syntheses that the N^2-isobutyryl group of d(O^6Me)G is difficult to remove. We have, therefore, switched to use of the acetyl group for amino protection of d(O^6Me)G. Triacetylated deoxyguanosine can be prepared in good yield and readily undergoes our sulfonylation/displacement procedure for O^6-alkylation to give crystalline material in 40% yield. This procedure has the advantage of being absolutely specific for O^6-alkylation, in contrast to direct alkylation methods that must rely on selective reaction at the O^6 position. The acetyl group is removed quite readily from **2b**, as expected, but since the rate of cleavage decreases in an oligomer, we used a 3-day treatment with methanol/DBU to ensure complete deprotection. The DBU/methanol mixture does not appear to cause any degradation under the reasonably anhydrous conditions that we have used.

The oligomer was then purified by reversed-phase HPLC both before and after detritylation. The first HPLC purification serves mainly to remove the non-trityl failure sequences. The second is then able to give highly pure material. However, for the N = T and N = C sequences, there was an impurity that was not separated by the second semipreparative column, although it was resolved, barely, on the analytical column. For these sequences an additional HPLC purification using 0.1 M AA as the aqueous phase was necessary to obtain completely homogeneous material. We generally use 0.1 M TEAA in these separations, but 0.1 M AA can also be used. In this case we were able to take advantage of the fact that the separations obtained with these two buffers usually show slight differences in relative retentions among the different components present. The HPLC profiles obtained for the pure oligomers, together with enzymatic digests, are shown in Figure 1.

It is worth emphasizing that obtaining high purity, especially with O^6MeG present, requires great care during deprotection and rigorous chromatographic purification and analysis under the highest resolution conditions possible.

Stability. The melting temperature, T_m, of a double-helical molecule is proportional to the free energy difference, $\Delta G°$, between the initial duplex form and the final single-stranded states. Furthermore, the difference in T_m between two duplexes, ΔT_m, is proportional to the free energy difference, $\Delta\Delta G°$, between the two duplexes provided that both duplexes melt to thermodynamically equivalent final single-stranded states.

We have determined the T_m of each of the alkylated sequences d[CGNGAATTC(O^6Me)GCG], N = A, C, T, or G, over a range of concentrations (ca. 10^{-4}–10^{-6} M in single strands). The resulting data are shown in Figure 2 as a plot of the inverse melting temperature, T_m^{-1}, against the logarithm of the concentration, ln C_T. Such a plot allows comparison of the melting temperatures of different molecules at the same concentration. T_m data for the "parent" molecule d(CGCGAATTCGCG) and the G·T "mismatch" molecule d(CGTGAATTCGCG) also have been included in Figure 2 for the purpose of comparison (Marky et al., 1981). The T_m's, ΔT_m's, and $\Delta\Delta G°$'s derived from these plots are listed in Table I for a concentration of 10^{-5} M.

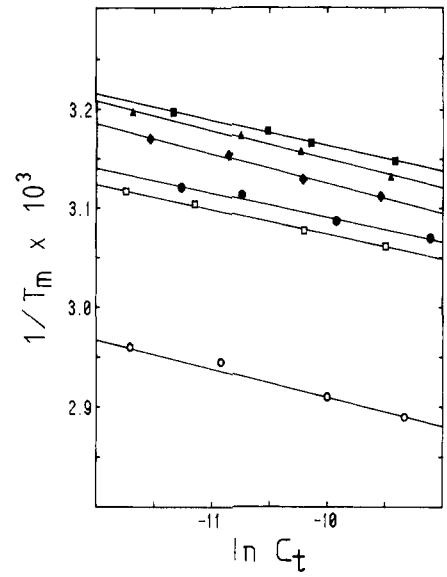


FIGURE 2: Plots of T_m^{-1} vs. ln C_T in 10 mM phosphate buffer, 1 M NaCl, and 1 mM EDTA at pH 7.0 for d[CGNGAATTC(O^6Me)GCG] [N = C (●); N = A (♦); N = G (▲); N = T (■)], d(CGCGAATTCGCG) (○), and d(CGTGAATTCGCG) (□) over a concentration range of 8–110 μM.

Table I: T_m, ΔT_m, and $\Delta\Delta G°$ of the Helix-to-Coil Transition at 10^{-5} M

compound	T_m (°C)	ΔT_m (°C)	$\Delta\Delta G°$ (cal/mol)[a]
d[CGNGAATTC(O^6Me)GCG]			
N = C	46.5	18.5	426
N = A	42.3	22.7	523
N = G	39.9	25.1	578
N = T	39.1	25.9	596
d(CGCGAATTCGCG)	65.0		
d(CGTGAATTCGCG)	48.2	16.8	387

[a] $\Delta\Delta G° = R(\Delta T_m) \ln C_T^{-1}$.

The destabilization caused by guanine O^6-methylation is evidenced by the substantially lower T_m's seen for each of the alkylated sequences. The corresponding $\Delta\Delta G°$ values are all fairly similar and are larger than that for the G·T mismatch sequence. These $\Delta\Delta G°$ values are, however, small enough to suggest that stacking is maintained in the alkylated molecules, as it was with the G·T mismatch (Patel et al., 1982). Similar results have been reported recently for other single base-pair mismatches (Tibanyenda et al., 1984). Surprisingly, it is the sequence with d(O^6Me)G opposite T that has the lowest T_m (largest $\Delta\Delta G°$) of them all, even though it appears that O^6MeG "codes" for T(U), while the N = C sequence has the highest T_m (smallest $\Delta\Delta G°$). Apparently in this case replication is not based on a more favorable free energy of O^6MeG·T "pairs" relative to other possible O^6MeG pairs.

From the data in Figure 2 it is possible to calculate van't Hoff enthalpy and entropy if a two-state melting transition can be assumed. However, if this assumption is not valid, the ΔH obtained will be lower than the true, calorimetric ΔH. In fact, neither the parent nor G·T mismatch molecules behave in a two-state fashion, and their van't Hoff and calorimetric enthalpies differ widely (Patel et al., 1982). Moreover, the size of the cooperative unit for these alkylated sequences may well be different from that of the parent. Thus, until these melting transitions have been characterized calorimetrically

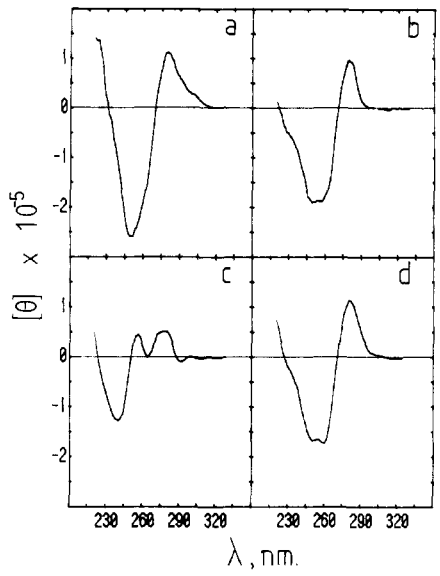


FIGURE 3: Circular dichroism spectra in 10 mM phosphate buffer, 1 M NaCl, and 1 mM EDTA at pH 7.0 and at 10 °C for d-[CGNGAATTC(O^6Me)GCG], where N = C (a), N = A (b), N = G (c), and N = T (d), at a concentration of 10 μM.

(in progress) enthalpy calculation and comparison would be premature.

The T_m values, however, are much less model sensitive and can be confidently compared with the T_m's of the parent and mismatch molecules. Yet the question of sequence dependence remains. While these dodecamers differ at only one residue, the T_m's observed reflect the melting behavior of the entire molecule, not just a particular O^6MeG·N pair. In order to be able to account properly for such sequence effects, we are synthesizing a different series of oligomers with varied nearest neighbors. Nevertheless, from the data available so far we can suggest that one effect of O^6-methylation on DNA is to create regions of localized instability in the duplex, regardless of the O^6MeG·N pair present. Whether this is due to an absence of effective base pairing, or to some distortion generated because of such pairing, remains to be determined. NMR studies should help to clarify this point; these are under way. The fact that a clear preference for formation of O^6MeG·T(U) pairs during replication has been reported may indicate that pairing between O^6MeG and T(U) does occur, while the low T_m we find for the N = T sequence may reflect some distortion of the duplex caused by such pairing.

Circular Dichroism. The CD spectra of the four d(O^6Me)G sequences in 1 M NaCl at low temperature are shown in Figure 3. The N = A, C, and T sequences display essentially similar CD's while that of the N = G sequence is different, perhaps because of the run of three consecutive guanine residues. Each is a semiconservative CD spectrum consistent with a "B" conformation.

Summary

The physical characterization of these d(O^6Me)G-containing molecules shows that the misincorporation of thymidine by d(O^6Me)G results in a net destabilization of the duplex, at least for the specific sequences studied. This result is consistent with our earlier work with the hexanucleoside pentaphosphates d[CGC(O^6Me)GCG] and d[CGT(O^6Me)GCG] for which very similar T_m's were observed, with the T sequence being slightly lower (Kuzmich et al., 1983). This destabilization is also consistent with the fact that the d(O^6Me)G·T pair is recognized as a mismatch by the *E. coli* mismatch correction system. Thus, in addition to resulting in a G·C → A·T mutation, O^6-methylation of mature DNA creates a region of localized instability that will persist unless or until the O^6MeG in the parent strand is demethylated by a methyltransferase. In fact, the low levels of methyltransferase present in certain tissues correlate well with the tissue-specific tumorigenicity of alkylating agents (Lewis & Swenberg, 1980). It may be that this localized instability, perhaps in the form of a kink or twist, disrupts some regulatory process or event within the cell that could be as significant as or more significant to the oncogenic process than is the mutation itself. A recent study of mutation and transformation by MNNG noted that these two events occur in markedly different time frames (Bignami et al., 1984). While maximal induction of point mutations is completed in a relatively short time, maximal neoplastic transformation requires a much longer time of exposure to MNNG, suggesting that the chemically induced transformation is the result of a more complex event than a single gene mutation. Further physical evaluations of these molecules and of others now being synthesized should help to clarify the effect on DNA of guanine O^6-alkylation.

Acknowledgments

We thank Dr. Kenneth J. Breslauer for helpful discussions and for use of his laboratory facilities (GM 23509) and M. Senior and X. Gao for excellent technical assistance.

References

Abbott, P. J., & Saffhill, R. (1979) *Biochim. Biophys. Acta 562*, 51–61.

Bignami, M., Ficorella, C., Dogliotti, E., Norman, R. L., Kaighn, M. E., & Saffiotti, U. (1984) *Cancer Res. 44*, 2452–2457.

Cairns, J. (1981) *Nature (London) 289*, 353–357.

Chow, F., Kempe, T., & Palm, G. (1981) *Nucleic Acids Res. 9*, 2807–2817.

Gaffney, B. L., & Jones, R. A. (1982a) *Tetrahedron Lett. 23*, 2253–2256.

Gaffney, B. L., & Jones, R. A. (1982b) *Tetrahedron Lett. 23*, 2257–2260.

Gaffney, B. L., Marky, L. A., & Jones, R. A. (1984) *Tetrahedron 40*, 3–13.

Gerchman, L. L., & Ludlum, D. B. (1973) *Biochim. Biophys. Acta 308*, 310–316.

Goth, R., & Rajewsky, M. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 639–643.

Hall, J. A., & Saffhill, R. (1983) *Nucleic Acids Res. 11*, 4185–4193.

Karran, P., & Marinus, M. G. (1982) *Nature (London) 296*, 868–869.

Kuzmich, S., Marky, L. A., & Jones, R. A. (1983) *Nucleic Acids Res. 11*, 3393–3404.

Lewis, J. G., & Swenberg, J. A. (1980) *Nature (London) 288*, 185–187.

Marky, L. A., Kozlowski, S., & Breslauer, K. J. (1981) *Biophys. J. 33*, 318a.

McBride, L. J., & Caruthers, M. H. (1983) *Tetrahedron Lett. 24*, 2953–2956.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Dallas, J., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 437–444.

Reddy, E. P., Reynolds, R. K., Santos, E., & Barbacid, M. (1982) *Nature (London) 300*, 149–152.

Santos, E., Reddy, E. P., Pulciani, S., Feldman, R. J., & Barbacid, M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4679–4683.

Singer, B., & Grunberger, D. (1983) in *Molecular Biology*

of Mutagens and Carcinogens, Plenum Press, New York.
Sukumar, S., Notario, V., Martin-Zanca, D., & Barbacid, M. (1983) *Nature (London) 306*, 658–661.
Tabin, C. J., Bradley, S. M., Bargmann, C. I., Weinberg, R. A., Papageorge, A. G., Scolnick, E. M., Dhar, R., Lowy, D. R., & Chang, E. H. (1982) *Nature (London) 300*, 143–149.
Taparowsky, E., Suard, Y., Fasano, O., Shimizu, K., Goldfarb, M., & Wigler, M. (1982) *Nature (London) 300*, 762–765.
Ti, G. S., Gaffney, B. L., & Jones, R. A. (1982) *J. Am. Chem. Soc. 104*, 1316–1319.
Tibanyenda, N., De Bruin, S. H., Haasnoot, C. A. G., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1984) *Eur. J. Biochem. 139*, 19–27.
Toorchen, D., & Topal, M. D. (1983) *Carcinogenesis 4*, 1591–1597.

Isolation of a Human cDNA for α_2-Thiol Proteinase Inhibitor and Its Identity with Low Molecular Weight Kininogen†

Iwao Ohkubo, Kotoku Kurachi,* Toshihide Takasawa, Hiroyuki Shiokawa, and Makoto Sasaki

ABSTRACT: A λgt11 cDNA library containing DNA inserts prepared from human liver mRNA has been screened with an antibody to human α_2-thiol proteinase inhibitor that was isolated from fresh plasma. Eighteen positive clones were isolated from one million phage, and each was plaque purified. The cDNA insert of one of these phage was sequenced and shown to code for α_2-thiol proteinase inhibitor as identified by a partial amino acid sequence of the light chain of α_2-thiol proteinase inhibitor. This cDNA insert contained 1529 base pairs coding for the complete α_2-thiol proteinase inhibitor. It included 45 base pairs of 5′ noncoding sequence, 1281 base pairs that code for pre α_2-thiol proteinase inhibitor, a stop codon, 160 base pairs of 3′ noncoding sequence, and 40 base pairs of poly(A) tail. The noncoding sequence on the 3′ end contained a potential recognition site (AATAAA) for processing and polyadenylation of precursor messenger RNA. The amino acid sequence of α_2-thiol proteinase inhibitor deduced from the cDNA showed a striking similarity (overall homology at 74%) to that of bovine low molecular weight (LMW) kininogen, including two internally repeated sequences and a nonapeptide sequence of bradykinin. These data clearly indicated that α_2-thiol proteinase inhibitor and LMW kininogen are identical. This was further supported by immunological cross-reactivity between α_2-thiol proteinase inhibitor and LMW kininogen. When the amino acid sequence of α_2-thiol proteinase inhibitor was compared with those for several low molecular weight thiol proteinase inhibitors, including human and chicken cystatins, rat liver thiol proteinase inhibitor, human stefin, and rat epidermal thiol proteinase inhibitor, it was obvious that there are clearly homologous structures, including two potential reactive site sequences of Gln-Val-Val-Ala-Gly.

Thiol proteinase inhibitors with high and low molecular weights have been isolated from plasma and various tissues and extensively studied (Sasaki et al., 1977, 1983; Järvinen, 1979; Ryley, 1979; Hirado et al., 1981; Lenney et al., 1982; Wakamatsu et al., 1982; Katunuma et al., 1983). These inhibitors inactivate a number of different thiol proteinases, including cathepsins, ficin, papain, and calpain. Accordingly, their physiological function appears to involve the regulation of the thiol proteinases. We have previously characterized two high molecular weight thiol proteinase inhibitors from human plasma and designated them as α_1-thiol proteinase inhibitor and α_2-thiol proteinase inhibitor (Sasaki et al., 1981). The α_1-thiol proteinase inhibitor is a glycoprotein with a minimum molecular weight (M_r) of about 60 000, as determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis under reduced conditions, while the α_2-thiol proteinase inhibitor is a glycoprotein composed of a heavy chain (app M_r 66 000 or 65 000) and a light chain (M_r 4200). The α_2-thiol proteinase inhibitor with the heavy chain of M_r 66 000 has also been designated α_2-thiol proteinase inhibitor$_1$, while the inhibitor with the heavy chain of M_r 65 000 was called α_2-thiol proteinase inhibitor$_2$. α_1-Thiol proteinase inhibitor and α_2-thiol proteinase inhibitors are closely related proteins since they show immunological cross-reactivity.

Several of the low molecular weight thiol proteinase inhibitors from mammalian tissues and serum have also been shown to be highly homologous by amino acid sequence analysis (Takio et al., 1983, 1984; Turk et al., 1983; Brzin et al., 1983, 1984; Machleidt et al., 1983).

In the present studies, we describe the isolation and characterization of a cDNA coding for human α_2-thiol proteinase inhibitor. Furthermore, the amino acid sequence predicted from the cDNA indicates that human α_2-thiol proteinase inhibitor is identical with LMW kininogen.[1] Cloning and sequencing of cDNAs for bovine LMW kininogen and HMW kininogen have recently been reported by Nawa et al. (1983)

† From the Department of Biochemistry, Nagoya City University Medical School, Nagoya 467, Japan (I.O. and M.S.), the Department of Biochemistry, University of Washington, Seattle, Washington 98195 (K.K.), and the Division of Biochemistry, Institute of Immunological Science, Hokkaido University, Sapporo 060, Japan (T.T. and H.S.).
Received August 22, 1984. This investigation was supported in part by the Ministry of Education, Science and Culture of Japan (Research Grant 59370013 to M.S.), the Naito Foundation (I.O.), and the National Institutes of Health (Research Grant HL 31511 to K.K. and Research Grant HL 16919 to Earl W. Davie).

[1] Abbreviations: LMW kininogen, low molecular weight kininogen; HMW kininogen, high molecular weight kininogen; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid.

0006-2960/84/0423-5691$01.50/0