Review Article: Research in Progress

Aqueous Hydration of Nucleic Acid Constituents: Monte Carlo Computer Simulation Studies

D.L. Beveridge, P.V. Maye, B. Jayaram, G. Ravishanker and M. Mezei
Chemistry Department,
Hunter College of the City University of New York,
695 Park Avenue,
New York, N.Y. 10021.

Abstract

Monte Carlo computer simulations were performed on dilute aqueous solutions of thymine, cytosine, uracil, adenine, guanine, the dimethyl phosphate anion in the gauche-gauche conformation and a ribose and deoxyribose derivative. The aqueous hydration of each molecule was analysed in terms of quasi-component distribution functions based on the Proximity Criterion, and partitioned into hydrophobic, hydrophilic and ionic contributions. Color stereo views of selected hydration complexes are also presented. A preliminary discussion of the transferability of functional group coordination numbers is given. The results enable to comment on two current problems related to the hydration of nucleic acids: a) the theory of Dickerson and coworkers on the role of water in the relative stability of the A and B forms of DNA and b) the idea of water bridges and filaments emerging from the computer simulation results on the hydration of DNA fragments by Clementi.

I. Introduction and Background

The structures of nucleic acids are known to be sensitive to the degree of hydration and other environmental effects. The relative stability of the A and B forms of DNA depends on the relative humidity of the sample fiber, and the sequence dependent conformational transitions to the newly discovered Z-form of DNA are induced by added salt, alcohol and concomitant alterations in water activity (1). A fundamental understanding of these and other important environmental effects in the structural chemistry of nucleic acids can be facilitated by a detailed description of the hydration of individual constituents: the nucleotide bases, phosphate group, and ribose and deoxyribose sugars. Information on the hydration of constituents can then be used as an aid in the interpretation of environmental effects in more complex systems.

With present generation digital computers, it has become feasible to approach the study of aqueous hydration of a dissolved molecule in a rigorous manner by means...
of statistical thermodynamic computer simulation. Here, a large assembly of water molecules, \( \text{O}(100) \), is treated with the interactions between all particles included and the condensed phase environment of the system introduced via periodic boundary conditions. Extensive studies of liquid water have been carried out from this point of view, and the level of agreement between calculation and available experimental data have been extensively characterized (2). Computer simulation studies from this Laboratory have been reported on aqueous solutions of simple ionic (3,4) and hydrophobic solutes (5-8). Only a few studies on aqueous solutions of small hydrophilic solutes have appeared to date (9-11). The pioneering work of Clementi, Corongiu and coworkers (12-14) has demonstrated the feasibility of Monte Carlo computer simulation applied to water, ions and nucleic acids and provided novel results such as the proposed filamentous structure in water networks about the backbone and across the grooves of DNA. Computer simulation studies relevant to nucleic acid hydration problems have also recently been reported by Pohorille, Burt, Pratt and MacElroy (15), and Danilov, Tólokh, Poltev and Malenkov (16). Computer simulations on crystal hydrates of nucleic acids have been reported by Mezei, Beveridge, Berman, Goodfellow, Finney and Neidle (17) and by Kim, Corongiu and Clementi (18). Molecular dynamics studies of nucleic acids have just begun to appear (19-21); and studies including water are known to be in progress (22).

A general procedure for the analysis of local solution environment of a dissolved solute has recently been devised in this Laboratory (9), whereby solvent molecules in the various \( \text{N} \)-molecule configurations are assigned to the closest solute atom (the "proximity criterion"), a procedure formally equivalent to the partitioning of the space in the simulation cell on the basis of the Voronoi polyhedra of the isolated solute (23). The subsequent analysis is organized in the general framework of the theory of quasicomponent distribution functions (24). Using this procedure, the spatial extent of the first hydration shell of the solute can be well defined, and the analysis can proceed on a solute atom, functional group, or subunit basis. This approach has been recently used to study the hydration of the peptide bond (11), glycine zwitterion (25), benzene (8), and solvent effects on the conformational stability of the alanine dipeptide (26,27). Thus, a structured formalism has been produced to relate the results on constituents to those on the whole system.

We describe herein the results of Monte Carlo computer simulations carried out on the aqueous hydration of individual base, phosphate and sugar constituents of nucleic acids at 25°C, and analyzed by means of the proximity criterion. We focus in this article on the structure of the first hydration shell or "hydration complex" for each system studied. A full description of the structural and energetic aspects of each system considered here, presented in context of previous literature on each system and available experimental data will be forthcoming.

II. Calculations

Separate statistical thermodynamical (T,V,N) ensemble Monte Carlo simulations were carried out on dilute aqueous solutions of each of the nucleic acid constitu-
ents under consideration here. Common aspects of the individual calculations are collected in this section, and specific details on a given system are described along with the results. A modified Metropolis procedure incorporating the force bias method and preferential sampling was used for each simulation. A full description of our simulation procedure has been given by Mehrotra et al. (28). The system for study in each case consisted of 216 rigid molecules, one solute and 215 water molecules. All simulations were performed at a temperature of 25°C and a total volume determined from experimentally obtained molar volume of water and solute. The condensed phase environment of the system was provided by means of periodic boundary conditions with solute-solute interactions excluded, thus representing a system at infinite dilution. A typical simulation begins with random positions and orientations of particles in the system and proceeds for an equilibration period of 1000K configurations. Ensemble averages and structural characteristics are computed over the following 100K segment of realization. Convergence was monitored as described in Ref. 29.

The N-particle configurational energy of the system was in each case calculated using potential functions determined from quantum mechanical calculations. For the water-water interactions, the MCY-CI(2) potential developed by Matsuoka et al. (30) was used. For the solute-water interactions, a potential function constructed from the 12-6-1 functional form with the transferable parameters of Clementi and coworkers (31-38) were used. Net atomic charges for atoms of each nucleic acid constituent are those published by Clementi and coworkers or were determined by LCAO-SCF-MO calculations using GAUSSIAN-80 (39) with the basis set described by Matsuoka et al. (40) and consistent with the potential function determination. In the configurational energy calculation, all potential functions for water-water interactions were truncated at a spherical cutoff of 7.75 Å, and solute-water interactions were treated under the minimum image convention.

The analysis of results in each case was carried out by the proximity criterion. The spatial limits of the first hydration shell for each case was determined from the calculated radial distribution functions and thus for a given atom may vary from molecule to molecule. Four or five configurations were then extracted from the realization at well spaced intervals. The coordinates of the hydration complex, including the solute, the first hydration shell and part of the second shell, were transferred to the PROPHET system for graphic display. In the final stages, each of the water molecules in the hydration complex were assigned to solute atoms based on the proximity criterion and color coded based on the mode of hydration: ionic (red and purple), hydrophilic (aqua and green) or hydrophobic (yellow). Stereo views of the hydration complex for each case were produced using an Envision 230 color graphics terminal used in conjunction with the NIH PROPHET system.

III. Results

The structure for the hydration complex for each system studied is described in terms of the total number of water molecules found in the first hydration shell and
the statistical average number of water molecules assigned to each solute atom in the first shell. A stereo view of the hydration complex from a single configuration extracted from each simulation is also given. We caution that no single configuration can be taken as representative of the statistical state of aqueous hydration at ambient temperature, and one must not infer too much from these figures. With this caveat, however, we feel the individual structures are a helpful contribution to understanding the aqueous hydration complex. Fuller details will be available in the original papers on each of the systems under consideration.

**Nucleotide Base Hydration**

Simulations on dilute aqueous solutions of the five bases commonly encountered in nucleic acid structure: thymine (T), cytosine (C), uracil (U), adenine (A), and guanine (G). The first hydration shells of T, C and U were found to contain 23.7, 23.5, and 17.8 water molecules respectively. The statistical distribution of water molecules in the first hydration shell of T is given in Figure 1, for C in Figure 2 and for U in Figure 3. Color stereo views of the aqueous hydration complex of T, C and U are given in Figure 4. The solute-water hydrogen bonds, defined as intermolecular contacts < 3.0 Å, are denoted as dashed lines. Water-water hydrogen bonds under this definition are shown as dotted lines.

The first hydration shells of A and G were found to include 19.6 and 22.5 water molecules, respectively. The statistical distribution of water molecules in the first shell of A is given in Figure 5 and of G in Figure 6. The hydration complex of A and G is shown in Figures 7a and 7b, respectively.

**Phosphate Hydration**

The phosphate group in nucleic acids is found predominantly as a monoanion at physiological pH (41). A prototype molecule for the study of this group in numerous works to date (42) is the dimethylphosphate anion, CH3OPO2OCH3 (DMP−). We continue this practice here, although it has been noted that this molecule is not fully representative of the phosphate in nucleic acid since the 3'-5' polarity is not expressed and the methyl groups do not mimic precisely the interface of the OPO3-O moiety into the backbone. We consider herein the aqueous hydration of DMP− with the phosphodiester torsion angles both assigned gauche values, since the gg conformation is found to predominate in oligonucleotide structures (43) as well as in DNA and RNA. Preliminary results will also be given for [DMP]aq in the trans-trans (tt) conformation.

The calculated first shell hydration complex for DMP− in our simulation of [DMP−]aq was comprised of 25.9 water molecules. The distribution of first shell waters amongst the solute according to the proximity criterion analysis is shown in Figure 8. Of the total waters, 19.2 are assigned to the hydrophobic hydration of the methyl group, 1.4 to ester oxygens, and 5.4 to the hydration of the anionic oxygens. The structure of the hydration complex in a representative configuration of the simulation is
Figure 1. Statistical distribution of waters in the first hydration shell of thymine.

Figure 2. Statistical distribution of waters in the first hydration shell of cytosine.

Figure 3. Statistical distribution of waters in the first hydration shell of uracil.

Figure 4. Statistical distribution of waters in the first hydration shell of guanine.

Figure 5. Statistical distribution of waters in the first hydration shell of adenine.

Figure 6. Statistical distribution of waters in the first hydration shell of DMP⁻ in the gg conformation.
shown in Figure 7c, with the various modes of hydration color-coded. A notable feature of this structure is the PO\textsuperscript{-} hydration, shown in red. The waters here assume positions which can be identified with anionic zones of hydration predicted by Pullman et al. (44) based on the solvation site method. The electrostriction of anionic water structure vis-a-vis that in the hydrophobic hydration shells is clearly evident. The strength of the ionic attraction is thus the dominating force in anionic hydration of DMP\textsuperscript{-}, with N-particle interactions and statistical weight factors a minor contribution. These factors are however expected to be important in hydrophobic and hydrophilic regions, especially the latter.

At the time of this writing, simulations have just been completed on the gauche-gauche (gg), gauche-trans (gt) and the trans-trans (tt) conformers of DMP\textsuperscript{-}. In adiabatic intermolecular phosphodiester potential surfaces for DMP\textsuperscript{-}, the gg, gt, and tt conformers are found to be relatively close in energy (45,46). Our calculations, in conjunction with the solvation free energies calculated using the hydration shell model (47,48), indicate the tt form to be significantly destabilized by hydration, a possible explanation for the fact that only gg and gt have been observed in oligonucleotide crystal hydrate structures. A hydration complex for the DMP\textsuperscript{-} in the tt form is shown on Figure 9a. Full details of these calculations are currently being prepared for publication.

Sugar Hydration

Dilute aqueous solutions of 1'-amino-4'-methyl derivatives of C3'-endo ribose and C2'-endo deoxyribose sugars were studied by computer simulation. The hydroxy, methyl and amino groups were positioned by energy minimization carried out with the MM2 molecular mechanics procedures available on PROPHET. The C3'-endo ribose derivative was found to have 24.2 waters in the first shell hydration complex. The hydration complex of the C2'-endo deoxyribose derivative contained 27.1 waters. The statistical distribution of first shell waters among solute atoms is shown in Figure 10 for the ribo and Figure 11 for the deoxyribo sugar form. Selected structures which have been obtained for the hydration complex of the ribo and deoxyribo systems are shown in Figure 9b and 9c, respectively.

IV. Discussion

The results on the hydration complex of nucleic acid constituents presented here permit us to take up in preliminary form the question of transferability, i.e., the extent to which the first shell coordinations of a given functional group are similar

Figure 4. The Left Panel. Color stereo view of a selected first hydration shell for thymine (top), cytosine (middle) and uracil (bottom).

Figure 7. The Middle Panel. Color stereo view of a selected first hydration shell for adenine (top), guanine (middle) and DMP\textsuperscript{-} in gg conformation (bottom).

Figure 9. The Right Panel. Color stereo view of a selected first hydration shell for DMP\textsuperscript{-} in tt conformation (top), ribose derivative (middle) and deoxy-ribose derivative (bottom).
in different molecules. Table I shows the minimum, maximum and average coordination numbers for the functional groups $-\text{CH}_3$, $>\text{CH}_2$, $>\text{CH}$, $>\text{CH}_2$, $>\text{O}$, $>\text{OH}$, $>\text{CO}$, $=\text{N}$, $=\text{NH}_2$, $>\text{NH}$, and $=\text{N}$ . The amount of data to work with is insufficient for a statistical analysis and the spread in the results is indicated here simply by the ratio of the largest and smallest values found, so the closer this quantity is to unity, the better. An idea of the numerical uncertainty (noise) in the calculations follows from the comparison of the two symmetry related methyl groups of the DMP$^-$ anion. Here atomic coordination numbers may indicate a noise level of 10-20% but the functional group coordination number is good to 5-10%. Since all of the simulations were similar in length, we can expect this assessment to be valid for other functional groups as well. The functional groups $-\text{CH}_3$, $>\text{O}$ and $>\text{OH}$ show a reasonable degree of transferability with an average of $\bar{K} = 9.0, 2.0$ and $0.8$, respectively. We next explored whether the spread in the coordination numbers

![Figure 10. Statistical distribution of waters in the first hydration shell of ribose derivative.](image)

![Figure 11. Statistical distribution of waters in the first hydration shell of deoxy-ribose derivative.](image)

### Table I

Comparison of functional group coordination numbers

<table>
<thead>
<tr>
<th>Functional group</th>
<th>$\langle K \rangle$</th>
<th>$K_{\min}$</th>
<th>$K_{\max}$</th>
<th>$K_{\max}/K_{\min}$</th>
<th>$K'_{\min}$</th>
<th>$K'_{\max}$</th>
<th>$K'<em>{\max}/K'</em>{\min}$</th>
<th>$N_{gr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\text{CH}_3$</td>
<td>9.10</td>
<td>8.52</td>
<td>9.74</td>
<td>1.14</td>
<td>8.52</td>
<td>9.89</td>
<td>1.16</td>
<td>5</td>
</tr>
<tr>
<td>$-\text{CH}_2$</td>
<td>3.91</td>
<td>3.91</td>
<td>3.91</td>
<td>1.00</td>
<td>3.91</td>
<td>3.91</td>
<td>1.00</td>
<td>1</td>
</tr>
<tr>
<td>$&gt;\text{CH}$</td>
<td>1.65</td>
<td>0.97</td>
<td>1.95</td>
<td>2.01</td>
<td>1.56</td>
<td>1.86</td>
<td>1.19</td>
<td>7</td>
</tr>
<tr>
<td>$&gt;\text{CH}_2$</td>
<td>4.26</td>
<td>3.37</td>
<td>5.48</td>
<td>1.63</td>
<td>2.54</td>
<td>7.70</td>
<td>3.03</td>
<td>6</td>
</tr>
<tr>
<td>$&gt;\text{O}$</td>
<td>0.80</td>
<td>0.68</td>
<td>0.98</td>
<td>1.44</td>
<td>0.69</td>
<td>1.12</td>
<td>1.62</td>
<td>4</td>
</tr>
<tr>
<td>$&gt;\text{OH}$</td>
<td>1.98</td>
<td>1.67</td>
<td>2.20</td>
<td>1.32</td>
<td>1.61</td>
<td>2.23</td>
<td>1.36</td>
<td>3</td>
</tr>
<tr>
<td>$&gt;\text{CO}$</td>
<td>3.52</td>
<td>2.33</td>
<td>4.56</td>
<td>1.96</td>
<td>2.50</td>
<td>4.60</td>
<td>1.84</td>
<td>5</td>
</tr>
<tr>
<td>$=\text{NH}_2$</td>
<td>3.54</td>
<td>2.12</td>
<td>5.40</td>
<td>2.55</td>
<td>3.18</td>
<td>3.99</td>
<td>1.23</td>
<td>5</td>
</tr>
<tr>
<td>$=\text{NH}$</td>
<td>2.84</td>
<td>1.27</td>
<td>3.97</td>
<td>1.59</td>
<td>5.12</td>
<td>3.22</td>
<td>1.81</td>
<td>8</td>
</tr>
<tr>
<td>$=\text{N}$</td>
<td>1.87</td>
<td>1.23</td>
<td>2.32</td>
<td>1.89</td>
<td>1.32</td>
<td>2.39</td>
<td>1.81</td>
<td>6</td>
</tr>
</tbody>
</table>

Legend: a. The subscripts min and max refer to the smallest and largest value found for the functional group; b. $<>$ represents average over the $N_{gr}$ occurrence of the functional group in the systems studied.
could be reduced by factoring out the variations in the volume of the first solvation shell that is due to both the variations in conformation and the different cutoffs used by computing the adjusted coordination number $\bar{K}'$ defined as the coordination number $\bar{K}$ divided by the volume of the first shell assigned to that group and multiplied by the average first coordination shell volume. These further results, also shown in Table I, can be summarized as follows: a) The $\bar{K}'$ values for the $\geq \text{CH}$ and $\geq \text{NH}_2$ groups appear to be transferable as opposed to the $\bar{K}$ values for these groups; b) the behaviour of $\bar{K}'$ values is essentially the same as the $\bar{K}$ values for the groups $\geq \text{CH}_3$, $\geq \text{O}$, $\geq \text{OH}$, $\geq \text{CO}$ and $\geq \text{N}$ groups; and c) the $\bar{K}'$ values for the $\geq \text{CH}$ group show significantly larger spread than the $\bar{K}$ values. Overall, we can conclude that all groups where transferability was observed contain a heavy atom in the $\text{sp}^3$ state and that none of the groups containing a $\pi$-bond show transferability for these nucleic acid constituents. It remains to be seen if the variation of the functional group coordination numbers can be correlated with other quantities, like atomic charge, binding energy or neighbour effects.

In conclusion, we wish to comment briefly on two current problems related to the hydration of nucleic acids: a) the theory of Dickerson and coworkers on the role of water in the relative stability of the A and B forms of DNA, and b) the idea of water bridges and filaments emerging from the computer simulation results on the hydration of DNA fragments by Clementi. In a recent article, Dickerson et al. (49) attribute the relative stability of the B form of DNA relative to the A form to a “spine of hydration” appearing as ordered water in the minor groove in the dodecamer $d(\text{CGCGAATTCCGG})$. We only wish to point out that the structure is expected to be extensively hydrated in the major groove as well, although the water in this region turned out to be crystallographically disordered. From an energetic point of view, the fact that waters of hydration are disordered does not necessarily diminish their potential to interact with the nucleic acid in an energetically favorable way. Without direct evidence to the contrary, one must allow for the fact that the disordered water may be contributing as much or more to the stability of the structure as ordered water, and that the spine of hydration found in the minor groove is only part of the story. Moreover, from a free energy point of view, the spine of hydration is in an entropically unfavorable state. Thus, disordered water proximal to a dissolved molecule in a crystal hydrate could contribute more stability to the free energy of the system than ordered water. This is speculative, of course, but we feel it to be worth noting.

The Monte Carlo simulation of Clementi, Corongiu and coworkers on systems of water, ions and DNA (14) has carried the dimensionality of systems treated by computer simulation to new levels of achievement. One of the important observations in their analysis of results was the idea of “filaments” of water in the aqueous hydration of nucleic acids. We feel a point of clarification is in order here. “Filaments”, to us, implies one-dimensional order and preferential stability in this direction. However, in the case of Clementi’s filaments, we note that the water-water interactions are actually less than the corresponding interaction in bulk water (c.f. Figure 9 of Ref. 50). The reason for this situation is fairly clear. The
waters of hydration are interacting strongly with the solute, and are oriented to optimize this interaction at the expense of water-water interactions. The apparent filamentous structure of the water originates as a reflection of quasi-linear order in the solute, not from preferred stability with the water network.

Water bridges between solute atoms have been considered as stabilizing features in nucleic acids since the early work of Lewin (51) on this subject. In examining simulation results on dissolved molecules in water, one frequently finds bridges in individual configurations. However, they rarely persist as the simulation evolves. Generally, with sufficient water available, a solute prefers to make two sequential hydrogen bonds to distinct water molecules rather than a bridge, presumably since the latter is entropically unfavorable. This aspect undoubtedly warrants more detailed study, but for the moment we urge caution in proposals involving water bridges in aqueous hydration unless firm supporting evidence is available.

V. Acknowledgement

These studies have been supported by NIH grant # GM-24914, NSF grant #CHE-8203501 and a series of CUNY Faculty Research Awards. Generous allocations of time on CUNY computer facilities made these calculations possible. Access to the NIH PROPHET system and the help of H. Baig and D. Berry of Bolt, Beranek and Newman on color graphics and molecular mechanics are gratefully acknowledged. Our research design has been influenced considerably by participation in the 1983 CECAM workshop on nucleic acids at the University of Paris at Orsay, organized by W. Olson and C. Moser.

References and Footnotes


Date Received: June 20, 1984

Manuscript transmitted electronically over telephone at 1200 Baud, ASCII.