Theoret. Chim. Acta (Berl.) 50, 317–325 (1979)

THEORETICA CHIMICA ACTA

Hydration Scheme of the Purine and Pyrimidine Bases and Base-Pairs of the Nucleic Acids

Bernard Pullman, Stanislav Miertus* and David Perahia

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

The hydration scheme of the bases of the nucleic acids, leading to a representation of their first hydration shell, was computed using the overlap multipole procedure. The shell involves five, four, four and three "bound" water molecules in G, A, C and T respectively. The formation of the base pairs displaces one water molecule in the A–T pair and four water molecules in the G–C pair from the hydration shell. The hydration produces a destabilization of the pairing energy in comparison to the binding in vacuo, greater for the G–C pair than for the A–T pair. There remains nevertheless an appreciable residual affinity for inter-base hydrogen bonding in water.

Key words: Nucleic acid bases and base pairs, hydration scheme of ~

1. Introduction

We have presented previously [1–4] a quantum-mechanical exploration of the *monohydration* scheme of the purine and pyrimidine bases of the nucleic acids. This study enabled to determine the main topographical and energetical characteristics of the interaction of a water molecule with these substances and indicated, in particular, the preferred sites of water binding to them.

In the present work we wish to extend this study to the determination of the *polyhydration* scheme of the bases resulting from the simultaneous interaction of a number of water molecules with these substrates. We propose to explore in this way the nature of the first solvation shell around these bases with the particular aim of determining the possible differences in their degree of hydration. At the same time, by comparing the results for the individual bases with similar results obtained recently in our laboratory [5] for the polyhydration scheme of the complementary base pairs of the nucleic acids, adenine-thymine and guanine-cytosine, we shall try to evaluate the effect of hydration on the dissociation equilibrium of the base pairs.

---

* Permanent address: Polymer Institute, Slovak Academy of Sciences, Bratislava, Czechoslovakia.

0040-5744/79/0050/0317/$01.80

2. Method

The calculations have been carried out with the overlap multipole procedure (OMTP) developed in this laboratory for the computation of the electrostatic interaction between molecules [4, 6]. Its essential feature is the use of a "polycentric" multipole expansion of the electron density of each of the interacting molecules which serves to evaluate the electrostatic interaction between them (see Ref. [4] for details). A minimization technique, a version of the simplex method [7], has been coupled to this computational procedure in order to determine the optimum positions of the water molecules around the ligands. This combination has been shown [4] to reproduce satisfactorily the hydration schemes of molecules obtained by full SCF *ab initio* supermolecule calculations.

The multipole expansions are obtained directly from SCF *ab initio* wave functions of the individual molecules. For adenine, thymine and guanine we used the wave functions of Clementi *et al.* [8]. The wave function for cytosine was recomputed in the same basis set for its experimental geometry.

In order to avoid false zones of apparent stability due to the neglect of the exchange repulsion, a distance of closest approach of 2.85 Å has been fixed between the oxygen of water O(W) and the first-row atoms, with the exception of the distance between O(W) and pyridine type nitrogens which was fixed at 2.95 Å [4]. The investigation was limited to the in-plane location of water molecules: the oxygen of water in the plane of the bases. *Ab initio* SCF computations indicate [4, 9] that these are by far the most important interactions. This is in contrast to results obtained by the CNDO method [4, 9, 10] which are obviously artifacts of that method.

The minimization technique used optimizes the interaction energy of a given water molecule with all its surroundings (ligand + the other water molecules), changing the geometrical position of the water molecule considered. This version enables to find the polyhydration scheme of the ligand taking into account all the possible interactions between the molecules of the model. For the initial positions of the water molecules we used the results of the monohydration studies indicated previously [1]. The number of water molecules was then increased progressively until the completion of the first solvation shell, the interaction energy of the water molecules being optimized one by one in succession in the presence of all the other water molecules involved. This procedure was then repeated, restarting with the first optimized water molecule and so on until the computed total interaction energy did not change significantly.

3. Results and Discussion

Results of the evaluation of the most stable arrangement(s) of the first (and obviously unique) hydration shell of the nucleic acid bases, together with the interaction energies involved, are indicated in Figs. 1–4. Before discussing in more detail the structure of these hydration shells it may be useful to indicate that the energies of water–water interaction, computed within the same systematics, amount to -6.7 kcal/mole in the "best" water dimer and to -5.1 kcal/mole in an arrangement in which a central water molecule is surrounded by four others. We may use one or the other of these two values as a lower limit for distinguishing the water molecules "bound" to the bases and also for improving the estimation of the "hydration" energy of the base by substracting from the sum of the base-water interaction energies, the corresponding water–water interaction energies. The interaction energies indicated in Figs. 1–4 are all greater than -5.1 kcal/mole. Consequently with this limit all the indicated water molecules belong to the hydration shell of the bases. With the limit of -6.7 kcal/mole some of these water molecules should not be considered as belonging to this shell.

We shall first discuss the results obtained by adopting -5.1 kcal/mole as the value distinguishing "bound" water molecules and indicate later the modifications introduced by adopting the value -6.7 kcal/mole.

3.1. Adenine (A)

Two structures are possible for the hydration shell of this molecule, illustrated in Fig. 1a and b. They have a similar positioning of three water molecules W_1, W_2 and W_3 but differ in the possible structure of the shell in the N_1–NH_2 region. In Fig. 1a, one water molecule, W_4, occupies a bridge position between N_1 and NH_2 with an interaction energy of $-10,2$ kcal/mole. In Fig. 1b, two water molecules W_4 and W_5 fit within this region, with interaction energies with the base of the order of -6.5 to -7 kcal/mole (and linked together between themselves by a hydrogen bond of -5.7 kcal/mole). In Fig. 1a the total interaction energy of adenine with the four water molecules amounts to -38.8 kcal/mole, in Fig. 1b the interaction energy with the five water molecules amounts to -42.0 kcal/mole.

Taking into consideration the energy necessary to split water–water bonds for the sake of establishing base–water interactions, the "hydration" energies are estimated

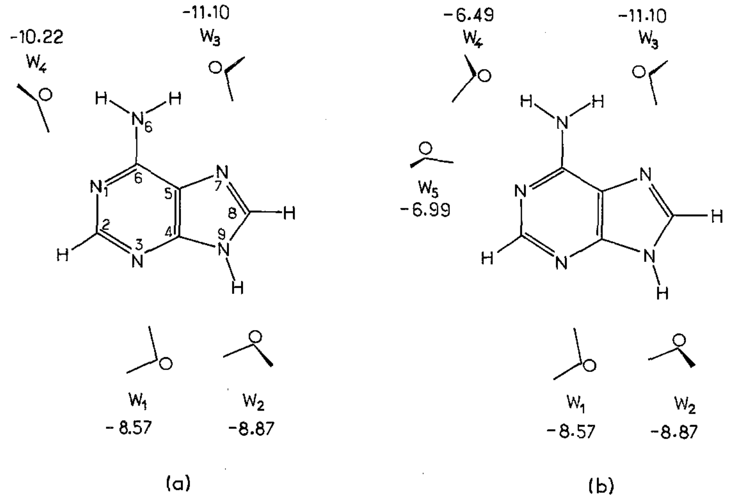

Fig. 1. Hydration scheme of adenine: a) with four water molecules, b) with five water molecules

to be -18.4 kcal/mole for the system $A+4H_2O$ and -16.5 kcal/mole for the system $A+5H_2O$. The system $A+4H_2O$ seems the most advantageous from that point of view.

3.2. Thymine (T)

The results of the computations are illustrated in Fig. 2. The best possible arrangements for hydration of this base imply three (Fig. 2a) or four (Fig. 2b) water molecules. Note two possible positions for W_3 in Fig. 2a, differing little in energy of interaction with the base (in future discussion we shall consider only the position with the greatest value of the energy). In Fig. 2b, W_2 and W_3 are hydrogen-bonded with an energy of -5.1 kcal/mole. The total energy of interaction of the water molecules with the base amounts to -26.7 kcal/mole in Fig. 2a and to -30.4 kcal/mole the Fig. 2b. When account is taken, as in the case of adenine, of the energy necessary for the splitting of the corresponding water–water interactions, the "hydration" energies come out equal to -11.4 kcal/mole for the system T + 3 water molecules and -10.0 kcal/mole for the system T + 4 water molecules. The system T $+3H_2O$ seems the most advantageous.

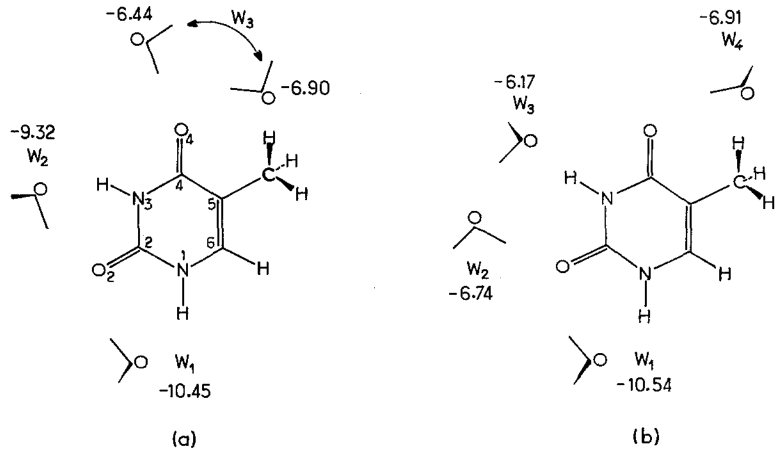

Fig. 2. Hydration scheme of thymine: a) with three water molecules, b) with four water molecules

3.3. Guanine (G)

The hydration shell of guanine (Fig. 3) implies five water molecules with two possible arrangements in the O_6–N_1H–NH_2 region, illustrated in Fig. 3a and 3b. In both figures W_4 and W_5 are hydrogen-bonded between themselves, with an energy of -5.7 kcal/mole in Fig. 3a and of -6 kcal/mole in Fig. 3b. The total interaction energies of the five water molecules with the base amount to -49.6 kcal/mole in the arrangement of Fig. 3a and to -48.8 kcal/mole in the arrangement of Fig. 3b. When account is taken of the breaking of the water–water interactions, the hydration energies come out as equal to -24.1 or -23.3 kcal/mole for the

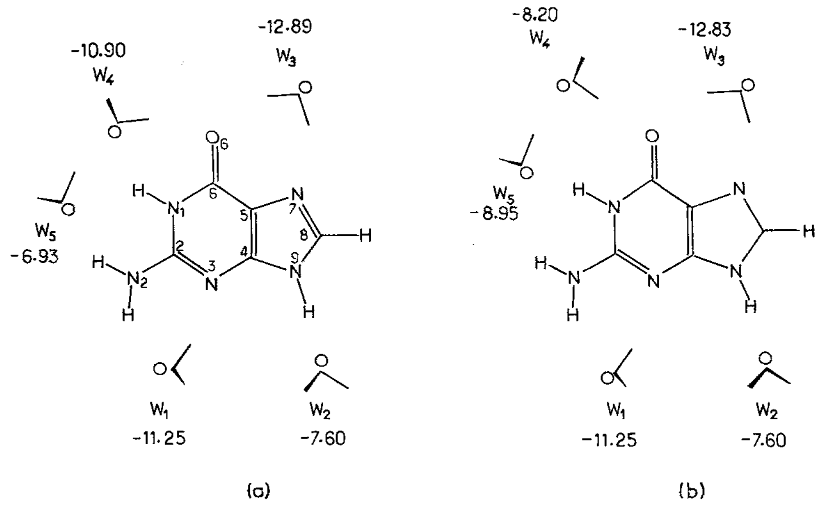

Fig. 3. Hydration scheme of guanine: a) and b) two possibilities with five water molecules

arrangement of Fig. 3a and 3b respectively. The two hydration schemes being nearly equivalent, a dynamic equilibrium should exist between them.

3.4. *Cytosine (C)*

This base yields a single polyhydration scheme, presented in Fig. 4. The hydration shell involves four water molecules. Their energy of interaction with the ligand amounts to -33.1 kcal/mole. Taking into account the expense of energy produced in breaking the water–water interactions, the hydration energy amounts to -12.7 kcal/mole.

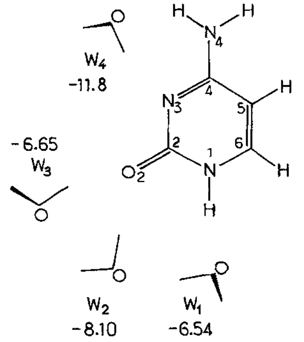

Fig. 4. Hydration scheme of cytosine

Altogether, the "best" polyhydration scheme for the four bases corresponds to the order:

$$G(5H_2O)-24.1 > A(4H_2O)-18.4 > C(4H_2O)-12.7 > T(3H_2O)-11.4$$

Table 1. Energetics of hydration of bases and base pairs (kcal/mole) (W···W = −5.1)

Ligand	Number of water molecules	Interaction energy with water	Hydration energy	
A	4	−38.8	−18.4	
A	5	−42.0	−16.5	
T	3	−26.7	−11.4	
T	4	−30.4	−10.0	
G	5	−48.8	−23.3	
G	5	−49.5	−24.1	
C	4	−33.1	−12.7	
A+T	7	−65.4	−29.7	
G+C	9	−82.7	−36.8	
A−T	6	−54.6	−24.0	
G−C	5	−43.5	−18.0	
(A+T)-(A−T)			−5.7	
(G[+]−C)-(G−C)			−18.8	
	Interbase energy		Total interaction energy (interbase + hydration)	Experimental in vacuo interbase interaction[a]
A−T	Gas phase	−14.6	−38.6	−13.0
G−C	”	−22.9	−40.9	−21.0
(A−T)-(A+T)			−8.9	
(G−C)-(G+C)			−4.1	

[a] Soukhodub, L. F., Yanson, J. K.: Nature **264**, 425 (1976).

These results are summarized in the upper part of Table 1. The upper part of Table 2 presents the corresponding results obtained by using −6.7 kcal/mole for the lower limiting value of the interaction energy for "bound" water. In this case the "best" polyhydration scheme corresponds to the order:

$$G(5H_2O)-16.1 > A(4H_2O)-12.0 > T(3H_2O)-6.6 > C(2H_2O)-6.4$$

The results are qualitatively similar in both cases: the hydration energies of the purines are greater than those of the pyrimidines, that of guanine being the greatest. The two pyrimidines have close hydration energies.

4. Hydration and Stability of Base Pairs

We may go a step farther and consider the incidence of hydration upon the stability of the complementary base pairs of DNA. In fact we have computed recently [5] the polyhydration scheme of the adenine-thymine and guanine-cytosine pairs, similarly to the procedure used here for the separated bases. The results indicate that when using the value of −5.1 kcal/mole as a lower limit for "bound" water the first hydration shell of the A–T pair involves six water molecules, with an interaction energy between them and the ligand of −54.6 kcal/mole, and the first hydration shell of the G–C pair five water molecules, with an interaction energy of −43.5 kcal/mole. If we consider the optimum number of water molecules around

Table 2. Energetics of hydration of bases and base-pairs (kcal/mole) (W···W = −6.7 kcal/mole)

Ligand	Number of water molecules	Interaction energy with water	Hydration energy	
A	4	−38.8	−12.0	
T	3	−26.7	−6.6	
G	5	−48.8	−15.3	
G	5	−49.5	−16.1	
C	2	−19.8	−6.4	
A+T	7	−65.5	−18.6	
G+C	7	−69.3	−22.4	
A−T	5	−48.0	−14.5	
G−C	3	−30.9	−10.8	
(A+T)−(A−T)		−17.5	−4.1	
(G+C)−(G−C)		−38.4	−11.6	
	Interbase energy		Total interaction energy (interbase +hydration	Experimental in vacuo interbase interaction[a]
A−T	Gas phase	−14.6	−28.9	−13.0
G−C	,,	−22.9	−33.7	−21.0
(A−T)−(A+T)			−10.3	
(G−C)−(G+C)			−11.3	

[a] Soukhodub, L. F., Yanson, J. K.: Nature **264**, 425 (1976).

the separated bases as indicated in the present study to be, under the same circumstances, four for A, three for T, five for G and four for C, we can see that the formation of the A–T pair displaces one water molecule and the formation of the G–C pair four water molecules from the first hydration shell of the ligands. The effect is thus much more pronounced in the case of the G–C pair. If we take into account the energy of splitting the water–water bonds for the formation of such bonds with the ligand, in the way in which this was done for the separated bases, we come out with a hydration energy of the A–T pair equal to −24.0 kcal/mole and that of the G–C pair equal to −18.0 kcal/mole. These numbers may be compared with the sum of the hydration energies of the two separated constituents of each pair: they are −29.7 kcal/mole for A+T and −36.8 kcal/mole for G+C. From the sole point of view of hydration energies we observe thus a loss of 5.7 kcal/mole when passing from A+T to A−T and of 18.8 kcal/mole when passing from G+C to G−C.

These results are summarized in the corresponding part of Table 1. Table 2 indicates the results obtained by using the value of −6.7 kcal/mole as a lower limit for "bound" water. Qualitatively the two sets of results are similar with differences in details. In particular in Table 2 the difference in the degree of hydration between the A–T and G–C pairs is of two molecules of water instead of one in Table 1.

For a more complete balance of the overall operation of the formation of the complementary pairs we need to take into account also the values of the interaction energies due to the hydrogen bonding between the bases. For this sake we have computed also these interactions within the same OMTP procedure used in this paper for the calculation of the base–water interactions. We find for these inter-base interactions, at the optimal distances taken from the literature [11], −14.6 kcal/mole for the A–T pair and −22.9 kcal/mole for the G–C pair. These values are greater, in particular the value for the A···T pair, than those obtained previously by other theoretical treatments (see e.g. [12–14]) or measured in non-polar solvents (for a recent review see e.g. [15]). They are close, however, to the recent experimental results of mass spectrometric studies of binding energies of the bases *in vacuo* [16], the values found being 13 kcal/mole for A–T and 21.0 kcal/-mole for G–C.

From the combination of these inter-base interaction energies with the results of the hydration energies given above we are led to the conclusions indicated in the bottom part of Tables 1 and 2, namely that: 1) the hydrated G–C pair is altogether more stable than the hydrated A–T pair, the overall interaction energies, inter-base + hydration, being −40.9 and −38.6 kcal/mole, respectively in the approximations of Table 1 and −33.7 and −28.9 kcal/mole in the approximations of Table 2; 2) the hydrated A–T pair is stable with respect to the hydrated separate bases by 8.9 kcal/mole and the hydrated G–C pair is stable with respect to the hydrated separate bases by 4.1 kcal/mole in the approximations of Table 1, the two values becoming 10.3 and 11.3 kcal/mole, respectively, in the approximations of Table 2.

These results indicate that hydration produces a destabilization of both base pairs in comparison to their binding *in vacuo*, this destabilization being particularly strong for the G–C pair. This result is different from that obtained by Sinanoğlu *et al.* [17] and by Rein *et al.* [18] who found that the hydration stabilizes the A–T pair and destabilizes the G–C pair and from the results of Danilov [19] who claims that hydration favours the formation of both base-pairs. Whatever the exact situation with respect to this particular point may be, it is manifest that there is still an appreciable affinity for interbase hydrogen bonding between A and T and between G and C in water solvent and that hydrogen bonding may therefore contribute to the conformational stability of nucleic acids and polynucleotides in aqueous solutions. A confirmation of this viewpoint may be found in recent results of infrared spectroscopy studies [20] on the association of derivatives of uracil and adenine in chloroform solution in the presence of dissolved water which have shown that the extent of association is substantially unchanged by the presence of water.

This result is not in contradiction with the well-known tendency of the bases to stack in water solution (see e.g. [15]). Unfortunately we cannot compare the two tendencies within the approximations of the present computation. These computations evaluate essentially the electrostatic component of interaction energies. They are satisfactory for the study of hydrogen-bonded interactions in which this component is known to be dominant (see e.g. [1–4]). In the stacking interactions, dispersion forces become important [12–14] and have to be taken explicitly into consideration. We are working on the extension of our procedure to include this last component.

References

1. Port, G. N. J., Pullman, A.: FEBS Letters **31**, 70 (1973)
2. Pullman, A., Pullman, B.: Quart. Rev. Biophys. **7**, 505 (1975)
3. Pullman, A., in: Structure and conformation of nucleic acids and protein-nucleic acid interactions, p. 457. M. Sundaralingam and S. T. Rao Eds., Baltimore: University Park Press 1975
4. Pullman, A., Perahia, D.: Theoret. Chim. Acta (Berl.) **48**, 29 (1978)
5. Goldblum, A., Perahia, D., Pullman, A.: FEBS Letters, in press
6. Dreyfus, M.: Thèse de 3è cycle, University of Paris (1970)
7. Chandler, J. P.: Simplex method, Q.C.P.E., Dept. Chemistry, Indiana University, Bloomington, Indiana (1969)
8. Clementi, E., André, J. M., Klint, M. C., Hahn, D.: Acta Phys. Acad. Sci. Hungarica **27**, 499 (1969)
9. Kwiatkowski, J. S., Perahia, D., Pullman, B.: Intern. J. Quantum Chem. **Q.B.S.5.**, in press.
10. Yamaguchi, T., Nagata, Ch.: J. Theoret. Biol. **69**, 693 (1977)
11. Voet, D., Rich, A.: Progress Nucleic Acid Res. Mol. Biol. **10**, 183 (1970)
12. Pullman, A., Pullman, B.: Advan. Quantum Chem. **4**, 267 (1968)
13. Pullman, B., Pullman, A.: Progr. Nucleic Acid Res. Mol. Biol. **9**, 327 (1969)
14. Rein, R., in: Intermolecular interactions: from diatomics to biopolymers, p. 307. B. Pullman, Ed., New York: Wiley Interscience 1978
15. Ts'o, P. O. P., in: Basic principles in nucleic acid chemistry, Vol. 1, p. 453. P. O. P. Ts'o Ed., New York: Academic Press 1974
16. Sukhodub, L. F., Yanson, I. K.: Nature **264**, 245 (1976)
17. Sinanoğlu, O., Abdulnur, S., Kestner, N. R., in: Electronic aspects of biochemistry, p. 301. B. Pullman, Ed., New York: Academic Press 1964
18. Rein, R., Cocckelenbergh, Y., Egan, J. T.: Intern. J. Quantum Chem. Quantum Biol. Symp. **24**, 145 (1975)
19. Danilov, V. I.: Mol. Biol. Repts. **2**, 263 (1975)
20. D'Albis, A., Wickens, M. P., Gratzer, W. B.: Biopolymers **14**, 1423 (1975)

Received July 5, 1978