Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 101, No. 1, February 1989, pp. 55–63.

Electrostatic potential mapping of nucleic acid constituents: I. Guanine, cytosine and the G–C base pair

ANIL KUMAR and P C MISHRA*
Department of Physics, Banaras Hindu University, Varanasi 221 005. India

MS received 13 April 1987; revised 22 November 1988

Abstract. Electrostatic potential maps of guanine, cytosine and the G–C base pair in the Watson–Crick and Hoogsteen configurations have been studied using a mapping dipole of length 1 Å and strength 1 Debye. Net charges on the atoms of the molecules were obtained by the MNDO molecular orbital method. Several closest distances of approach (CDA) between the mapping dipole and atomic sites of the molecules were considered and potentials calculated using interactions of net charges keeping the dipole in the molecular plane in each case. While on one hand this work supports the existence of potential minima obtained by monopole isopotential mapping, on the other it yields additional useful information, e.g. electric field directions which correspond to the directions of probable hydrogen bonds with other species. A deep localized potential region is observed around N_4 of cytosine in the Hoogsteen configuration while it is not present in the Watson–Crick base pairing.

1. Introduction

Mapping of molecular electrostatic potentials has been shown to yield useful information regarding the binding sites of reactants in many reactions including some complex ones (Scrocco and Tomasi 1978; Politzer *et al* 1985). In the prevalent method of potential mapping (Scrocco and Tomasi 1978; Politzer *et al* 1985), a positive monopole is used as the mapping device. As this device does not have a spatial charge distribution and orientational capability, the potentials calculated with its help may not in general correspond to intermolecular interactions, particularly to those between two polar species. Thus a monopole isopotential map yields no information regarding the directionality of certain kinds of interactions e.g. hydrogen bonding. Two further limitations of the monopole isopotential maps are clear. First, in this approach, the calculated potential energies, both the negative and positive ones, are too large in magnitude to correspond to real intermolecular interactions. Second, this kind of map yields no information regarding the probable paths of approacn of two interacting molecular species to come closer to each other.

In the procedure of potential mapping adopted by us, the above mentioned shortcomings are partly rectified. A dipole appears to be a suitable device to study biomolecular electrostatic potentials since the "recognition" of such species occurs via long range interactions where the dipole–dipole contributions would be dominant.

* For correspondence

2. Method of calculations

The approach adopted in an earlier work (Mishra and Tewari 1987) using a dipole as the mapping probe was characterised by the following: (i) Potential energies were calculated using the multipole interaction scheme (Price *et al* 1984) and (ii) only one value (2 Å) of the closest distance of approach (CDA) between the dipole and atomic sites of the molecules studied was considered. Though the results obtained were reasonable and in satisfactory agreement with experiment, certain difficulties were experienced. For example, effects of the neglected higher order terms of the multipole interaction scheme could not be estimated, and even at the octupole–dipole interaction level, at certain points of the chosen CDA surface, convergence of the series could not ensured. However, as the CDA value increases, this problem progressively becomes less serious. But as the most important potential features of molecules are expected to occur near their periphery, use of large CDA values alone, in general, may not be appropriate. In this situation, an approach using interactions of the net charges would be preferable over one based on the multipole interaction formalism. Therefore, in the present work, potential energies of interaction between the mapping dipole and the molecule in question were evaluated by considering interactions of the net charges.

In addition to electrostatic potentials, there is another important property that should be considered i.e. the steric accessibility (Pullman and Pullman 1981) of an atomic site. This is intended to indicate how readily, in terms of steric hindrance, an active site can be approached by another species. This aspect is automatically included in our method as a chosen CDA value is a basic quantity considered for each point where the potential is calculated. The van der Waal radii of the atoms are also considered to appropriately distinguish the different sites from one another. The two ends of the mapping dipole carrying the positive and negative charges were assigned a van der Waal radius of 1 Å each.

Net charges for the molecules were obtained using the MNDO semiempirical molecular orbital method (Dewar and Thiel 1977) and their experimentally determined geometries (Voet and Rich 1970; Arnott *et al* 1969; Hoogsteen 1963). The mapping dipole having a length 1 Å and strength 1 Debye was allowed to move on a geometrical surface enclosing the molecule in question such that CDA was fixed as per choice e.g. (sum of the van der Waal radii of the closest two sites, one in the molecule under study and the other in the mapping dipole + a chosen constant distance); different values of CDA were taken and for each value the potentials were calculated. At each chosen point on the CDA surface, the dipole was allowed to orient itself along the direction of minimum interaction energy; these energies were taken as the characteristic electrostatic potential energies of those points. Thus the present method of potential mapping has one common feature with electric field mapping (Brooks and Lawley 1962; Bonaccorsi *et al* 1972) i.e. the direction of the mapping dipole which represents the electric field direction at that point. But while in the present work the quantity evaluated is the potential energy of interaction, which is usually the quantity of main interest, in the electric field mapping the quantity obtained is that of the electric field.

3. Results and discussion

The calculated net charge distributions for G, C and the G–C base pairs are presented in table 1. Potential maps for cytosine (C), guanine (G) and the G–C hydrogen-bonded

Table 1. Calculated net charges.

Atom*	Net charge			
	Cytosine	Guanine	G–C pair (W–C)†	G–C pair (Hoogsteen)
N_1(C)	−0·347		−0·380	−0·410
C_2(C)	0·419		0·433	0·479
N_3(C)	−0·401		−0·467	−0·515
C_4(C)	0·308		0·342	0·188
C_5(C)	−0·265		−0·285	−0·172
C_6(C)	0·178		0·178	0·110
H_1(C)	0·219		0·322	0·295
O_2(C)	−0·350		−0·459	−0·434
N_4[NH_2 of (C)]	−0·324		−0·319	−0·348
H[NH_2 of (C)]+	0·197		0·194	0·391
H′[NH_2 of (C)]	0·211		0·265	0·107
H_5(C)	0·084		0·087	0·080
H_6(C)	0·072		0·074	0·055
N_1(G)		−0·378	−0·381	−0·372
C_2(G)		0·390	0·403	0·404
N_3(G)		−0·353	−0·373	−0·353
C_4(G)		0·151	0·141	0·147
C_5(G)		−0·228	−0·215	−0·167
C_6(G)		0·447	0·480	0·431
N_7(G)		−0·144	−0·201	−0·232
C_8(G)		0·054	0·050	0·100
N_9(G)		−0·232	−0·286	−0·283
H_1(G)		0·196	0·246	0·204
N_2(NH_2 of G)		−0·352	−0·355	−0·355
H(NH_2 of G)		0·227	0·212	0·226
H′(NH_2 of G)		0·205	0·245	0·207
O_6(G)		−0·338	−0·405	−0·285
H_8(G)		0·123	0·122	0·137
H_9(G)		0·232	0·331	0·365

* When in parentheses, C stands for cytosine and G for guanine.
\+ This atom is H_3 in the Hoogsteen configuration of the G–C base pair.
† W–C – Watson–Crick.

pair in the Watson–Crick as well as the Hoogsteen configurations obtained with the mapping dipole in the molecular planes are presented in figures 1 to 4 respectively. In figures 1 and 2 corresponding to G and C four values of CDA ~2, 5, 7·5 and 10 Å were considered whereas for the G–C base pair in figures 3 and 4 only two values of CDA ~2 and 5 Å were taken. In figures 1 to 4, the minimum potential energy orientations of the mapping dipole and the corresponding potential energies of interactions (in the unit of 0·1 kcal/mole) are given at the chosen points on the CDA surfaces. Potential energies for those points which are not shown in these figures may be obtained approximately by interpolation with the help of the neighbouring values. We observe the following features for the different cases.

3.1 *Cytosine*

The potential map for cytosine is presented in figure 1. We find from this map that for the CDA surface closest to the molecule, the minimum potential energy of interaction

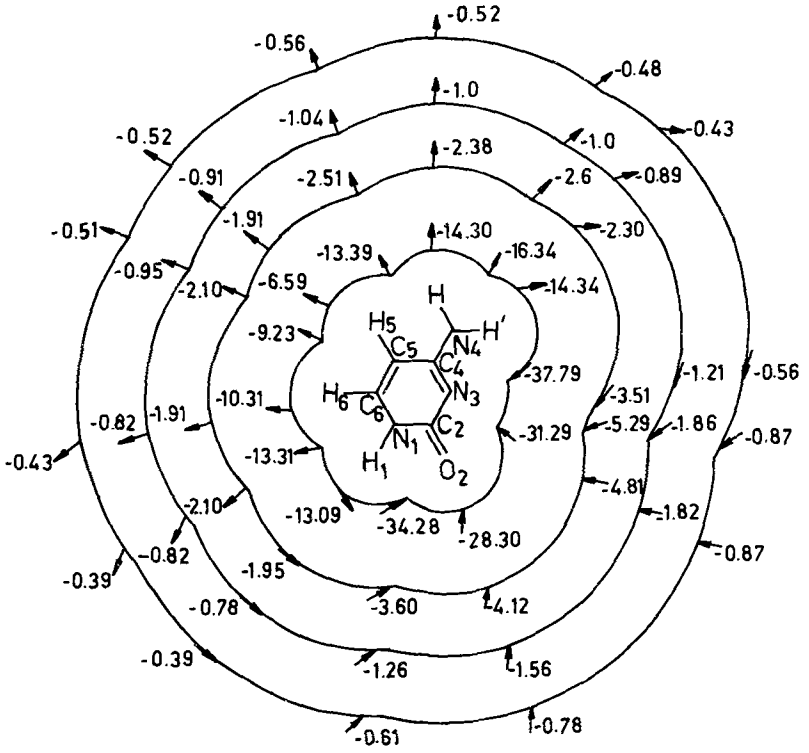

Figure 1. Dipole potential map of cytosine, the dipole being in the molecular plane.

lies near N_3. There is a fairly extended region around O_2 which is also associated with large negative potentials. On going to the surfaces of higher CDA values, magnitudes of the potentials get reduced as expected but broadly the same features as mentioned above are retained. The potential map of figure 1 shows that even if a dipole is brought near the molecule to near the $C_5 = C_6$ bond, it may eventually drift either to near N_3 or near O_2. However, in doing so, it may encounter small barriers of upto 0·2 kcal/mole, in a liquid environment these barriers may be easily overcome due to fluctuations at the usual temperatures such as those in biological systems.

The potential map obtained with the multipole interaction scheme (Mishra and Tewari 1987) showed broadly the same features as those of figure 1. However, from the quantitative point of view, the multipole interaction-based map (Mishra and Tewari 1987) shows the region near N_3 to be much more attractive than that near O_2, in comparison to the present map of figure 1. In the *ab initio* monopole isopotential map of cytosine (Bonaccorsi *et al* 1972; Politzer *et al* 1985), the primary minimum is found to be near N_3, whereas a secondary minimum occurs near O_2. Several experimental studies on the binding of protonating and alkylating agents with cytosine and its derivatives have been performed (Lawley 1961; Brooks and Lawley 1962; Jardetzky *et al* 1963). These include metal ions, e.g. Cu(II), Mn(II) etc. as well as complexes e.g. (glycylglycinato) (cytidine), Cu(II). These studies reveal that N_3 of cytosine is the main site involved in binding (Szalda *et al* 1975) but in certain systems O_2 of cytosine is found to be the binding site (Aoki 1976a). Thus the main features

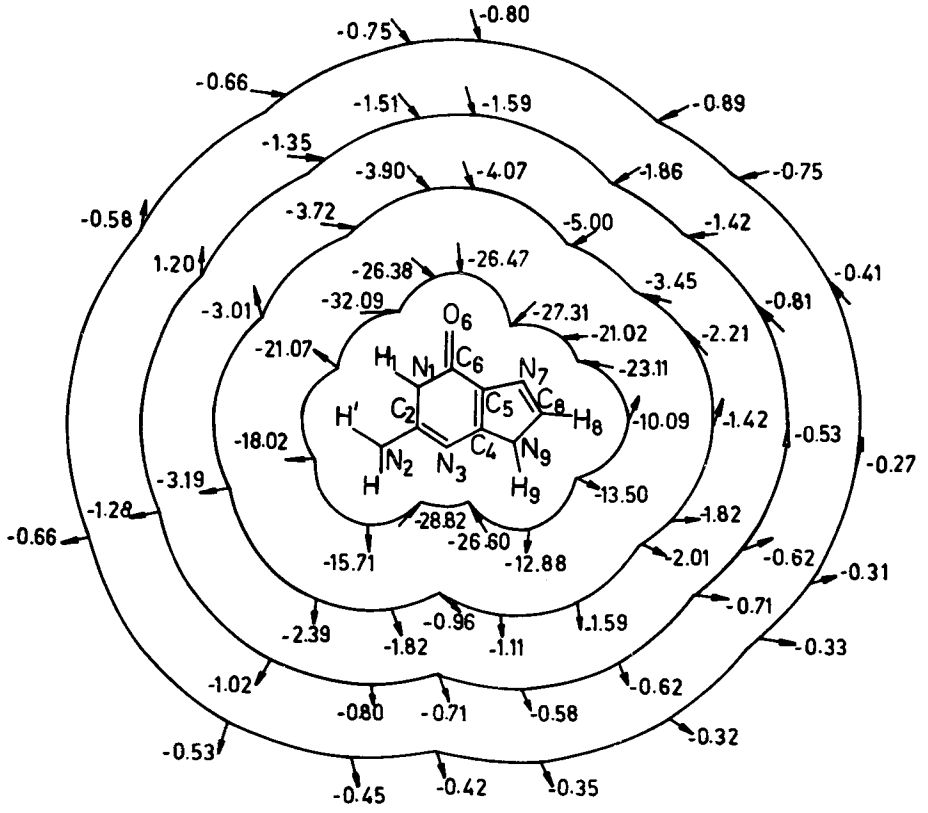

Figure 2. Dipole potential map of guanine, the dipole being in the molecular plane.

of the present results given in figure 1 are in agreement with those of the *ab initio* monopole isopotential maps and experimental observations.

3.2 *Guanine*

The potential map of guanine obtained with the mapping dipole lying in the molecular plane is presented in figure 2. There are three potential minima in this map, one each near N_3, N_7 and O_6. Let us first consider the potentials at the CDA surface closest to the molecule. We find that in this case the minimum near O_6 is the deepest. The attractive region around O_6 is quite extended. The potential minimum region near N_7 has two closely situated components separated by a low barrier; one of them may be considered to be associated with O_6 and N_7 jointly. The minimum potential energies near N_3, O_6 and N_7 are not too widely different but they follow the order $O_6 > N_3 > N_7$. The three minima are obtained in the multipole interaction calculation also but in that case the minimum jointly associated with both O_6 and N_7 is the deepest (Mishra and Tewari 1987). The *ab initio* monopole isopotential map of guanine also exhibits the three minima (Scrocco and Tomasi 1978), the potential energies being in the order $N_7 > O_6 > N_3$.

The mapping dipole directions and potential energies in front of N_3 on the various

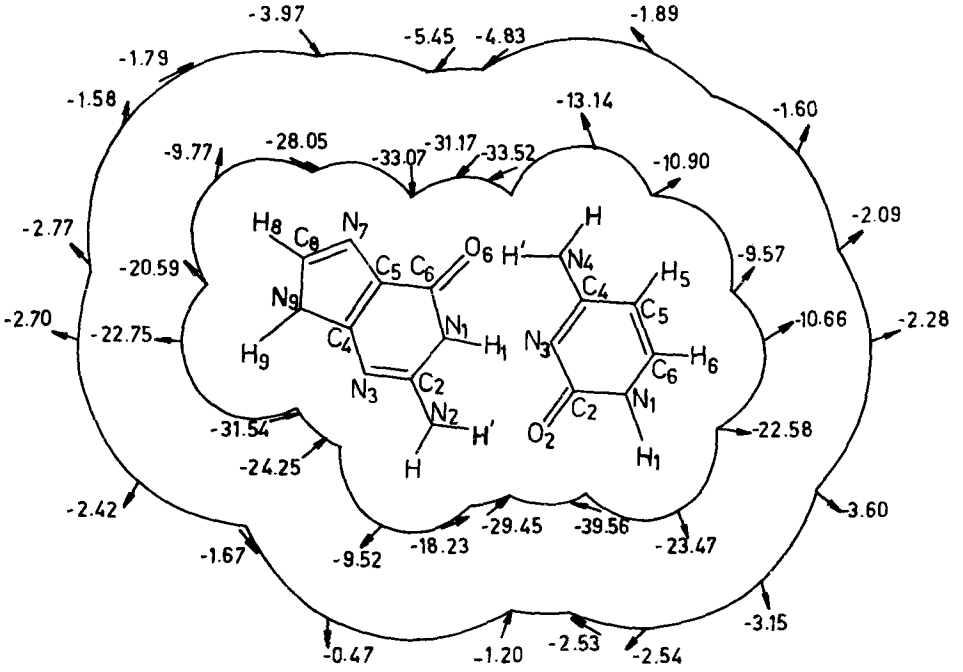

Figure 3. Dipole potential map of the G–C base pair in the Watson–Crick configuration, the dipole being in the molecular plane.

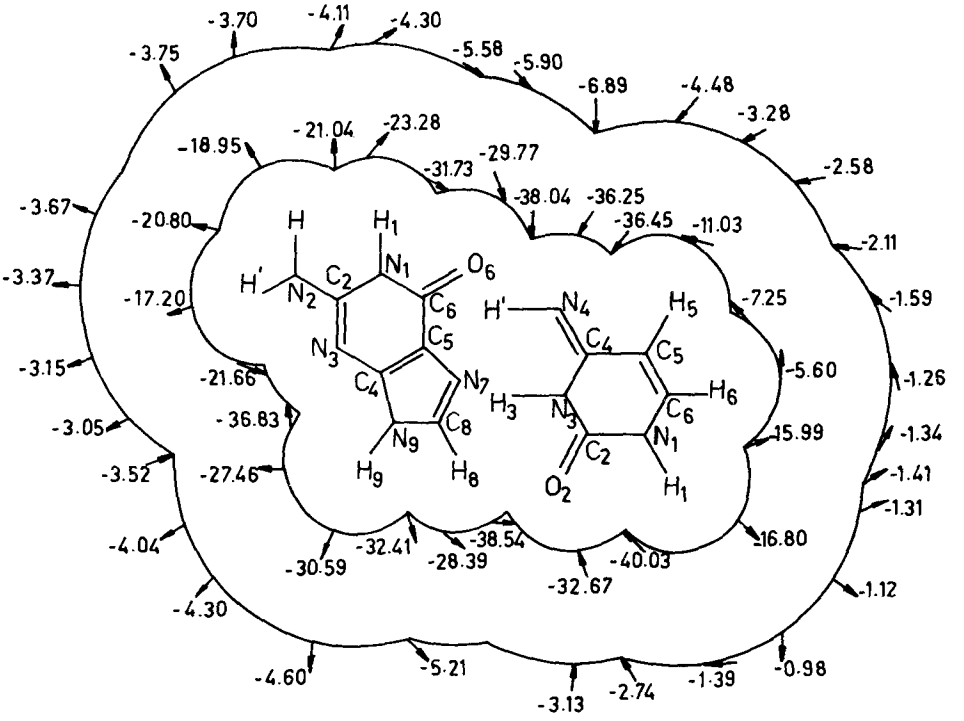

Figure 4. Dipole potential map of the G–C base pair in the Hoogsteen configuration, the dipole being in the molecular plane.

CDA surfaces in figure 2 show that the corresponding attractive potential region is strongly localized in contrast to those associated with O_6 and N_7 which are quite extended. It may be expected that such localized or extended nature of potential wells may respectively reduce or enhance the probability of binding of external species at the corresponding sites of molecules.

Binding of several metal cations and molecular species with guanine has been studied experimentally (Lawley 1957; Pal 1962; Fiskin and Beer 1965; Loveless 1969; Lawley and Thatcher 1970; Gellert and Bau 1975; Goodgame *et al* 1975; Millard *et al* 1975; Aoki 1976b; Dehand and Jordanov 1976; Sletten and Lie 1976). Some of these, for example *cis*-dichlorodiammine platinum(II), act as anticarcinogenic drugs (Goodgame *et al* 1975; Millard *et al* 1975; Dehand and Jordanov 1976). In these studies, N_7 or both O_6 and N_7 of guanine have been found to be involved in binding with the drug (Goodgame *et al* 1975; Millard *et al* 1975; Dehand and Jordanov 1976). These results are in agreement with the potential map features discussed above. Further, it is noted that orientations of the complexes bound to O_6 and N_7 of guanine (Gellert and Bau 1975; Aoki 1976b; Sletten and Lie 1976) are in accordance with the electric field direction in this region of the molecule.

Figure 2 shows an attractive potential region with a shallow minimum near C_8 and N_9 of guanine. Existence of this minimum is seen at all the CDA surfaces of figure 2. It is known that C_8 of guanine is attacked by certain carcinogens and therefore a favourable binding site for such species in the vicinity of C_8 is expected (Kriek 1980). It is to be noted that the monopole isopotential map of guanine (Scrocco and Tomasi 1978; Politzer *et al* 1985) shows this region in the molecular plane to be totally repulsive and hence such a potential map provides no explanation for the observed reactivity of the C_8 site of guanine.

3.3 *G–C base pair*

A study of the potential maps of the G–C base pair is more relevant to biological systems than to isolated bases. Further, it is desirable to investigate how the potential features are modified in going from the isolated bases to the base pair. The potential map of the G–C base pair in the Watson–Crick as well as in Hoogsteen configurations are presented in figures 3 and 4. The following observations are made.

3.3a *Watson–Crick configuration:* The distribution of net charges in the G–C base pair presented in table 1 shows that both the H_9 of guanine and the H_1 of cytosine lose considerable amounts of electronic charge ($\sim 0{\cdot}1e$) while N_7, O_6, N_9 of guanine and O_2, N_3 of cytosine gain the same as compared to the isolated bases.

It is obvious that the N_3 of cytosine in the G–C base pair (Watson–Crick) is blocked for attack by other species in the molecular plane. The potential map of figure 3 exhibits four potential energy minima, the corresponding negative potential energies being in the order $O_2(C) > O_6(G) \simeq N_7(G) > N_3(G)$. The monopole isopotential map of the base pair also shows four potential energy minima (Sokalski 1987). The attractive potential region near the O_2 of cytosine (figure 3) is larger than that in isolated cytosine (figure 1). The attractive potential energy region near the N_3 of guanine persists in having almost the same potential energy values as in isolated guanine, but the potential minimum region gets somewhat shifted from near the NH_2 group to near the H_9 of the molecule. The electric field directions vary rapidly in

going from the region near O_2 of cytosine to that near the NH_2 group of guanine and then to that near the N_3 of guanine in the G–C base pair.

Now let us consider the region of figure 3 where the amino group of cytosine is hydrogen-bonded with the O_6 of guanine. We see that the potential values are modified considerably in this region as compared to the isolated bases. The electric field directions in the G–C base pair in this region in front of the C_4–N_4 bond of cytosine are noticeably modified in comparison to the isolated bases. The region of potential minimum associated with O_6 and N_7 of guanine remains quite extended and deep in the base pair also as found in isolated guanine; this result is in accordance with earlier studies (Lavery and Pullman 1982). Figure 3 shows that negative potential energy values near the C_8–N_9 of guanine and near the N_1–C_6 of cytosine are much larger than in the isolated bases. Therefore, these sites of the bases would have significantly higher reactivities in the base pair than in the isolated bases. The reactivity of C_8 of guanine has been discussed earlier. C_6 of cytosine is known to be involved in certain addition reactions (Budowsky 1976).

3.3b *Hoogsteen configuration*: From the net charge distribution, presented in table 1, it is noted that the H_9 of guanine and the H_3, H_1 and C_5 of cytosine lose electronic charges appreciably while the N_7 of guanine and the H′, N_3, C_4 and O_2 of cytosine gain electronic charges. Surprisingly, the net electronic charge at N_4 of cytosine remains almost the same as in isolated cytosine.

The potential map of the G–C base pair in the Hoogsteen configuration is shown in figure 4. There are four minimum potential energy regions observed near the O_2 and N_4 of cytosine, and the N_3 and O_6 of guanine in this case. The potential energy values are in the order $O_2(C) > N_4(C) \simeq N_3(G) > O_6(G)$. In this case the N_7 of guanine is blocked for attack by other species in the molecular plane. The negative potential energy near the O_2 of cytosine is larger than in isolated cytosine and this region (near O_2 of cytosine) is somewhat more extended in comparison to the G–C base pair in the Watson–Crick configuration (figure 3). The secondary attractive potential energy minimum is associated with N_4 of cytosine and this potential well is localized. The electric field directions are drastically changed in this region as compared to those in the corresponding region of the Watson–Crick configuration (figure 3), as well as in comparison to those near N_4 in isolated cytosine (figure 1). The region near the C_8 and N_9 of guanine is more attractive in this configuration of the G–C base pair than in Watson–Crick configuration.

Acknowledgements

This work was partly supported by project grants from UGC, New Delhi for which PCM expresses his thanks. AK thanks CSIR, New Delhi, for a fellowship.

References

Aoki A 1976a *J. Chem. Soc., Chem. Commun.* 748
Aoki A 1976b *Acta Cryst.* **B32** 1454
Arnott S, Dover S D and Wonacott A J 1969 *Acta Cryst.* **B25** 2192
Bonaccorsi R, Pullman A, Scrocco E and Tomasi J 1972 *Theor. Chim. Acta* **24** 51

Brooks P and Lawley P D 1962 *J. Chem. Soc.* 1348
Budowsky E I 1976 *Prog. Nucleic Acid Res. Mol. Biol.* **16** 125
Dehand J and Jordanov J 1976 *J. Chem. Soc., Chem. Commun.* 598
Dewar M J S and Thiel W 1977 *J. Am. Chem. Soc.* **99** 4899
Fiskin A M and Beer M 1965 *Biochem. J.* 1289
Gellert R W and Bau R 1975 *J. Am. Chem. Soc.* **97** 7379
Goodgame D M L, Jeeves I, Phillips F L and Skapski A C 1975 *Biochim. Biophys. Acta* **378** 153
Hoogsteen K 1963 *Acta Cryst.* **16** 907
Jardetzky O, Pappas P and Wade N G 1963 *J. Am. Chem. Soc.* **85** 1657
Kriek E 1980 in *Carcinogenesis: Fundamental mechanism and environmental effects* (eds) B Pullman, P O P T'So and H Gelboin (Boston: D Reidel) p. 103
Lavery R and Pullman B 1982 *Nucleic Acid Res.* **10** 4383
Lawley P D 1957 *Proc. Chem. Soc. (London)* 290
Lawley P D 1961 *Prog. Nucleic Acid Res. Mol. Biol.* **5** 89
Lawley P D and Thatcher G J 1970 *Biochem. J.* **116** 643
Loveless A 1969 *Nature (London)* **223** 206
Millard M M, Macquet J P and Theophanides T 1975 *Biochem. Biophys. Acta* **402** 166
Mishra P C and Tewari R D 1987 *Int. J. Quantum Chem.* **32** 181
Pal B C 1962 *Biochemistry* **1** 558
Politzer P, Laurence P R and Jayasuriya K 1985 *Environ. Health Perspec.* **61** 191
Price S L, Stone A J and Alderton M 1984 *Mol. Phys.* **52** 987
Pullman A and Pullman B 1981 *Q. Rev. Biophys.* **14** 289
Scrocco E and Tomasi J 1978 *Adv. Quantum Chem.* **11** 115
Sokalski W A 1987 *J. Mol. Struct. (Theochem.)* **150** 235
Sletten E and Lie B 1976 *Acta Cryst.* **B32** 3301
Szalda D J, Marzilli L G and Kistenmacher T J 1975 *Biochem. Biophys. Res. Commun.* **63** 601
Voet D and Rich A 1970 *Prog. Nucleic Acid Res. Mol. Biol.* **10** 183