Self Base Pairing in a Complementary Deoxydinucleoside Monophosphate Duplex: Crystal and Molecular Structure of Deoxycytidylyl-(3′–5′)-deoxyguanosine†

W. B. T. Cruse,* E. Egert, O. Kennard, G. B. Sala, S. A. Salisbury, and M. A. Viswamitra

ABSTRACT: The molecular structure of ammonium deoxycytidylyl-(3′–5′)-deoxyguanosine, crystallized from aqueous acetone near pH 4, was determined for X-ray diffraction data. The crystals were tetragonal, space group $P4_32_12$ with $a = b = 11.078$ (1) Å and $c = 45.826$ (4) Å. The structure was solved by tangent expansion of phases based on a derived phosphorus position and refined to $R = 0.060$ by full matrix least squares. Molecules related by a 2-fold symmetry axis are connected by hydrogen bonds between the bases and form parallel right-handed duplexes. Pairs of cytosines share a proton at N(3) and are joined by three hydrogen bonds: N(4)-H···O(2), O(2)···H-N(4), and N(3)-H···N(3). Guanines are joined by two hydrogen bonds: N(2)-H···N(3) and N(3)···H-N(2). Base-stacking interactions within the duplex are weak with the cytosine and guanine ring planes inclined at 24° to each other in each monomer. Despite the unusual arrangement of the molecules, the sugar phosphate backbone has the g^-g^- conformation normally associated with right-handed double helical structures. Conformational parameters of the nucleosides are also typical with both sugars C(2′)-endo and glycosidic torsion angles 55° for cytidine and 94° for guanosine. The bonding geometry of the bases is influenced by hydrogen bonding and charge-transfer networks in the crystal lattice. The solvent molecules interact with the dimer in three fused circular hydrogen bonding domains with a single disordered ammonium cation per d(CpG) dimer. Parallels with the formation of self base pairs and their implications in molecular biology are discussed.

High-resolution X-ray diffraction analyses of small segments of DNA give precise descriptions of their conformational and hydrogen-bonding properties, which can in turn lead to a better understanding of the structural properties of larger oligonucleotides and localized regions of DNA itself. To date the dinucleoside phosphate deoxycytidylyl-(3′–5′)-deoxyguanosine, d(CpG), has only been crystallized in the presence of intercalating drugs such as proflavine (Shieh et al., 1980) and a terpyridyl–platinum complex (Wang et al., 1978). In both structures Watson–Crick type hydrogen bonding was observed between two antiparallel strands of d(CpG), and the conformational properties of the helical segment were considerably influenced by the presence of the intercalator.

Recently a number of studies of longer oligonucleotide structures containing cytosine and guanine have been published. All contain self-complementary helical segments. The tetramer d(CpGpCpG) (Crawford et al., 1980) crystallizes as left-handed Z-DNA, similar to the hexamer, d(CpGpCpGpCpG), which has a 2 base pair repeat unit (Wang et al., 1979). The high salt form of the tetramer d(CpGpCpG) also forms a left-handed double helix, the Z′ helix, which differs from the low-salt Z form in the sugar pucker of the deoxyguanosines (Drew et al., 1980). Watson–Crick type C·G base pairs were observed in the dodecamer d(CpGpCpGpApApTpTpCpGpCpG) which crystallizes as a right-handed B-DNA helix with slightly more than 10 base pairs per turn (Wing et al., 1980) and in oligonucleotides forming A-DNA helices (Shakked et al., 1981; Conner et al., 1982; Wang et al., 1982).

During the investigation of the crystallization properties of d(CpG), large crystals of the ammonium salt were obtained readily only near pH 4. We report the first X-ray structural analysis of this dinucleoside phosphate in the absence of an intercalator. In the structure the molecules are not linked by complementary Watson–Crick hydrogen bonds but form self-base-paired duplexes. The structure is of interest as a model of self-pairing of bases which, at the polynucleotide level, could lead to errors of recognitition during replication and transcription.

Experimental Procedures

d(CpG) was synthesized in the laboratory, using the diester method, and purified by DEAE-cellulose column chromatography (Weber & Khorana, 1972). Tetragonal plates were grown by layering acetone onto a 25 mM unbuffered aqueous solution (pH 4) of the ammonium salt until an acetone to water ratio of 2:1 was obtained.

The density of the crystals, measured in a bromoform/xylene gradient (1.48 g·cm^{-3}), suggested the presence of eight formula units of $NH_4^+ \cdot C_{19}H_{24}N_8O_{10}P^- \cdot 3H_2O$ in the unit cell. Crystal data are given in Table I.

A crystal of dimensions 0.25 × 0.25 × 0.15 mm was mounted in air on a Syntex $P2_1$ diffractometer. Cell dimensions were calculated from 15 automatically centered reflections.

Three-dimensional intensity data ($2\theta < 116°$, 2367 unique reflections) were collected at room temperature with graphite monochromatized Cu Kα radiation. The intensities were measured by the $\theta/2\theta$ method. No significant reduction was observed in the intensities of three monitor reflections during the 2 weeks of data collection. Strong high-order reflections were profile fitted. Intensities were corrected for Lorentz and polarization factors but not for absorption ($\mu = 14.2$ cm^{-1}); 1908 reflections with $F > 3\sigma$ were treated as observed.

The structure was solved by a combination of Patterson and direct methods. The position of the phosphorus atom was determined from a sharpened Patterson map. Phases derived from the phosphorus position were used as input to the tangent formula expansion (Karle, 1968). An E map, constructed with

† From the University Chemical Laboratory, Cambridge CB2 1EW, United Kingdom. *Received September 21, 1982.* This work was supported by a long-term grant of the Medical Research Council, including a Visiting Fellowship to M.A.V. E.E. was the recipient of a Postdoctoral Fellowship of the Deutsche Forschungsgemeinschaft.

0006-2960/83/0422-1833$01.50/0

Table I: Crystal Data

stoichiometry	$NH_4^+ \cdot C_{19}H_{24}N_8O_{10}P^- \cdot 3H_2O$
space group	$P4_32_12$
Z	8
$a = b$	11.078 (1) Å
c	45.826 (4) Å
d_c	1.48 g·cm^{-3}
d_m	1.48 (1) g·cm^{-3}
μ (Cu Kα)	14.2 cm^{-1}

Table II: Atom Coordinates (×10^4)

	x/a	y/b	z/c
N(1C)	7692 (5)	7994 (5)	7736 (1)
C(2C)	8629 (8)	8765 (7)	7732 (2)
O(2C)	9187 (5)	9063 (5)	7955 (1)
N(3C)	8988 (5)	9255 (5)	7467 (1)
C(4C)	8384 (8)	8996 (8)	7222 (1)
N(4C)	8764 (7)	9505 (7)	6979 (1)
C(5C)	7392 (7)	8206 (7)	7231 (1)
C(6C)	7083 (7)	7733 (7)	7485 (2)
C(1′C)	7217 (6)	7584 (6)	8016 (1)
C(2′C)	5998 (6)	8123 (6)	8089 (1)
C(3′C)	5487 (6)	7191 (6)	8293 (1)
C(4′C)	5989 (6)	6025 (6)	8171 (1)
O(1′C)	7040 (4)	6330 (4)	8000 (1)
C(5′C)	5096 (7)	5343 (7)	7978 (1)
O(5′C)	4590 (5)	6100 (5)	7763 (1)
O(3′C)	5981 (4)	7357 (4)	8582 (1)
P(1)	5292 (2)	8165 (2)	8816 (0.4)
O(1P)	4004 (5)	7832 (5)	8822 (1)
O(2P)	5992 (5)	8066 (4)	9089 (1)
O(5′G)	5328 (4)	9524 (4)	8696 (1)
C(5′G)	6443 (7)	10204 (6)	8717 (1)
C(4′G)	6320 (7)	11408 (6)	8567 (1)
O(1′G)	6166 (4)	11191 (4)	8256 (1)
C(3′G)	5244 (8)	12178 (6)	8663 (2)
O(3′G)	5652 (5)	13386 (5)	8652 (1)
C(2′G)	4313 (6)	11924 (7)	8426 (1)
C(1′G)	5091 (6)	11782 (6)	8159 (1)
N(1G)	5554 (6)	10778 (5)	7094 (1)
C(2G)	6131 (7)	11640 (6)	7258 (2)
N(2G)	6987 (7)	12268 (6)	7126 (1)
N(3G)	5900 (5)	11830 (5)	7537 (1)
C(4G)	5004 (6)	11110 (6)	7637 (1)
C(5G)	4362 (6)	10255 (6)	7485 (2)
C(6G)	4627 (7)	10072 (7)	7186 (2)
O(6G)	4169 (5)	9336 (5)	7016 (1)
N(7G)	3536 (6)	9702 (6)	7668 (2)
C(8G)	3727 (7)	10207 (7)	7920 (2)
N(9G)	4591 (5)	11090 (5)	7920 (1)
O(1W)	2725 (5)	9522 (5)	9153 (1)
O(4W)	3771 (6)	5247 (5)	8777 (1)
O(2W)	2219 (8)	8432 (8)	8373 (1)
N/O(W3)	2440 (7)	7362 (8)	7825 (2)

Table III: Bond Lengths (Å)

C(2C)–N(1C)	1.344 (11)	C(6C)–N(1C)	1.363 (10)
C(1′C)–N(1C)	1.462 (9)	O(2C)–C(2C)	1.239 (10)
N(3C)–C(2C)	1.390 (10)	C(4C)–N(3C)	1.339 (10)
N(4C)–C(4C)	1.316 (11)	C(5C)–C(4C)	1.405 (12)
C(6C)–C(5C)	1.324 (11)	C(2′C)–C(1′C)	1.512 (12)
O(1′C)–C(1′C)	1.406 (9)	C(3′C)–C(2′C)	1.503 (11)
C(4′C)–C(3′C)	1.513 (11)	O(3′C)–C(3′C)	1.445 (8)
O(1′C)–C(4′C)	1.443 (9)	C(5′C)–C(4′C)	1.527 (12)
O(5′C)–C(5′C)	1.407 (10)	P(1)–O(3′C)	1.592 (6)
O(1P)–P(1)	1.474 (6)	O(2P)–P(1)	1.475 (6)
O(5′G)–P(1)	1.602 (6)	C(5′G)–O(5′G)	1.450 (9)
C(4′G)–C(5′G)	1.507 (12)	O(1′G)–C(4′G)	1.454 (8)
C(3′G)–C(4′G)	1.532 (13)	C(1′G)–O(1′G)	1.431 (9)
O(3′G)–C(3′G)	1.414 (10)	C(2′G)–C(3′G)	1.526 (12)
C(1′G)–C(2′G)	1.506 (11)	N(9G)–C(1′G)	1.447 (9)
C(2G)–N(1G)	1.374 (10)	C(6G)–N(1G)	1.359 (11)
N(2G)–C(2G)	1.322 (11)	N(3G)–C(2G)	1.321 (9)
C(4G)–N(3G)	1.353 (10)	C(5G)–C(4G)	1.374 (11)
N(9G)–C(4G)	1.373 (9)	C(6G)–C(5G)	1.416 (12)
N(7G)–C(5G)	1.383 (11)	O(6G)–C(6G)	1.236 (11)
C(8G)–N(7G)	1.299 (11)	N(9G)–C(8G)	1.369 (10)

Table IV: Bond Angles (deg)

C(6C)–N(1C)–C(2C)	120.5 (7)	C(1′C)–N(1C)–C(2C)	119.0 (6)
C(1′C)–N(1C)–C(6C)	119.8 (7)	O(2C)–C(2C)–N(1C)	123.0 (8)
N(3C)–C(2C)–N(1C)	118.6 (7)	N(3C)–C(2C)–O(2C)	118.4 (8)
C(4C)–N(3C)–C(2C)	120.5 (7)	N(4C)–C(4C)–N(3C)	117.2 (8)
C(5C)–C(4C)–N(3C)	120.1 (7)	C(5C)–C(4C)–N(4C)	122.8 (8)
C(6C)–C(5C)–C(4C)	118.2 (8)	C(5C)–C(6C)–N(1C)	122.0 (8)
C(2′C)–C(1′C)–N(1C)	113.1 (6)	O(1′C)–C(1′C)–N(1C)	108.1 (6)
O(1′C)–C(1′C)–C(2′C)	106.1 (6)	C(3′C)–C(2′C)–C(1′C)	101.6 (6)
C(4′C)–C(3′C)–C(2′C)	102.6 (6)	O(3′C)–C(3′C)–C(2′C)	109.9 (6)
O(3′C)–C(3′C)–C(4′C)	108.0 (6)	O(1′C)–C(4′C)–C(3′C)	107.3 (6)
C(5′C)–C(4′C)–C(3′C)	113.5 (7)	C(5′C)–C(4′C)–O(1′C)	108.9 (6)
C(4′C)–O(1′C)–C(1′C)	108.3 (6)	O(5′C)–C(5′C)–C(4′C)	111.6 (7)
P(1)–O(3′C)–C(3′C)	120.6 (5)	O(1P)–P(1)–O(3′C)	109.7 (4)
O(2P)–P(1)–O(3′C)	106.1 (4)	O(2P)–P(1)–O(1P)	118.2 (4)
O(5′G)–P(1)–O(3′C)	106.6 (3)	O(5′G)–P(1)–O(1P)	105.5 (4)
O(5′G)–P(1)–O(2P)	110.3 (4)	C(5′G)–O(5′G)–P(1)	119.1 (5)
C(4′G)–C(5′G)–O(5′G)	110.7 (7)	O(1′G)–C(4′G)–C(5′G)	108.2 (6)
C(3′G)–C(4′G)–C(5′G)	115.5 (7)	C(3′G)–C(4′G)–O(1′G)	106.5 (6)
C(1′G)–O(1′G)–C(4′G)	109.1 (6)	O(3′G)–C(3′G)–C(4′G)	105.5 (7)
C(2′G)–C(3′G)–C(4′G)	102.6 (7)	C(2′G)–C(3′G)–O(3′G)	111.4 (7)
C(1′G)–C(2′G)–C(3′G)	102.3 (7)	C(2′G)–C(1′G)–O(1′G)	105.7 (6)
N(9G)–C(1′G)–O(1′G)	108.2 (6)	N(9G)–C(1′G)–C(2′G)	116.9 (7)
C(6G)–N(1G)–C(2G)	125.5 (7)	N(2G)–C(2G)–N(1G)	116.7 (7)
N(3G)–C(2G)–N(1G)	123.5 (8)	N(3G)–C(2G)–N(2G)	119.8 (7)
C(4G)–N(3G)–C(2G)	112.1 (7)	C(5G)–C(4G)–N(3G)	127.9 (7)
N(9G)–C(4G)–N(3G)	125.0 (7)	N(9G)–C(4G)–C(5G)	107.1 (7)
C(6G)–C(5G)–C(4G)	118.8 (8)	N(7G)–C(5G)–C(4G)	110.0 (7)
N(7G)–C(5G)–C(6G)	131.2 (7)	C(5G)–C(6G)–N(1G)	112.1 (8)
O(6G)–C(6G)–N(1G)	119.5 (8)	O(6G)–C(6G)–C(5G)	128.3 (8)
C(8G)–N(7G)–C(5G)	103.8 (7)	N(9G)–C(8G)–N(7G)	114.9 (8)
C(4G)–N(9G)–C(1′G)	125.3 (6)	C(8G)–N(9G)–C(1′G)	130.3 (6)
C(8G)–N(9G)–C(4G)	104.2 (7)		

the expanded phases, yielded the location of 18 atoms. The positions of the remaining non-hydrogen atoms, including those of four solvent molecules, were determined from the difference electron density maps. The resultant structure was refined isotropically to $R = 0.13$, at which stage hydrogen atoms, covalently bonded to the carbon (clearly visible in difference maps), or ring nitrogen atoms, were included at calculated positions. Further refinement with anisotropic thermal parameters for the non-hydrogen atoms and isotropic parameters for the hydrogen atoms, riding at a covalent bond distance of 0.99 Å, converged at $R = 0.08$. A difference map at this stage yielded three peaks which could be interpreted as hydrogen atoms; two were bonded to the water oxygen O(W1) and the third to N(2) of guanine. With this information it was possible to construct an acceptable hydrogen-bonding scheme and assign the remaining hydrogen atoms, assuming covalent bond distances of 0.8 Å and linear hydrogen bonds. Only the "ammonium" nitrogen atom N/O(W3) proved difficult to fit into the scheme as it had four possible hydrogen bond acceptor contacts <3.0 Å and a fifth short contact (2.8 Å) with a symmetry related N/O(W3) atom. It seemed likely that the ammonium ion was part of a disordered $NH_3 \cdots H^+ \cdots H_2O$ system. It was given the form factor for N during the remaining calculations. The refinements converged at $R_w = 0.061$ ($R = 0.060$). In the final cycle the quantity $\sum W\Delta^2$ was minimized with $W = 1/[\sigma^2(F) + 0.0045F^2]$. The final difference map contained four peaks in the range of ±0.3 e·Å^{-3} in the vicinity of the solvent molecules O(W2) and N/O(W3). All computations were performed with the program SHELX76 (Sheldrick, 1976).

Results and Discussion

The coordinates of the non-hydrogen atoms are given in Table II. Bond lengths, bond angles, and important non-

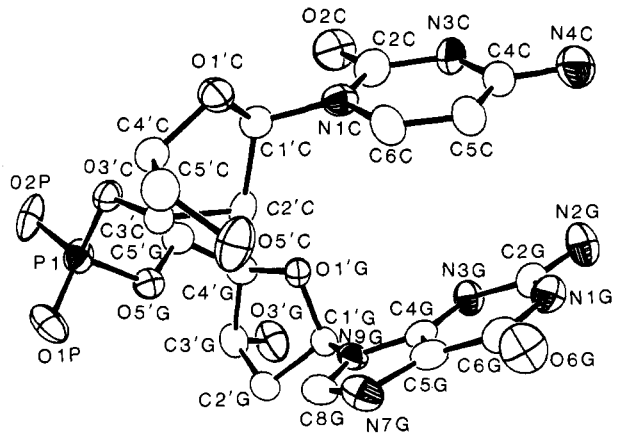

FIGURE 1: Atom numbering scheme for dCpG.

Table V: Selected Nonbonded Distances (Å)[a]

N(3C)-O(2C)[b]	3.365	N(4C)-O(2C)[b]	2.830
N(3C)-N(3C)[b]	2.730	O(2P)-N(4C)[c]	3.168
N/O(W3)-O(5'C)	2.825	O(5'C)-O(5'C)[d]	2.663
O(W1)-O(1P)	2.797	O(W4)-O(1P)	2.884
O(W2)-O(1P)	2.932	N(1G)-O(2P)[e]	2.787
N(2G)-O(2P)[e]	2.845	O(W4)-O(3'G)[f]	2.981
N(3G)-N(2G)[b]	2.868	N/O(W3)-O(6G)[d]	2.720
O(W1)-O(6G)[c]	2.839	N/O(W3)-N(7G)	2.897
N/O(W3)-N(7G)[d]	3.351	O(W4)-O(W1)[g]	2.702
O(W2)-O(W4)[h]	2.795	N/O(W3)-O(W2)	2.834
N/O(W3)-N/O(W3)[d]	2.824		

[a] Key to symmetry operations relating designated atoms to reference atoms at (x, y, z): (b) $2-y, 2-x, 1.5-z$; (c) $-0.5+y, 1.5-x, 0.25+z$; (d) $1-y, 2-x, 1.5-z$; (e) $1-y, 1-x, 1.5-z$; (f) $1.5-y, 0.5+x, -0.25+z$; (g) $x, -1+y, 1.75-z$; (h) $0.5-x, -0.5+y, 1.75-z$.

bonded contacts are given in Tables III, IV, and V. Tables of thermal parameters and hydrogen atom parameters are available (see paragraph at end of paper regarding supplementary material).

The asymmetric unit of the crystal structure consists of one molecule of d(CpG), which is illustrated in Figure 1 together with the numbering scheme used. The molecule is folded about the O(3'C)–P and P–O(5'G) bonds, resulting in a substantial superposition of the base and sugar residues of cytidine with the guanine base.

Base Pairing. Two molecules of d(CpG) are related by the crystallographic 2-fold symmetry axis to form a parallel self-base-paired right-handed helical segment (Figure 2) which is in contrast to other structures containing C and G referred to in the introduction.

In the present structure half of the cytosine bases are protonated at N(3) as shown in Figure 3 (top), and pairs are linked by three relatively short hydrogen bonds: two between O(2C) and N(4C) of length 2.83 (1) Å and the third between the ring atoms N(3C) which are 2.73(1) Å apart. A peak of 0.22 e.Å^{-3} was observed on the final difference map midway between these two atoms. N(4C) is also linked by a relatively long hydrogen bond (3.17 Å) to O(2P).

Protonated C·C base pairs with triple hydrogen bonds were reported for cytosine-5-acetic acid (Marsh et al., 1962) and in the 1:1 complex of proflavine and the ribodinucleoside phosphate CpA (Westhof & Sundaralingam, 1980).

The two guanine bases are linked by hydrogen bonds N(2G)-H···N(3G)* (asterisks denote symmetry-related atoms) of length 2.87 Å illustrated in Figure 3 (bottom). This figure also shows the atoms hydrogen bonded to N(1), N(2), and O(6) which would normally form Watson–Crick hydrogen bonds with the O(2), N(3), and N(4) atoms of cytosine.

The G·G hydrogen-bonding scheme has been observed in other structures but only as a secondary feature in the presence of conventional G·C base pairs. In the 1:1 complex of 9-ethylguanine and 1-methylcytosine (O'Brien, 1967), for example, the asymmetric unit of the structure is a Watson–Crick base pair, but adjacent guanines are linked by a long (3.01 Å) hydrogen bond between N(2) and N(3)*, across a center of symmetry. Similar hydrogen bonding was observed as part of the overlap contact between adjacent guanine bases of the B-DNA dodecamer d(CpGpCpGpApApTpTpCpGpCpG) (Dickerson & Drew, 1981).

Figure 2 is a stereoview of the parallel right-handed helical duplex where the two molecules are related by a 2-fold rota-

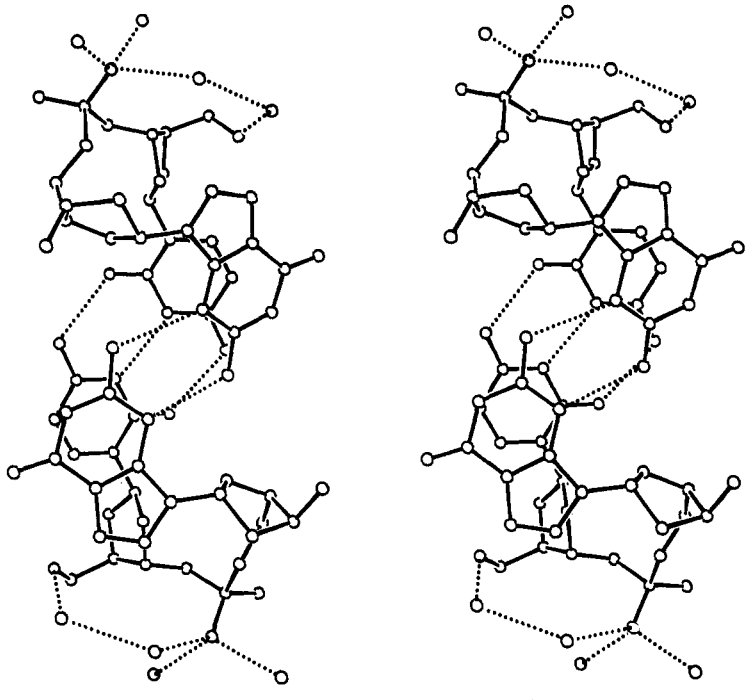

FIGURE 2: Stereoscopic view of the dCpG dimer down the crystallographic 2-fold axis.

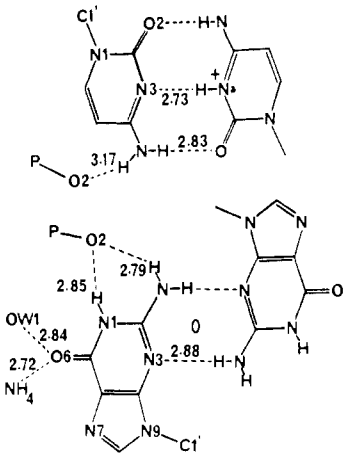

FIGURE 3: (Top) Hydrogen bonding pattern for the C·CH$^+$ pair. (Bottom) Hydrogen bonding pattern for the G·G pair.

Table VI: Comparison of Cytosine Base Geometries

	neutral[a]	dCpG[b]	CYACET[c]	proton-ated[a]
	Bond Lengths (Å)			
N(1)–C(2)	1.399	1.344	1.351	1.381
C(2)–N(3)	1.356	1.390	1.365	1.387
N(3)–C(4)	1.334	1.339	1.354	1.352
C(4)–C(5)	1.426	1.405	1.426	1.413
C(5)–C(6)	1.337	1.324	1.364	1.341
C(6)–N(1)	1.364	1.363	1.365	1.362
C(2)–O(2)	1.237	1.239	1.250	1.211
C(4)–N(4)	1.337	1.316	1.323	1.313
	Bond Angles (deg)			
C(6)–N(1)–C(2)	120.6	120.5	122.0	121.5
N(1)–C(2)–C(3)	118.9	118.6	117.7	114.9
C(2)–N(3)–C(4)	120.0	120.5	121.9	125.1
N(3)–C(4)–C(5)	121.8	120.1	120.7	117.5
C(4)–C(5)–C(6)	117.6	118.2	115.6	118.5
C(5)–C(6)–N(1)	121.0	122.0	122.1	122.5
N(1)–C(2)–O(2)	119.2	123.0	121.2	123.5
N(3)–C(2)–O(2)	121.9	118.4	121.1	121.6
N(3)–C(4)–N(4)	117.9	117.2	117.3	119.5
C(5)–C(4)–N(4)	120.3	122.8	122.0	123.0

[a] Mean values from Taylor & Kennard (1982a). [b] This work. [c] Cytosine-5-acetic acid (Marsh et al., 1962).

tion. The diameter of the helix, i.e., the C(1′)–C(1′) distance, is 9.92 Å at the cytidines and 7.77 Å at the guanosines. Both are shorter than the 10.8 Å found between C and G in a B-DNA helix. The helical turn per residue, i.e., the angle between the vectors C(1′C)–N(1C) and C(1′G)–N(9G), is 66°. The propellor twist angles are 22.5° for C·C* base pair and 28.6° for the G·G* base pair.

Molecular Geometry. An interesting feature of the cytosine geometry in the present structure is the similarity of the bonding pattern to that of a conjugated network of isolated double bonds rather than the delocalized aromatic systems found in other cytosine derivatives. This is illustrated in Table VI where the geometries of d(CpG) and cytosine-5-acetic acid, both hemiprotonated at N(3) are compared with mean values for neutral and N(3) protonated structures (Taylor & Kennard, 1982a).

One might expect that with half the d(CpG) molecules protonated at N(3C) and with a statistical and probably dynamic disorder between neutral and charged species in the C·CH$^+$ pair, the observed geometry would be intermediate between the two averaged values. This, however, is not the case, and moreover, the observed bond distances for N(1C)–C(2C) and C(5C)–C(6C) are below the minimum reported values for 17 protonated and 14 neutral residues (in Taylor & Kennard, 1982b). Although abnormal bond distances in crystal structures can be rationalized in terms of neglect of systematic errors (i.e., due to absorption, thermal liberation, improper weighting schemes during refinement, or an incorrect model), the chemical resonableness of the observed geometry in all parts of this structure suggests that the bonding pattern is highly significant. Comparisons with averages neglect the relative contributions from available valence bond forms which vary from structure to structure.

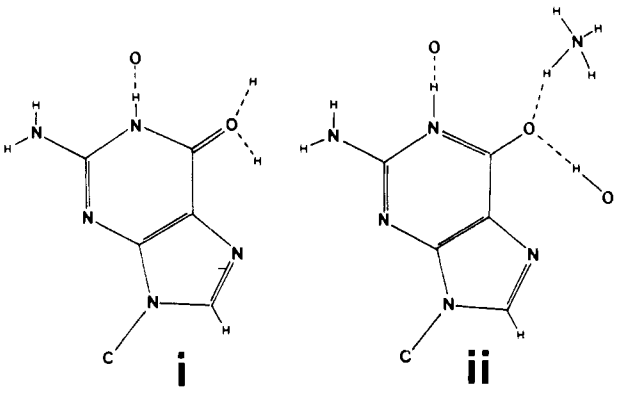

FIGURE 4: Prominent valence bond forms for guanine in dCpG. The N(1)–C(6) distance is shortened from values of 1.390, 1.390, and 1.391 Å (guanine monohydrate and guanosine dihydrate, two molecules) to 1.359 (11) Å in dCpG.

Previous studies of nucleotide structures (Voet & Rich, 1970; Viswamitra et al., 1971) have indicated that a charge on N(3) in cytosine affects the bond angles at C(2), N(3), and C(4). In d(CpG), the angle C(2)–N(3)–C(4) [120.5 (7)°] is closer to the value for neutral cytosine whereas the angle N(1)–C(2)–O(2) [123.0 (8)°] is closer to the value in the protonated base. The C(2C)–N(3C) distance of 1.390 (10) Å is longer than the value of 1.356 Å for the neutral base (Table VI) and is chemically consistent with bond lengthening due to repulsion between the partial positive charges on the carbonyl carbon atom, C(2) and N(3).

In d(CpG), if the guanine ring is uncharged, we would expect its bonding geometry to agree with values determined for structures containing neutral guanine residues, such as guanine monohydrate (Thewalt et al., 1971) and guanosine dihydrate (Thewalt et al., 1970). This is indeed the case with the exception of the N(1G)–C(6G) bond which is shortened due to the additional contribution of valence structure (II) (Figure 4) which is stabilized by hydrogen bonding in the crystal structure. A hydrogen atom is donated from N(1G) to the phosphate oxygen atom O(2P), and two H bonds are accepted by O(6G), one from the disordered ammonium ion N/O(W3) and the other from the water molecule O(W1). The chain H_3N^+-H···O(6G)–C(6G)–N(1G)-H···O$^-$(2P) forms a charge-transfer network, neutralizing the partial negative charge of O(2P).

Table VII contains an analysis of the planarity of the base and sugar residues. Atoms of the cytosine base are coplanar, except for O(2C). The anomeric carbon atom C(1′C) is displaced by 0.17 Å from the mean plane of the base, resulting in a partially pyramidal geometry around N(1C). A similar deviation is observed for the anomeric carbon atom C(1′G) (0.12 Å). Additionally, in the guanine ring, N(1G) and C(8G) are significantly displaced from the mean ring plane in the same direction.

Figure 5 illustrates the relation between the pyrimidine and purine bases. The inclination between the mean planes is 24°. The stacking interactions are weak with mean plane to ring

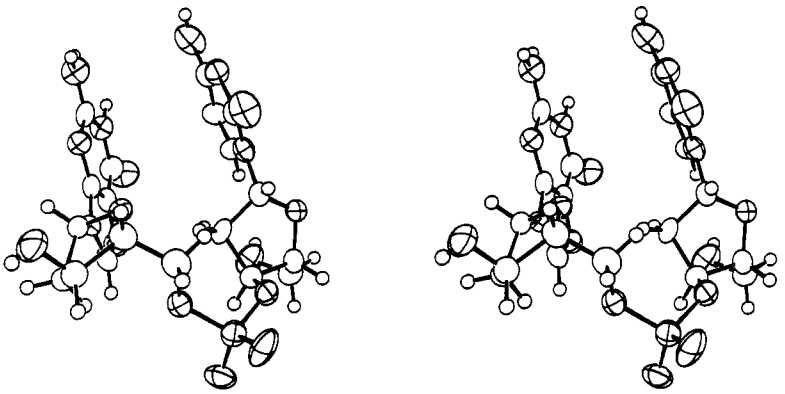

FIGURE 5: Stereoscopic view of the d(CpG) monomer showing the relative inclination of the two bases.

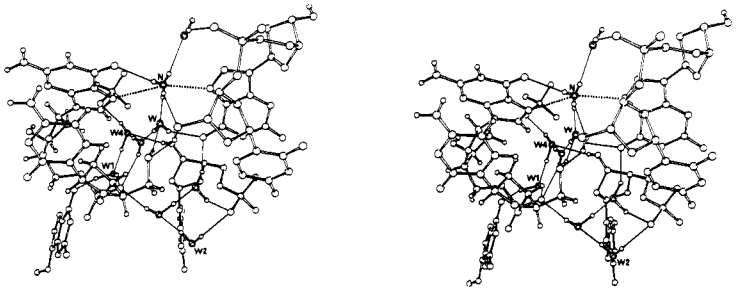

FIGURE 6: Stereoscopic view of the solvent–dCpG H-bonding interactions. The close contacts between N/O(W3) and N(7G) are indicated by dotted lines.

atom distances ranging from a minimum of 3.30 Å to a maximum of 5.06 Å.

Both deoxyribose residues have the C(2′)-endo conformation, and in each the C(2′) atom is out-of-plane by 0.52 Å. The glycosidic angles (χ)[1] are 54.6° for cytidine and 94.3° for guanosine. The latter is somewhat larger than that generally observed in nucleosides.

The phosphate shows no unusual features. The two P–O ester bonds are comparatively long, while the two other P–O bonds are short and precisely equal, suggesting that the negative charge is fully delocalized. The backbone has the g^-g^- conformation typical of right-handed helical structures with Watson–Crick base pairing although in d(CpG), the right-handed helical fragment consists of two parallel strands linked by non-Watson–Crick hydrogen bonds. The values of the torsion angles are α = 205.5°, β = 174.5°, γ = 53.2°, δ = 144.8°, ϵ = −156.0°, and ζ = −66.4°.

Solvent Network. The stereo diagram of Figure 6 shows the unique contacts less than 3.4 Å between the d(CpG) molecule and the solvent network. The latter consists of isolated eight-membered chains linked by hydrogen bonds. The crystallographic 2-fold symmetry axis passes through the midpoint. Each chain must contain one positive charge to preserve electrical neutrality in the crystal lattice.

Two solvent molecules O(W2) and N/O(W3) are, as discussed earlier, disordered. It is likely that the positive charge is shared between the two solvent molecules on either side of the symmetry axis, forming a disordered H_3N^+-H···OH_2 system. The alternative interpretations of a disordered H_2O^+-H···OH_2 or H_3N^+-H···NH_3 pair are less likely on chemical grounds.

The NH_4^+ species is surrounded by six potential H-bond acceptors, in an approximately octahedral arrangement. As a result, a variety of alternative linear and bifurcated hydrogen

Table VII: Description of Mean Planes for Base and Sugar Residues[a]

atom	dev from mean plane (Å)	atom	dev from plane (Å)
plane I		plane II	
N(1)C*	−0.007	N(1)G*	0.038
C(2)C*	0.012	C(2)G*	−0.003
O(2)C	0.052	N(2)G	0.012
N(3)C*	−0.009	N(3)G*	−0.009
C(4)C*	0.000	C(4)G*	−0.018
N(4)C	−0.002	C(5)G*	−0.018
C(5)C*	0.005	C(6)G*	−0.015
C(6)C*	−0.002	O(6)G	0.013
C(1′)C	0.173	N(7)G*	−0.010
		C(8)G*	0.030
		N(9)G*	0.005
		C(1′)G	0.120
plane III		plane IV	
C(1′)C*	0.005	C(1′)G*	0.002
C(2′)C	−0.563	C(2′)G	−0.561
C(3′)C*	−0.004	C(3′)G*	−0.002
C(4′)C*	0.007	C(4′)G*	0.004
C(5′)C	−1.188	C(5′)G	−1.166
O(1′)C*	−0.007	O(1′)G*	−0.004
O(3′)C	1.364	O(3′)G	1.353
O(5′)C	2.403	O(5′)G	−2.418
N(1)C	−0.795	N(9)G	−0.706

Equations of Planes

plane I: $-0.6142x_0 + 0.7725y_0 + 0.1610z_0 = 7.3234$
plane II: $0.6795x_0 - 0.6913y_0 + 0.2458z_0 = 3.8782$
plane III: $0.5691x_0 - 0.1163y_0 + 0.8140z_0 = 33.4711$
plane IV: $0.5244x_0 + 0.8283y_0 - 0.1975z_0 = 6.3802$

[a] Atoms with asterisks were used in the calculation of the mean plane.

[1] In the present paper the torsion angles are defined by using the nomenclature recommended by the IUB–IUPAC JCBN (Davies & Dixon, 1980). According to this α = O(3′)–P–O(5′)–C(5′), β = P–O(5′)–C(4′), γ = O(5′)–C(5′)–C(4′)–C(3′), δ = C(5′)–C(4′)–C(3′)–O(3′), ϵ = C(4′)–C(3′)–O(3′)–P, ζ = C(3′)–O(3′)–P–O(5′), χ = O(4′)–C(1′)–N(9)–C(4) for G and O(4′)–C(1′)–N(1)–C(2) for C. The concordance with the notation by Sundaralingam (1969) is $\alpha = \omega$, $\beta = \phi$, $\gamma = \psi$, $\delta = \psi'$, $\epsilon = \phi'$, and $\zeta = \omega'$.

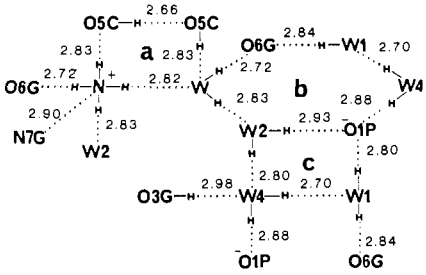

FIGURE 7: Schematic diagram of the three unique circular H-bonding domains. Ring a is bisected by a 2-fold axis. Donor–acceptor distances (in angstroms) are indicated.

bonds may be formed with the NH_4^+ ion free to tumble.

Figure 7 is a schematic illustration of the extended, fused network of circular hydrogen bonds which confers high stability on the structure. The networks fall into three domains a, b, and c containing 4, 6, and 4 H bonds, respectively. Ring a is bisected by the crystallographic 2-fold axis, and thus both the H-bonding scheme as depicted and the related one with N and W interchanged occur equally in the extended crystal structure. This duality, however, does not affect the H-bond pattern in domains b and c. Only one of the phosphate oxygen atoms, O(1P), makes H-bonded contacts with the solvent molecules. The second phosphate oxygen atom O(2P) is H bonded to the N atoms of the bases (Figure 3).

Summary and Conclusion

This paper reports the first X-ray analysis of the deoxydinucleoside phosphate d(CpG) which could only be satisfactorily crystallized under conditions critically related to the protonation of cytosine. The structure contains a self-base-paired duplex in which the unconventional hydrogen bonding between the bases is probably the most important stabilizing factor. It illustrates the potential for self-pairing in a nucleotide dimer in the absence of stabilization by intercalating drugs and so in this respect contrasts with the crystal structure of the CpA–proflavine complex (Westhof & Sundaralingam, 1980).

Evidence for self-pairing, particularly of cytosine, is not confined to the solid state. In solution, CD and ORD spectra of d(CpG) and d(GpC) indicate that acidification results in substantial changes in base stacking (Warshaw & Tinoco, 1966), which may arise from interactions similar to those observed in the present structure. At the polymer level too, duplex formation is well described; it is a characteristic property of hemiprotonated poly(C) and poly(dC) (Ts'o, 1974) and extends to alternating copolymers (Gray et al., 1981). Recently it has been shown also that poly(dG) can associate into a duplex structure (Dugaiczyk et al., 1981).

Geometrical dissimilarity with Watson–Crick base pairs is a general disqualification which excludes both C·CH^+ and G·G as observed in this structure from consideration as mechanisms by which mismatches can occur in molecular biology (Drake & Baltz, 1976; Topal & Fresco, 1976a,b). Preliminary model buiilding suggests, however, that despite its considerable differences C·CH^+ can be accommodated within a somewhat distorted, right-handed double helix (data not shown). Whatever the basis by which they are discriminated against when genetic information is replicated, transcribed, or translated, misincorporations attributable particularly to C·CH^+ are apparently avoided most successfully (Yarus, 1979; Topal & Fresco, 1976a). Self-pairing of cytosine is nevertheless a widely observed property and, when the small but significant extent to which the base is protonated at physiological pH (cytidine $pK_a \sim 4.4$) is taken into account, at least potentially a source of mutations and errors in protein synthesis.

Supplementary Material Available

Hydrogen atom coordinates and a list of anisotropic thermal parameters for the non-hydrogen atoms (2 pages). Ordering information is given on any current masthead page.

Registry No. Ammonium deoxycytidylyl-(3′–5′)-deoxyguanosine trihydrate, 84624-84-0.

References

Conner, B. N., Takano, T., Tanaka, S., Itakura, K., & Dickerson, R. E. (1982) *Nature (London) 295*, 294–299.

Crawford, J. L., Kolpak, F. J., Wang, A. H. J., Quigley, G. J., van Boom, J. H., van der Marel, G., & Rich, A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4016–4020.

Davies, D. B., & Dixon, H. B. F. (1980) *J. Chem. Biochem. Nomencl. 14* (2), 1980.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Drake, J. W., & Baltz, R. H. (1976) *Annu. Rev. Biochem. 45*, 11–38.

Drew, H., Takano, T., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 286*, 567–573.

Dugaiczyk, A., Robberson, D. L., & Ullrich, A. (1980) *Biochemistry 19*, 5869–5873.

Gray, D. M., Vaughan, M., Ratliff, R. L., & Hayes, F. N. (1980) *Nucleic Acids Res. 8*, 3695–3707.

Karle, J. (1968) *Acta Crystallogr., Sect. B B24*, 182–186.

Marsh, R. E., Bierstedt, R., & Eichhorn, E. L. (1962) *Acta Crystallogr. 15*, 310–316.

O'Brien, E. J. (1967) *Acta Crystallogr. 23*, 92–106.

Shakked, Z., Rabinovich, D., Cruse, W. B. T., Egert, E., Kennard, O., Sala, G., Salisbury, S. A., & Viswamitra, M. A. (1981) *Proc. R. Soc. London, Ser. B B213*, 478–487.

Sheldrick, G. M. (1976) SHELX76. *Program for Crystal Structure Determination*, University of Cambridge, Cambridge, England.

Shieh, H. S., Berman, H. M., Dabrow, M., & Neidle, S. (1980) *Nucleic Acids Res. 8*, 85–97.

Sundaralingham, M. (1969) *Biopolymers 7*, 821–860.

Taylor, R., & Kennard, O. (1982a) *J. Am. Chem. Soc. 104*, 3209–3212.

Taylor, R., & Kennard, O. (1982b) *J. Mol. Struct. 78*, 1–28.

Thewalt, U., Bugg, C. E., & Marsh, R. E. (1970) *Acta Crystallogr., Sect. B B26*, 1089–1101.

Thewalt, U., Bugg, C. E., & Marsh, R. E. (1971) *Acta Crystallogr., Sect. B B27*, 2358–2363.

Topal, M. D., & Fresco, J. R. (1976a) *Nature (London) 263*, 285–289.

Topal, M. D., & Fresco, J. R. (1976b) *Nature (London) 263*, 289–293.

Ts'o, P. O. P. (1974) in *Basic Principles of Nucleic Acid Research* (Ts'o, P. O. P., Ed.) Vol. 2, pp 306–465, Academic Press, New York.

Viswamitra, M. A., Reddy, B. S., Lin, G. H.-Y., & Sundaralingam, M. (1971) *J. Am. Chem. Soc. 93*, 4565–4573.

Voet, D., & Rich, A. (1970) *Prog. Nucleic Acid Res. Mol. Biol. 10*, 183–265.

Wang, A. H. J., Nathans, J., van der Marel, G., van Boom, J. H., & Rich, A. (1978) *Nature (London) 276*, 471–474.

Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G., & Rich, A. (1979) *Nature (London) 287*, 680–685.

Wang, A. H. J., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968–3972.

Warshaw, M. M., & Tinoco, I. (1965) *J. Mol. Biol. 13*, 54–64.
Weber, H., & Khorana, H. G. (1972) *J. Mol. Biol. 72*, 219–249.
Westhof, E., & Sundaralingam, M. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1852–1856.
Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755–758.
Yarus, M. (1979) *Prog. Nucleic Acid Res. Mol. Biol. 23*, 195–225.

Purification of a HeLa Cell High Molecular Weight Actin Binding Protein and Its Identification in HeLa Cell Plasma Membrane Ghosts and Intact HeLa Cells†

Robert R. Weihing

ABSTRACT: The high molecular weight protein (HMWP) which was previously observed to be a major component of the actin based gels formed by incubating cytoplasmic extracts of HeLa cells at 25 °C [Weihing, R. R. (1977) *J. Cell Biol. 75*, 95–103] has now been purified by gel filtration of 0.6 M KCl extracts of precipitated gels. A few hundred micrograms of HMWP, which is about 90% pure, can be isolated from 4 × 10^9 cells. HMWP can gel muscle actin and cross-link it into filament bundles. Its subunit molecular weight is 250 000, its Stokes radius is 125 Å, and its sedimentation coefficient is 9 S. A native molecular weight of 480 000 was calculated by using the latter two parameters, and therefore the native molecule is a dimer. Its amino acid analysis is nearly indistinguishable from that of macrophage actin binding protein and of mammalian and avian filamins. All of these findings indicate that HMWP is homologous to the latter proteins. However, HeLa cell HMWP and avian filamin must differ in their primary sequences because their partial peptide maps are distinct and because an antiserum against HMWP reacts only weakly with filamin. For studies on the intracellular location of HMWP, a goat antiserum against purified HMWP was prepared and characterized and then used to localize HMWP in suspension grown cells. The technique of immunoblotting revealed that the antiserum reacted virtually exclusively with the high molecular weight polypeptide that comigrates with HMWP in cell lysates and in $ZnCl_2$-stabilized plasma membrane ghosts prepared from HeLa cells [Gruenstein, E., Rich, A., & Weihing, R. R. (1975) *J. Cell Biol. 64*, 223–234] and that it did not react with rabbit myosin heavy chain, microtubule proteins (MAPS and tubulin) from HeLa cells and calf brain, or the proteins of human erythrocyte ghosts including spectrin. Suspension-grown cells which were stained with the antiserum by the technique of indirect immunofluorescence showed bright fluorescence at the rim of the cells and less intense generalized fluorescence. If preimmune serum or immune serum treated with HMWP was substituted for the immune serum, then staining at the rim was not observed, but the generalized fluorescence was only slightly reduced; unpermeabilized cells were not stained. These results indicate that HMWP is a component of the cortical cytoplasm of HeLa cells. Possible functions of cortical HMWP are discussed briefly.

Eukaryotic cells contain a variety of actin binding proteins that are believed to influence the organization and function of the actin-based microfilament system [reviewed in Schliwa (1981) and Weeds (1982)]. This laboratory has investigated an actin binding protein of HeLa cells through studies of the actin-based gelation of cytoplasmic extracts of HeLa cells. These studies produced three lines of evidence suggesting that gelation could be explained, at least in part, by cross-linking of actin into a three-dimensional network by a protein designated HMWP.[1] First, electrophoretic analysis showed that actin and HMWP are prominent polypeptide components of isolated gels (Weihing, 1976a,b, 1977). Second, dilution of cell extract (Weihing, 1977) or treatment of cell extract with micromolar concentrations of cytochalasin B (Weihing, 1976b) inhibits gelation and concomitantly prevents the increase in sedimentability of actin and HMWP which always accompanies gelation. Finally, heavy meromyosin inhibits gelation and concomitantly prevents the increase in sedimentability of HMWP (Weihing, 1977), possibly by competing with HMWP for sites on actin. I now report the purification and partial characterization of HMWP from HeLa cells, its identification in plasma membrane ghosts, and its localization in situ in cells grown in suspension.

Experimental Procedures

General Methods and Reagents. Growth and harvesting of cells, protein analysis, testing of gelation, gel electrophoresis, and preparation of myosin were done as described previously (Weihing, 1977). Muscle actin was purified from an acetone

† From the Cell Biology Group, Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545. *Received June 3, 1982; revised manuscript received November 10, 1982.* This work was supported by U.S. Public Health Service Research Grants CA-15229 (to R.R.W.), Specialized Cancer Center Grant CA-P30-12708 (to Thoru Pederson) from the National Cancer Institute, and a Cancer Research Scholarship from the American Cancer Society (Massachusetts Division) (to R.R.W.). Portions of the work have been presented at the meeting on The Cytoskeleton: Membranes and Movement at Cold Spring Harbor, NY, May 1979, and at the 66th annual meeting of the Federation of American Societies for Experimental Biology at New Orleans, LA, April 1982.

[1] Abbreviations: HMWP, high molecular weight protein; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid; NaDodSO$_4$, sodium dodecyl sulfate; PMSF, phenylmethanesulfonyl fluoride; TLCK, tosyllysine chloromethyl ketone; DTE, dithioerythritol; DE-52, microgranular diethylaminoethylcellulose, manufactured by Whatman; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/83/0422-1839$01.50/0