Does a Stacked DNA Base Pair Hydrate Better than a Hydrogen-Bonded One?: An ab Initio Study

D. Sivanesan,[†] K. Babu,[‡] Shridhar R. Gadre,*,[‡] V. Subramanian,*,[†] and T. Ramasami[†]

Chemical Laboratory, Central Leather Research Institute, Adyar, Chennai, 600 020, India, and Department of Chemistry, University of Pune, Pune 411 007, India

A systematic comparative study has been carried out on the explicit hydration of the cytosine base (C), cytosine–cytosine hydrogen-bonded (H-bonded) base pair (CC) and cytosine–cytosine stacked dimer (C/C). An electrostatics-based model, electrostatic potential for intermolecular complexation (EPIC) has been used to generate initial geometries of the hydrated complexes for ab initio calculations. For this purpose, first the topography of the molecular electrostatic potential (MESP) for C, CC and C/C is explored. Several geometries of the complexes, C...$(H_2O)_n$, CC...$(H_2O)_n$ and C/C...$(H_2O)_n$ ($n = 1$, 2 and 3) are investigated with EPIC followed by constrained optimization at the HF/6-31G** level. Further, single point (SP) energy calculations have been performed at the MP2/6-31G** level to assess the role of electron correlation contribution in the hydration process. This study reveals that C/C stacked dimer hydrates better than the hydrogen-bonded (H-bonded) CC pair. Energetics of these systems show a clear-cut additional stability of 1 to 2.5 kcal/mol at the HF/6-31G** level and more than 3 kcal/mol at the MP2/6-31G**//HF/6-31G** level for C/C...$3H_2O$ complex as compared to CC...$3H_2O$. The present study thus confirms that the stacked base pair hydrates better than the corresponding H-bonded base pair.

Introduction

Nucleic acids provide exciting and difficult challenges for chemists and biochemists alike. These molecules are inherently important, yet highly complex to study. It is well-known that the two main factors responsible for the stabilization of DNA double helix are stacking and hydrogen bonding (H-bonding) interactions between the nucleic acid bases.[1–5] The solvent effect is also important in determining the physicochemical characteristics of nucleic acid bases.[6] A knowledge of energetics of solvated bases is essential for understanding the biological role of DNA. It has been observed experimentally and theoretically that the structure of DNA is highly sensitive to the humidity and adopts various types of conformations, depending upon the relative humidity.[7]

Many experimental studies suggest that in nonpolar solvents and in gas phase, the nucleic acid bases associate mainly by H-bonding whereas in water medium, stacked configurations are preferred.[8,9] The free energy perturbation/molecular dynamics studies have revealed that the stacked complexes are much better solvated by the water than the respective H-bonded pairs.[10] A considerable number of theoretical studies have been published on the subject of base association in the presence of solvent.[11] Quantum chemical ab initio calculations have provided a relatively consistent picture of Watson–Crick and noncanonical base pairs, in both the gas and solvent.[12] It is known that the H-bonding interactions are specific whereas stacking interactions are nonspecific and flexible.[13] The water-assisted interactions (specific solvation) are known to be significant in DNA base pair and stacking interactions.[11,14,15] The location of preferred hydration sites around DNA has also been the subject of a number of experimental and theoretical studies.[7a,b,16]

Recent crystal structure studies[17] emphasize the importance of specific H-bonding between the nucleotide bases and tightly bound water molecules, leading to the stabilization of intramolecular and intermolecular structures of double helix. Guerra et al.[18] have emphasized the importance of the environmental effects on the H-bonds in DNA base pairs. Ts'O et al.[19] have provided evidence that base pairing and stacking interactions in nucleic acids usually occur in competition with solvent interactions with the participating groups.

Cytosine base in particular has been the subject of several experimental and theoretical studies.[20] These studies have revealed that the H-bonded base pair is the most stable in gas phase. Recently, Kurita et al.[20d] have investigated the efficiency of the molecular orbital method using a Slater-type basis-set and nonlocal density functional formalism for describing the stacking energy of cytosine dimer. They found that the energies obtained by employing density functional theory (DFT) are comparable with those from the MP2 methods. Recently, Kobayashi[20f] studied the relative stabilities of the tautomers of cytosine using CCSD(T) method with cc-pvtz(-f) basis. These were found to be in agreement with experimental observations.[20] Sambrano et al.[20g] have carried out the calculations on the various tautomers of the cytosine base in aqueous media. This study demonstrated that the inclusion of electrostatic solvent effect yields an appreciable change in the stability order of cytosine tautomers. Morpurgo et al.[20h] have found that the inclusion of polarized and diffuse functions as well as incorporation of electron correlation are necessary in the description of cytosine molecule due to the nonplanarity of the amino group.

The ab initio calculations have been performed on protonated cytosine stacked and H-bonded pairs in a polar solvent.[21] Sponer et al.[22] evaluated the interaction energies of stacked cytosine dimer by the ab initio methods. Subramanian et al.[23] have studied the effect of solvent on the stacking properties of

---

† Central Leather Research Institute.
‡ University of Pune.

cytosine dimer using Onsager SCRF model. Recently, Alemán[16b] showed that cytosine base is more hydrophilic than thymine. The free energy of solvation for cytosine has been computed by different groups employing a variety of methods.[24] A systematic study on the hydration patterns of bases has also been carried out.[7c,25] Experimental evidence[7c,26] has provided the following order of hydrophilicity: guanine > cytosine > adenine > uracil > thymine.

The ab initio quantum chemical studies devoted to solvation of stacked bases are rather rare compared to those on H-bonded base pairs. Possible reasons are: the complex nature of the potential energy surface of stacked pairs (as compared to that of H-bonded pairs) and the predominant role-played by electron correlation in the stacked pairs. The proper inclusion of solvent effect into the calculations is perhaps the most important though exceptionally difficult task of contemporary quantum chemistry. There are several self-consistent reaction field (SCRF) methods available for modeling the solvent effects (nonspecific solvation). However, they treat the solvent as a continuum characterized by its dielectric constant[27] and the electronic effects associated with specific solute−solvent interactions are not explicitly dealt with. A theoretical study on cytosine tautomers in water medium has been carried out using continuum models.[28] This study brings out the importance of electrostatics and the influence of the environment conditions. Claverie et al.[29] proposed the use of combined discrete/SCRF methods to predict energies and electronic properties in solvent environment for the nucleic acid bases. Alemán[16b,c,30] has made investigations on the hydration of cytosine base using both combined discrete/ SCRF and SCRF models. These studies prove that the combined discrete/SCRF methods provide more useful information for understanding the properties of bases than either discrete or SCRF methods.

The electrostatic interactions determine the strength of H-bonding between the nucleic acid bases, hydration of biomolecules and binding of polyvalent cation to DNA bases.[22a,31] However, an analysis of the binding sites in terms of complete electrostatic description of bases and base pairs is conspicuous by its absence from the earlier literature. Several models have been developed using complementary nature of molecular electrostatic potential to probe guest−host interaction and molecular recognition.[32a] Gadre and co-workers have recently developed a novel docking model, viz. electrostatic potential for intermolecular complexation (EPIC).[32b] The appropriateness of this model has recently been demonstrated for exploring the interactions in DNA base pairs and trimers.[33] Importance of electrostatics in molecular biology has been highlighted by Warshel and Russel.[34] Since base pairing and stacking involve different kind of interactions, understanding the corresponding solvation patterns posseses a challenging problem.

The present study is aimed at answering the following questions: (i) Which hydrates better: H-bonded pair or stacked pair? (ii) What is the role of electron correlation effect on the hydrated complexes? A comparative study on the specific solvation effects on nucleic acid base, base pair and stacked pair has not yet been carried out either experimentally or theoretically to the best of the authors' knowledge. In view of the importance of this problem, we have undertaken a case study of explicit hydration on cytosine base, CC base pair and C/C stacked dimer using the EPIC model, followed by an ab initio level treatment.

Computational Methodology

In the present investigation, specific solvation effects are taken into account by the explicit inclusion of water molecules in the first coordination sphere around the cytosine molecule using EPIC. This model has been used to generate guess geometries of the solute...nH$_2$O complexes for a subsequent ab initio investigation. EPIC uses the rich topographical features of the molecular electrostatic potential (MESP). The MESP $V(\mathbf{r})$, at a point $\mathbf{r}$ due to nuclear charges $\{Z_A\}$ located at $\{\mathbf{R_A}\}$ and the electronic charge density $\rho(\mathbf{r})$ is defined as

$$V(\mathbf{r}) = \sum_{A}^{N} \frac{Z_A}{|\mathbf{r} - \mathbf{R_A}|} - \int \frac{\rho(\mathbf{r}') \, d^3\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}$$

$V(\mathbf{r})$ can assume positive as well as negative values and can provide useful information regarding electron-rich sites.[35] The first and second terms in the above equation refer to the bare nuclear potential and electronic contributions to the MESP, respectively. The gradient vector field characterizes topological features of a function of many variables. Topological analysis of $V(\mathbf{r})$ involves the identification and characterization of the critical points (CPs), the points at which $\nabla V(\mathbf{r}) = 0$.[36] The nondegenerate CPs of $V(\mathbf{r})$ are characterized as one of the four possible types: (3,+3), (3,−3), (3,+1), (3,−1). The CP corresponding to (3,+3) refers to a minimum, (3, −3) stands for a maximum and the (3,+1) and (3,−1) CPs represent saddle points. Such a topographical analysis has not yet been applied in the study of DNA base-pair−water interactions, though a glancing reference has been made.[33]

Thus, in the present study, the CPs are located and characterized[37] for the C, CC H-bonded pair and stacked C/C dimer. The MESP-driven atom centered point charges are obtained for C, CC, C/C and water molecules using the program GRID.[38] The MESP isosurfaces for C, CC and C/C displayed in Figure 1 have been generated with the help of the graphics program UNIVIS.[39] Exploiting the complementary nature of MESP value at the CPs, several guess geometries for the hydrated complexes have been generated. Subsequently, the electrostatic interaction energy of the intermolecular complexes is minimized according to the following equation using EPIC.[32b]

$$E = \frac{1}{2} \sum_{A,B}^{N} V_{A,i} q_{B,i}$$

During the interaction energy minimization, the internal geometries of cytosine and water molecules are kept fixed. The water molecule is moved around the cytosine by giving all the possible rotational and translational degrees of freedom. The molecules are prevented from getting too close to each other by defining definite van der Waals envelopes for A and B. The same procedure has been applied in evaluation of interaction energies for the other complexes, involving more than two molecules.

The EPIC-generated geometries are used as starting point for further ab initio optimization. The ab initio optimizations are done at the HF/6-31G** level using both G94[40] and GAMESS[41] package. The single point calculations have also been made on the hydrated cytosine base pair and stacked dimer at the MP2/ 6-31G** level to include the correlation contributions. The initial geometries for the C and CC base pair optimized at the HF/6-31G** level. Since C/C stacked dimer is not a global minimum in the PES at the HF level, the geometry obtained based on empirical potential optimization has been used.[21b] The geometry of C/C stacked dimer is kept fixed during optimization of the complexes.

The HF and MP2 level interaction energies are evaluated as the difference between energy of the complex and the sum of
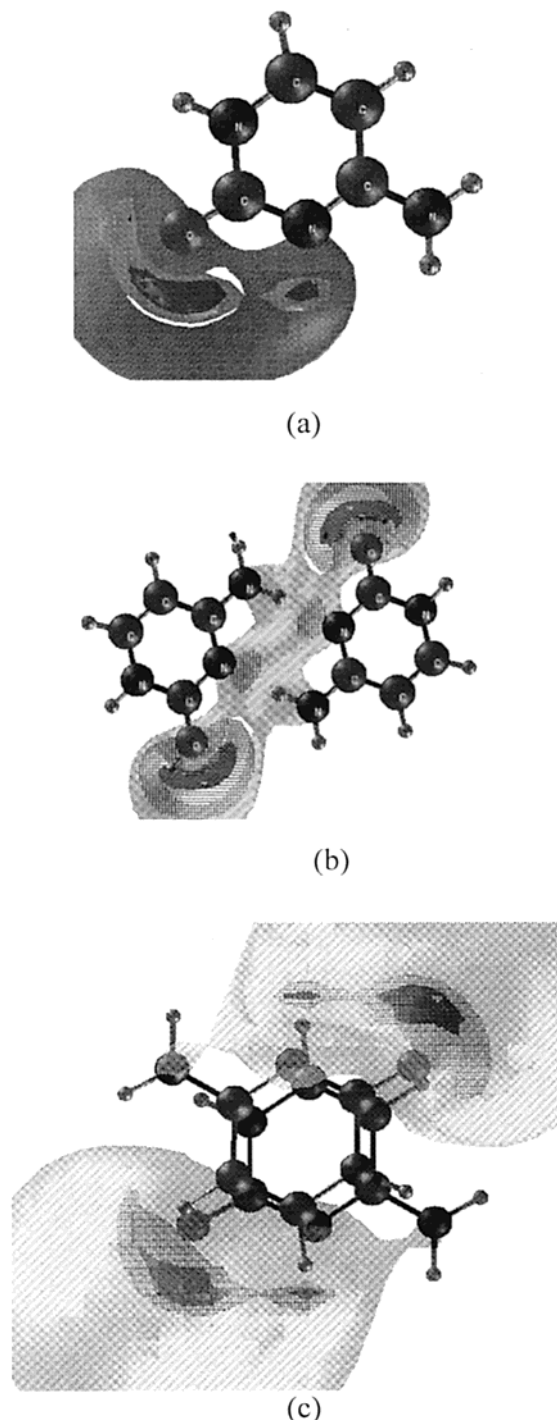
Hydration in a Stacked DNA Base Pair

J. Phys. Chem. A, Vol. 104, No. 46, 2000 **10889**

Figure 1. MESP isosurfaces of values (going inward from outside): (a) −28.24, −61.50, and −69.03 kcal mol^{-1} for cytosine, (b) −62.75, −34.51, and −21.96 kcal mol^{-1} for CC H-bonded pair, and (c) −65.26, −56.48, and −9.41 kcal mol^{-1} for C/C dimer. See text for details.

monomer energies.

$$\Delta E_{int} = E(C...nH_2O) - \{E(C) + nE(H_2O)\}$$

All calculations have been carried out on the PARAM 10000 supercomputer and SGI Origin 200 workstation. Attempt has also been made on some model complexes to assess how inclusion of diffuse functions and basis set superposition correction (BSSE) would influence the rank order of ΔE_{int} predicted by our calculations.

Results and Discussion

The topographical features as well as the pictorial representation of MESP in the form of isosurfaces are very useful for exploring molecular recognition phenomena and host−guest interactions. In the present case study, the topographical features provide potential sites on C, CC pair and C/C stacked dimer where the water molecules can be anchored. The MESP isosurfaces for isolated C, CC pair and C/C stacked dimer are depicted in Figure 1. The negative valued isosurface of MESP indicates the electrophilic sites on these molecules.

Interaction Energy and Geometry of Hydrated Cytosine Base. From Figure 1, it is evident that the charge concentration sites lie in the molecular plane of cytosine, which are primarily responsible for the binding with the hydrogens of water molecules. The most negative MESP minimum for the cytosine base is −83.2 kcal/mol. The MESP features reveal that ring N1, N3, O(C), and NH2−N(H) groups are able to make specific H-bonding interactions with water molecules. From Figure 1a, it can be predicted that the N3 and O(C) atoms would act as the primary binding sites for hydrogens of water complemeted by the hydrogens of cytosine binding with oxygens of corresponding water molecule. The (3,−3) CPs represent the potential sites for H-bonding whose strength is determined by the value of MESP at that CP.

Several guess geometries for the complex C...H$_2$O were generated by employing MESP complementarity of the interacting moieties. This involves positioning H of one molecule close to the most negative MESP CP of the other molecule and *vice-versa* leading to a lock-and-key type of MESP arrangement. These trial geometries are then docked with EPIC to minimize the interaction energy. The EPIC-minimized geometries of cytosine water complexes are then used as starting geometries for further ab initio optimization. Three distinct geometries for C...H$_2$O complexes, each having two hydrogen bonds, have been obtained at the HF/6-31G** level. Similar exercise performed on C...2H$_2$O and C...3H$_2$O yielded four and five distinct geometries, respectively at the HF/6-31G** level of calculation. Only the energetically favored geometries of these complexes are depicted in Figure 2.

The HF interaction energies of C...nH$_2$O (n = 1, 2, 3) complexes are shown in Table 1. The incremental interaction energies for the addition of second and third water molecules are −11.8 and −10.2 kcal/mol, respectively. It is interesting to note that all the geometries corresponding to C...3(H$_2$O) molecule have comparable energies due to rather similar H-bonding patterns. The geometry corresponding to C1W1 (cf. Figure 2a) was recently optimized at HF/6-31G(d,p) by Alemán[16b] and at the MP2/6-31G(d,p) level by Gorb et al.[42] Their geometrical parameters along with the present results are given in Table 2. As expected, the MP2 level bond lengths are consistently lower than the HF counterparts due to the electron correlation. The comparison of geometrical parameters at HF and MP2 levels indicates that the HF method is reasonable enough to carry out geometrical optimization of large H-bonded complexes. The donor acceptor distances predicted at the HF level are larger about 0.03−0.09 Å as compared to their MP2 counterparts. Further, it has been found that the intermolecular angles in the C...H$_2$O compex are very similar at both HF and MP2 levels.

Comparison of Interaction Energies of Hydrated H-Bonded and Stacked Cytosine Bases. It may be noticed from Figure 1 that the MESP of C/C stacked dimer is more negative as compared to that of the CC base pair. The most negative MESP minimum for C/C is −80.2 kJ/mol whereas it is −76.3
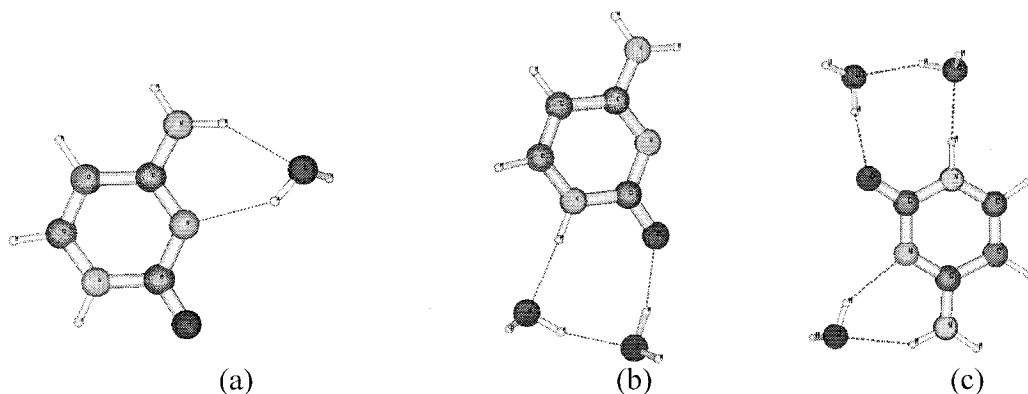
Figure 2. Ab initio optimized structures of (a) C...H$_2$O, (b) C...2H$_2$O, and (c) C...3H$_2$O. See text and Table 1 for details and the corresponding interaction energy values obtained using the HF/6-31G** level.

TABLE 1: Interaction Energies (kcal mol^{-1}) of Cytosine–Water Complexes at the HF/6-31G Level Using ab Initio and EPIC Methods**

molecular geometry	ΔE_{EPIC}	ΔE_{HF}
C1W1	−11.4	−11.7
C2W1	−11.4	−10.9
C3W1	−8.4	−8.3
C1W2	−20.7	−23.5
C2W2	−18.6	−22.3
C3W2	−20.4	−19.7
C4W2	−24.1	−22.3
C1W3	−28.8	−33.6
C2W3	−31.1	−33.7
C3W3	−33.3	−33.3
C4W3	−29.8	−33.5
C5W3	−31.7	−33.7

W1, W2 and W3 stand for one, two and three water complexes, respectively.

TABLE 2: Selected Intermolecular Geometrical Parameters (Å) for Cytosine–H$_2$O Complex C1W1 (see Figure 2a)

geometrical params	HF/6-31G** [a]	HF/6-31G(d)[b]	MP2/6-31G(d)[c]
H$_9$...O$_w$	2.048	2.051	1.908
O$_1$...H$_w$	1.983	1.979	1.888
N$_2$...O$_w$	2.921	2.938	2.827
O$_1$...O$_w$	2.843	2.836	2.785

[a] Present values. [b] Taken from ref 16b. [c] Taken from ref 42.

kJ/mol in the case of CC pair. Hence we may expect stronger electroststic interactions for C/C with water molecules in contrast to the CC pair. The initial geometries of hydrated CC and C/C complexes obtained by EPIC were further optimized at the HF/6-31G** level. The interaction energies calculated at HF/6-31G** and MP2/6-31G**//HF/6-31G** levels for CC and C/C pairs are given in Tables 3 and 4. Selected geometries for CC...nH$_2$O and C/C...nH$_2$O (n = 1, 2, 3) complexes are displayed in Figures 3 and 4, respectively. Both the geometries of CC...H$_2$O have the same energy at the MP2/6-31G** level, whereas at the HF/6-31G** level, CCP2W1 is lower in energy than CCP1W1 geometry (cf. Figure 3a). The correlation energy contribution to the interaction energy for the CCP1W1 complex is 3.2 kcal/mol, which is higher than that for CCP2W1 complex. Although the electron correlation is important in the stabilization of CC pair, the dominant interaction is electrostatic in nature. The position of water molecules in the hydration process would significantly influence these two interactions. In the case of CCP1W1, the water molecule is hydrogen bonded to carbonyl oxygen and proton of N1 of one cytosine. In the case of CCP2W1 complex, the water molecule is present between the paired cytosines. This intervening water molecule modulates

TABLE 3: Hartree–Fock (ΔE_{HF}) and EPIC (ΔE_{EPIC}) Optimized Interaction Energies (kcal mol^{-1}) at the HF/6-31G Level for CC...nH$_2$O Complexes[a]**

molecular geometry	ΔE_{EPIC}	ΔE_{HF}	ΔE_{MP2}(SP)
CCP1W1	−11.8	−11.4	−14.6
CCP2W1	−10.3	−12.1	−14.6
CCP1W2	−29.3	−25.1	−30.7
CCP2W2	−18.6	−22.3	−28.0
CCP3W2	−18.5	−22.8	−28.1
CCP4W2	−20.3	−23.6	−30.6
CCP5W2	−24.3	−22.9	−29.4
CCP6W2	−21.3	−23.2	−30.1
CCP7W2	−20.1	−23.1	−29.8
CCP1W3	−30.4	−34.3	−42.6
CCP2W3	−29.7	−35.6	−44.7
CCP3W3	−33.1	−34.5	−43.7
CCP4W3	−35.5	−34.9	−43.5
CCP5W3	−32.6	−34.4	−43.1
CCP6W3	−33.2	−34.2	−42.4
CCP7W3	−33.2	−34.0	−41.7
CCP8W3	−34.3	−35.5	−43.6
CCP9W3	−35.2	−33.4	−43.7

[a] ΔE_{MP2}(SP) single point interaction energies at the MP2/6-31G** level (see text for details).

TABLE 4: Hartree–Fock (ΔE_{HF}) and EPIC (ΔE_{EPIC}) Optimized Interaction Energies (kcal mol^{-1}) at the HF/6-31G Level for CC...nH$_2$O Complexes[a]**

molecular geometry	ΔE_{EPIC}	ΔE_{HF}	ΔE_{MP2}(SP)
CCS1W1	−17.2	−12.6	−15.7
CCS2W1	−11.9	−11.7	−14.7
CCS1W2	−24.7	−25.1	−31.7
CCS2W2	−25.3	−24.9	−30.5
CCS3W2	−27.2	−26.5	−32.1
CCS4W2	−24.1	−23.9	−29.5
CCS5W2	−22.7	−23.8	−29.4
CCS1W3	−32.6	−33.8	−41.8
CCS2W3	−31.8	−36.3	−45.7
CCS3W3	−33.5	−35.5	−45.2
CCS4W3	−36.2	−37.6	−46.1
CCS5W3	−37.5	−36.2	−44.8
CCS6W3	−35.5	−36.5	−46.0
CCS7W3	−35.2	−34.2	−45.2

[a] ΔE_{MP2}(SP) single-point interaction energies at the MP2/6-31G** level (see text for details).

both the electrostatic and dispersion interactions and hence the difference in the energetics of the hydrated CC pair.

The CCP1W2 emerges as the energetically most favored geometry among doubly hydrated CC complexes. This is because both the water molecules are bound to the most negative valued MESP CPs of CC pair. In the case of interaction of three water molecules with CC, it may be noticed that geometrically
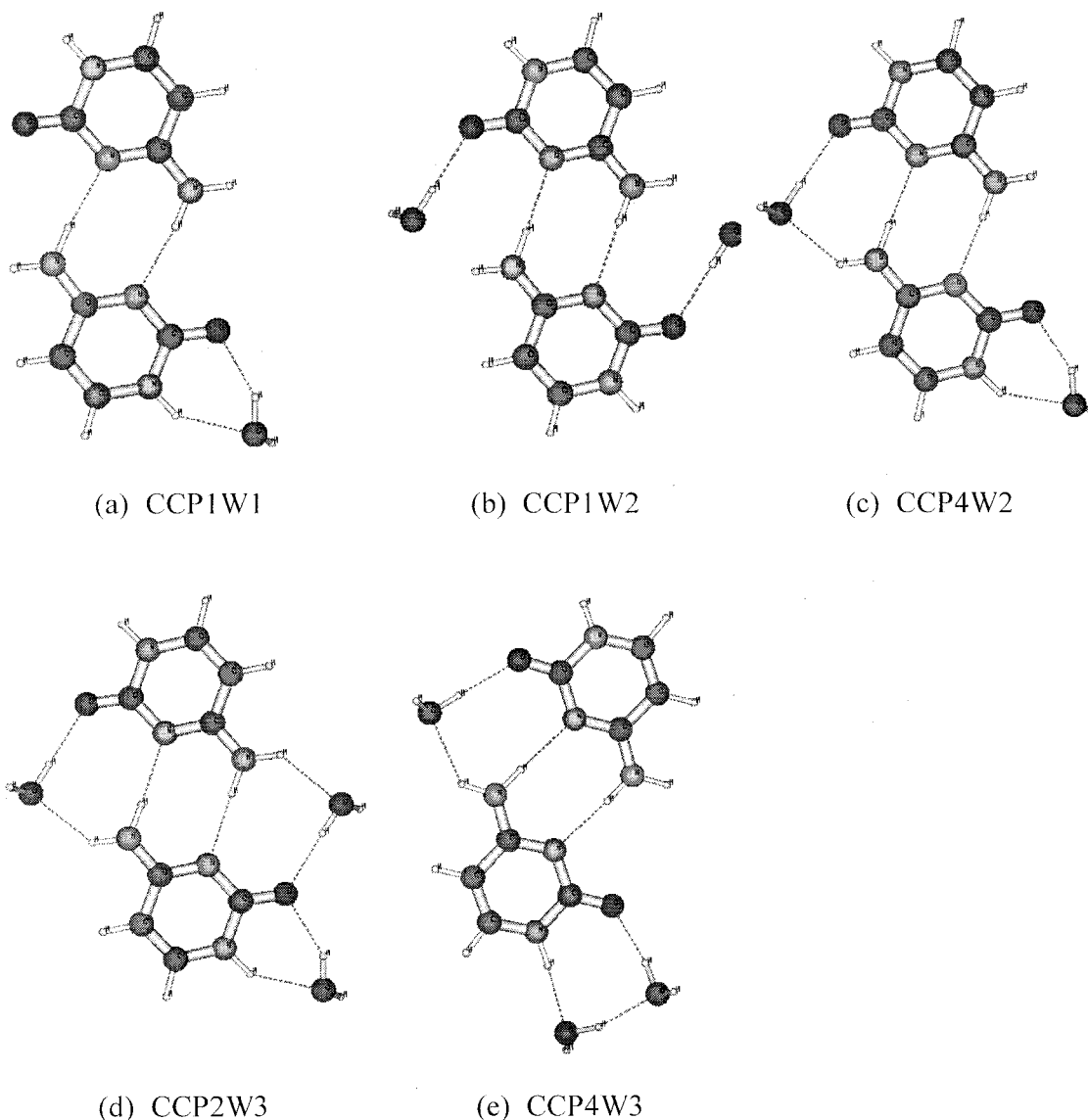
Hydration in a Stacked DNA Base Pair

J. Phys. Chem. A, Vol. 104, No. 46, 2000 **10891**

(a) CCP1W1   (b) CCP1W2   (c) CCP4W2

(d) CCP2W3   (e) CCP4W3

Figure 3. Ab initio optimized structures of (a) CC...H_2O, (b−c) CC...$2H_2O$, and (d−e) CC...$3H_2O$. See text and Table 3 for details and the corresponding interaction energy values obtained using the HF/6-31G** level.

similar complexes like CCP3W3 and CCP4W3 have similar energies. It is observed that the CCP2W3 (cf. Figure 3d) has the numerically highest interaction energy as compared to other complexes. This may be due to more number of H-bonds, which are bonded in the electron rich sites contributing to additional strength.

The geometries of C/C...H_2O, C/C...$2H_2O$ and C/C...$3H_2O$ are displayed in Figure 4. As seen from Table 4, the CCS1W1 (Figure 4a) is energetically more favored than the CCS2W1 complex. This may be due to complementary nature of the electrostatic interaction involved in the stabilization of the complexes. The geometries CCS3W2 (cf. Figure 4c) and CCS4W3 (cf. Figure 4e) emerge as the most favored ones for C/C...$2H_2O$ and C/C... $3H_2O$ at HF and MP2 levels. It can be brought out from Tables 3 and 4, that the interaction energies for one, two and three water molecules with CC and C/C complexes follow an additive behavior. It should be noted that in the hydrated complexes of C/C, water molecules bind to both the cytosines highlighting that cooperativity plays a crucial role in the hydration of stacked pairs.

The interaction energy of hydrated C/C stacked dimer (constrained optimized: coordinates of C/C were kept frozen) is numerically higher by about 1 to 2.5 kcal/mol at HF compared

to hydrated CC base pair (fully optimized). The complete optimization on hydrated C/C would lead to a further stabilization increasing the above difference. The MP2/6-31G** constrained optimization (the coordinates of CC and C/C were kept frozen) have been performed on CCP8W3 geometry of CC...$3H_2O$ pair and CCS4W3 geometry of C/C...$3H_2O$ which are the most stable geometries at the HF level. After a reasonable optimization the interaction energies turned out to be −44.9 kcal/mol and −48.6 kcal/mol for CCP8W3 and CCS4W3 complexes, respectively with a difference of 3.7 kcal/mol. A look at the structure CCP8W3 reveals that most of the primary sites are hydrated and hence increment in the interaction energy for a further addition of three water molecules will not be numerically greater than 44.9 kcal/mol. On the other hand, the increment in the interaction energy for adding three more water molecules to the corresponding best hydrated stacked pair structure (viz. CCP4W3) is expected to be comparable to the ΔE of triply hydrated C/C (viz. −47.8 kcal/mol). Hence, the difference (at the MP2 level) between the hexahydrated CC and C/C is expected to be at least 7 kcal/mol in favor of the later. This difference guarantees that the stacked dimer is better hydrated than H-bonded base pair. It is also apparent that the higher interaction energies predicted for the C/C stacked water
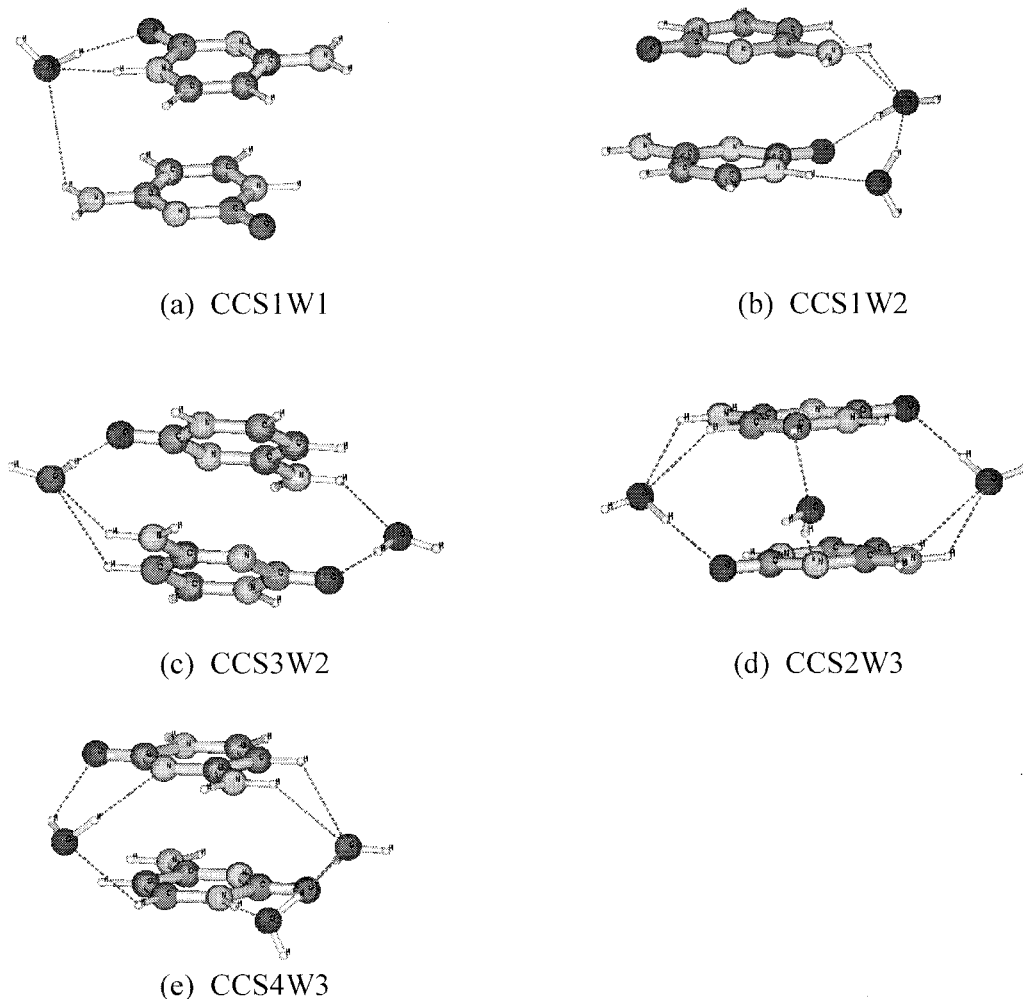
(a) CCS1W1   (b) CCS1W2

(c) CCS3W2   (d) CCS2W3

(e) CCS4W3

Figure 4. Ab initio optimized structures of (a) C/C...H_2O, (b−c) C/C...$2H_2O$, and (d−e) C/C...$3H_2O$. See text and Table 4 for details and the corresponding interaction energy values obtained using the HF/6-31G** level.

complexes are due to the cooperative nature of the interaction. That is, in the case of C/C...nH_2O complexes, both cytosine are involved in H-bonding interaction with water molecules, whereas in hydrated CC geometries, both the cytosine are H-bonded with each other, and hence they do not compete for H-bonding with water molecules. The presence of more binding sites for water molecules in stacked C/C dimer leads to C/C...nH_2O having more number of hydrogen bonds than CC...nH_2O complexes. Further, it is also clear from the MESP isosurface that C/C stacked dimer (cf. Figure 1c) has more negative MESP region than CC H-bonded pair (cf. Figure 1b). Our results agree qualitatively well with the experimental evidence, which shows that base stacked pair solvates better than H-bonded one.[43] From Tables 3 and 4 it can be inferred that the correlation contribution to the interaction energy is uniformly higher for all the C/C...$3H_2O$ geometries compared to the CC...$3H_2O$ counterparts. This may be due to the fact that increasing the number of interacting atoms increases the dispersion contribution relative to the role-played by the electrostatic interactions. An unconstrained optimization (upto a gradient of 0.001) of CC2W3 at the MP2/6-31G** level leads to a considerable pyramidalization of the −NH_2 group with the amino hydrogens bending away from the other cytosine molecule. The interaction energy also increases to −52.0 kcal/ mol. Since all the geometries (including C/C) studied in this investigation are not optimized at the MP2 level the changes in the interaction energy trends brought out by pyramidalization

is not appropriate to compare here. However, the basic binding patterns of water molecules CC2W3 remain the same.

To assess the influence of inclusion of diffuse functions in the basis set, HF/6-31++G(d,p) calculations have been carried out on cytosine−water complexes (C1W1 and C2W2). The interaction energies of these complexes at this level turn out to be −10.2 and −9.4 kcal/mol, respectively which are typically 87% of their HF/6-31G(d,p) counterparts viz. −11.7 and −10.9 kcal/mol. Although the inclusion of diffuse functions does alter the value of interaction energy, there may not be a change in the trend predicted by the HF/6-31G(d,p) level of theory. The calculations have also been performed on CCS1W3 and CCS4W3 to study the importance of basis set superposition error (BSSE) in the prediction of interaction energy. The BSSE-corrected interaction energies of CCS1W3 and CCS4W3 are −27.2 and −30.9 kcal/mol respectively as against uncorrected values of −33.8 and −37.6 kcal/mol at the HF/6-31G(d,p) level of computation. The scaling factors ($\Delta E_{corrected}/\Delta E_{uncorrected}$) turn out to be 0.80 and 0.82 for CCS1W3 and CCS4W3 complexes. To estimate the BSSE correction at the MP2 level, calculations were performed on the same complexes. The respective scaling factors are 0.67 and 0.72. It can be noted that the incorporation of BSSE correction in the prediction of interaction energies does bring down the numerical value of the interaction energy but does not seem to alter the rank order predicted by HF/6-31G-(d,p) as well as the MP2/6-31G** levels of calculations. These

Hydration in a Stacked DNA Base Pair

J. Phys. Chem. A, Vol. 104, No. 46, 2000 **10893**

findings are also in accordance with previous investigation on hydration of uracil carried out by Gadre et al.[44]

It is gratifying to note that electrostatics-based EPIC model is able to clearly distinguish various sites toward their interactions with water molecules. The number of water molecules included in this study is limited to three. These three water molecules in the first hydration shell are expected to bring out the differences in the interaction energies in the CC and C/C pairs, since they are most tightly bound to the DNA base. The addition of more number of water molecules would be computationally more involved and would require substantially large computational power.

Concluding Remarks

This study reveals that there are specific patterns of water clusters, which add to the stability of the complex. These identified patterns can be used to construct the geometries of complexes with large number of water molecules. The topography-based approach seems to offer insight into the primary interaction patterns of base–water formation. The EPIC interaction energies are quite comparable to their ab initio counterparts. EPIC also provides good starting geometries for ab initio optimization and hence reduces the CPU time to a large extent.

The interaction energy predicted for the hydrated C/C dimer is numerically higher than that predicted for hydrated CC pair and the difference in energy ranges from 1 to 2.5 kcal/mol at the HF level. This is consistently seen at all the levels of calculation reported here. The higher interaction energy predicted in the case of hydrated C/C dimer may be due to the cooperative binding of water molecules to both the cytosines. It is shown from the MP2 constrained optimization calculations that the enhancement in interaction energy for C/C...$3H_2O$ complex is more than 3.7 kcal/mol when compared to CC...$3H_2O$ complex. This reveals that although the differnce in interaction energy between CC and C/C is small, it may be magnified further upon full MP2 optimization. Hence, it is apt to state on the basis of the present findings that the stacked pair hydrates better than base pair, which confirms the experimental evidence.[43]

It is observed that even though the correlation contribution to the interaction energy is higher (0.6 to 2.6 kcal/mol) in the case of C/C...nH_2O as compared to CC...nH_2O, electrostatics seem to be the deciding factor for trends in interaction energies shown by both these complexes. It has also been observed that the inclusion of diffuse funcions in the basis set does not seem to alter the trends predicted by our calculations. Incorporation of the BSSE correction scales down the interaction energies without altering the rank order observed in the present investigation. We believe that the patterns observed in the present study could be gainfully employed for further probing of complete first and second hydration shells of DNA bases and base pairs. Electrostatics indeed seems to play a vital role in such hydration processes.

References and Notes

(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(2) Rajski, S. R.; Williams, R. M. *Chem. Rev.* **1998**, *98*, 2723.

(3) Frier, S. M.; Alkema, D.; Sinclair, A.; Neilson, T.; Turner, D. H. *Biochemistry* **1985**, *24*, 4533.

(4) Frier, C.; Kierzek, R.; Caruthers, M. H.; Neilson, T. *Biochemistry* **1986**, *25*, 3209.

(5) Frier, D.; Sugimoto, S. M.; Sinclair, A.; Alkema, D.; Noeilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H. *Biochemistry* **1986**, *25*, 3214.

(6) Orozco, M.; Luque, F. J. *Biopolymers* **1993**, *33*, 1851.

(7) (a) Franklin, R. E.; Gosling, R. G. *Acta Crystallogr.* **1953**, *6*, 673. (b) Poltev, V. I.; Teplukhin, A. V.; Malenkov, G. G. *Biochemistry* **1992**, *42*, 1499. (c) Schneider, B.; Cohen, D.; Berman, H. M. *Biopolymers* **1992**, *32*, 725.

(8) (a) Binford, J. S.; Holloway, D. M. *J. Mol. Biol.* **1968**, *31*, 91. (b) Yanson, B.; Teplitsky, I. K.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.

(9) (a) Kyogoku, Y.; Lord, R. C.; Rich, A. *J. Am. Chem. Soc.* **1967**, *89*, 496. (b) Schweizer, D M. P.; Broom, A. D.; Ts'O, P. O. P.; Hollis D. P. *J. Am. Chem. Soc.* **1968**, *90*, 1042.

(10) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734.

(11) (a) Langlet, J.; Claverie, P.; Caron, F.; Boenve, J. C. *Int. J. Quantum Chem.* **1981**, *19*, 229. (b) Poltev, V I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3* 739. (c) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302. (d) Pohorille, A.; Burt, S. K.; MacKlroy, R. D. *J. Am. Chem. Soc.* **1984**, *106*, 402. (e) Weiner, S. J.; Kollman, P. A.; Nguygen, D.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

(12) (a) Sivanesan, D.; Subramanian, V.; Nair, B. U.; Ramasami, T. *Indian J. Chem.* **2000**, *39A*, 132. (b) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965 and references therein. (c) Brameld, K.; Dasgupta, S.; Goddard, W. A. *J. Phys. Chem. B* **1997**, *101*, 4851. (d) Sponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263.

(13) (a) Spöner, J.; Kypr, J. *J. Mol. Biol.* **1991**, *217*, 201. (b) Spöner, J.; Kypr, J. In *Theoretical Biochemistry & Molecular Biophysics*; Beveridge, D. L., Lavery, R., Eds.; Adenine Press: Guilderland, NY, 1990; p 271. (c) Sponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 27.

(14) (a) Saenger, W. *The Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983. (b) Spöner, J.; Hobza, P.; Leszczyski, J.; In *Computational Chemistry, Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific Publisher: Singapore, 1996; p 185. (c). Kwiatkowski, J. S.; Pearson, W. B. In *Theoretical Biochemistry and Molecular Biophysiscs*; Beveridge, D. L., Lavery, R., Eds.; Adenine Press: Guilderland, NY, 1990; Vol. 1, p 153.

(15) Rapport, Z., Ed. *The Chemistry of Enols*; Wiley: New York, 1990.

(16) (a) Wolf, B.; Hanlon, S. *Biochemistry* **1973**, *14*, 1661. (b) Alemán, C. *Chem. Phys. Lett.* **1999**, *302*, 461. (c) Alemán, C. *Chem. Phys.* **1999**, *244*, 151.

(17) (a) Holbrook, S. R.; Cheng, C.; Tinoco, I.; Kim, S.H. *Nature* **1991**, *353*, 579. (b) Correll, C. C.; Freeborn, B.; Moore, P. B.; Steitz, T. A. *Cell* **1997**, *91*, 705.

(18) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2842 and references therein.

(19) Ts'O, P. O. P. *Biol. Macromol.* **1970**, *4*, 49.

(20) (a) Sponer. J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365 and references therein. (b) Shinoda, T.; Shima, N.;. Tsukada, M. *J. Theor. Biol.* **1991**, *151*, 433. (c) Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792. (d) Kurita, M.; Araki, K.; Nakao, K.; Kobayashi, K. *Chem. Phys. Lett.* **1999**, 313, 693. (e) Brameld, K.; Dasgupta, S.; Goddard, W. A. *J. Phys. Chem. B* **1997**, *101*, 4851. (f) Kobayashi, K. *J. Phys. Chem. A* **1998**, *102*, 10813. (g) Sambrano, J. R.; De Souza, A. R.; Queralt, J. J.; Andres, J. *Chem. Phys. Lett.* **2000**, *317*, 437. (h) Morpurgo, S.; Bossa, M.; Morpurgo, G. O. *Chem. Phys. Lett.* **1997**, *102*, 10813.

(21) (a) Zhanpeisov, N. V.; Leszczynski, J. *J. Phys. Chem. B* **1998**, *102*, 9109. (b) Spöner, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590. (c) Spöner, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1990**, 13, 695. (d) Spöner, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1994**, *12*, 671. (e) Spöner, J.; Hobza, P. *Int. J. Quantum Chem.* **1995**, *57*, 959.

(22) (a) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, *17*, 841 and references therein.

(23) Subramanian, V.; Sivanesan, D.; Ramasami, T. *Chem. Phys. Lett.* **1998**, *290*, 189.

(24) (a) Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74. (b) Gao, J. *Biophys. Chem.* **1994**, *51*, 253. (e) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811 and references therein.

(25) Berman, H. M.; Sown, A.; Ginell, S.; Beveridge, D. L. *J. Biomol. Struct. Dyn.* **1988**, *5*, 1101.

(26) Cullis, P. M.; Wolfenden, R. *Biochemistry* **1981**, *20*, 3024.

(27) (a) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213 and see references therein. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027. (c) Luque, F. J.; Zhang, Y.; Alemán, C.; Gao, J.; Orozco, M. *J. Phys. Chem.* **1996**, *100*, 4269. (d) Foresman, J. B.; Keith, K. B.; Soonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098 and references therein.

10894 *J. Phys. Chem. A, Vol. 104, No. 46, 2000*

Sivanesan et al.

(28) Sambrano, J. R.; De Souza, A. R.; Queralt, J. J.; Andres, J. *Chem. Phys. Lett*. **2000**, *317*, 437.

(29) Claverie, P.; Daudey, J. P.; Langlet, J.; Pullman, B.; Piazzola, D. *J. Phys. Chem.* **1978**, *82*, 405.

(30) Alemán, C. *Chem. Phys*. **2000**, *253*, 13.

(31) (a) Spŏner, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn*. **1996**, *13*, 695. (b) Soliva, R.; Orozco, M.; Luque, F. J. *J. Comput. Chem*. **1997**, *18*, 980. (c) Anwander, E. H. S.; Probst, M. M.; Rode, B. M. *Bioploymers* **1990**, *29*, 757. (d) Burda, J. V.; Spŏner, J.; Hobza, P. *J. Phys. Chem*. **1996**, *100*, 7250 and references therein. (e) Burda, J. V.; Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B*. **1997**, *101*, 9670 and references therein.

(32) (a) See for a summary of models: Gadre, S. R.; Bhadane, P. K.; Pundlik, S. S.; Pingale, S. S. In *molecular electrostatic potentials: Concepts and applications*; Murray, J. S., Sen, K. D., Eds.; Elsevier: Amsterdam, 1996; pp 219−255, and references therein. (b) Pundlik, S. S. Development of an Electrostatic Model for Weak Intermolecular Complexation (EPIC). Ph.D. Thesis, University of Pune, India, 1998.

(33) (a) Gadre, S. R.; Pundlik, S. S. *J. Phys. Chem.* **1997**, *101*, 3298. (b) Pundlik, S. S.; Gadre, S. R. *J. Phys. Chem*. **1997**, *101*, 9657.

(34) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.

(35) Politzer, P.; Truhlar, D. G. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Plenum: New York, 1981. Tomasi, J.; Bonaccorsi, R.; Cammi, R. In *Theoretical methods of chemical bonding*; Maksic, Z. B., Ed.; Springer: New York, 1990; Vol. 3.

(36) (a) Stewart, I. *Sci. Am.* **1991**, *264,* 123. (b) Bader, R. F. W.; Nguyen, T. T.; Dang, Y. A. *Phys. Rev.* **1981**, *44*, 893.

(37) Shirsat, R. N.; Bapat, S. V.; Gadre, S. R.; the package UNIPROP see for details. *Chem. Phys. Lett*. **1992**, **200**, 373.

(38) Chipot, C.; GRID. The FORTRAN Program for Fitting Charges to Molecular Electrostatic Potentials and Fields; University de Nancy I: France, 1992.

(39) Limaye, A. C.; Inamdar, P. V.; Dattawadkar, S. M.; Gadre, S. R. UNIVIS: a visualization package. *J. Mol. Graph*. **1996**, 14.

(40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stafanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Bingley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez; Pople, J. A. *Gaussian94*; Gaussian Inc.: Pittsburgh, PA, 1995.

(41) The package GAMESS. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; together with Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(42) Gorb, L.; Leszczynski, J. *Int. J. Quantum Chem*. **1998**, *70*, 855.

(43) Nakano, N. I.; Igarashi, S. J. *Biochemistry* **1970**, *9*, 577.

(44) Gadre, S. R.; Babu, K.; Rendell, A. P. *J. Phys. Chem. A*, in press.