QUANTUM-MECHANICAL STUDIES OF ENVIRONMENTAL EFFECTS ON BIOMOLECULES

II. HYDRATION SITES IN PURINES AND PYRIMIDINES*

G.N.J. PORT and A. PULLMAN

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique
associé au C.N.R.S., 13, rue P. et M. Curie, Paris 5è, France

Received 21 January 1973

1. Introduction

We have recently undertaken a theoretical study of the effect of the water environment on the structure and properties of fundamental biomolecules and polymers. In this as yet unexplored area the first step is the determination of the most probable sites of fixation and of the orientation of the water molecules. A thorough investigation of this problem in the case of the peptide unit has been reported in paper I of this series [1]. The present note reports the results of a similar study for the fundamental base components of DNA, adenine, guanine, cytosine and thymine.

2. Method

The size of the molecules involved in the present case precluded the utilization of the *ab initio* SCF supermolecule approach used in the study of the formamide hydrates. It has, however, been shown, both in paper I and in previous studies on the hydrogen bond, that the electrostatic component of the interaction energy plays the dominant role in hydrogen bonding [2–5], particularly in determining the relative orientations of the two partners at equilibrium distance [1]. We have therefore computed only this electrostatic component using a procedure recently developed in our laboratory [6], the essential steps of which are as follows:

i) This procedure requires the knowledge of a wave function for the isolated substrate and for water. In the present case *ab initio* SCF wave functions were available both for the nucleic acid bases [7] and for water [8] in a gaussian basis set of single – ζ reasonable accuracy (see [1] for details).

ii) Each molecular orbital φ_i composing the wave function being written as a linear combination of atomic orbitals:

$$\varphi_i = \sum_i c_{ir} \chi_r,$$

the ground state electron density distribution is:

$$\rho = \sum_i \sum_{r,s} 2\, c_{ir}\, c_{is}\, \chi_r\, \chi_s$$

It appears as a summation of one-center (r, s on the same atom) and two-center (r, s on different atoms) overlap distributions.

iii) Every overlap product $\chi_r \chi_s$ is replaced by a multipole expansion around an appropriate point: for a one-center product the point is the corresponding nucleus, for a two-center product it is the middle of the segment joining the nuclei carrying χ_r and χ_s, respectively. A first program computes for a given molecule with n atoms all the required terms up to the quadrupoles at each of the $(n(n+1))/2$ centers.

iv) The interaction energy between the two sets of multipoles is computed in the next step using the appropriate formula derived from classical electrostatic

---

* This work was supported by the A.T.P. No A655-2303 of C.N.R.S.

theory. Values obtained by this procedure have been shown to be a very good approximation of the electrostatic interaction energy computed *exactly* from the density distributions of the separate partners in the case of various conformations of hydrogen-bonded molecules [3, 4, 6, 8].

In the present work we have used the multipole-procedure for a rather complete exploration of the interaction energy hypersurface for each base–water complex in the following fashion: first the water molecule was allowed to turn completely around the periphery of the substrate, with its oxygen in the plane of the base, its orientation being optimized by rotation about three coordinate axes at each position. A constant distance of 2.85 Å was maintained between the oxygen of water and the atoms of the base with which a hydrogen bond might be formed (carbons included), and in the course of the various movements, the approach of any two non-hydrogen-bonded atoms closer than the sum of their Van der Waals radii was forbidden. This precaution eliminates false zones of apparent stability due to the neglect of exchange repulsions important at short distances. A further exploration of the possibilities of hydration above (or below) the molecular plane was made in the most interesting areas.

3. Results

3.1. *"In-plane" interaction*

Figs. 1 to 4 give a pictorial representation of the results obtained for adenine, guanine, cytosine and thymine for the arrangements in which the oxygen of water was maintained in the plane of the base. We shall call these arrangements the "in-plane" configurations, among which we shall distinguish, however, the coplanar conformations in which the water molecule is coplanar with the base, and the perpendicular conformations in which the oxygen of water remains in the plane of the base but the hydrogens are located above and below this plane, in a plane perpendicular to it. The "in-plane" configurations appear in all cases by far the most favorable ones: the largest value of the electrostatic energy for "out-of-plane" interaction never reaches over 50% of the maximal values of the "in-plane" interaction.

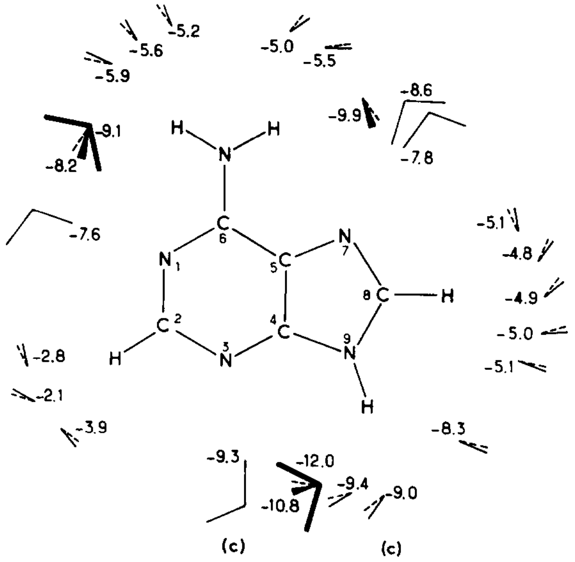

Fig. 1. Hydration sites in adenine. Energies in kcal/mole. Heavy lines: preferred hydration sites. Full lines: coplanar arrangement of water and base. Half-dashed: perpendicular arrangement of water with respect to the plane of the base. (c) refers to configurations akin to conventional hydrogen-bonding.

3.1.1. Adenine

Three regions of strong interaction are observed around the periphery of adenine (fig. 1).

The deepest minimum occurs when the water molecule is coplanar with the base and forms a bridging interaction between N_3 and N_9H, its oxygen turned towards the NH group and one of its hydrogens pointing towards N_3: about 30% of the stabilization energy is gained by displacing the water molecule from the "classical" hydrogen-bonded positions (marked (c) in fig. 1, "classical" H bond with N_3 or NH alone) to this bridging position.

The second stable region of association involves N_7 and one hydrogen of the amino group. Here, the configuration NH...O is nearly linear and the attraction of N_7 results in the rotation of the plane of water so as to turn its hydrogens towards this atom. This hydration site is more sterically crowded so that the water molecule prefers to lie perpendicularly to the base plane.

Finally a third favorable region is situated between the other amino hydrogen and N_1. The site is less

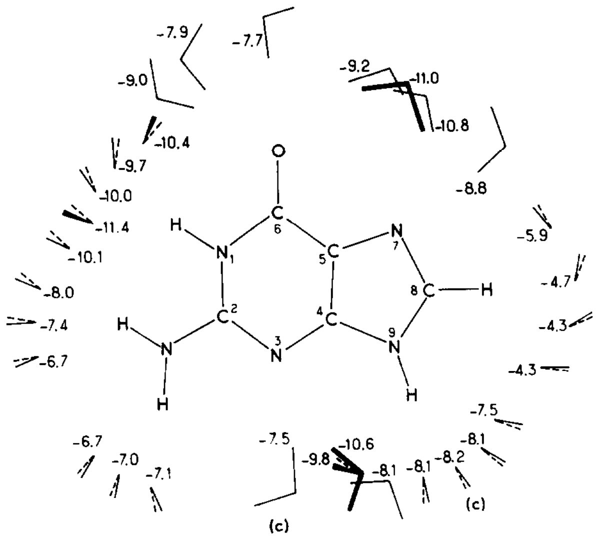

Fig. 2. Hydration sites in guanine. Energies in kcal/mole. Heavy lines: preferred hydration sites. Full lines: coplanar arrangement of water and base. Half-dashed: perpendicular arrangement of water with respect to the plane of the base. (c) refers to configurations akin to conventional hydrogen-bonding.

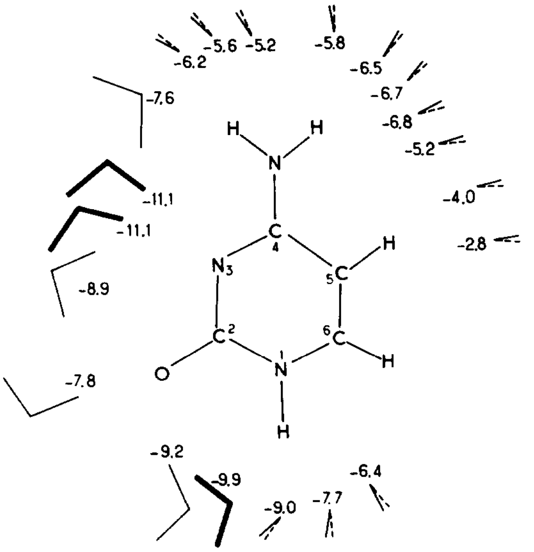

Fig. 3. Hydration sites in cytosine. Energies in kcal/mole. Heavy lines: preferred hydration sites. Full lines: coplanar arrangement of water and base. Half-dashed: perpendicular arrangement of water with respect to the plane of the base.

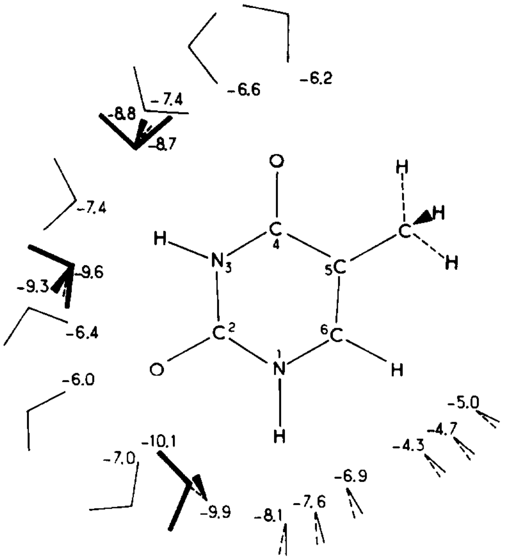

Fig. 4. Hydration sites in thymine. Energies in kcal/mole. Heavy lines: preferred hydration sites. Full lines: coplanar arrangement of water and base. Half-dashed: perpendicular arrangement of water with respect to the plane of the base.

sterically crowded and the coplanar conformation is slightly favored over the perpendicular one by 1 kcal/mole.

Although the three above-defined regions are well separated, it appears that the electrostatic interaction is nowhere totally repulsive for the in-plane water approach, the C_2H area being the least favorable from this point of view (it is possible that exchange repulsions may there overcompensate the attractive term).

3.1.2. Guanine

The in-plane interactions depicted in fig. 2 resemble those described for adenine in many features, namely the absence of regions of repulsion, the appearance of well-defined favorable positions for the water molecule with preference for bridging interactions when possible, preference for coplanar arrangement of the two molecules whenever the site is not too crowded.

Again three essential hydration sites may be distinguished.

Replacement of the NH_2 group of adenine by a carbonyl group provides the possibility of an in-plane bridging configuration between the oxygen and N_7 which is more favorable than the perpendicular arrangement found in this region for adenine.

The second minimum is exactly analogous to the deepest one in adenine, occurring between the N_3 and N_9H atoms with the all-planar arrangement (−10.6 kcal/mole) here only slightly favored over the perpendicular one (−9.8).

Finally a large attractive area is seen on the left side of the molecule with two distinct minima, one between N_1H and the amino group, the other between N_1H and the oxygen: the water molecule prefers there the perpendicular arrangement around N_1H with two favored orientations, one clearly influenced by the presence of the neighbour NH proton-donor of the amino group, the other taking advantage of the oxygen attraction.

3.1.3. Cytosine

Fig. 3 indicates that the N_3 atom provides the most favorable hydration site in cytosine. The binding appears in this case more in line with the classical image of hydrogen bonding than in the previous cases. The relative geometrical positions of N_3 and the oxygen are not favorable for the formation of a strong bridge like the one observed between O and N_7 in guanine.

A second hydration site is situated on the other side of the carbonyl group, where the water molecule is shared between the oxygen and N_1H: there is no marked difference in stability whether H_2O is in the plane of the base (−9.9 kcal/mole) or perpendicular to it (−9.5).

There is no other region of comparably strong attraction in cytosine but we observe a smaller minimum facing the second amino nitrogen where water sits perpendicularly to the base plane, acting as proton acceptor in a conventional hydrogen bond.

3.1.4. Thymine

Fig. 4 shows the occurrence of two symmetrically placed hydration sites in thymine on each side of the C_2O_2 carbonyl oxygen, each taking advantage of the attraction of the neighbour NH group for the oxygen of water, the planar arrangement being only very slightly favored over the perpendicular one.

On the other hand another favorable site exists between N_3H and the C_4O_4 carbonyl oxygen where again the coplanar and perpendicular conformations have very similar bonding energies.

3.2. *Out-of-plane interaction*

As already mentioned, the in-plane interactions are much more favorable than any out-of-plane interaction.

The situation is illustrated by calculations for a water molecule placed above the NH_2 group of cytosine or adenine, the most favorable arrangement for an out-of-plane interaction [1]:

Hydration with one OH-bond of water hanging directly over the nitrogen yields a maximum interaction of only 4.4 kcal/mole in cytosine, a value which may be increased to −4.9 by bending the molecule towards N_3. The gain is slightly larger for a bridge with N_7 in adenine (from −3.5 to −4.8 kcal/mole). These values are much smaller than for the favorable "in-plane" interactions.

4. Conclusions

This study determines the most favorable positions for hydration of the four bases. They are located, as described above, around the base periphery with the oxygen of water in the molecular plane. There seems in fact to be a tendency of the water molecule to rest entirely in the same plane as the base if possible: however where such coplanarity leads to steric crowding, water may assume a perpendicular configuration with one H above and one H below the plane of the base, with very little loss of stabilization energy.

Out-of-plane monohydration appears unlikely in these molecules. A similar conclusion was reached in the case of formamide [1] where the result of the electrostatic approximation was confirmed by SCF computations. It must be stressed that this conclusion does not exclude, however, the possibility of out-of-plane hydration bridging two molecules (in stacks or in crystal packing) or two parts of a composite molecule as in ribosides or ribotides. In fact, the present "in-plane" results show how effective a second proton-donor or proton-acceptor can be in determining the most stable position of water already hydrogen-bonded to a first atom.

Acknowledgements

We gratefully acknowledge the communication by Dr. E. Clementi of the coefficients of the atomic orbitals necessary to compute the multiple expansion for the four bases.

One of us (G.N.J.P.) acknowledges a fellowship from the Royal Society.

References

[1] G. Alagona, A. Pullman, E. Scrocco and J. Tomasi, Intern. J. Peptide Protein Chem., in press.

[2] B. Pullman and A. Pullman, Progr. Nucleic Acid Res. Mol. Biol. 9 (1969) 327.

[3] M. Dreyfus and A. Pullman, Compt. Rend. 271 (1970) 457.

[4] M. Dreyfus and A. Pullman, Theoret. Chim. Acta 19 (1970) 20.

[5] R. Bonaccorsi, C. Petrongolo, E. Scrocco and J. Tomasi, Theoret. Chim. Acta 20 (1971) 331.

[6] M. Dreyfus, Thèse 3è Cycle, University of Paris, 1970.

[7] E. Clementi, J.M. André, M.Cl. André, D. Klint and D. Hahn, Acta Phys. Acad. Sci. Hung. 27 (1969) 493.

[8] H. Berthod and A. Pullman, unpublished results.