Ab Initio Study of the Crystallographic Solvation Pattern of the Cytosine-Guanine Base Pair in DNA

Federico Moroni, Antonino Famulari,* and Mario Raimondi

Dipartimento di Chimica Fisica ed Elettrochimica and Centro CNR - CSRSRC, Università degli Studi di Milano, via Golgi 19-20133 Milano, Italy

Received: August 9, 2000; In Final Form: October 5, 2000

The influence of hydration on the Watson−Crick cytosine−guanine base pair was investigated, testing the ability of the self-consistent field for molecular interaction (SCF-MI) ab initio method to reproduce the hydration pattern present in a real system (the base pair in the DNA framework). The positions of hydration sites around the base pair predicted by a knowledge-based approach employing crystallographic data were compared to the ab initio optimized structures. The SCF-MI method was applied to perform basis set superposition error (BSSE)-free geometry optimization. The hydration shell taken into account—comprising five water molecules, three on guanine and two on cytosine—"saturates" the base pair, engaging all of the available hydrogen bond donors/acceptors. The interaction between water and the base pair was also analyzed from the energetic viewpoint, highlighting the role of water in the pair stabilization.

Introduction

The presence of water plays an important role in the conformation and interactions of nucleic acids.[1] In particular, a shell of tightly bound water molecules whose properties differ from the bulk water was detected in the DNA.[2] Recently it was also suggested that these water molecules mark the positions of binding sites at protein−DNA interfaces.[3]

The X-ray crystallography of DNA fragments not only elucidates the system conformation but also allows to discover the hydration structure around them.[1] However, to determine a general hydration pattern and understand the underlying mechanisms, the individual single-crystal structure of a DNA oligomer determining the position of water molecules in specific sites is of scarce meaning due to sequence and conformation variability and also to crystal packing effects. Schneider developed a method of density representation of spatial distribution to characterize the hydration shell around DNA bases.[4] Given a DNA conformational type, the positions of water molecules around a given kind of base resulting from various oligomer structures taken from the Nucleic Acid Database[5] are collected and superimposed. The density of these joint water molecules could be analyzed by a pseudo-crystallographic method giving positions, occupancies and distributions of the averaged hydration sites.[6]

A discrete amount of ab initio calculations has been published on the interaction of explicit water molecules with different nucleic acid bases and base pairs.[7−15] The most addressed issue is the study of the effects of solvent molecules on relative proton affinity[8] and tautomer stability.[10−14] The nucleic acid bases are all characterized by several possible hydrogen bond acceptors and donors. In an explicit inclusion of the solvating water, care must be taken in the choice of number and positions of the solvent molecules. This choice has to be consistent with the system one is modeling. For instance, the preferred hydration site for isolated cytosine involves the O2 atom and the N1 hydrogen of the base.[9,10,15] However, due to the sugar backbone covalently bonded to the N1 atom, this site is forbidden in the nucleic acid. The majority of the secondary hydration sites have to be, obviously, discarded in the case of paired bases which engage from four to six hydrogen bond acceptors/donors in this interaction. Moreover, the variation of the number of water molecules could affect noticeably the geometry of a base pair. In the study of the hydration of the isocytosine−cytosine complex (a model of the CG pair), Zhanpeisov et al.[12] found an optimized structure of the fully hydrated system, containing six water molecules, with a strong buckling between the two bases (more than 30°), while the pair remains almost planar when only four water molecules are included. The same strong nonplanarity was obtained for the CG pair in the presence of eight interacting water molecules.[14]

In a series of preceding papers, we have studied the structures and energetics of various DNA related systems at the ab initio level. These calculations included several canonical and non-canonical isolated base pairs,[16] and the effects of various (free and hydrated) mono and divalent metal cations on the stability of the cytosine−guanine WC base pair.[17] To ensure the absence of the basis set superposition error (BSSE), the self-consistent field for molecular interaction (SCF-MI) approach[18,19] was used for all calculations. The SCF-MI approach provided quite accurate predictions of the properties of these substrates, in very good agreement with the standard counterpoise (CP) corrected SCF results determined with large basis sets. In particular, it was demonstrated that the SCF-MI method gives reliable results even with the 3-21G basis.[16,17,20]

In this work, we compare a SCF-MI study of the interaction of cytosine−guanine base pair with water to the averaged hydration structure obtained by Schneider for this base pair in B-DNA.[6] It is to be noted that in the real system, the presence of contiguous bases and of charged phosphate groups provides a network of interactions for the water molecules absent in the system considered in our calculations. Correspondence between the two approaches would be an index of strong base identity dependence of the hydration pattern.

* Corresponding author.

10.1021/jp0028195 CCC: $20.00
Published on Web 01/31/2001

We have performed full geometry optimization of the cytosine-guanine base pair with water molecules initially located in the crystallographic hydration sites.[6] As already emphasized above, the number of water molecules which compose the hydration shell is important. Schneider characterized three and five averaged hydration sites for cytosine and guanine, respectively. However, the average number of water molecules per base is less than these values and the occupancies of some of the hydration sites are rather small. We have considered only sites with occupancies greater than 0.6: this led to three water molecules for guanine and two for cytosine, values to be compared with the "stoichiometric" occupancies of the two bases, 2.14 and 1.95, respectively.[6] In an attempt to decompose the energy of the hydrated base pair, we performed calculations with fewer water molecules: cooperativity and anticooperativity of the hydration process have also been analyzed.

We used the convention of Seeman,[21] whereby hydration sites on the minor groove side are labeled "S" and those on the major groove "W". An increasing number is added to identify them univocally, and those relating to cytosine are primed to be distinguished from the other base sites. In the rest of the paper, we shall use the term "site" to indicate the crystallographic averaged hydration sites, and water molecule positions to mean the actual ab initio computed geometries.

Methods

All the ab initio calculations were performed with the SCF-MI procedure implemented in the GAMESS-US package.[22] On the basis of the satisfactory results for similar systems,[16,17,20] we employed the standard split valence 3-21G basis set, with all of the electrons considered explicitly. The structures were fully gradient optimized without any constraint.

The use of the SCF-MI method allows to compute geometry optimization for weakly interacting systems on a BSSE-free potential energy surface (PES). A brief introduction to the most relevant elements of the SCF-MI algorithm is reported in the appendix; for a more detailed account of the theory see references 18 and 19. Within the SCF-MI approach, the total binding energy of a supermolecule composed of K interacting fragments is expressed in a simple way:

$$E_{\text{int}} = E_{\text{SCF-MI}} - \sum_{k=1}^{K} E^{k}_{\text{SCF}}$$

where the SCF energies are determined at the monomer optimized geometries, taking properly into account geometry relaxation effects. For this reason when comparing the SCF-MI interaction energy with other BSSE corrected ones, the deformation energy of the monomers has to be included. The addition of this term brings, for example, the value of the 3-21G SCF-MI interaction energy for the C−G pair within 1 kcal/mol of the HF/6-31G** and of the MP2/6-31G*(0.25)//HF/6-31G* values.[23]

When dealing with systems composed of more than two fragments, it could be of interest to further decompose the total binding energy. However, owing to the presence of many-body and geometry relaxation contributions, this decomposition cannot be made in an unequivocal way. To evaluate the hydration energy of the base pair solvated by n water molecules, we employed the following expression:

$$E_{\text{idr}} = E_{\text{SCF-MI}} - E^{\text{C-G}}_{\text{SCF-MI}} - nE^{\text{H}_2\text{O}}_{\text{SCF}}$$

where the first term is the total SCF-MI energy and the last two terms are the energies of the corresponding optimized fragments. This quantity, containing water−base and water−water interaction terms, provides a direct estimate of the solvation energy of the substrate; it lacks direct information about the modification of the base pair interaction caused by the presence of water.

The deformation energy of the base pair is defined as the difference between the total energies of the isolated pair at different geometries

$$E_{\text{def}} = E^{\#\text{C-G}}_{\text{SCF-MI}} - E^{\text{C-G}}_{\text{SCF-MI}}$$

where the superscript # in the first term indicates that the energy refers to the geometry of the base pair in the hydrated system. The E_{def} is always positive and it is a direct index of geometry modification of the pair in the presence of water. Some authors[12−14] report the binding energy of the isolated complex at the geometry of the hydrated system as a measure of the base pair interaction energy; namely

$$E^{\#\text{C-G}} - E^{\text{C}} - E^{\text{G}} = E_{\text{def}} + E^{\text{C-G}}_{\text{int}}$$

It is to be noted that this "interaction energy" is, by definition, always smaller than the binding energy of the isolated base pair, and the many-body effects are not properly taken into account. We define the interaction, or binding, energy of the pair in the presence of water in a different way:

$$E_{\text{pair}} = E_{\text{SCF-MI}} - E^{\#\text{C-}n\text{H}_2\text{O}}_{\text{SCF-MI}} - E^{\#\text{G-}n\text{H}_2\text{O}}_{\text{SCF-MI}} + E^{\#n\text{H}_2\text{O}}_{\text{SCF-MI}}$$

where the # always means that the energies are computed at the geometry of the hydrated system. The two terms with negative sign subtract the individual interaction of the base with water and water−water interactions from the total energy; the last term is needed to correct for the twice removal of water−water interactions. The contributions to E_{pair} are the cytosine−guanine two-body interaction and all the higher order terms which include both the bases. In this definition, the waters are treated as a single fragment to limit the unmanageable growth of n body terms at the increasing number of fragments. As regards this definition of E_{pair}, it could be argued that the true binding energy would be obtained by moving the two hydrated bases away while allowing the systems to relax. However, our aim was to define an index of the C−G interaction energy in the presence of water molecules. On the contrary, the true binding energy may contain also strong water−water interaction terms (see Figure 5).

In the rest of this paper, the values of all the energy terms presented will be reported with inverted sign. In this way, positive and negative values will indicate attractive and repulsive interactions, respectively.

Results and Discussion

Hydration of Cytosine. Two hydration sites are considered for the pyrimidinic base: S1′ and W1′. In the first one, the water molecule is near the cytosine oxygen atom acting as hydrogen bond acceptor, whereas the second one lies in the vicinity of the amino group of the base that plays the role of the hydrogen bond donor. The optimized geometry of the tetramer resulting from the occupation of these sites is presented in Figure 1, where the positions of the average hydration sites are marked with dark circles. The distances of all the heteroatoms involved in water−base and base−base hydrogen bonds for this system (S1′W1′) are reported in Table 1. In the case of the W1′ water, accordance with the experimental position is almost perfect,

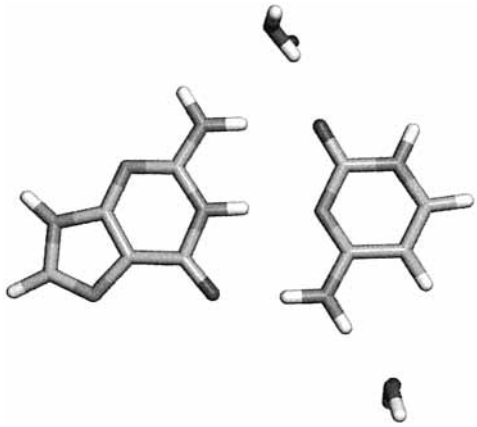

Figure 1. Ab initio optimized structure of the S1′W1′ hydrated cytosine−guanine pair. Dark circles represent average crystallographic hydration sites.

instead the S1′ hydration site distance from cytosine oxygen (2.63 Å) is shorter with respect to our computed value (2.93 Å). The hydration site and the position of the water molecule are located out of the plane of the base pair but on opposite sides, the first below (0.67 Å) and the latter above (0.34 Å). However, our theoretical result appears more acceptable considering that the value of S1′−O2′ distance reported by Schneider for B-DNA[6] is smaller than those distances presented in his previous work for different DNA conformations.[4] Besides this, the standard oxygen−oxygen distance for hydrogen bonds involving carbonyl groups is even greater than 3 Å.[24] In our modeling system, due to the almost perfect planarity of the base pair, the different location relative to the base plane—characterized by a remarkably small energy difference— is of minor relevance. Due to the presence of phosphate groups and adjacent bases, things are obviously different in the DNA framework. However, for another conformational type (Z-DNA), S1′ hydration sites are found both above and below the plain of the base pair.[4]

The presence of the two waters on the cytosine molecule does not affect the base pair conformation in an appreciable way. The larger interbase variation is 0.08 Å for the N2−O2′ distance and the intramolecular bond lengths are within the third decimal digit with respect to the isolated pair.

To better analyze water−CG interactions, we performed also the optimization of the two monohydrated systems (S1′ and W1′). The final position of the water molecules in these trimers does not differ significantly from that of the S1′W1′. Energy analysis of all the systems is reported in Table 2. The strongest hydrogen bond is formed by the W1′ water which act as acceptor for the H−N2′: 6.66 kcal/mol versus 4.99 of the S1′. It is interesting to note that the relative strength of the hydrogen bonds is in accordance with the different occupation numbers of the two sites determined by Schneider: 0.90 for W1′ and 0.75 for S1′.[6] The sum of the hydration energies of the monohydrated systems is 0.34 kcal/mol smaller than that of the tetramer. This difference might indicate a cooperative character of the interaction of the two water molecules with the base pair.

The small value of the deformation energy, only 2% of the isolated pair binding energy, reflects the light variations of base pair conformation already discussed. Our estimate of the base pair binding energy in the presence of the two water molecules is 25.05 kcal/mol, ∼10% greater than that of isolated cytosine−guanine pair. One of the major contributions to this value could be explained by the three-body term generated by the S1′ water−cytosine−guanine subsystem, where the water molecule, bound to the cytosine oxygen, is also close to the purinic amino group ($O_{S1'}$−N2 distance is 3.13 Å).

To analyze the differences between the hydration of the pair and that of an isolated base, we accomplished the geometry optimization of cytosine alone with a water molecule in the W1′ site. This choice was due to the fact that among the two possible sites, this one leads to the same hydrogen bond pattern in both the CG pair and isolated cytosine. The binding (hydration) energy for the isolated base is only 6.26 kcal/mol, 0.4 kcal/mol smaller than that of the base pair. This effect, in opposition to the expected less acidic character of the amino group involved in two hydrogen bonds, can be explained by the three-body interaction. It is to be noted that the binding energy obtained for this isolated base−water system compares well with the value obtained by Aleman:[9] 6.2 kcal/mol at the CP corrected MP2/6-31G(d) level.

To show the accuracy of the SCF-MI interaction energies we also performed a single point calculation on the monohydrated S1′ system. HF/6-31G** and MP2/6-31G** energy evaluations on the SCF-MI geometry were performed. The corresponding CP-corrected binding energies are 30.09 and 30.58 kcal/mol respectively, demonstrating the negligible effect of electron correlation on the determination of the interaction energy for this system. These values seem larger with respect to the SCF-MI binding energy (27.52 kcal/mol) but these calculations do not include the deformation energy of the monomers. By adding this contribution, the HF value is lowered by at least 2 kcal,[23] and the effect on the correlated binding energy could be even greater.[25]

Hydration of Guanine. The studied guanine hydration sites lie one on the minor and two on the major groove side. The first one, S1, is almost equidistant from the N3 atom and the H(N2) hydrogen, suggesting an interaction with both the atoms. The major groove sites, W1 and W2, lie near the O6 and N7 at a relative distance compatible with water−water hydrogen bond. The optimized structure of the trihydrated pair is reported in Figure 2 and selected heteroatom distances in Table 1. At first sight, accordance with the crystallographic hydration sites is worse than for the cytosine. The two major groove water molecules are shifted away toward cytosine, with one of them located almost at the midpoint of the W1 and W2 hydration sites. Better agreement is obtained for the S1 water acting simultaneously as hydrogen bond acceptor for the guanine amino proton and as donor for the N3 atom. While two of the hydration sites are displaced from the plane of the bases (S1 0.88 Å above and W2 0.68 below), all three water oxygens are found almost coplanar with the pair by the SCF-MI calculation.

The minor groove water molecule is arranged as to make a planar cyclic-like structure with the N3−C2−N2−H fragment of the base, the nonbonded water hydrogen lying out of this plane. This conformation is analogous to the water trimer equilibrium structure,[26] the base fragment mimicking two water molecules. On the major groove side, the water molecule located between the two hydration sites is hydrogen bonded with the N7 of the base. The same molecule acts as hydrogen bond acceptor for the second water which forms a bifurcated bond including the guanine oxygen. Formation of these bifurcated structures is not uncommon and explains why some organic compounds crystallize as hydrates with a number of hydrogen bond donors far less than the number of acceptors.[27] The two cytosine amino hydrogens are near to this second water; the O_{W1}−H(N4′) distance is ∼2.5 Å. However, this distance and the O_{W1}−H−N4 angle are far from standard values for hydrogen

TABLE 1: Selected Intermolecular Distances (Å) for the Cytosine−Guanine Pair at Different Hydration Levels

system	O6−N4′	N1−N3′	N2−O2′	Os1′−O2′	Ow1′−N4′	Os1−N3	Ow1-O6	Ow2-N7
isolated cg	2.94	3.04	2.99					
isolated cg[a]	2.94	3.05	2.99					
S1′W1′	2.95	3.08	3.07	2.93	2.98			
S1W1W2	2.95	3.03	2.98			3.00	3.13	3.22
penta-hydrated[a]	3.01	3.07	2.98	2.98	2.86	3.02	2.97	3.07

[a]Methylated bases.

TABLE 2: Energy Analysis (kcal/mol) for the Cytosine-Guanine Pair at Different Hydration Levels (for a definition of the various terms see the Methods section)

system	E_{int}	E_{idr}	E_{def}	E_{pair}
isolated cg	22.53			22.53
isolated cg[a]	22.45			22.45
s1′	27.52	4.99		
w1′	29.20	6.66		
s1′w1′	34.53	11.99	−0.49	25.05
s1	28.81	6.27		
wx	31.13	8.60		
w1w2	40.07	17.54		
s1w1w2	46.14	23.60	−0.67	26.71
penta-hydrated	68.59	46.06		
penta-hydrated[a]	64.34	41.88	−1.34	26.97

[a] Methylated bases.

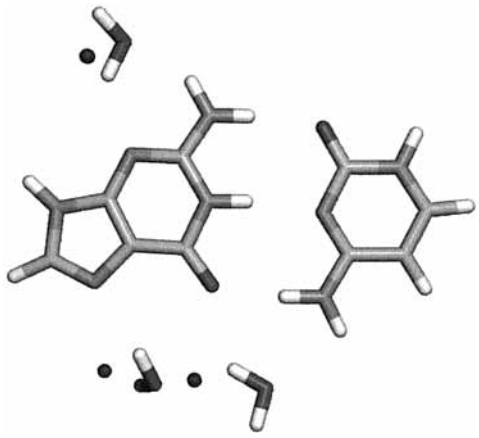

Figure 2. Ab initio optimized structure of the S1W1W2 hydrated cytosine−guanine pair. Dark circles represent average crystallographic hydration sites.

bonds. As in the case of cytosine, the effect of hydration on the base pair conformation is negligible.

Geometry optimization of the S1 monohydrated pair leads to the same water arrangement as in the trihydrated case. Despite the double hydrogen bond interaction with the base, E_{idr} is only 6.27 kcal/mol; a possible explanation is that the corresponding Y··H−X angles are far from the optimal linear arrangement. Irrespective of the initial position (W1 or W2 site), a single water molecule placed on the major groove side gives the same optimized structure (Figure 3) which is labeled Wx. This water makes two two-center bonds with the N7 and O6 atoms, resulting in a remarkably strong interaction with the base pair (E_{idr} = 8.60 kcal/mol). The addition of a second water molecule gives a conformation similar to that of the trihydrated system and it is also characterized by a high hydration energy (17.54 kcal/mol). The strength of the water−base interaction in the guanine major groove side is justified by two factors: the vicinity of three hydrophilic groups, namely the O6 and N7 atoms of guanine and the cytosine amino group, and the interaction with the strong dipole moment of the purinic base oriented toward the water molecules.

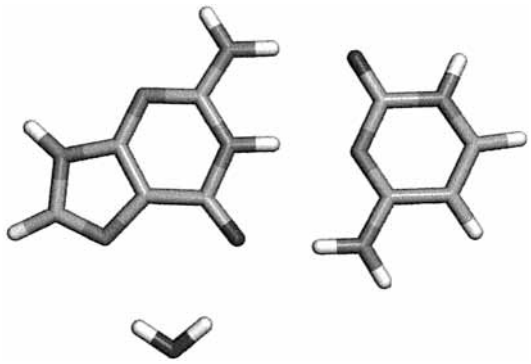

Figure 3. Ab initio optimized structure of the Wx hydrated cytosine−guanine pair.

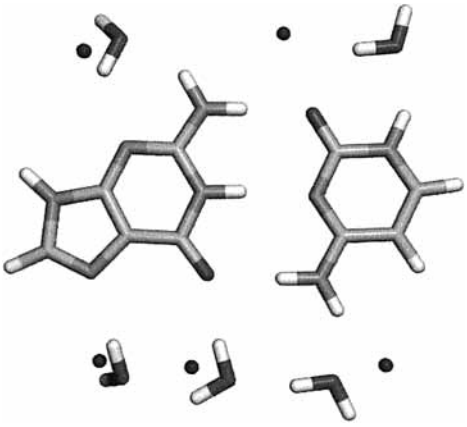

Figure 4. Ab initio optimized structure of the penta-hydrated cytosine−guanine pair. Dark circles represent average crystallographic hydration sites.

As regards the trihydrated complex, it is to be noted that E_{idr} (23.60 kcal/mol) is 0.21 less than the sum of the separated hydration energy of S1 and W1W2 systems (see Table 2), indicating a small anticooperativity effect. The deformation energy of the base pair is 35% greater in magnitude than that for the hydration of cytosine. The striking base pair binding energy (E_{pair}) enhancement (Table 2), more than 4 kcal/mol, is due to the three-body term arising from the second major groove water molecule bridging the two bases. Comparison of the hydration energy of guanine and cytosine (7.87 and 6.00 kcal/mol per water molecule, respectively) shows a net preference for the hydration of the purinic base.

Penta-Hydrated System. A problem arises when optimizing the structure of the system with concurrent hydration of the two bases. The S1′ water molecule departs from the corresponding hydration site to act also as hydrogen bond acceptor for the H−N1′ hydrogen (Figure 4). The orientation of this water molecule corresponds to the conformation of the isolated cytosine−water global minimum.[9,15] However, in the DNA framework, the N1′ is not available for hydrogen bond formation as it is bonded to the sugar backbone. To circumvent this problem, we performed calculations on N1′ and N9 methylated bases to simulate the anchoring to the backbone. Differences

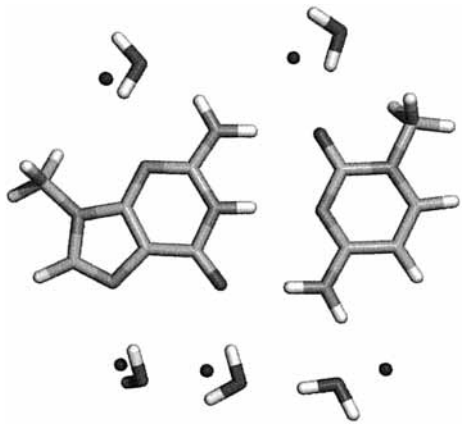

Figure 5. Ab initio optimized structure of the penta-hydrated cytosine−guanine pair with methylated bases. Dark circles represent average crystallographic hydration sites.

between standard and methylated CG pairs are almost irrelevant: the binding energy difference is less than 0.4% and the intermolecular distances vary in the order of one hundredth of angstrom.

The optimized structure of the methylated penta-hydrated base pair is reported in Figure 5. The correspondence between water positions and hydration sites with previous intermediate results (Figures 1, 2) is enhanced in some cases and diminished in others. In particular, the major groove water molecules of guanine move toward the corresponding W1 and W2 sites due to the presence of the W1′ water which shifts away from its site to interact with the purinic water. The distances between guanine water oxygens and their corresponding crystallographic hydration sites lie in the range 0.9−1.2 Å, a value to be considered acceptable owing to the diversity of experimental and theoretical approaches. On the cytosine fragment, the distances are 1.45 Å for the W1′ site and 1.93 Å for the S1′. The last value is particularly large and arises from the different location with respect to the pair plane: the S1′ site lies ∼0.7 Å below this plane, the corresponding calculated position ∼0.7 above. As previously noted, difference in energy between these two arrangements is minimal, and discarding the component of the distance in the direction perpendicular to the base plane would reduce the S1′ water oxygen/hydration site distance to 1.28 Å.

The hydrogen bonding pattern established between the water moiety and the base pair corresponds to that of the individual hydrations of the two bases with, in addition, an extra hydrogen bond formed between W1 and W1′ water molecules. The arrangement of water molecules around the CG pair is such that all of the base hydrogen bond donors/acceptors are involved in bonds, allowing to conclude that five water molecules "saturate" the first hydration shell. The hydration energy of this system is 41.88 kcal/mol, 6.29 greater than the sum of E_{idr} for the individual hydrated bases: the difference is mainly justified by the formation of the new hydrogen bond between the W1 and W1′ water molecules.

Deformation of the base pair is minimal, even if in this case the corresponding deformation energy rises to −1.34 kcal/mol. It is to be mentioned that the value of E_{def} seems to be almost linearly related to the number of surrounding water molecules. The base pair binding energy for the penta-hydrated pair is 26.97 kcal/mol. The enhancement in base pairing for this system (4.51 kcal/mol) is smaller than one would expect from the values of individual base hydrations, 2.52 and 4.18 kcal/mol for cytosine and guanine, respectively. A justification of this lies in the different positions of the water molecules which almost bridged the two bases in the previous systems: in particular, the W1 water, responsible of the base pairing enhancement in the S1W1W2 system, is shifted away from the pyrimidinic amino group to interact with W1′ water, but reducing the C−G-W1 three-body component of the energy. The same occurs on the minor groove side where the S1′ water departs from the purinic amino group, reducing the C−G-S1′ term.

Conclusions

The influence of hydration on the Watson−Crick cytosine−guanine base pair was investigated, testing the ability of the SCF-MI ab initio method to reproduce the hydration pattern present in a real system (the base pair in the DNA framework). The positions of hydration sites around the base pair predicted by a knowledge-based approach employing crystallographic data were compared to the ab initio optimized structures. To prevent the formation of biologically unrealistic hydrogen bonds in the full (penta) hydrated system, the use of the methylated analogues of the two bases was necessary.

The distance between calculated positions of water molecules and crystallographic averaged hydration sites are within 1.5 Å. This value is remarkably good, keeping in mind the difference between experimental and computational approaches. The crystallographic data refer to a base pair in the DNA framework in solid phase, while the ab initio study is performed on the pair solvated with five water molecules simulating the gas phase. Besides this, the statistical nature of the averaged hydration sites should be emphasized: the average root-mean-square deviation between the crystallographically determined water positions around a particular base pair and the predicted hydration sites is 1.0 Å[3]. The correspondence between averaged crystallographic data and our "isolated" system corroborates that the hydration pattern of bases in B-DNA strongly depends on the chemical nature of the bases themselves.[6]

Energetic analysis of the base−water interactions shows a net preference for the hydration of the purinic base. In particular, a region characterized by remarkably high base−water binding energies was found on the guanine major groove side. The inclusion of five water molecules allows all of the available hydrogen bond donors/acceptors of the CG pair to be involved in at least one hydrogen bond. This provides further evidence that only five ordered water molecules may compose the first hydration shell of base pairs in DNA.[6] "Saturation" of this first hydration shell resulted in negligible variations of the conformation of the pair. Despite this, the hydration causes more than 4 kcal/mol enhancement in the CG binding energy, emphasizing the important role of the surrounding water in base pair matching and mismatching. On the basis of the MP2 calculations performed on hydrated cytosine, the effects of correlation was found negligible.

The next aim will be to extend these studies to other systems of biological interest including the adenine−thymine complex and modified base pair analogues.

Appendix

A supersystem of K closed shell interacting fragments $a_1 ... a_K$ containing 2N electrons ($N = N_1 + N_2 + ... N_K$) is described by the one determinant SCF-MI wave function

$$\Psi(1....2N) = \boldsymbol{A}[\varphi_{1,1}(1)\bar{\varphi}_{1,1}(2)...\varphi_{K,N_K}(2N-1)\bar{\varphi}_{K,N_K}(2N)]$$

where $\boldsymbol{A}$ is the total antisymmetrizer operator. The method is based on the partitioning of the total basis set

$$\chi = (\chi_1|\chi_2| \ldots \chi_K)$$

so as MOs of different fragments are expanded in different subsets ($M = M_1 + \ldots + M_K$ is the basis set size), they are free to overlap. Accordingly, the N_k doubly occupied molecular orbitals of the fragment k, $\varphi_k = (\varphi_{k,1} \ldots \varphi_{k,N_k})$,

$$\varphi_k = \chi_k \mathbf{T}_k$$

are expanded in the set $\chi_k = (\chi_{k,1} \ldots \chi_{k,M_k})$, where $\mathbf{T}_k$ is an $M_k \times N_k$ matrix and M_k is the number of basis orbitals centered on the fragment k. The total ($M \times N$) matrix of the partitioned molecular orbital coefficients **T**, defined as

$$\varphi = (\varphi_1|\varphi_2| \ldots \varphi_K) \qquad \varphi = \chi \mathbf{T}$$

has a block diagonal form where the diagonal blocks are the $\mathbf{T}_1 \ldots \mathbf{T}_K$ matrices while the other blocks are null matrices.

The total energy of the SCF-MI wave function

$$E = \mathrm{Tr}[\mathbf{D} \cdot \mathbf{h}] + \mathrm{Tr}[\mathbf{D} \cdot \mathbf{F}(\mathbf{D})]$$

is written in term of the usual Fock (**F**) and one-electron integral (**h**) matrices expressed in the atomic orbitals basis set; the density matrix **D**

$$\mathbf{D} = \mathbf{T}(\mathbf{T}^\dagger \mathbf{S} \mathbf{T})^{-1} \mathbf{T}^\dagger$$

satisfies the general idempotency condition (**DSD** = **D**).

The appearance of the BSSE is avoided by assuming and maintaining the orbital coefficient variation matrix in a block diagonal form. The stationary condition $\delta E = 0$ is equivalent to K secular problems

$$\begin{bmatrix} \mathbf{F}'_k \mathbf{T}'_k = \mathbf{S}'_k \mathbf{T}'_k \mathbf{L}_k \\ \mathbf{T}'^\dagger_k \mathbf{S}'_k \mathbf{T}'_k = \mathbf{1}^{N_k} \end{bmatrix}$$

in terms of effective Fock and overlap matrixes $\mathbf{F}'_k$ and $\mathbf{S}'_k$. These matrixes are Hermitian and possess the correct asymptotic behavior: in the limit of infinite separation of the fragments, $\mathbf{F}'_k$ and $\mathbf{S}'_k$ become the Fock and overlap matrices of the individual systems.

The SCF-MI binding energy can be expressed as

$$\Delta E_{\mathrm{SCF-MI}} = E_{\mathrm{SCF-MI}} - \sum_{k=1}^{K} E^k_{\mathrm{SCF}}$$

taking properly into account geometry relaxation effects. The validity of the method extends from the long range to the region of the minimum and of short distances.

Following the scheme proposed by Gerratt and Mills,[28] see also Pulay[29] and Yamaguchi et al.,[30] the calculation of first and second derivatives was implemented[19] and inserted into GAMESS−US package.[22] Level shifting[31] and DIIS[32] techniques have been adopted to increase the convergence performances.

References and Notes

(1) Berman, H. M. *Curr. Opin. Struct. Biol.* **1991**, *1*, 423−427. Berman, H. M. *Curr. Opin. Struct. Biol.* **1994**, *4*, 345−350.

(2) Wang, J. H. *J. Am. Chem. Soc.* **1955**, *77*, 258−260. Tunis, M.-J. B.; Hearst, J. E. *Biopolymers* **1968**, *6*, 1345−1353. Kubinec, M. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 8739−8740.

(3) Woda, J.; Schneider, B.; Patel, K.; Mistry, K.; Berman, H. M. *Biophys. J.* **1998**, *75*, 2170−2177.

(4) Schneider, B.; Cohen, D. M.; Schleifer, L.; Srinivasan, A. R.; Olson, W. K.; Berman, H. M. *Biophys. J.* **1993**, *65*, 2291−2303.

(5) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751−759.

(6) Schneider, B.; Berman, H. M. *Biophys. J.* **1995**, *69*, 2661−2669.

(7) Aleman, C. *Chem. Phys.* **1999**, *244*, 151−162. Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **1999**, *103*, 8853−8860. Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V. *J. Phys. Chem. A* **1999**, *103*, 7912−7917. van Mourik, T.; Price, S. L.; Clary, D. C. *J. Phys. Chem. A* **1999**, *103*, 1611−1618. Brandl, M.; Meyer, M.; Suhnel, J. *J. Am. Chem. Soc.* **1999**, *121*, 2605−2606. Gadre, S. R.; Babu, K. *J. Phys. Chem. A* **2000**, *104*, 8976−8982.

(8) Maes, G.; Smets, J.; Adamowicz, L.; McCarthy, W.; Vanbael, M. K.; Houben, L.; Schoone, K. *J. Mol. Struct.* **1997**, *410*, 315−322. Chandra, A. K.; Nguyen, M. T.; Zeegers-Huyskens, T. *J. Mol. Struct. (THEOCHEM)* **2000**, *519*, 1−11.

(9) Aleman, C. *Chem. Phys. Lett.* **1999**, *302*, 461−470.

(10) Gorb, L.; Leszczynski, J. *Int. J. Quantum Chem.* **1998**, *70*, 855−862.

(11) Gorb, L.; Podolyan, Y.; Leszczynski, J. *J. Mol. Struct. (THEOCHEM)* **1999**, *487*, 47−55. Gorb, L.; Leszczynski, J. *J. Am. Chem. Soc.* **1998**, *120*, 5024−5032.

(12) Zhanpeisov, N. U.; Leszczynski, J. *J. Phys. Chem. B* **1998**, *102*, 9109−9118.

(13) Zhanpeisov, N. U.; Leszczynski, J. *J. Phys. Chem. A* **1998**, *102*, 6167−6172.

(14) Zhanpeisov, N. U.; Leszczynski, J. *J. Mol. Struct. (THEOCHEM)* **1999**, *487*, 107−115.

(15) van Mourik, T.; Benoit, D. M.; Price, S. L.; Clary, D. C. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1281−1290.

(16) Famulari, A.; Raimondi, M.; Sironi, M. Hartree−Fock study of hydrogen bonded systems in the absence of basis set superposition error: The nucleic acid base pairs. In *Progress in Theoretical Chemistry & Physics*; Kluwer Academic Publishers: Dordrecht, 2000; in press.

(17) Pelmenschikov, A.; Zilberberg, I.; Leszczynski, J.; Famulari, A.; Sironi, M.; Raimondi, M. *Chem. Phys. Lett.* **1999**, *314*, 496−500. Famulari, A.; Moroni, F.; Sironi, M.; Gianinetti, E.; Raimondi, M. *J. Mol. Struct. (THEOCHEM)* **2000**, *529*, 209−217. Famulari, A.; Moroni, F.; Sironi, M.; Raimondi, M. *Comput. Chem.* **2000**, *24*, 341−349. Raimondi, M.; Famulari, A.; Gianinetti, E.; Sironi, M.; Moroni, F. Modification of Roothan equations for the ab initio calculation of interactions in large molecular systems in the absence of basis set superposition erros. In *Large-Scale Scientific Computations of Engineering and Environmental Problems II*; Griebel, M., Margenov, S., Yalamov, P., Eds.; Vieweg: Wiesbaden, 2000; pp 335−343.

(18) Famulari, A.; Gianinetti, E.; Raimondi, M.; Sironi, M.; Vandoni, I. *Theor. Chem. Acc.* **1998**, *99*, 358−365. Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. Self-Consistent Field Theory of Weakly Bonded Systems. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2000; Vol. 5, pp 141−170. Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251−266.

(19) Famulari, A.; Gianinetti, E.; Raimondi, M.; Sironi, M. *Int. J. Quantum Chem.* **1998**, *69*, 151−158.

(20) Raimondi, M.; Famulari, A.; Gianinetti, E. *Int. J. Quantum Chem.* **1999**, *74*, 259−269.

(21) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804−808.

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

(23) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974.

(24) Dimitrova, Y.; Peyerimhoff, S. D. *J. Phys. Chem.* **1993**, *97*, 12731−12736.

(25) Sponer, J.; Hobza, P. *J. Phys. Chem. A* **2000**, *104*, 4592−4597.

(26) Mo, O.; Yanez, M. *J. Chem. Phys.* **1992**, *97*, 6628−6638.

(27) Desiraju, G. R. *J. Chem. Soc., Chem. Commun.* **1991**, 426.

(28) Gerratt, J.; Mills, I. M. *J. Chem. Phys.* **1968**, *49*, 1719.

(29) Pulay, P. *Mol. Phys.* **1969**, *17*, 197. Pulay, P. *Adv. Chem. Phys.* **1987**, *69*, 241.

(30) Yamaguchi, Y.; Osamura, Y.; Goddard, J. D.; Schaefer, H. F. *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*; Oxford University Press: Oxford (UK), 1994.

(31) Hillier, I. H.; Saunders: V. R. *Int. J. Quantum Chem.* **1970**, *4*, 503.

(32) Pulay, P. *J. Comput. Chem.* **1982**, *3*, 556−560.