REPORTER MOLECULES AS PROBES OF DNA CONFORMATION: STRUCTURE OF A CRYSTALLINE COMPLEX CONTAINING 2-METHYL-4-NITRO-ANILINE ETHYLENE DIMETHYLAMMONIUM HYDROBROMIDE --5-IODOCYTIDYLYL(3'-5')GUANOSINE

Nand K. Vyas\*, Meenakshi N. Vyas\*, Shri C. Jain and Henry M. Sobell

Department of Radiation Biology and Biophysics and the Cancer Center, The University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

Department of Chemistry,
The University of Rochester, River Campus Station,
Rochester, New York 14627

Received April 13, 1984

2-Methyl-4-nitroaniline ethylene dimethylammonium hydrobromide forms a crystalline complex with the self-complementary dinucleoside monophosphate, 5-iodocytidylyl(3'-5')guanosine. The crystals are tetragonal, with  $\underline{a} = \underline{b} = 32.192$  Å and  $\underline{c} = 23.964$  Å, space group P4<sub>3</sub>2<sub>1</sub>2. The structure has been solved to atomic resolution by Patterson and Fourier methods, and refined by full matrix least squares. 5-Iodocytidylyl(3'-5')guanosine molecules are held together in pairs through Watson-Crick base-pairing, forming an antiparallel duplex structure. Nitroaniline molecules stack above and below guanine-cytosine pairs in this duplex structure. In addition, a third nitroaniline molecule stacks on one of the other two nitroaniline molecules. The asymmetric unit contains two 5-iodocytidylyl(3'-5')guanosine molecules, three nitroaniline molecules, one bromide ion and thirty-one water molecules, a total of 160 atoms. Details of the structure are described.

In his series of classic studies, Gabbay (1-4) synthesized a large number of organic compounds that contain different aromatic ring systems linked to a variety of side chains to investigate their ability to bind to DNA. These molecules -- called "reporter" molecules -- were synthesized to study dynamic aspects of DNA structure and, in particular, to investigate the relationship between DNA breathing phenomena and drug intercalation. For this purpose, Gabbay synthesized three different classes of intercalators. The first -- called "partial" intercalators -- bind between partially unstacked base-pairs in DNA. Examples include several nitroaniline derivatives, aromatic polypeptides and steroidal diamines. The second -- correspond to "classic" intercalators -- molecules that, like ethidium and acridine orange, intercalate fully into DNA without necessitating the transient disruption of hydrogen bonds connecting base-pairs. Examples include a series of phenanthro-

<sup>\*</sup>Current address: Department of Biochemistry, Rice University, Houston, Texas 77251

Figure 1. Chemical structure of the 2-methyl-4-nitroaniline ethylene dimethylammonium ion.

line and naphthylimide derivatives. The third -- which we call "breather" intercalators -- provide probes that relate DNA breathing motions with the intercalation process. These molecules contain bulky side chain substituents whose presence might be expected to interfere with the intercalation process (one example is N,N-[N-benzyl-N,N-dimethylethylammonium]-1,8,4,5-naphthyldimide). For steric reasons, intercalation by this class of compounds must be preceded by the transient rupture of hydrogen-bonds connecting base-pairs -- events termed, DNA breathing. The ability of these compounds to intercalate into DNA provides convincing evidence that drug intercalation and DNA breathing are related phenomena.

Here, we describe the results of an X-ray crystallographic study of a complex containing 2-methyl-4-nitroaniline ethylene dimethylammonium hydrobromide (MNAED, shown in Figure 1) and 5-iodocytidylyl(3'-5')guanosine (iodoCpG). The structure is of interest since it demonstrates an unusual type of association between this reporter molecule and the nucleic acid component. Other interesting features of this crystal structure are presented.

# Materials and Methods

2-Methyl-4-nitroaniline ethylene dimethylammonium hydrobromide (MNAED) was a gift from Professor Edward J. Gabbay, and was used without further purification. The dinucleoside monophosphate, cytidylyl(3'-5')guanosine (CpG), was purchased as the ammonium salt from Sigma Chemical Company and converted to the iodinated form using methods described previously (5). Deep yellow pyramidal shaped crystals were obtained by slow evaporation of equimolar mixtures of MNAED and iodoCpG. The crystal density was estimated by the floatation method to be 1.545 gm/cm<sup>3</sup>.

Crystals were initially characterized by rotation and precession photographs using nickel filtered  $\text{Cu}\text{K}\alpha$  radiation, and the cell parameters then refined on a CAD-4 automatic diffractometer. Unit cell dimensions are: a = b = 32.192(10) Å and c = 23.964(8) Å, space group P4\_2\_12. Data were collected at low temperature (2°C) by the Molecular Data Corporation on a CAD-4 system out to a 20 Bragg angle of 85°, using monochromatic CuK $\alpha$  radiation. A total of 4,942 unique reflections were collected, of which 2,002 were significantly above background (>3 $\sigma$ I). Three standard reflections were monitored periodically during the course of data collection and no significant changes in intensity were observed. The data were corrected for Lorentz and polarization effects. In addition, absorption corrections were applied based on a series of  $\psi$  scans. Structure factors were put on

an absolute scale and the overall temperature factor estimated by Wilson's method. These were then converted into quasi-normalized structure factors using the method of Karle and Hauptman (6).

An  $(E^2-1)$  Patterson map was computed and this revealed the positions of two iodine atoms. A Fourier map based on these heavy atom positions showed another heavy peak, interpreted to be a bromide ion. Subsequent attempts to gain additional meaningful chemical information by Fourier methods were not successful. However, atomic superposition based on the positions of the heavy atoms gave about eighty peaks which, when combined with Fourier and sum-Fourier maps, eventually gave the complete structure. The asymmetric unit contains two iodoCpG molecules, three MNAED molecules and 31 water molecules. a total of 160 atoms. This was then refined by full matrix least squares, using rigid groups for purine and pyrimidine bases and phosphate groups. Isotropic temperature factors were used for light atoms, while anisotropic temperature factors were used for the heavy atoms. The final residual based on 2,002 observed reflections is 11.4%. Coordinates and temperature factors are shown in Table 1.

# Results

Figures 2a and b show the complex viewed approximately parallel to and perpendicular to the quanine-cytosine base-pairs and MNAED molecules. IodoCpG forms a miniature Watson-Crick duplex, with two MNAED molecules stacking immediately above and below. This 2:2 structure possesses approximate 2-fold symmetry -- the bromide ion lying very close to this 2-fold axis. In addition, a third MNAED molecule stacks on MNAED(1) with its side chain oppositely oriented. The overall complex is electrically neutral. Each MNAED molecule carries a positive charge and these are counterbalanced by the negatively charged phosphate groups and bromide ion.

The conformation of the iodoCpG base-paired structure is of particular interest for several reasons. The separation between least squares planes of the base-pairs is 3.7 Å -- a distance considerably larger than the normal separation (3.4 Å). This reflects an unusual buckling present within each guanine-cytosine base-pair. Least squares planes computations indicate that guanine and cytosine rings form a dihedral angle of about 19° -- hydrogen bonds connecting base-pairs are therefore bent this amount. The sugar-phosphate backbone conformation is C3' endo (3'-5') C3' endo in iodoCpG(1) and C3' endo (3'-5') C2' endo in iodoCpG(2). The angular twist between the basepairs is about 26° -- this is due in part to the mixed sugar puckering present in one of the two chains. Dihedral angles that define the overall conformation are summarized in Table 2.

Duplexes are related by 2-fold symmetry and are held together by hydrogen bonds (see Figures 3a and b). These involve the 2-amino- group and the N(3) ring nitrogen on one guanosine residue in the first duplex interacting with the O(2) carbonyl- and O(2)' hydroxyl- group on a cytidine residue in the second symmetry related duplex. In addition, the 2-amino- group of the other

Table 1. Final coordinates and isotropic temperature factors of MNAEDiodoCpG complex after least squares refinement. Occupancy factors for disordered water structure are indicated

		Tacto	15 101	u 1501	dereu	nucci scie	ic cur	c uic	murca	UC G		
NO.	ATOM	X/A	Y/F	Z/C	В	NO.	ATOH	X/A	Y/B	Z/C	B	
				5-	IODOCYTID	YLYL(3'-5')GUANC	SINE					
		101	DO-CPG(1)					101	DO-CPO(2)			
1	15 (1	0.0920	0.4318	-0.1049	6.5							
2	N1 C1	0.0920	0.4318	-0.1129	6.2	42 43	15 C2 N1 C2	0.0110	0.3935 0.2916	0.2072 0.3051	6.8	
3	C2 C1 O2 C1	0.1626	0.3021	-0.0559	8.4	44 45	C2 C2	0.1038	0.2764	0.2525	4.5	
5	N3 C1	0.1486	0.3284	-0.0498 -0.0173	4.1	46	02 C2 N3 C2	0.1270	0.2974	0.2085	7.1	
6	C4 C1	0.1271	0.3639	-0.0256	7.7	47	C4 C2	0.0625	0.3312	0.2146	6.4	
7 8	N4 C1 C5 C1	0.1143	0.3847	0.0169 -0.0843	5.7 9.7	48	N4 C2 C5 C2	0.0480	0.3507	0.2680	7.2	
9	C6 C1	0.1355	0.3509	-0.1242	2.7	50	C6 C2	0.0455	0.3236	0.3119	5.0	
10	C1'C1 C2'C1	0.1787 0.1561	0.2939	-0.1578 -0.1763	7 · 1 8 · 5	51 52	C1'C2 C2'C2	0.1062	0.2620	0.3518	4.7	
12	C3'C1	0.1311	0.2726	-0.2274	8.8	53	C3.C2	0.1442	0.3026	0.4166	6.5	
13	C4'C1	0.1606	0.3035	-0.2525	4.8	54	C41C2	0.1050	0.2856	0.4440	7.4	
14	01 1C1 C51C1	0.1835 0.1418	0.3202 0.3386	-0.2043 -0.2855	5.6 11.5	55 56	01'C2 C5'C2	0.0799	0.2682	0.3989 0.4764	7.3	
16	C5 'C1	0.1168	0.3611	-0.2513	10.9	57	05 ° C 2	0.0711	0.3533	0.4421	10.8	
17 18	02'C1	0.1849 0.1148	0.2244	-0.1960 -0.2613	6.4	59 59	02°C2	0.1693	0.2343	0.3906	8.7 5.6	
19	F1	0.0676	0.2244	-0.2610	9.9	60	PZ	0.2161	0.3376	0.4402	6.1	
20	D1 P1	0.0375	0.2565	-0.2434	13.0	61	01 P2	0.2025	0.3745	0.4109	8,5	
21 22	02 P1 05'G1 C5'G1	0.0648	0.2046 0.1891	-0.3159 -0.2145	15.5	62 63	02 P2 05'G2	0.2376	0.3402	0.4950	7.9 7.9	
23	C5'G1	0.0962	0.1551	-0.2169	9.9	64	C5′G2	0.2596	0.2742	0.4115	5.4	
24	C4'G1 C3'G1	0.0972	0.1326	-0.1407	7.6	65 66	C41G2 C31G2	0.2811	0.2546	0.3615	6.1	
26	C2'61	0.0667	0.1261	-0.0715	5.6	67	C21G2	0.2778	0.3085	0.2964	11.2	
27 28	C1'G1 01'G1	0.1033	0.1574	-0.0694 -0.1246	5.8	69 69	C1'62 O1'62	0.2682	0.2783 0.2546	0.2730	6-6 5.8	
29	02 'G1	0.0838	0.0882	-0.1248	12.8	70	01'62	0.2527	0.2346	0.3161	17.3	
30	D3'G1	0.0352	0.0891	-0.1536	11.2	71	03'62	0.3520	0.2586	0.3261	13.2	
31 32	N1 B1 C2 G1	0.0958	0.2567	0.0961	6.1	72 73	N1 G2 C2 G2	0.1794	0.3081	0.0952	4.1	
33	N2 G1	0.1345	0.2045	0.1365	3.2	74	N2 02	0.2148	0.2506	0.0621	3.1	
34 35	N3 G1 C4 G1	0.1157 0.0947	0.1959 0.2148	0.0444	4.3 3.5	75 76	N3 G2 C4 G2	0.2269	0.2692	0.1524	4.5	
36	C5 G1	0.0741	0.2515	0.0030	5.1	77	C5 G2	0.1904	0.3305	0.1884	8.5	
37 38	C6 G1	0.0733 0.0568	0.2757	0.0513	5.5	78 79	C6 82	0.1689	0.3391	0.1340	4.5	
39	N7 G1	0.0540	0.2588	-0.0469	5.8 4.4	80	D6 G2 N7 G2	0.1428 0.1888	0.3651 0.3559	0.1243	3.9 8.3	
40	C8 G1	0.0648	0.2252	-0.0771	3.2	81	CB G2	0.2146	0.3354	0.2699	7.7	
41	N9 G1	0.0901	0.1974	-0.0505	5.9	82	N9 G2	0.2327	0.3004	0.2447	7.3	
		MNAED NO	OLECULE (1	)				MNAED M	OLECULE (2	)		
83	C1 D1	-0.0180	0.2327	0.0943	7.3	99	C1 B2	0.2422	0.4133	0.1069	4.4	
84 85	C2 D1	0.0024	0.1928	0.0964	6.2	100	C2 D2	0.2677	0.3794	0.0968	6.3	
85 86	C3 B1 C4 D1	0.0213	0.1806	0.1436	5.2 3.7	101 102	C3 D2 C4 D2	0.2492	0.3614	0.0425	5.1 4.5	
87	C4 D1 C5 D1	0.0002	0.2469	0.1894	6.7	103	C5 D2	0.2194	0.4118	0.0071	3.2	
86 89	C6 D1 C7 D1	-0.0193 0.0026	0.2606	0.1412	7.4 7.2	104 105	C6 D2 C7 D2	0.2166	0.4305	0.0597	5.8 7.3	
90	N4 D1	0.0398	0.1946	0.2417	6.3	106	N4 D2	0.2469	0.3580	-0.0564	8.8	
91 92	NG101 NG201	0.0581	0.1614	0.2436	9.4 9.5	107 108	NO1D2 NO2D2	0.2721	0.3299	-0.0615 -0.0943	11.3	
93	NR D1	-0.0364	0.2466	0.0403	6.0	109	NB B2	0.2379	0.4317	0.1573	7.5	
94 95	C9 D1	-0.0608	0.2852	0.0331	6.8	110	C9 D2	0.2147	0.4699	0.1680	6.6	
96	N11D1	-0.0378 -0.0146	0.3261	0.0203	4.4 7.1	111 112	C10D2 N11D2	0.1676	0.4657 0.440B	0.1802 0.2274	10.1 B.4	
97	C12D1	-0.0448	0.3187	-0.0769	14.6	113	C12D2	0.1230	0.4292	0.2371	6.7	
98	C1301	0.0011	0.3625	-0.0368	7.6	114	C13D2	0.1778	0.4545	0.2811	12.5	
					MNAET	MOLECULE(3)						
115	BR	0.0287	0.4440	0.0902	0.5	124	NO1D3	-0.0579	0.0929	0.0928	21.5	
116	C1 D3 C2 D3	-0.0240 -0.0081	0.1151	0.2914	7.2 13.9	125 126	NO2D3 NO D3	-0.0976 -0.0060	0.1447	0.1177 0.3584	21.8	
118	C3 D3	-0.0239	0.0876	0.1978	16.7	127	C9 D3	-0.0296	0.1168	0.4122	18.7	
119	C4 D3 C5 D3	-0.0544	0.1187	0.1831	15.4	126	C10D3	-0.0061	0.1321	0.4649	28.3	
120		-0.0687 -0.0532	0.1440	0.2221	10.7	129 130	N11D3 C12D3	0.0371	0.1276	0.4550	26.3	
122	C6 D3 C7 D3	0.0230	0.0535	0.2647	6.4	131	C1303	0.0482	0.1550	0.4110	38.0	
123	N4 D3	-0.0729	0.1165	0.1254	16.3							
					SOLVENT	MOLECULE ATOMS						
132 133	0₩1 0₩2	0.4770	0.3924	0.4768	27.7	147 148	0W14R 0W15A	0.3275 -0.0781	0.3275	0.5000	21.0	0.9
133	OW2 OW3	-0.0812 0.4347	0.2009	0.5092	17.6	148	0W15A	-0.0781 -0.0557	0.2363	0.6100 0.6538	7.9 5.9	0.
135	084	-0.1105	0.2823	0.8391	18.6	150	DW16A	0.1239	0.4155	0.4178	19.1	0.
136	0#5 0#6	0.3263	0.4860	0.3938	22.4	151 152	OW168 OW17A	0.1394	0.4318	0.4619	27.5 8.5	0.
130	OW7	0.4002	0.1258	0.4143	19.9	153	DW178	0.3278	0.0279	0.3576	15.1	٥.
139	OMS	0.0105	0.0289	0.0471	27.6	154	DW170	0.3582	0.0423	0.3803	6.0	٥.
	DW9 DW10	0.1101	0.1519	0.6342	13.2 27.6	155 156	OWIER	-0.0257 0.0297	0.3953	0.7993 0.8134	10.7	٥.
140		0.1562	0.2594	0.5817	19.1	157 158	DW19A DW19B	0.0966	0.4731	0.3715	18.0	o.
141 142	0911	0 4						0.4263	0.0199	0.4618	23.9	
141 142 143	0W11 0W12 0W13A	0.1753	0.1862	0.2955	10.8 0.		0W20A	0,0285		0.7497	9.1	0.
141 142	OW12 OW13A OW13B	0.1753 -0.1292 -0.1012 -0.1254		0.2933 0.6663 0.6431 0.7512	10.8 0.	50 159 50 160		0,0285 -0,0145	0.4653 0.4813	0.7497 0.6789	9.3	0.

guanosine residue in the first duplex hydrogen bonds to the furanose ring oxygen on its symmetry related counterpart. A total of six hydrogen bonds connect these symmetry related duplexes.

Figure 4 shows a projection of the structure down the  $\underline{c}$  axis. Bromide ions and iodine atoms cluster together around the 4-fold screw axis, forming

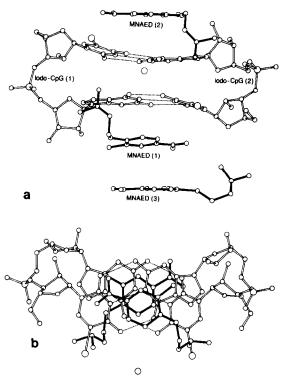


Figure 2. A portion of the MNAED-iodoCpG complex viewed approximately parallel to (a) and perpendicular to (b) base-pairs and MNAED molecules.

Table 2. Torsional angles describing the conformations of the sugarphosphate chains in the MNAED-iodoCpG crystalline complex

Torsional Angle	iodoCpG(1)	iodoCpG(2)	
01'C-C1'C-N1C-C6C	13	12	
01'G-C1'G-N9G-C8G	49	77	
05'C-C5'C-C4'C-C3'C	61	53	
C5'C-C4'C-C3'C-03'C	80	79	
C4'C-C3'C-O3'C-P	221	200	
C3'C-O3'C-P-O5'G	273	283	
03'C-P-05'G-C5'G	305	299	
P-05'G-C5'G-C4'G	165	175	
05'G-C5'G-C4'G-C3'G	42	63	
C5'G-C4'G-C3'G-03'G	89	143	
C4'C-01'C-C1'C-C2'C	8	6	
01'C-C1'C-C2'C-C3'C	-27	-27	
C1'C-C2'C-C3'C-C4'C	37	36	
C2'C-C3'C-C4'C-01'C	-34	-34	
C3'C-C4'C-01'C-C1'C	17	18	
C4'G-01'G-C1'G-C2'G	- 7	-22	
01'G-C1'G-C2'G-C3'G	-20	33	
C1'G-C2'G-C3'G-C4'G	36	-30	
C2'G-C3'G-C4'G-01'G	-40	20	
C3'G-C4'G-01'G-C1'G	31	1	

The torsional angle is defined in terms of 4 consecutive atoms, ABCD, the positive sense of rotation is clockwise from A to D while looking down the BC bond.

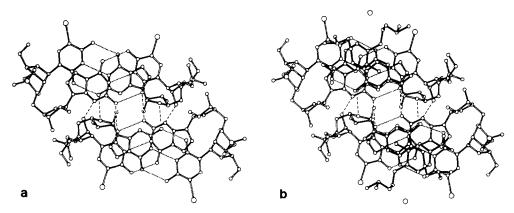


Figure 3. Hydrogen bonding between 2-fold symmetry related iodoCpG duplexes (a) and these duplexes with stacked MNAED molecules shown above and below (b). See text for discussion.

a helical columnar arrangement in the  $\underline{z}$  direction. MNAED molecules cluster around a 2-fold screw axis, themselves related by additional 2-fold rotational symmetry along the diagonal. Hydrogen bonded pairs of 3:2 MNAED-iodoCpG complexes pack together in a zig-zag fashion in the crystalline lattice. The complex is heavily hydrated with water molecules hydrogen bonded to sugar

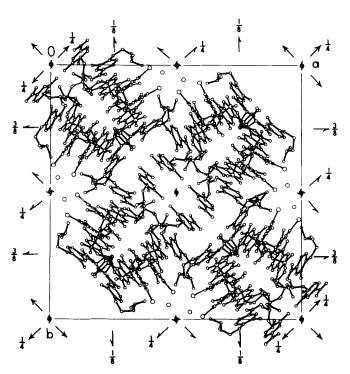


Figure 4. Lattice figure showing the crystal structure of the MNAED-iodoCpG complex viewed down the  $\underline{c}$  axis. For simplicity, water structure is omitted. See text for discussion.

hydroxyl- groups and phosphate oxygen atoms. Additional hydrogen bonds form between water molecules and other hydrophillic groups on base-pairs and MNAED molecules.

#### Discussion

2-Methyl-4-nitroaniline ethylene dimethylammonium hydrobromide is one in a series of nitroaniline reporter molecules synthesized by Gabbay that intercalate -- either partially or fully -- into DNA structure. This class of molecules is of particular interest, since they resemble naturally occuring aromatic amino- acids such as tyrosine and phenylalanine. Their ability to intercalate into DNA raises the possibility that aromatic side chains in proteins might also intercalate when binding to DNA (7-9).

Recent experimental evidence indicates the presence of nuclease hypersensitive sites in eukaryotic DNA, many of these located at 5' ends of genes (10,11). These same sites are sensitive to cleavage by a 1,10-phenanthrolinecopper(I) complex, a known intercalating agent. These data suggest the existence of an altered DNA conformation in these regions recognized by nucleases and intercalators by a common intercalative mechanism (12).

The current study was carried out to investigate the ability of 2-methyl-4-nitroaniline ethylene dimethylammonium hydrobromide to intercalate -- either partially or fully -- into the self-complementary dinucleoside monophosphate, iodoCpG. Although intercalation in the classic sense has not been observed, this nitroaniline derivative stacks above and below base-pairs in the miniature Watson-Crick type duplex structure. The absence of intercalation could reflect the smaller stacking energies associated with the aromatic ring system in MNAED and the base-pairs, along with other factors determining the energetics of the complex. We have repeatedly observed that molecules containing larger conjugated ring systems form intercalation complexes with this same dinucleoside monophosphate (13-15). It is possible that longer oligonucleotides would provide better model systems to study the interaction of MNAED with DNA, and we are pursuing this further at the present time.

## Acknowledgements

This work has been supported in part by the National Institutes of Health and the Department of Energy. The paper has been assigned report no. UR-2384 at the Department of Energy Project, the University of Rochester.

## References

- Gabbay, E.J. (1977) Bioorganic Chemistry, 3, Macro- and Multimolecular Systems, pp. 33-70, Academic Press, Inc., New York and San Francisco.
- Gabbay, E.J., DeStefano, R. and Baxter, C.S. (1973) Biochem. Biophys. 2.
- Res. Commun. <u>51</u>, No. 4, 1083-1089. Gabbay, E.J., Glaser, R. and Gaffney, B.L. (1970) Ann. N.Y. Acad. Sci. 3. 171, 810-826.

- 4. Gabbay, E.J., Scofield, R.E. and Baxter, C.S. (1973) J. Amer. Chem. Soc. 95, No. 23, 7850-7857.
- Tsai, C.-C., Jain, S.C. and Sobell, H.M. (1977) J. Mol. Biol. 114, 301-315.
- 6. Karle, J. and Hauptman, H. (1953) Acta Cryst. 6, 473-476.
- Helene, C. (1971) Nature New Biol. 234,  $120-12\overline{1}$ .
- Coleman, J.E., Anderson, R.A., Ratcliffe, R.G. and Armitage, I.M. (1976) Biochemistry 15, 5419-5430.
- Sobell, H.M. (1979) Biological Regulation and Development 1, ed. R.F. Goldberger, pp. 171-199, Plenum Publishing Company, New York.
- 10. Jessee, B., Gargiulo, G., Razvi, F. and Worcel, A. (1982) Nucleic
- Acids Res. 10, No. 19, 5823-5834.
  Cartwright, I.L. and Elgin, S.C.R. (1982) Nucleic Acids Res. 10, No. 19, 5835-5852. 11.
- Banerjee, A. and Sobell, H.M. (1983) J. Biomol. Structure and Dynamics  $\underline{1}$ , No. 1, 253-262. 12.
- 13. Jain, S.C., Tsai, C.-C. and Sobell, H.M. (1977) J. Mol. Biol. 114, 317-331.
- Reddy, B.S., Seshadri, T.P., Sakore, T.D. and Sobell, H.M. (1979) 14. J. Mol. Biol. 135, 787-812.
- 15. Jain, S.C., Bhandary, K.K. and Sobell, H.M. (1979) J. Mol. Biol. 135, 813-840.