
The molecular electrostatic potential, steric accessibility and hydration of Dickerson's B-DNA dodecamer d(CpGpCpGpApApTpTpCpGpCpG)

Richard Lavery and Bernard Pullman

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique, associé au C.N.R.S.,
13, rue Pierre et Marie Curie, 75005 Paris, France

Received 15 June 1981

ABSTRACT

The paper presents the results of computation of the electrostatic potential and steric accessibility of the B-DNA self-complementary dodecamer CGCGAATTCGCG following the geometry of the recent single crystal structure of Dickerson et al. This structure shows significant variations from classical B-DNA; their influences on the calculated properties are discussed. The results are related to general features of hydration of the crystal. A particularly significant general finding concerns the greater negative potential in the center of the oligonucleotide helix than at its extremities. This will be a general feature of such short helices, independent of their base sequence. It may have important implications for the reactivity of DNA oligomers.

INTRODUCTION

In a recent and very stimulating series of publications Dickerson and coworkers (1-4) have presented the single crystal structure of a B-DNA dodecamer with the self-complementary base sequence CGCGAATTCGCG. This structure, which reveals many significant variations from classical B-DNA geometry (as established from sequence averaged fibre diffraction studies), heralds a new era in nucleic acid research.

We have carried out recently in our laboratory a series of theoretical studies on the effect of the macromolecular structure of nucleic acids upon their biochemical behaviour, the two outstanding features investigated being the molecular electrostatic potential and the accessibility to reactive sites or regions (5-12). The availability of Dickerson's results has now incited us to explore the effect of the observed structural variations along the nucleotide chain on the calculated properties and their implication for the interactions of this species with surrounding or attacking molecules e.g. the water of hydration present in the crystal. From a broader point of view we wish also to distinguish among these structural variations those which may be due to the specific sequencing of this oligonucleotide and those which are essentially determined by its finite dimensions and the positioning of the

bases or base pairs with respect to the center or the extremities of this relatively short double helix. Such a distinction is expected to avoid erroneous extrapolation of results to infinite chains of DNA and to indicate possibly some general features predictable for oligonucleotide helices.

METHOD

Our technique for calculating the electrostatic potentials of macromolecules has been described in detail in previous articles (5, 7, 13) and will not be restated here. We note, however, one complexity with Dickerson's dodecamer, namely that all its phosphates and sugars have somewhat different conformations. Following our technique, the electrostatic potential of the dodecamer will be calculated as the superposition of the potentials of its individual subunits, the bases, sugars and phosphates. These potentials are evaluated from the multipole expansion of the charge densities of these subunits, which are in turn calculated from their ab initio wavefunctions. As it would be prohibitively expensive to compute the complete set of different wavefunctions for the 22 phosphates and 24 sugars of the dodecamer we employ the simplification of representing closely grouped conformations by a single multipole expansion, as was done in our studies of tRNA^{Phe} (9). Thus four representative sugar conformations (C2'-endo, C1'-exo, O1'-endo and C3'-endo) and three representative phosphate conformations (with ω , ω' , torsion angles : - 46°, - 96° (g^-g^-), - 85°, - 45° (g^-g^-) and - 72°, 172° (g^-t)) were employed.

The calculation of static steric accessibility of the atoms of the dodecamer was carried out using the methodology also described previously (1, 8-9). Accessibilities are calculated as areas, expressed in Å², accessible on the van der Waals spheres of the atoms concerned toward an attacking sphere of 1.2 Å radius. Hydrogens were included in the dodecamer using standardized bond lengths.

RESULTS AND DISCUSSION

A schematic representation of the dodecamer studied is shown in fig. 1, which also gives the numbering of the bases (the same as that employed by Dickerson et al.), sugars and phosphates.

We shall center our attention on the molecular electrostatic potentials associated with sites on the nucleic acid bases susceptible to electrophilic attack. These sites may be divided (5) into two principal groups : those lying in the plane of the bases, associated primarily with pyrimidine-like nitrogens (e.g. N7 of guanine, notation N7(G)) and their carbonyl oxygens (e.g. O6(G)) and those lying symmetrically above and below the plane of the bases, associa-

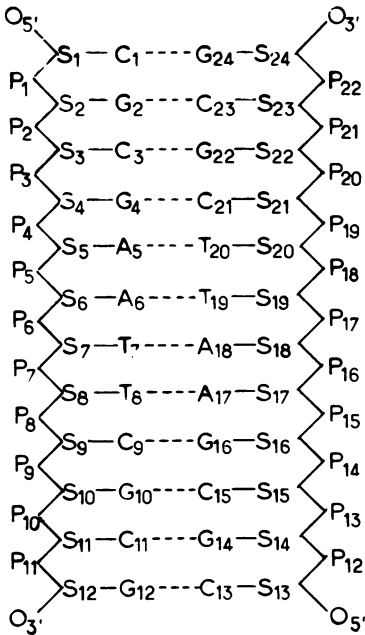


FIGURE 1. Schematic representation of the dodecamer with the numbering employed.

ted principally with the amino groups and C8 of the purines. In the latter case only one result is presented corresponding to the more negative of the two potentials calculated.

The potentials presented correspond to an unscreened dodecamer and are thus strongly negative due in particular to the influence of the 22 anionic phosphate groups. Mg^{2+} counteractions are known to be present in the crystal but their positions are disordered (3) and it is therefore difficult to include their effect explicitly in the present study. We have shown in previous publications concerning the influence of counteractions on B-DNA potentials (14, 15), that when either Na^+ , Mg^{2+} or $CH_3NH_3^+$ cations are site bound to the nucleic acid phosphates, although the magnitudes of the base site potentials are strongly reduced, their ordering is only slightly modified. We currently expect a similar situation to prevail for Dickerson's dodecamer, despite its local deviations from the typical regular B-DNA conformation. Discussion of the effect of the second type of counterion present in the crystal, namely spermine, will be taken up in a later publication.

The site potentials for the different guanines of the dodecamer are indicated in the first part of table I. Their distribution shows three principal features. Firstly, the order of the potentials of the sites remains relatively

TABLE 1. POTENTIAL AND ACCESSIBILITIES FOR GUANINES

BASE	SITE POTENTIAL (kcal/mole)						ATOM ACCESSIBILITY (Å ²)					
	N1	N2	N3	N7	C8	O6	N1	N2	N3	N7	C8	O6
2	-508	-522	-540	-542	-492	-528	0.0	0.1	0.5	1.9	0.2	1.6
4	-581	-593	-629	-631	-576	-611	0.0	0.2	1.0	2.8	0.3	1.2
10	-542	-568	-612	-584	-532	-566	0.0	0.0	0.2	5.1	1.0	3.5
12	-457	-455	-477	-496	-447	-470	0.9	1.8	1.9	4.9	1.6	3.4
14	-511	-522	-533	-544	-491	-534	0.0	0.3	0.4	2.9	0.0	1.9
16	-585	-613	-632	-630	-573	-609	0.0	0.0	0.6	3.3	0.6	1.8
22	-515	-555	-596	-585	-527	-571	0.0	0.0	0.6	4.0	0.4	2.5
24	-454	-459	-483	-502	-446	-472	0.7	1.5	1.7	4.7	1.2	4.8

constant in all bases. The strongest potentials are associated either with N3 or N7, O6 has an intermediate potential and N1, N2 and C8 have the weakest ones. Note that the more negative sites occur in the plane of the bases and the weaker ones out of plane. Secondly, the guanines which would be equivalent by symmetry in a regular classical B-DNA dodecamer, that is bases 2 and 14, 4 and 16, 10 and 22, 12 and 24, remain very similar to each other in the dodecamer. Thirdly, as a direct effect of the finite length of the DNA segment, the potentials associated with the bases closer to the center of the dodecamer (e.g. guanines 4 and 16) are much more negative than those associated with the bases located at the ends of the dodecamer (guanines 12 and 24). This difference in potential may be as great as 150 kcal/mole for equivalent sites and represents therefore a more important effect than the variations in the potential within a given base (typically of the order of 50 kcal/mole maximum).

The site potentials for cytosines, adenines and thymines, given in the first parts of tables 2, 3 and 4 respectively, show similar features to those noted for guanine. For these bases the order of site potentials is, however, constant for each species and, as for guanine, the strongest potentials are associated with the in-plane sites (O2 of cytosine, N3 and N7 of adenine, O2 and O4 of thymine). The ordering of the potentials for each of the four types of base may be summarized thus as :

Guanine : N3 = N7 > O6 = N2 > N1 = C8
 Cytosine : O2 > N4 > C5
 Adenine : N3 > N7 > N6 > C8
 Thymine : O2 > O4 > N3 > C5

TABLE 2. POTENTIALS AND ACCESSIBILITIES FOR CYTOSINES

BASE	SITE POTENTIAL (kcal/mole)			ATOM ACCESSIBILITY (Å ²)		
	N4	C5	O2	N4	C5	O2
1	- 460	- 446	- 493	1.4	2.5	3.2
3	- 560	- 543	- 562	0.1	0.8	1.1
9	- 570	- 555	- 629	0.0	0.2	1.0
11	- 512	- 493	- 544	0.1	0.5	1.1
13	- 436	- 410	- 478	2.1	3.3	1.6
15	- 545	- 535	- 572	0.3	0.8	0.9
21	- 569	- 555	- 608	0.0	0.1	1.1
23	- 513	- 459	- 524	0.1	0.4	1.1

TABLE 3. POTENTIALS AND ACCESSIBILITIES FOR ADENINES

BASE	SITE POTENTIAL (kcal/mole)				ATOM ACCESSIBILITY (Å ²)			
	N3	N6	N7	C8	N3	N6	N7	C8
5	- 645	- 601	- 635	- 567	0.8	0.1	2.1	0.6
6	- 667	- 592	- 633	- 568	0.7	0.1	2.0	0.0
17	- 669	- 597	- 640	- 597	0.5	0.1	2.1	0.5
18	- 679	- 596	- 640	- 579	0.7	0.1	2.1	0.0

TABLE 4. POTENTIALS AND ACCESSIBILITIES FOR THYMINES

BASE	SITE POTENTIAL (kcal/mole)				ATOM ACCESSIBILITY (Å ²)			
	N3	N6	N7	C8	N3	N6	N7	C8
7	- 580	- 568	- 655	- 605	0.0	0.0	0.6	2.5
8	- 589	- 565	- 658	- 603	0.0	0.0	0.6	1.6
19	- 586	- 566	- 658	- 613	0.0	0.0	0.8	1.9
20	- 593	- 568	- 644	- 609	0.0	0.0	1.0	1.2

The cytosines also show clearly the grouping of the pseudo-equivalent pairs (numbers 1 and 13, 3 and 15, 9 and 21, 11 and 23) and the notably stronger potentials for the more central bases. For adenine and thymine these two effects are much less significant as these bases are closely grouped in the center of

the dodecamer and the four adenines and four thymines are thus similar in terms of their site potentials.

The variation of base site potentials along the dodecamer axis is summarized in fig. 2 where each of the twenty four bases is represented by a vertical line whose ends are the most and least negative site potentials calculated for that base. The result is a clear demonstration of the much more negative potentials found in the center of the oligonucleotide. The two strands of the double helix are seen to be very similar and for each of them the potentials of the central adenines and thymines are 150-200 kcal/mole more negative than those of the terminal guanines or cytosines. This is an important effect which will be found in any finite segment of helical DNA and whose implications we shall consider at the end of this section.

The atomic accessibilities of the dodecamer bases have been calculated for the same nucleophilic centers of the dodecamer and are presented in the second parts of tables 1, 2, 3 and 4 for the guanines, cytosines, adenines and thymines, respectively.

As for the base site potentials, the order of base atom accessibilities remains largely unchanged for each type of base, namely.

Guanine : N7 > O6 > N3 ≈ C8 > N2 ≈ N1

Cytosine : O2 > C5 > N4

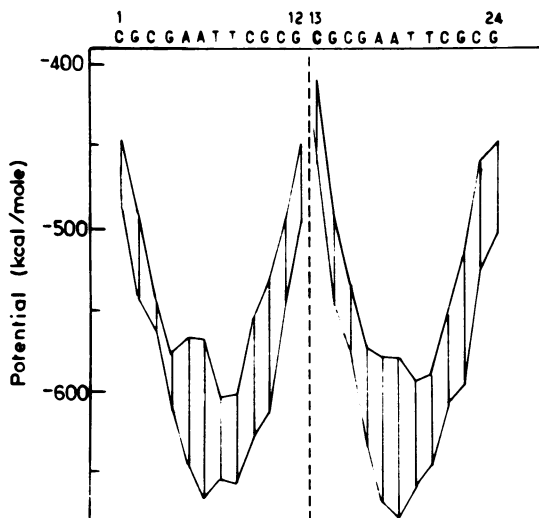


FIGURE 2. Variations of site potential on the bases along the dodecamer.

Adenine : $N7 > N3 \approx C8 > N6$

Thymine : $O4 > O2 > C5 \approx N3$

This order is virtually unchanged from that calculated in our previous studies for B-DNA with classical geometry (6). Nevertheless, the similarity of bases which would be symmetry equivalent in a regular B-DNA is more perturbed in terms of accessibilities than it was for site potentials. This is understandable due to the higher sensitivity of static accessibilities to small, local variations in conformation. Thus e.g. such local variation occurs at base pair G10-C15 where the local twist angle is the highest for the dodecamer and guanine 10 is almost unstacked from cytosine 11 (3). This leads to particularly high accessibilities calculated for N7 and O6 of guanine 10. A related situation between bases 22 and 23 has a similar effect on the accessibility of these atoms on guanine 22.

On the other hand, the influence of the finite length of the helix leads to an increased accessibility of the terminal bases guanines 12 and 24 and cytosines 1 and 13. This effect may be seen clearly in figure 3 where the largest and smallest accessibilities associated with each base along the dodecamer are shown in a manner similar to that for the site potentials in figure 2.

We now turn to another representation of the electrostatic potential of the dodecamer which gives a broader picture of its overall distribution. This is the potential on the surface envelope surrounding the dodecamer (c.f. 7). We shall consider two different views of this envelope, the first of which

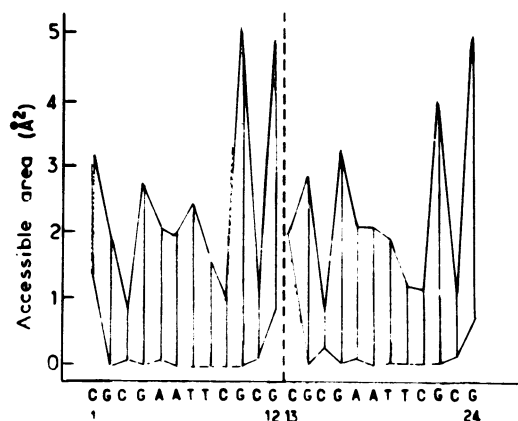


FIGURE 3. Variations of atom accessibility of the bases along the dodecamer.

overlooks the minor groove of the central AATT sequence and the major grooves of the GCGC end sequences, while the second overlooks the major groove of the AATT sequence and the minor grooves of the GCGC ends. For the first view, the molecular graphic of the dodecamer, surrounded by the profile of its surface envelope is given in figure 4 (cytosine 1 may be seen in the top left of this graphic) and the corresponding surface potentials are given in figure 5. These potentials are represented by six grades of shading, darker shades implying more negative values (Table 5). Each grade represents a range of approximately 50 kcal/mole, the overall range of potential on the envelope being slightly more than 300 kcal/mole.

Figure 5 shows with great clarity the much more negative potentials closer to the center of the dodecamer and particularly their concentration in a narrow central band, which by references to figure 4 can be seen to correspond to the minor groove of the AATT base sequence. The letter M marks the most negative potential on the envelope viewed in this direction, which lies on the

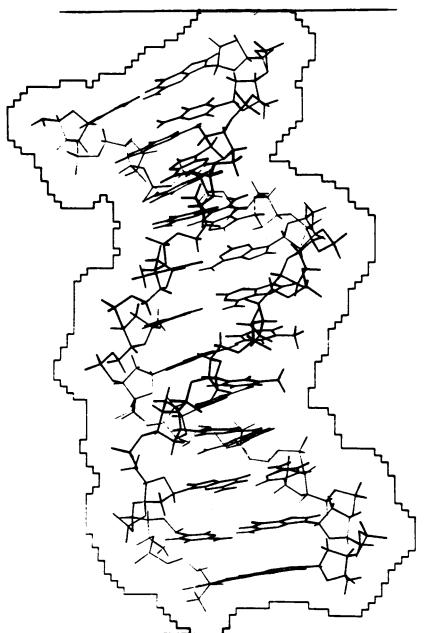


FIGURE 4. Molecular diagram of the dodecamer viewing the side of the AATT minor groove.

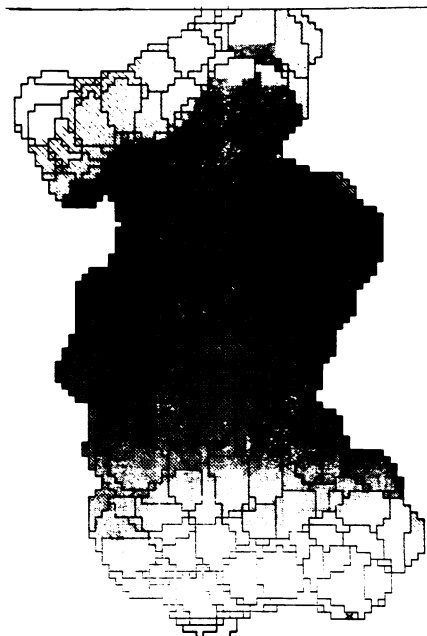








FIGURE 5. Envelope potential of the dodecamer on the side of the AATT minor groove.

TABLE 5. SHADINGS USED IN REPRESENTING THE SURFACE ENVELOPE POTENTIALS

SHADING	POTENTIAL RANGE (kcal/mole)
	- 314
	↓ - 366
	↓ - 418
	↓ - 471
	↓ - 523
	↓ - 575
	↓ - 627

atomic sphere of O2 of thymine 19 (- 623 kcal/mole).

The view from the other side of the surface envelope is given by the molecular graphic in figure 6, (cytosine 1 is visible in the top right of this graphic) and by the surface potentials in figure 7. These present a somewhat different distribution : although a central zone of more negative potentials may again be seen, it is much broader than in figure 5 and less deep. The minimum (M) on this view, now lies on the surface of a phosphate anionic oxygen and not on a base atom as in the AATT minor groove and its magnitude, - 590 kcal/mole, is some 30 kcal/mole less than that in the preceding view.

It is obvious that the more negative potentials in the central zone of the dodecamer visible both on the surface envelope and for base sites are a direct effect of the finite length of this DNA segment. The more negative

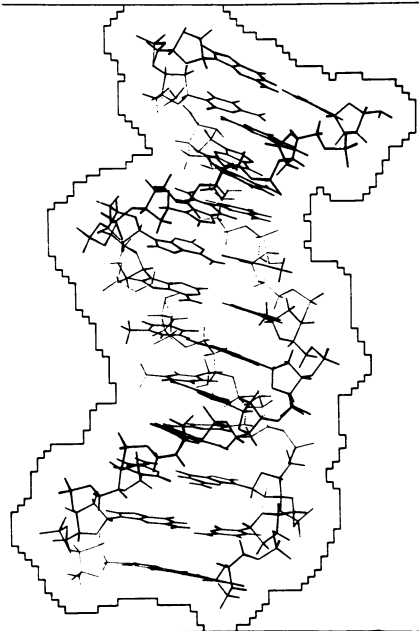


FIGURE 6. Molecular diagram of the dodecamer viewing the side of the AATT major groove.

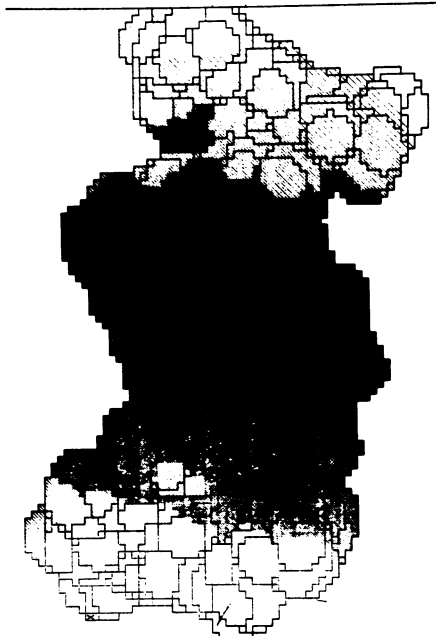


FIGURE 7. Envelope potential of the dodecamer on the side of the AATT major groove.

potential in the minor groove of the AATT sequence compared to its major groove is a general feature of AT rich sequences which has been noted in our previous studies of B-DNA (6, 7) and is in contrast to GC rich sequences where the major groove contains the deeper potentials. The shortness of the double helix leads thus to an altogether very deep potential well in the central AATT segment.

As a preliminary attempt to show the possible practical significance of the calculations we may consider their possible implication for interpreting the interaction of the dodecamer with other molecules. In this respect Dickerson and coworkers provided data on the location of 72 water molecules as well as one spermine in the dodecamer crystal (4). We shall consider these interactions in detail in a future publication and would like to make here only a few general observations on the hydration scheme.

Following Drew and Dickerson (4) the most significant features related to this water distribution are the following : 1) most of the ordered waters are

found in the groove of the double helix, while waters associated with the phosphate groups are rather disordered ; 2) the minor groove shows a clear regular structure (a "spine") with bidentate waters binding to adenine N3 and thymine O2 ; up to four layers of hydration are observed in the central AATT sequence, which presents thus the highest affinity for ordered hydration ; the minor groove becomes drier and drier as it winds from the AATT center toward either CGCG end ; 3) more sparse monodentate waters are found in the major groove with no regular structure.

An overall correlation of the calculated potentials with these results is observed. We may note that the highest concentration of waters occurs in the AATT minor groove where the most negative potentials are found. These potentials decline rapidly towards the ends of the dodecamer, as do the number of ordered water molecules observed. The narrowness of the most negative potential zone in the minor groove is also a factor which would be expected to contribute to the formation of a regular structure, in contrast to the disordered monodentate waters found in the major groove where there is a broad zone of somewhat less negative potential.

It may also be remarked that the principal electrophilic atoms of the bases observed to be involved in hydrogen bonding to waters in the minor groove, N3(A) and O2(T), or in the major groove O6(G), N7(G), N7(A) and O4(T) are all quite accessible.

A complete understanding of the hydration of the dodecamer which obviously implies a network of suitable hydrogen bonds with the oligonucleotide substrate would require calculation of the full energy of interaction between the oligonucleotide and the water molecules and between the water molecules themselves as well as taking into consideration the counterions present and the interactions between adjacent dodecamers in the crystal. It may be remarked, however, that the greater possibilities of the minor groove of AT sequences of B-DNA (in comparison with their major groove or with the two grooves in GC sequences) for hydrogen-bonding to water molecules is evident already in our computations on the hydration scheme of the complementary base pairs (16). The distribution of the potential, as presented here, may constitute an element of explanation complementary to this classical one. It may be worth underlining that, on the contrary, considerations based solely on accessible surface areas for base atoms in the major and minor grooves of the AT or GC sequences of the dodecamer (obtainable by summation of the accessibilities of the atoms) as developed by Alden and Kim (17, 18), would lead to the erroneous suggestion of a greater water structuration in the major groove of both these sequences

and would not predict the predominant hydration of the minor groove of the central AT sequence.

We may note that one other interaction observed in the dodecamer crystal, namely, the binding of $\text{cis-Pt}(\text{NH}_3)_2\text{Cl}_2$ in the vicinity of the N7 and O6 atoms of guanines 4 and 16 (1) can be correlated with the calculated high accessibilities of these two atoms and their electrostatic potentials, which are most negative for these two guanines, closest to the center of the dodecamer.

CONCLUSION

Electrostatic molecular potentials and atomic accessibilities have been calculated for the self complementary dodecamer CGCGAATTCGCG using the single crystal geometry solved by Dickerson and coworkers. The results show that the local variations from classical B-DNA conformation have only a slight effect on the calculated base site potentials. The influence on the atom accessibilities, which are more sensitive to local structure, is somewhat greater.

A particularly significant observation concerns, however, the altogether greater negative potentials in the center of this finite segment of DNA compared to its extremities. This situation, which is independent of the base sequence, may have important implications for the interpretation of the reactivity of such DNA oligomers. It should be taken into account before any general conclusions on the influence of base sequence on DNA interactions are drawn from studies of oligonucleotides.

ACKNOWLEDGEMENT

The authors wish to thank Prof. R.E. Dickerson for Communicating to them manuscripts prior to publication.

REFERENCES

- (1) R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, *Nature* 287 (1980) 755-758
- (2) H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, *Proc. Natl. Acad. Sci. USA* 78 (1981) 2179-2183
- (3) R.E. Dickerson and H.R. Drew, *J. Mol. Biol.* (1981) in press
- (4) H.R. Drew and R.E. Dickerson, *J. Mol. Biol.* (1981) in press
- (5) B. Pullman, D. Perahia and D. Cauchy, *Nucl. Acids Res.* 6 (1979) 3821-3829
- (6) D. Perahia, A. Pullman and B. Pullman, *Intern. J. Quant. Chem. Quant. Biol. Symp.* 6 (1979) 353-362
- (7) R. Lavery, A. Pullman and B. Pullman, *Nucl. Acids Res.* 8 (1980) 1061-1079
- (8) K. Zakrzewska, R. Lavery, A. Pullman and B. Pullman, *Nucl. Acids Res.* 8 (1980) 3917-3932
- (9) R. Lavery, A. Pullman, B. Pullman and M. de Oliveira, *Nucl. Acids Res.* 8 (1980) 5095-5112
- (10) R. Lavery, A. Pullman and B. Pullman, *Theoret. Chim. Acta* 57 (1980) 233-243

- (11) R. Lavery and B. Pullman, *Int. J. Quant. Chem., Quant. Biol. Symp.* 8, in press
- (12) K. Zakrzewska, R. Lavery and B. Pullman in *Biomolecular Stereodynamics*, ed. R.H. Sarma, 1981, Pergamon Press, in press
- (13) A. Pullman and B. Pullman in "Molecular Electrostatic Potentials in Chemistry and Biochemistry" eds. P. Politzer and D.G. Truhlar, Plenum, New York (1980) pp. 381-405
- (14) R. Lavery, D. Cauchy, O. de la Luz Rojas and A. Pullman, *Int. J. Quant. Chem. Quant. Biol. Symp.* 7 (1980) 323-330
- (15) D. Cauchy, R. Lavery and B. Pullman, *Theoret. Chim. Acta* 57 (1980) 323-327
- (16) A. Goldblum, D. Perahia and A. Pullman, *FEBS Letters* 91 (1978) 213-215
- (17) C.J. Alden and S.H. Kim, *J. Mol. Biol.* 132 (1979) 411-434
- (18) C.J. Alden and S.H. Kim, in *Stereodynamics of Molecular Systems*, ed. R.H. Sarma, Pergamon Press (1979) pp. 331-350