Hydration of Nucleic Acid Fragments: Comparison of Theory and Experiment for High-Resolution Crystal Structures of RNA, DNA, and DNA–Drug Complexes

Gerhard Hummer,* Angel E. García,* and Dikeos Mario Soumpasis‡

*Theoretical Biology and Biophysics Group T-10, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and ‡Biocomputation Group, Max Planck Institute for Biophysical Chemistry, D-37018 Göttingen, Germany

ABSTRACT A computationally efficient method to describe the organization of water around solvated biomolecules is presented. It is based on a statistical mechanical expression for the water-density distribution in terms of particle correlation functions. The method is applied to analyze the hydration of small nucleic acid molecules in the crystal environment, for which high-resolution x-ray crystal structures have been reported. Results for RNA [r(ApU) · r(ApU)] and DNA [d(CpG) · d(CpG) in Z form and with parallel strand orientation] and for DNA–drug complexes [d(CpG) · d(CpG) with the drug proflavine intercalated] are described. A detailed comparison of theoretical and experimental data shows positional agreement for the experimentally observed water sites. The presented method can be used for refinement of the water structure in x-ray crystallography, hydration analysis of nuclear magnetic resonance structures, and theoretical modeling of biological macromolecules such as molecular docking studies. The speed of the computations allows hydration analyses of molecules of almost arbitrary size (tRNA, protein–nucleic acid complexes, etc.) in the crystal environment and in aqueous solution.

INTRODUCTION

The reactions as well as the intrinsic stability of biological macromolecules depend crucially on interactions with water. As a ubiquitous solvent, water forms an integral part of biological systems. Correspondingly, enormous experimental and theoretical effort has been devoted to elucidating the role of water in biomolecular interactions. Concerning the structural hydration of biological macromolecules, a considerable amount of experimental information has been accumulated, particularly from x-ray and neutron diffraction studies of crystals (see, for instance, Savage and Wlodawer, 1986; Dickerson, 1992) and more recently from multidimensional nuclear magnetic resonance (NMR) measurements in solution (Otting et al., 1991; Liepinsh et al., 1992; Kubinec and Wemmer, 1992). Results for the hydration of nucleic acids have been widely reviewed (Texter, 1978; Saenger, 1984, 1987; Buckin, 1987; Westhof and Beveridge, 1990; Berman, 1991, 1994). An extensive analysis of the crystallographic data for the hydration of DNA oligomers was presented recently (Schneider et al., 1992a, 1993). Nevertheless, our understanding of biomolecular hydration phenomena is still rather limited (Levitt and Park, 1993).

The theoretical modeling of hydrated biomolecules relies mostly on the use of computer simulation methods. Monte Carlo and molecular dynamics simulations offer versatile and generally applicable tools for the study of systems containing biomolecules in solution. The necessary interaction potentials are reasonably well developed to describe adequately various properties of hydrated biomolecules, if technical difficulties are handled with the required care.

The major problem in computer simulation studies of biomolecular hydration phenomena is the rather inefficient statistical sampling of quantities such as the local water density. This is a consequence of the large system sizes involved and the slow relaxation times of many structural quantities in aqueous phases, resulting in enormous computing times. In addition, the rather open liquid-water structure with a low bulk-water particle density of $\rho_0 \approx 33$ nm^{-3} and the correspondingly poor density statistics further hamper studies of the structural hydration of solvated macromolecules. Therefore, it is important to develop alternative theoretical tools for studying hydration phenomena. These methods have to emphasize the computational efficiency to allow the biologically important analysis of large systems such as protein–nucleic acid complexes or large numbers of mutations of proteins (Zhang and Matthews, 1994).

Here we describe an efficient and accurate method to calculate the detailed structural organization of water near biological macromolecules. It is based on the potentials-of-mean-force (PMF) formalism applied earlier to describe ionic effects on nucleic acids (Soumpasis, 1993). The method presented is orders of magnitude faster than computer simulations and allows us to study the solvation of macromolecules of almost arbitrary size. An outline of the method in the context of the ice–water interface was given previously (Hummer and Soumpasis, 1994a). Some results for DNA oligomers (Hummer and Soumpasis, 1994b) and preliminary results for the structural hydration of proteins (García et al., 1994) and A-, B-, and Z-DNA (Hummer et al., 1994) have been described. For ideal B-DNA oligomers, a comparison with published simulation results by Forester and McDonald (1991) was recently presented (Hummer and Soumpasis, 1994c).

In this paper we present a description of the theoretical framework, putting particular emphasis on discussing the approximations and limitations involved. Careful testing of the method is crucial for appraising its practical value. This is accomplished by comparison with high-resolution experimental data for the structural hydration. Among the currently available experimental techniques, x-ray crystallography provides the most detailed information about the solvent structure, achieving resolutions in the 1-Å range for atom positions. We will compare the calculated water densities with the water positions in highly refined crystal structures of nucleic acids fragments, for which atomic coordinates are available in the nucleic acid database (Berman et al., 1992).

We discuss results for a broad set of nucleic acids: r(ApU) · r(ApU) and d(CpG) · d(CpG) in Z form and with parallel strand orientation and d(CpG) · d(CpG) with the drug proflavine intercalated. The comparison with high-resolution experimental data allows us to assess the quality of the method presented and the validity of its approximations. It also demonstrates an important application, i.e., the theoretical support of the assignment of water molecules in protein and nucleic acid crystals, and it develops the physical basis for future investigations of the hydration of large molecules, for which lower-resolution NMR or x-ray crystal structures exist but do not contain detailed information about the solvent organization.

THEORY

The development of the theoretical framework is organized as follows. We express the water-density distribution near a macromolecule (in solution or in the crystal environment) in terms of particle correlation functions. This exact relation is then expanded to yield an expression using lower-order correlation functions. We will truncate this expansion after the three-particle level, i.e., using only pair- and triplet-correlation functions. The next step is the reduction of the number of correlation functions required by classifying the atom types as polar and nonpolar, resulting in a working formula for biomolecular hydration calculations.

Statistical mechanical treatment of the biomolecular hydration

The central quantity in the PMF hydration calculations is the water-density distribution. This function describes the structural equilibrium properties of the water phase around a biomolecule. For macromolecules in dilute solution, we expect the density distribution to approach the bulk-water density of 1 g/cm³ within approximately 1 nm from the molecular surface. At the molecular surface, the density distribution shows a distinct structure, reflecting the preferential arrangement of the water molecules at the interface. In a crystal environment, the density distribution is expected to show even more ordering as a consequence of the stronger inhomogeneity of the system.

To compute the conditional solvent density at a position r_1, we apply a statistical mechanical description of inhomogeneous systems, expressing the density as a configuration space integral. The inhomogeneity in the water phase is caused by the presence of the solvated macromolecule. To simplify the presentation, results for an atomic solvent will be given. However, the generalization to molecular solvents is straightforward.

We describe the solvated macromolecule by atomic coordinates r_{i_α} of N_α atoms of type α and M different types of atoms α. The set of atomic coordinates (obtained from x-ray crystallography, NMR, modeling, etc.) will be denoted by $\{r_{i_\alpha}\}$. We assume a rigid equilibrium structure of the solvated molecule. However, it is possible to relax this assumption in refined versions of the method by using representative ensembles of structures. These can, for instance, be obtained from structural data consistent with NMR constraints or by using the temperature factors from x-ray crystallography to describe the thermal motion.

The solvated macromolecule appears as an external field in the Boltzmann weighting factor of the configuration space integral. For a canonical ensemble of N solvent particles with coordinates $\{r_i\}$ we find for the conditional solvent density

$$\rho(r_1 \mid \{r_{i_\alpha}\}) =$$

$$N \frac{\int dr_2 \ldots dr_N \exp[-\beta U(\{r_i\}, \{r_{i_\alpha}\})]}{\int d\{r_i\} \exp[-\beta U(\{r_i\}, \{r_{i_\alpha}\})]}, \quad (1)$$

where $\beta = 1/k_B T$ and U is the total potential energy in a classical mechanical description. At this point, the computer simulation and the PMF expansion methods separate. The former attempts to calculate the conditional density distribution directly from the configuration space integral representation in Eq. 1 by means of statistical averages of the density for small volumes on a grid. The PMF expansion method, on the other hand, further simplifies Eq. 1 by expressing it in terms of structural correlation functions (Hummer and Soumpasis, 1994a, c):

$$\rho(r \mid \{r_{i_\alpha}\}) = \rho_0 \frac{g^{(1;\{N_\alpha\})}(r, \{r_{i_\alpha}\})}{g^{(\{N_\alpha\})}(\{r_{i_\alpha}\})}, \quad (2)$$

where $\rho_0 = N/V$ is the solvent density. Multiparticle correlation functions describe the molecular structure of equilibrium fluids and are well-known quantities in the statistical mechanical theory of fluid phases (Hansen and McDonald, 1986). In particular, pair correlation functions (and the corresponding structure factor obtained by Fourier transformation) have been extensively studied for numerous condensed matter systems by both experimental and theoretical means.

As the first step of the formal development, we have obtained an expression for the water-density distribution near a solute molecule in terms of particle correlation functions. The next step requires us to simplify this expression, using the PMF expansion.

Potentials-of-mean-force expansion of the water-density distribution

The higher-order correlations of Eq. 2 are complicated functions of their positional arguments, and they are not easily amenable to a direct calculation. Therefore, to obtain a numerically tractable method, we have to find an approximate representation in terms of lower-order correlation functions. This is accomplished by using the PMF expansion for the correlation functions (Kirkwood, 1935; Münster, 1969; Hummer and Soumpasis, 1994a). Applied to Eq. 2, we obtain an expansion for the density in terms of lower-order correlations (Hummer and Soumpasis, 1994c):

$$\rho(\mathbf{r} \mid \{\mathbf{r}_{i_\alpha}\}) = \rho_0 \left[\prod_{\alpha=1}^{M} \prod_{i_\alpha=1}^{N_\alpha} g^{(1;\alpha)}(\mathbf{r}, \mathbf{r}_{i_\alpha}) \right]$$

$$\left[\prod_{\alpha=1}^{M} \prod_{i_\alpha=1}^{N_\alpha} \prod_{\beta=\alpha}^{M} \prod_{i_\beta=1+\delta_{\alpha\beta}i_\alpha}^{N_\beta} \right. \tag{3}$$

$$\left. \frac{g^{(1;\alpha,\beta)}(\mathbf{r}, \mathbf{r}_{i_\alpha}, \mathbf{r}_{i_\beta})}{g^{(1;\alpha)}(\mathbf{r}, \mathbf{r}_{i_\alpha})g^{(\alpha,\beta)}(\mathbf{r}_{i_\alpha}, \mathbf{r}_{i_\beta})g^{(\beta;1)}(\mathbf{r}_{i_\beta}, \mathbf{r})} \right]...,$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol. The first term in the expansion is the bulk density of water. The second term is a product of pair correlations over all atoms of the solute molecule, and the third term gives the three-particle correction as a product over all distinct pairs of atoms on the solute molecule. The accuracy of the expansion is expected to increase with the number of terms retained. However, at least for water, the essential information is contained in the lowest-order terms (Hummer and Soumpasis, 1994a).

From a systematic expansion of the formally correct expression Eq. 2, we have now obtained an approximate representation of the water-density distribution as a product of lower-order correlation functions. The next step involves the reduction of the required database of correlation functions.

Atom classes—polar and nonpolar

The direct application of Eq. 3 to calculating the water-density distribution near a solvated macromolecule would require knowledge of a prohibitively large number of two- and three-particle correlations between atoms of the macromolecule and water oxygen. Therefore, it is essential to reduce the number of different atom types by grouping them into similar classes with respect to their interactions with water.

The dominant interactions between water and a hydrated macromolecule are caused by hydrogen bonding between polar and charged groups and water molecules. This observation allows us to simplify Eq. 3 with respect to the number of different atom types α used to model the macromolecule. The electronegative oxygen and nitrogen atoms in nucleic acids exhibit a strong tendency to form hydrogen bonds with water molecules. Geometrically, these hydrogen bonds are similar (in bond length and bond angle) to those formed between water molecules.

In the following, we will obtain a first-order picture by equating all DNA nitrogen and oxygen atoms to water oxygen with respect to their effect on ordering water. This yields the working formula for calculating the water-oxygen density around any configuration of electronegative atoms (oxygen and nitrogen for nucleic acids) at positions $\mathbf{r}_1,...,\mathbf{r}_n$:

$$\rho_0(\mathbf{r} \mid \mathbf{r}_1,..., \mathbf{r}_n) = \rho_0 \prod_{i=1}^{n} g_{OO}^{(2)}(\mathbf{r}, \mathbf{r}_i)$$

$$\prod_{j=1}^{n-1} \prod_{k=j+1}^{n} \frac{g_{OOO}^{(3)}(\mathbf{r}, \mathbf{r}_j, \mathbf{r}_k)}{g_{OO}^{(2)}(\mathbf{r}, \mathbf{r}_j)g_{OO}^{(2)}(\mathbf{r}_j, \mathbf{r}_k)g_{OO}^{(2)}(\mathbf{r}_k, \mathbf{r})}, \tag{4}$$

where ρ_0 is the bulk-water density. $g_{OO}^{(2)}$ and $g_{OOO}^{(3)}$ are the two- and three-particle correlation functions of water oxygens in bulk water, respectively.

The nonpolar carbon atoms are represented as excluded volume. Inside a sphere with radius 0.3 nm the water-oxygen density is set to zero. This corresponds to approximating the pair correlations in Eq. 3 by a unit step function and neglecting the three-particle correlations if they involve a nonpolar atom. The size of the exclusion sphere corresponds approximately to the distance of closest approach of a water oxygen to a solvated methane. For molecules with a large hydrophobic surface it might prove essential to refine this description. This can be accomplished along the lines suggested by Eq. 3, for instance, by using the correlation functions of methane in water to describe the effects of nonpolar atoms.

The molecular surface of nucleic acids is essentially hydrophilic. Therefore, we expect this first-order treatment to yield sufficiently accurate results. With respect to the phosphate groups, we expect the negative charge to be well represented by the four oxygens, each carrying a negative partial charge. We have also described a refined treatment modeling the two anionic oxygens on the PO_4^- group explicitly, resulting in excellent agreement of the phosphate hydration with computer simulation results (Hummer and Soumpasis, 1994c). We will in the following adopt the simple modeling of the phosphate groups with all four phosphate oxygens represented by water oxygens. Supported by the agreement with the experimental data, we consider this treatment to describe the phosphate hydration adequately.

With the classification of the atoms as polar and nonpolar, we have obtained a working model for biomolecular hydration studies. We will now discuss the required input information, i.e., the correlations database.

Water correlation functions

The local water density around a solute molecule is given by Eq. 4. Based on the assumptions described above, this formula requires the two- and three-particle correlation functions of water-oxygen atoms in bulk water. For "simple" (monoatomic) fluid systems with additive pair interactions depending only on the particle distance, the PMF density expansion up to the pair level (Kirkwood superposition ap-

proximation; Kirkwood, 1935) is expected to be quantitatively reliable. At least at high densities, the atomic structure of these liquids is determined mostly by packing effects. This leads to the success of perturbation theories of the structure and thermodynamics of "simple" liquids, and the structure caused by dense packing is described well by the Kirkwood superposition approximation.

However, for water with its strongly anisotropic interactions, the three-particle corrections are important. The oxygen atoms of three water molecules in contact preferentially form isosceles triangles with edges of approximately 0.275-, 0.275-, and 0.45-nm length rather than equilateral triangles with nearest-neighbor distances of 0.275 nm. This structure is characteristic for ice lattices and reflects the locally icelike, tetrahedral arrangement in the liquid-water phase. The triplet correlations correct for this hydrogen-bond structure. In a study of the interface of ice and water, we have shown that it is essential to retain the triplet term in the density expansion (Hummer and Soumpasis, 1994a). We have calculated the water-density distribution at the interface of ice Ih and water, using the PMF expansion method, and compared the data with extensive Monte Carlo computer simulations. At the two-particle correlations level we have observed considerable discrepancies. However, inclusion of the three-particle terms resulted in quantitative agreement of calculated and simulated density distributions.

Triplet correlations in water were first calculated by Soumpasis et al. (1991) for the ab initio NCC water model (Niesar et al., 1990). We subsequently calculated the oxygen pair and triplet correlation functions $g_{OO}^{(2)}$ and $g_{OOO}^{(3)}$ of the simple point charge (SPC) model of water (Berendsen et al., 1981) and of TIP3P water (Jorgensen et al., 1983). We observe little effect of the water model used on the water-density distributions obtained from the PMF expansion, because different water models give similar results for the two-particle correlations $g_{OO}^{(2)}$; also, the three-particle correlations $g_{OOO}^{(3)}$ show little variation. For this study we use the SPC correlations database.

The SPC water model has been used extensively in computer studies of solvated biomolecules. For a homogeneous and isotropic system, the pair and triplet correlation functions depend only on one and three pair distances, respectively. They were calculated previously from Metropolis Monte Carlo simulations at discrete intervals of 0.005 and 0.02 nm up to pair distances of 1.1 and 0.72 nm, respectively. The details of these calculations are described elsewhere (Hummer and Soumpasis, 1994a).

In nucleic acids, polar atoms are found as close as ~0.2 nm. Inasmuch as oxygen pair distances below 0.25 nm seldom occur in SPC water at room temperature and atmospheric pressure, the three-particle corrections for pair distances between 0.19 and 0.25 nm were calculated in separate constrained simulations (Hummer and Soumpasis, 1994c). The discrete two- and three-particle correlation functions are used along with linear and trilinear interpolation routines for the pair distances in Eq. 4.

RESULTS AND DISCUSSION

It is important to analyze the validity of the approximations in a new theoretical formalism by comparison with experimental data. The most detailed experimental results for structural hydration have been obtained by x-ray crystallography. These crystallographic studies report the atomic positions of macromolecules, water, and ions in the asymmetric unit of a crystal, derived from the observed electron density.

In our calculation of biomolecular hydration, we obtain a water-density distribution. At this stage of the development of our method, we are concerned mainly with correctly predicting sites of high probability of finding a water molecule, i.e., high-water-density regions. The identification of strongly localized water molecules at the biomolecular surface is of particular importance in biomolecular hydration studies. Correspondingly, the comparison of calculated and experimental data is centered around the question of whether the observed positions of water molecules in crystals agree with regions of high calculated water density.

To answer this question, we will first compare the calculated high-density regions graphically with the observed water sites. We compute the water density in the crystal environment on a Cartesian grid (0.02-nm width; 0.03 nm for the proflavine crystal). Grid points of highest density are then shown together with the crystal water sites. A threshold of five times the bulk-water density ρ_0 allows us to illustrate the characteristic features of the hydration structure. The value of the threshold is somewhat arbitrary. It was chosen such that regions of highest water density can clearly be identified without overburdening the graphical representation of the complex three-dimensional water-density distribution. Positional agreement between crystallographic and theoretical studies is achieved if the experimental water sites fall within the calculated high-density regions.

The next step of the comparison is quantitative. We compute the radial density distribution $g(r)$ around each crystallographic water site. $g(r)$ is defined as the average water-oxygen density (in units of the bulk density ρ_0) on a sphere with radius r around a given water site in the crystal. It is calculated by choosing 1000 points randomly distributed on spheres with radii in intervals of 0.01 nm. Agreement with experimental data is obtained if the density distributions $g(r)$ are peaked at or near $r = 0$.

We also calculate integrated hydration numbers $N(r)$, i.e., the average number of water molecules within a given distance from a crystal water site:

$$N(r) = 4\pi \rho_0 \int_0^r ds \, s^2 g(s). \tag{5}$$

$N(r)$ is expected to reach a value of 1 (one water molecule) within $r \approx 0.2$ nm.

We now proceed to compare our results with experimental data.

The RNA dinucleoside monophosphate r(ApU) · r(ApU)

Using Eq. 4, the water-oxygen density has been calculated in a crystal of RNA dinucleoside monophosphates r(ApU) · r(ApU) (Seeman et al., 1976). The characteristics of the crystal are compiled in Table 1. One of the terminal O5' atoms is found at two partly occupied positions. Results will be shown only for the dominant position of the terminal O5' atom (occupancy 0.73), because they agree closely with those computed for the structure with the weakly occupied O5' site. Here and in the following we use the coordinates deposited in the nucleic acid database (NDB) (Berman et al., 1992) in conjunction with the crystal-symmetry information.

Figs. 1 and 2 show major- and minor-groove views of the molecule. The grid points of highest density are marked by cyan spheres. Crystal water sites are shown as slightly larger magenta spheres. The PMF calculation shows a partly connected network of high-water-density regions around the dinucleoside. The water molecules of the crystal are located within these PMF high-water-density regions. This observation of a high-water-density network with the crystal sites embedded agrees with the crystallographic observations of Seeman et al. (1976). Those authors ascribe only qualitative value to the conventional model that uses Gaussian distributions of the solvent around lattice sites but rather propose to view the system as a solid–liquid interface showing thermal and statistical disorder in the liquid phase. The high-water-density channels observed in the PMF calculations are connecting the crystal sites and therefore reflect this inherent disorder in the solvent organization. The four crystal water sites described with partial occupancies fall within such large clusters of high PMF water density (Fig. 2). However, the PMF calculation also shows some distinctly localized water regions, although not with a strictly spherical (or ellipsoidal) shape (Fig. 1).

The water-oxygen sites in the crystal are at least within 0.1 nm of regions with PMF density larger than 5 ρ_0, except for water site W9 (0.12-nm distance), which is loosely coordinated to a sodium ion (water site W9 is located at the lower left of the sodium ion in the center of Fig. 2; the numbering scheme follows Seeman et al., 1976). Interestingly, several of the water molecules bound to the sodium ions (2 cations per asymmetric unit) are also reproduced as high-density regions, although the sodium ions are not considered in the PMF calculation. In particular, the two water molecules coordinated to the sodium cation located between two phosphate groups appear as distinct high-density regions (upper left quadrant of Fig. 2). However, at the sodium sites the PMF calculation does not show high water density, supporting the assignment of the two cations in the crystallographic study.

To permit a quantitative comparison of x-ray results for the hydration and PMF calculations, Fig. 3 shows the radial water-density distribution function $g(r)$ around each of the 14 water sites in the crystal. The $g(r)$ curves are grouped according to their maximum values. For water-oxygen sites W1–W7 and W10 we observe distinct peaks of the PMF density at $r = 0$. For three of the sites with partial occupancy (W11, W13, and W14) the peak in $g(r)$ is shifted to r between 0.05 and 0.1 nm. The fourth site with fractional occupancy (W12) shows a weakly peaked but broad $g(r)$. The sites W8 and W9 show an unstructured $g(r)$. Both the W8 and the W9 water sites are within 0.3 nm of a sodium cation, which was not considered in the calculations. This explains the failure of the theory to reproduce the two hydration sites. It also indicates that it would be important to take into account strongly localized ions (in particular, multivalent cations coordinated to DNA) to describe accurately the hydration in their vicinity.

Fig. 4 shows the $N(r)$ curves grouped as in Fig. 3. For most of the sites, $N(r)$ reaches a value of 1 for r close to 0.2 nm. Interestingly, the strongly peaked $g(r)$ curves around water sites W1, W2, and W7 do not contain enough water-oxygen density within the peak region. This indicates that the PMF calculation might underestimate the water density in these high-density regions. On the other hand, the water sites with a broad $g(r)$ contain one water molecule within roughly 0.2 nm, with most contributions coming from the tails of $g(r)$.

In essence, we observe a structured network of high water density with the crystal water sites embedded, reflecting thermal disorder in a solid–liquid-like interface. The radial density distributions $g(r)$ are peaked at or near the crystal sites, except for two water sites coordinated to a sodium ion, indicating positional agreement of theory and experiment.

The deoxydinucleoside monophosphate d(CpG) · d(CpG)

The crystal structure of the dinucleoside pair d(CpG) · d(CpG) has been determined from x-ray data by Ramakrishnan and Viswamitra (1988; see Table 1). Figs. 5 and 6 show major- and minor-groove views of the hydration of the molecule. Again, the PMF calculation yields a network of high

TABLE 1 Characteristics of the nucleic acid crystals used in the study

Composition*	Space Group	Resolution‡	R Factor	NDB Code§	Conformation	Reference
[r(ApU)]$_2$·[Na$^+$]$_2$·[H$_2$O]$_{12}$	P2$_1$	0.08	0.057	ARB002	A-RNA	(Seeman et al., 1976)
[d(CpG)]$_2$·[NH$_4{}^+$]$_2$·[H$_2$O]$_{13}$	P2$_1$2$_1$2$_1$	0.085	0.136	ZDB020	Z-DNA	(Ramakrishnan and Viswamitra, 1988)
[d(CpG)]$_2$·[Na$^+$]·[H$_2$O]$_7$	P2$_1$2$_1$2$_1$	0.086	0.041	UDB005	Parallel DNA	(Coll et al., 1987)
[d(CpG)]$_2$·[Pf]$_2$·[H$_2$O]$_{25}$	P2$_1$2$_1$2	0.083	0.15	DDB009	DNA/drug	(Shieh et al., 1980)

*Composition of the asymmetric unit (Pf is proflavine).
‡Crystallographic resolution in nanometers.
§Nucleic acid database (NDB) identification (Berman et al., 1992).
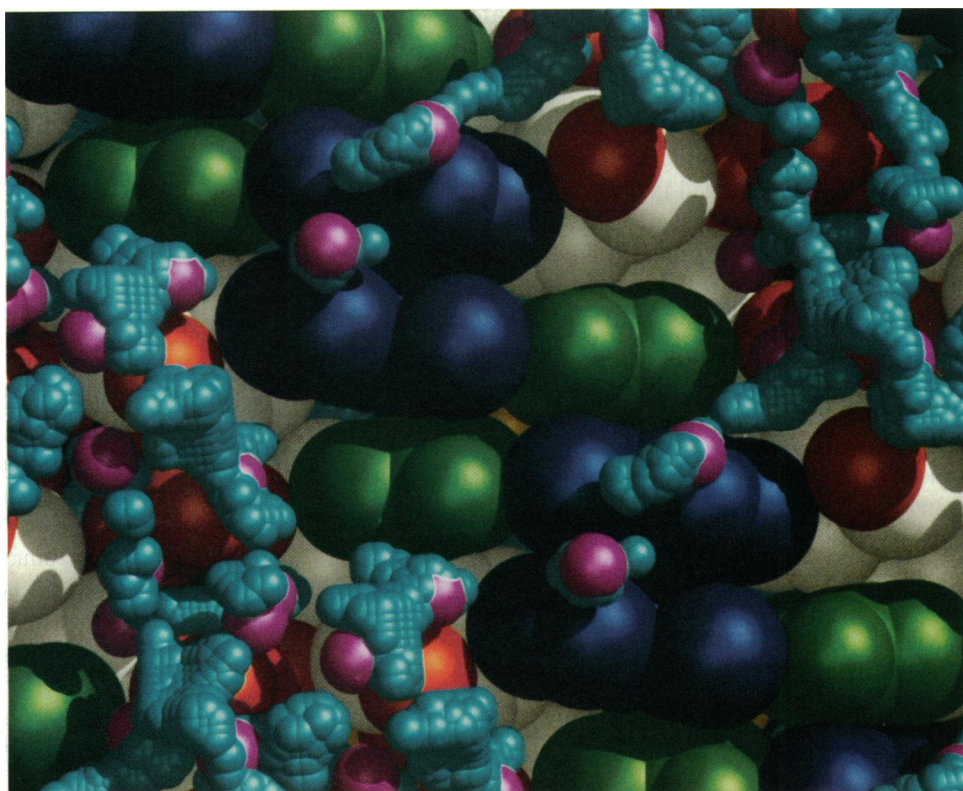
FIGURE 1   Major-groove view of the hydration of the RNA dinucleoside monophosphate r(ApU) · r(ApU) in the crystal environment. Purine bases (here: adenine) are shown in blue, pyrimidines (here: uridine) in green, sugars in white, and atoms of the PO_4^- group in red. Small magenta and yellow spheres (radius $R = 0.08$ nm) show the positions of water oxygens and ions (here: sodium) in the crystal (Seeman et al., 1976). Positions marked by small cyan spheres ($R = 0.05$ nm) indicate points of highest water density in the PMF expansion ($\geq 5\rho_0$).

FIGURE 2   Minor-groove view of the hydration of the RNA dinucleoside monophosphate r(ApU) · r(ApU) in the crystal environment. Details as in Fig. 1. The four water sites found to have partial occupancies in the crystal can be found in the lower left quadrant of the picture. They can be identified because the small 0.08-nm magenta spheres partially overlap.
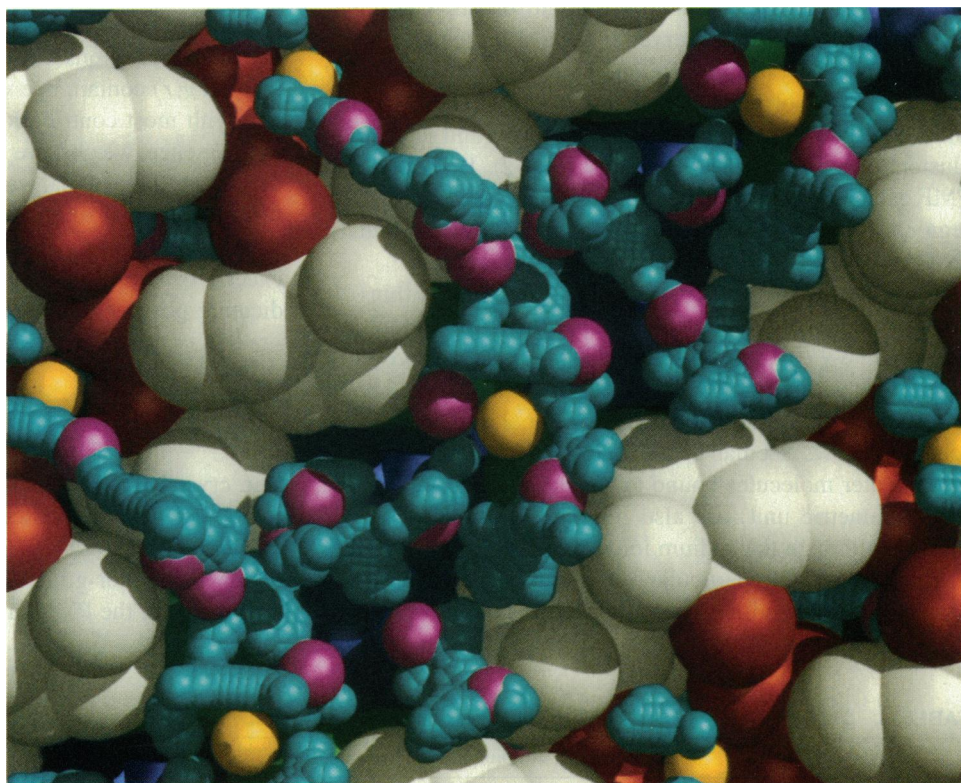
water density with the crystal water molecules embedded. The maximum distance of crystal water sites to regions with PMF density $\rho > 5\rho_0$ is 0.1 nm, indicating positional agreement of the PMF and crystal water sites.

Interestingly, the PMF computation also shows high water density at the ammonium-ion sites in the crystal (Fig. 6), suggesting that these sites might actually be occupied by water. In the crystallographic assignment it is difficult to
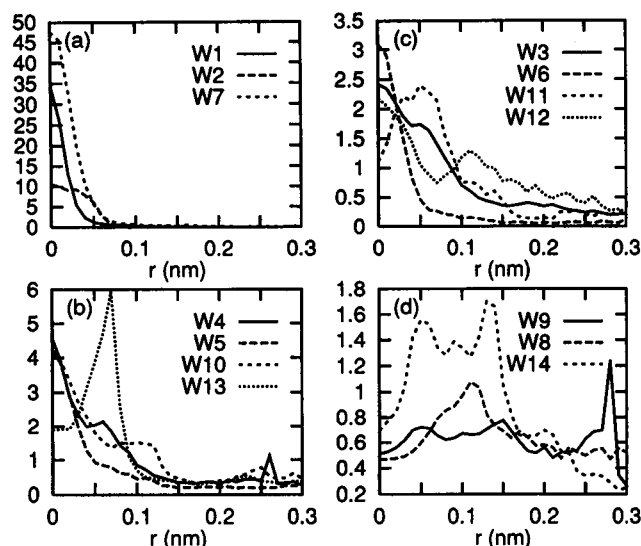
FIGURE 3   Radial distribution functions $g(r)$ of the water-oxygen density (in units of the bulk-water density ρ_0) around the 14 crystal water sites of the RNA dinucleoside monophosphate r(ApU) · r(ApU) (Seeman et al., 1976). The $g(r)$ curves are grouped into four panels according to their maximum values.
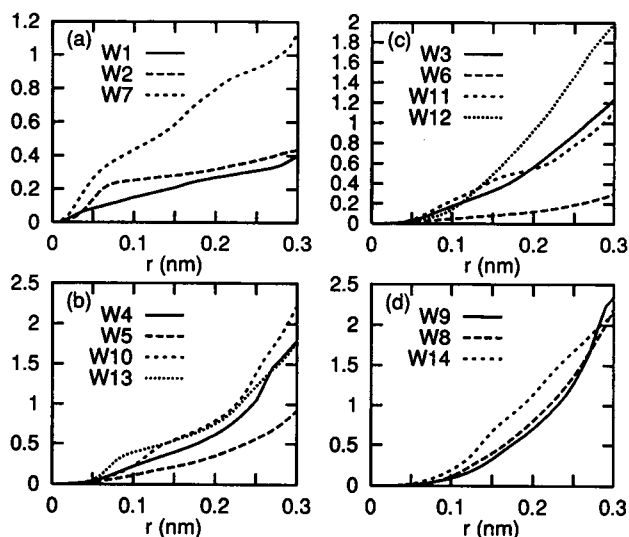

FIGURE 4   Hydration numbers $N(r)$ for the 14 crystal water sites of the RNA dinucleoside monophosphate r(ApU) · r(ApU) (Seeman et al., 1976). The $N(r)$ curves are obtained by integration of $g(r)$ from Fig. 3. The $N(r)$ curves are grouped into four panels as in Fig. 3.

distinguish the electron density of water and NH_4^+ cations. Ramakrishnan and Viswamitra (1988) have identified the ammonium ions based on the intermolecular interactions. The maximum PMF water density near their ammonium site N2 is approximately seven times the bulk-water density. The NH_4^+ site N1, on the other hand, has an even higher water density peak of ~13 in 0.1-nm distance. We conclude that at least site N1 might be partially occupied by water. This is further corroborated by the presence of high-water-density regions at the water sites in contact with the two NH_4^+ sites.

These water peaks are caused entirely by the DNA molecule, as the NH_4^+ ions are not considered in our calculation, and the PMF calculation shows that they are consistent with peaks at the NH_4^+ sites.

Figs. 7 and 8 show the radial water-oxygen density distribution $g(r)$ around each of the crystal water sites and the corresponding integrated hydration number $N(r)$, as defined in Eq. 5. Eight of the $g(r)$ curves for the 13 water sites are strongly peaked around $r = 0$ and decay rapidly within 0.05–0.1 nm. Fig. 7 (c) shows a group of rather broad $g(r)$ curves, where two sites (W10 and W11) show peaks shifted to $r \approx$ 0.05 nm. The $N(r)$ curves shown in Fig. 8 reach a value of 1 for r typically between 0.2 and 0.3 nm. Also included in Figs. 7 and 8 are the $g(r)$ and $N(r)$ curves for the two NH_4^+ sites. In both cases we observe a strongly peaked $g(r)$, with peak heights of ~4.5 at $r = 0$. $N(r)$ reaches 1 at $r \approx 0.25$ nm. This further supports the possibility that the two sites assigned to ions in the x-ray crystal structure might actually be occupied by water molecules.

Difficulties in the assignment of NH_4^+ ions were also reported in the x-ray crystallographic determination of the structure of the RNA dinucleoside r(GpC) · r(GpC) (Aggarwal et al., 1983). A subsequent computer simulation study of the crystal hydrate (Elliott and Goodfellow, 1987) showed significant deviations of the NH_4^+ ion positions between x-ray structure and simulation.

Interestingly, the highest $g(r)$ peaks in Fig. 7 are significantly lower than those found for the RNA dinucleoside (Fig. 3). However, it is important to keep in mind that the $g(r)$ curves involve angular averages over complex density distributions. Small deviations of the density peak from the water site in the crystal can result in a considerable reduction of the $g(r)$ peak height. The angular average may not resolve this positional discrepancy, because a strongly localized high-density region can be suppressed by larger regions with low density. A similar problem also arises in the fitting of x-ray electron densities with complex shapes by Gaussian distributions, where actual and fitted peak positions may differ.

Overall, we observe positional agreement between the crystal water sites and regions of high calculated water density for the d(CpG) · d(CpG) crystal. However, our analysis suggests that the ammonium-ion sites in the crystal might at least be partly occupied by water.

The deoxydinucleoside monophosphate d(CpG) · d(CpG) with parallel strand orientation

Two x-ray crystal structures were reported for the dinucleoside monophosphate duplex d(CpG) · d(CpG) with parallel strand orientation, which forms non-Watson–Crick G·G and protonated C·C base pairs (Cruse et al., 1983; Coll et al., 1987). Here, we use the crystal structure of the sodium salt by Coll et al. (1987) with a slightly lower R factor (see Table 1).

Fig. 9 shows the structural hydration of the molecule in the crystal environment. The PMF calculation of the water-

FIGURE 5 Major-groove view of the hydration of the deoxydinu-cleoside monophosphate d(CpG) · d(CpG) in the crystal environment (Ramakrishnan and Viswamitra, 1988). Details as in Fig. 1.
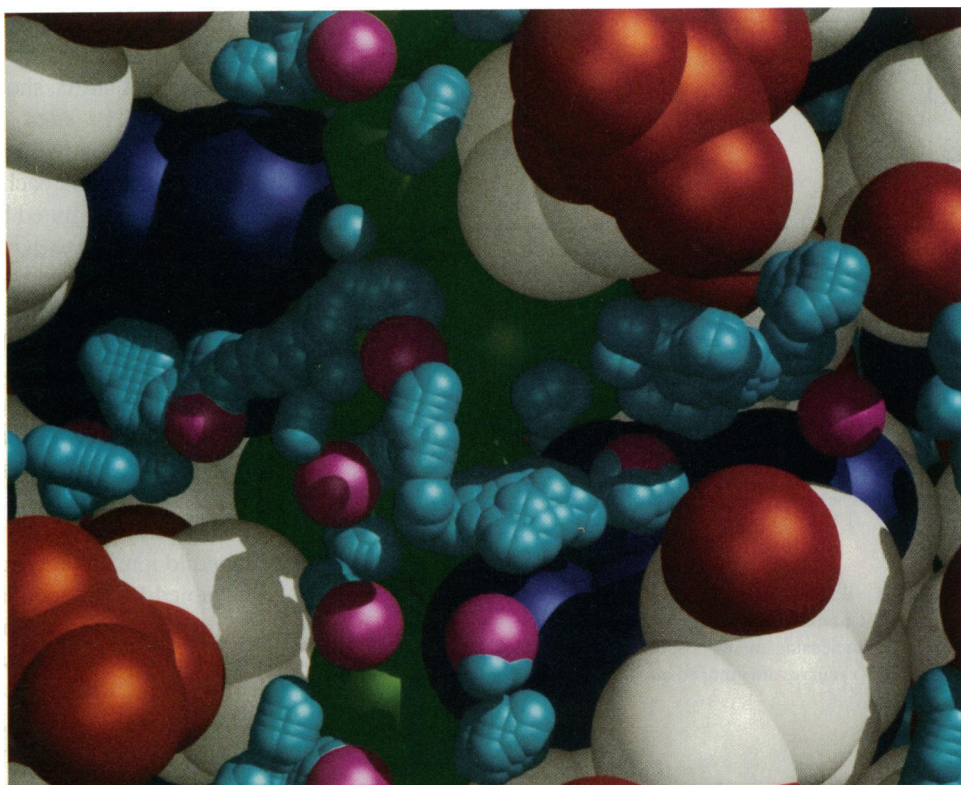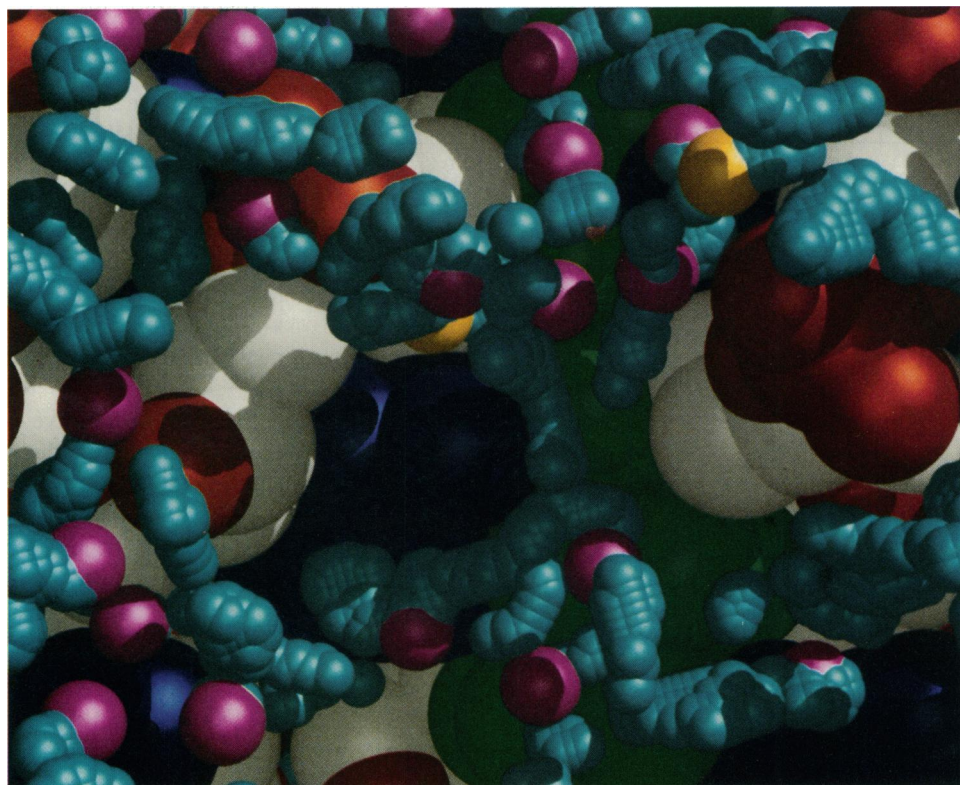
FIGURE 6 Minor-groove view of the hydration of the deoxydinu-cleoside monophosphate d(CpG) · d(CpG) in the crystal environment. Details as in Fig. 1.

density distribution shows a more strongly structured hydration than the network-type hydration of the previously studied molecules. The theoretical results show peaks at or close to five of the seven water sites in the crystal. We do not observe considerable water density at the Na^+ site of the crystal and at the two water sites hydrating the ion, supporting the crystallographic assignment. However, we observe two density peaks between the two water sites W2 and W3
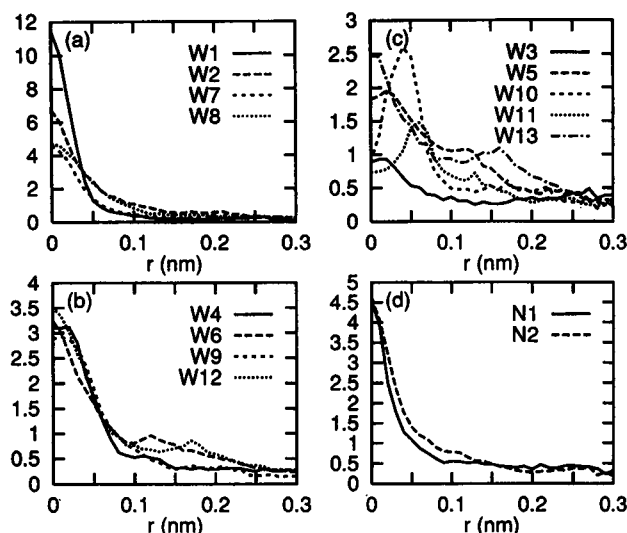
FIGURE 7 Radial distribution functions $g(r)$ (in units of ρ_0) of the water-oxygen density in the crystal around the 13 water sites $[(a)-(c)]$ and the two sodium-ion sites (d) of the deoxydinucleoside monophosphate $d(CpG) \cdot d(CpG)$ in the crystal (Ramakrishnan and Viswamitra, 1988). The $g(r)$ curves are grouped according to their maximum values.
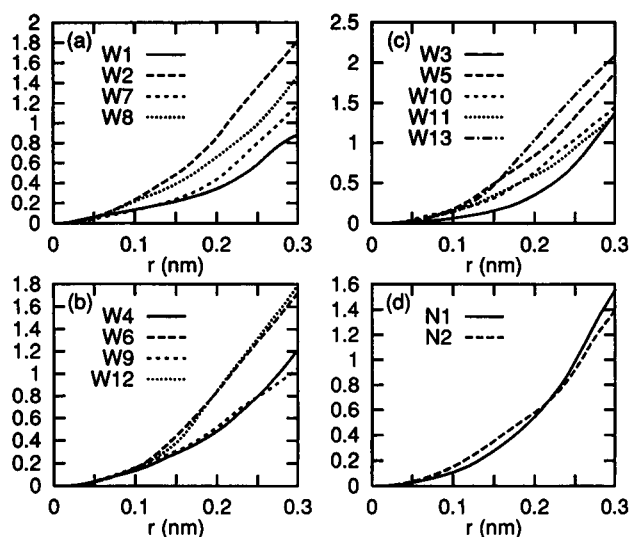

FIGURE 8 Hydration numbers $N(r)$ for the water and ion sites of the deoxydinucleoside monophosphate $d(CpG) \cdot d(CpG)$ (Ramakrishnan and Viswamitra, 1988). The $N(r)$ curves are grouped as in Fig. 7.

and the sodium cation, which could be occupied by water in the absence of the sodium ion. The two water molecules observed in the crystal are hydrating the sodium ion, which was not included in the PMF calculation. These water molecules are in a "frustrated" state, where they cannot fully satisfy the DNA hydration (as indicated by the PMF calculation) when the sodium ion is present.

Figs. 10 and 11 show the radial distribution function $g(r)$ and the corresponding hydration number $N(r)$ around each of the seven water sites and the sodium site in the crystal. The $g(r)$ curves of three of the seven water sites are strongly peaked at $r = 0$. One peak is somewhat weaker; and another

water site shows a strong $g(r)$ peak shifted to $r \approx 0.05$ nm. The two water sites hydrating the sodium cation in the crystal show little structure in their $g(r)$ curves, as does the sodium site.

Both the graphical representation of Fig. 9 and the $g(r)$ curves of Fig. 10 show positional agreement between theory and experiment (except for the two water molecules coordinated to ions). However, if we look at the integrated hydration numbers $N(r)$, the agreement is less satisfactory. Within 0.2 nm, the five water sites not in contact with the sodium ion contain only 0.2–0.35 water molecules. Therefore, although we correctly reproduce the positions of the water molecules, the density peaks appear to be too low.

The reason for this is expected to be the densely packed structure of this crystal. Both dinucleoside crystals $r(ApU) \cdot r(ApU)$ and $d(CpG) \cdot d(CpG)$ studied above contain approximately twice the number of water molecules per base pair. From the theoretical outline of the PMF hydration method it is clear that the method works best in a liquid-water environment. In the narrow channels formed in highly compact crystals, the calculations reproduce the correct positions for the density peaks as shown. But neglected higher-order correlation corrections (four and more particles) of short range may add to the peak intensity.

Interestingly, computer simulation studies of biomolecular crystals face a similar problem. In principle, a simulation at a constant chemical potential of water (e.g., grand-canonical Monte Carlo simulation) would permit determination of the number of water molecules per unit cell. However, this approach is seriously hampered by technical problems (low acceptance rates for water-molecule insertions and deletions). Therefore, the number of water molecules is, for instance, estimated from simulations with different particle numbers based on simple energetic arguments (Kim and Clementi, 1985a).

In summary, the hydration of the parallel dinucleoside is more structured than the network-type hydration of the previously studied molecules. Again, we observe positional agreement of theory and experiment, except for two water sites coordinated to a cation. The integrated water densities are too small, which we ascribe to the tightly packed structure of the crystal.

The deoxydinucleoside monophosphate $d(CpG) \cdot d(CpG)$ with the drug proflavine intercalated

The hydration structure of the 2:2 complex of proflavine and the deoxydinucleoside monophosphate $d(CpG) \cdot d(CpG)$ has been studied extensively by x-ray crystallography (Neidle et al., 1980; Shieh et al., 1980; Schneider et al., 1992b). In addition, a large number of computer simulation studies of this system analyzed energetic, dynamic, and structural aspects of the hydration (Mezei et al., 1983; Kim et al., 1983; Kim and Clementi, 1985a, b; Swaminathan et al., 1990; Herzyk et al., 1991). Here we use the crystal structure by Shieh et al. (1980; see Table 1).

FIGURE 9 The hydration of the deoxydinucleoside monophosphate d(CpG) · d(CpG) with parallel strand orientation in the crystal environment (Coll et al., 1987). Details as in Fig. 1.
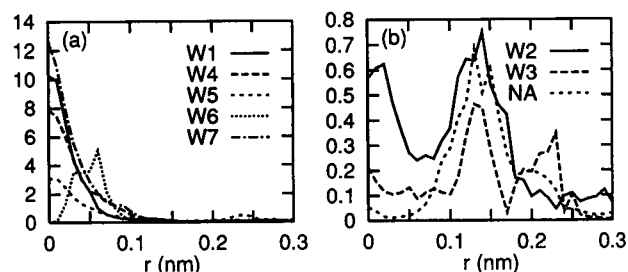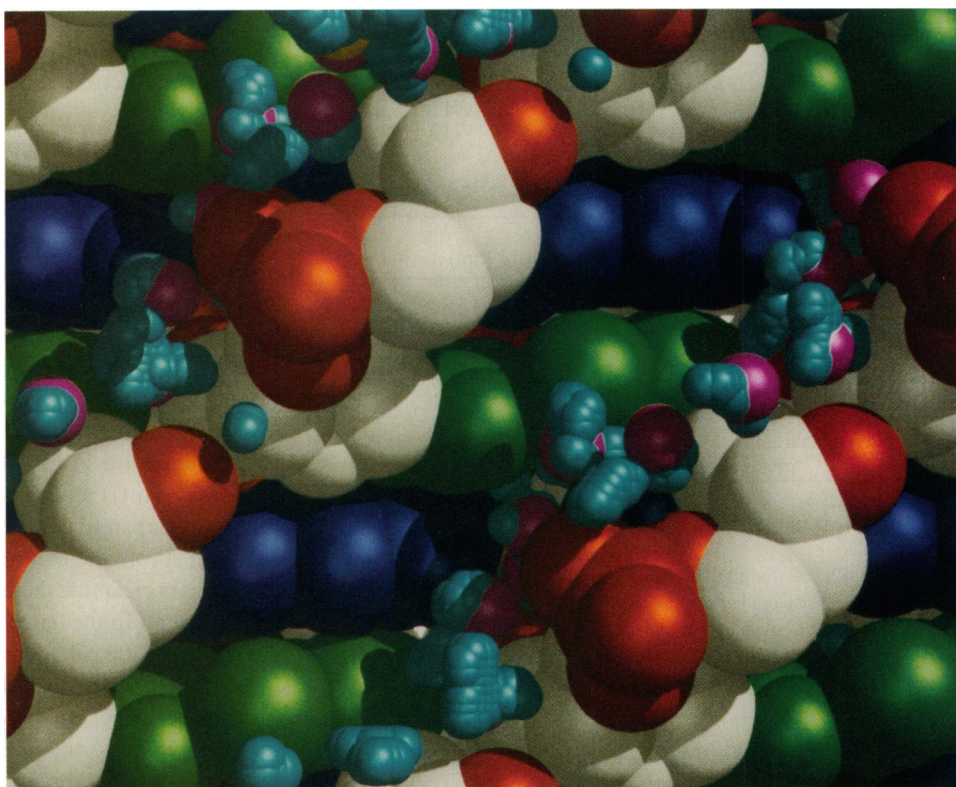

FIGURE 10 Radial distribution functions $g(r)$ (in units of ρ_0) of the water-oxygen density in the crystal around the seven water sites and the sodium ion of the deoxydinucleoside monophosphate d(CpG) · d(CpG) with parallel strand orientation in the crystal (Coll et al., 1987). (a) Water sites that are not coordinated to a sodium ion. (b) Water density $g(r)$ around the sodium site and the two neighboring water sites.
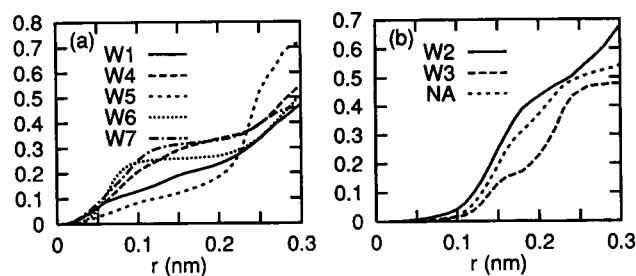
FIGURE 11 Hydration numbers $N(r)$ for the water and ion sites of the deoxydinucleoside monophosphate d(CpG) · d(CpG) with parallel strand orientation in the crystal (Coll et al., 1987). The $N(r)$ curves are obtained by integration of $g(r)$ from Fig. 10. The $N(r)$ curves are grouped into two panels as in Fig. 10.

In the PMF expansion method, the positively charged nitrogen on the proflavine has been accounted for by using the pair and triplet correlation functions (with water oxygens) of a single-charged cation dissolved in SPC water. (The results of this explicit treatment of the charged group are similar to those obtained by describing the nitrogen as a polar oxygen.) Figs. 12 and 13 show major- and minor-groove views of the hydration of the stacked molecules in the crystal. The small minor-groove cavity in the crystal is filled with a particularly characteristic heptagon network of water molecules (Fig. 13). This structure is reproduced very well by the PMF calculations of the water density. All crystal water sites fall within regions of high water density in the calculation. Two symmetry-related heptagon edges are formed by high-water-

density regions that connect the vertices. An additional high-density site is observed inside the heptagon. The high-density regions connecting the hydration sites indicate some structural flexibility in the water structure. This flexibility was also observed in molecular dynamics simulations of the crystal hydrate (Swaminathan et al., 1990), where the heptagon was present not at every instance but in a time-averaged sense. The density calculated from the PMF expansion is an approximate ensemble average of the local occupancy and does not imply the concurrent occupancy of all hydration sites.

The water sites of the pentagon network in the major-groove channel are only partly reproduced as high-density regions in the PMF calculations (Fig. 12). Interestingly, in previous computer simulation studies similar observations

FIGURE 12 Major-groove view of the hydration of the deoxydinucleoside monophosphate d(CpG) · d(CpG) with the drug proflavine intercalated in the crystal environment. The proflavine molecule is shown in gray. The smallest cyan spheres ($R = 0.02$ nm) indicate points with $3\rho_0 < \rho < 5\rho_0$. Details as in Fig. 1.
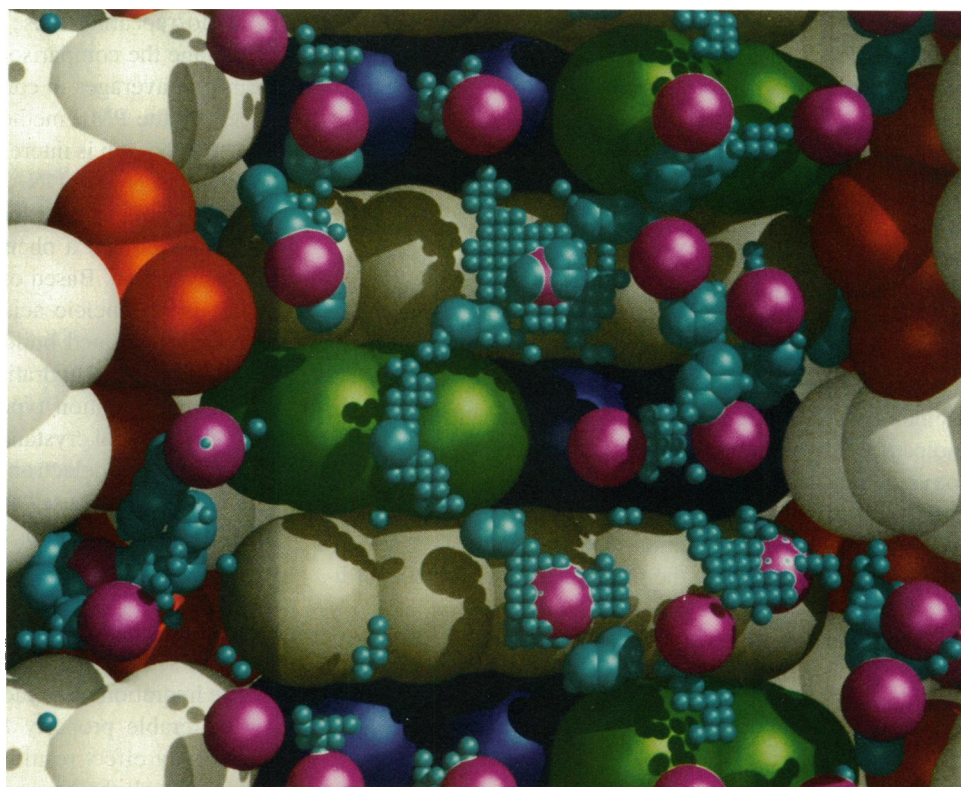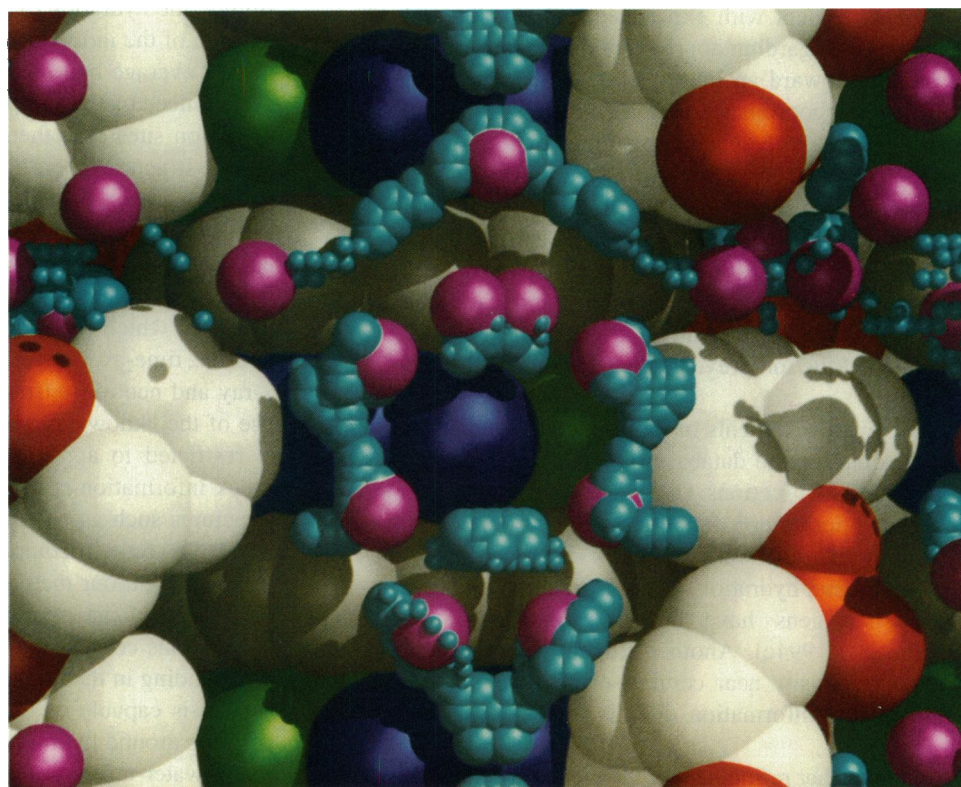

FIGURE 13 Minor-groove view of the hydration of the deoxydinucleoside monophosphate d(CpG) · d(CpG) with the drug proflavine intercalated in the crystal environment. Details as in Figs. 1 and 12.

were made, showing that the water heptagons are better established than the pentagon network (Mezei et al., 1983; Kim et al., 1983). Extensive crystallographic studies of the system at various temperatures also revealed some flexibility in the

pentagon network (Berman, 1986; Schneider et al., 1992b). The main reason for this appears to be that the major-groove cavity in the crystal containing the pentagon network is considerably larger. Therefore, fewer interactions of water mol-

ecules with the DNA–proflavine complex are possible, resulting in a reduced localization of water compared with that in the narrow minor-groove cavity.

CONCLUDING REMARKS

We have described a method to study the structural hydration of biological macromolecules. It is based on a statistical mechanical description in terms of particle correlation functions and involves three levels of approximations: (1) We assume that in the expansion of the water-density distribution using lower-order particle correlation functions it is sufficient to include two- and three-particle correlations; and (2a) water molecules interact most strongly with electronegative atoms (oxygen and nitrogen on nucleic acids), which (2b) can be approximately represented by water oxygens. Therefore, only water correlation functions are used in the density calculations. (3) We neglect the presence of mobile or coordinated ions.

Approximation (1) was tested previously by comparing calculated density distributions at the ice–water interface with extensive computer simulations, showing excellent agreement if the three-particle correlations are included (Hummer and Soumpasis, 1994a). Regarding approximations (2a) and (2b), we rely in part on the physical arguments discussed in the subsection Atom classes. However, the observed agreement with experimental data supports our assumptions. Regarding approximation (3), our main concern is directed toward an improved description of strongly localized (e.g., coordinated multivalent) cations. However, the effect of ions on the water structure is expected to be short-ranged. The results of this study for water sites coordinated to localized ions are ambiguous: Some were reproduced well despite the neglect of the ions (see the subsection on r(APU) · r(APU)); for others, the method clearly failed (see the subsections on r(APU) · r(APU) and parallel-stranded d(CpG) · d(CpG)), indicating a frustrated state where the presence of the ion prevents the occupation of the optimum positions.

Future improvements of the method are expected mainly from extending the database of correlations used. We currently see two directions: One can use a larger class of atoms, for which correlation functions are determined from computer simulations. An example is discussed in a study of DNA oligomer hydration, where correlation functions for the anionic oxygens have been calculated (Hummer and Soumpasis, 1994c). Another strategy involves calculating the water density near certain structural units (e.g., PO_4^-) and use this information directly in an expansion similar to Eq. 3.

In the further refinement of the method we hope to receive some guidance from direct comparisons with computer simulation data. However, as mentioned above, computer simulation studies of large biomolecules in aqueous solution or in the crystal environment are very time consuming and require large resources. Nevertheless, over the past 15 years many such studies have been reported in the literature. Therefore, to avoid repeating these computational studies, we would encourage the comparison of local water densities obtained from grid averages in computer simulations with those produced by the PMF method.

At this point it is interesting to discuss a recently proposed approach to predict DNA hydration (Schneider et al., 1993), which is similar to that of Pitt et al. (1993). Schneider et al. (1993) developed a phenomenological approach to predict DNA hydration. Based on the positional data of water molecules in 40 nucleic acid crystals, those authors extracted so-called "hydrated building blocks," which contain information about the hydration of nucleic acid bases in a particular conformation type (A-, B-, or Z-DNA). The information of several crystals is then averaged and interpreted in terms of a pseudoelectron density of water around each of the four base types. These "average building blocks" are used to predict DNA hydration. Aside from technical problems involved (e.g., small size of the database of water positions used), the assumption that each base is in A, B, or Z conformation is restrictive and neglects local variations in structure and sequence.

The hydration of an individual base in a given DNA is not a transferable property fixed to that base but a complex many-body effect resulting from the interactions of all the waters and all the bases present. (The same is true not only for DNA but for any multiunit molecule in solution.) More simply, the hydration of two bases in close proximity is not the sum of the individual hydrations of the bases (contained in the "average building block"). Some consequences have been noticed by the authors, e.g., "The method predicts more hydration sites than there are water positions in the actual crystal structures," or "... identifies hydration sites that are too close to be occupied simultaneously by two waters." These problems are related to the reinterpretation of the calculated pseudoelectron density in terms of actual water sites, whereas we base our analysis on the local water-density distribution, which corresponds directly to measurable quantities, i.e., time-averaged electron and hydrogen densities from x-ray and neutron diffraction, respectively. In addition, the use of the "knowledge-based" approach of Schneider et al. is restricted to applications for which extensive building block information is available (A-, B-, Z-DNA), excluding structures such as parallel DNA, triple helices, drug–DNA and DNA–protein complexes, and single strands.

In contrast, our method (which is firmly based on statistical physics) can cope with all these situations. The triplet correlations contain the essential information about hydrogen bonding in the liquid-water phase, and therefore the approach is capable of predicting the subtle water networks forming around hydrophilic sites owing to cooperative (i.e., many-water-molecule) effects without any x-ray database knowledge of water positions. Database analysis can never replace theories of complex systems; it is useful for testing the technical approximations involved in the numerical implementations of theories, as done in this paper.

In this paper we have presented calculations on high-resolution crystal structures of small nucleic acid molecules

to illustrate the applicability of the method. Agreement between the calculated high-water-density regions and positions of water molecules in the crystals has been observed. We have shown that PMF hydration calculations can help to assign water sites in the crystallographic analysis. In particular, the method may provide theoretical support if the experimental data are not sufficient to distinguish whether water molecules or ions occupy a certain site in the crystal. Complementary studies of the ion-density distribution using an analogous PMF description based on ionic correlation functions (Klement et al., 1991) might provide further insight. The identification of water and ion sites is of great importance for a better understanding of the biochemical activity of many enzymes or catalytically active RNA molecules and of how this activity is related to the biomolecular structure.

From its physical basis, the method is expected to work best in a fluid-water environment. Therefore, the method can help to distinguish between water sites merely caused by the crystal packing from those also expected in the solution environment (Zhang and Matthews, 1994). Moreover, what makes this type of approach particularly appealing is its great computational efficiency. For instance, the water-density calculation for a unit cell of the DNA–proflavine crystal (approximately 380,000 grid points) requires fewer than 20 min of CPU time on a Silicon Graphics Indigo workstation.

The PMF method makes it possible to analyze the hydration of very large molecules such as tRNA, analogous to the studies described here for small nucleic acid fragments and requiring only few hours of CPU time on a workstation computer. Therefore, investigations of the role of water in the biologically important interactions of proteins and nucleic acids or in antibody–antigen binding processes become feasible with the presented method, analyzing the function of water as a link between polar groups in the complex.

Furthermore, in many applications to large biomolecular systems the region of interest is small, e.g., the reaction center of a protein or a drug-binding area. This points out an important advantage of our approach: The hydration calculation is entirely local; i.e., in order to compute the density at a given position, no explicit information about the density at any other point is required. This simplifies the problem at hand greatly because no complicated boundary conditions need to be implemented, and the analysis can be restricted to only the region of interest.

Therefore, based on the accuracy and speed of the calculations, an interactive modeling of biomolecular hydration becomes feasible in future implementations guided toward crystallographic data analysis or molecular docking studies.

REFERENCES

Aggarwal, A., S. A. Islam, R. Kuroda, M. R. Sanderson, S. Neidle, and H. M. Berman. 1983. The structure of the ribodinucleoside monophosphate guanylyl-3', 5'-cytidine as its ammonium octahydrate salt. *Acta Crystallogr. B.* 39:98–104.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for water in relation to protein hydration. *In*

Intermolecular Forces: Proceedings of the 14th Jerusalem Symposium on Quantum Chemistry and Biochemistry. B. Pullman, editor. Reidel, Dordrecht, The Netherlands. 331–342.

Berman, H. M. 1986. Hydration of nucleic acid crystals. *Ann. N. Y. Acad. Sci.* 482:166–178.

Berman, H. M. 1991. Hydration of DNA. *Curr. Opin. Struct. Biol.* 1:423–427.

Berman, H. M. 1994. Hydration of DNA: Take 2. *Curr. Opin. Struct. Biol.* 4:345–350.

Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751–759.

Buckin, V. A. 1987. Experimental studies of DNA-water interaction. *Mol. Biol.* 21:615–629. [Mol. Biol. (USSR) 21:512–525].

Coll, M., X. Solans, M. Font-Altaba, and J. A. Subirana. 1987. Crystal and molecular structure of the sodium salt of the dinucleotide duplex d(CpG). *J. Biomol. Struct. Dyn.* 4:797–811.

Cruse, W. B. T., E. Egert, O. Kennard, G. B. Sala, S. A. Salisbury, and M. A. Viswamitra. 1983. Self base pairing in a complementary deoxydinucleoside monophosphate duplex: Crystal and molecular structure of deoxycytidylyl-(3'-5')-deoxyguanosine. *Biochemistry.* 22:1833–1839.

Dickerson, R. E. 1992. DNA structure from A to Z. *Meth. Enzymol.* 211:67–111.

Elliott, R. J., and J. M. Goodfellow. 1987. Monte Carlo simulations of nucleotide crystal hydrates and their counter-ions. *J. Theor. Biol.* 127:403–412.

Forester, T. R., and I. R. McDonald. 1991. Molecular dynamics studies of the behaviour of water molecules and small ions in concentrated solutions of polymeric B-DNA. *Mol. Phys.* 72:643–660.

García, A. E., G. Hummer, and D. M. Soumpasis. 1994. Theoretical description of protein hydration: A potential-of-mean-force calculation based on two- and three-particle correlation functions. *Biophys. J.* 66:A130.

Hansen, J.-P., and I. R. McDonald. 1986. Theory of Simple Liquids. Academic Press, London.

Herzyk, P., J. M. Goodfellow, and S. Neidle. 1991. Molecular dynamics simulations of dinucleoside and dinucleoside-drug crystal hydrates. *J. Biomol. Struct. Dyn.* 9:363–386.

Hummer, G., and D. M. Soumpasis. 1994a. Computation of the water density distribution at the ice-water interface using the potentials-of-mean-force expansion. *Phys. Rev.* E49:591–596.

Hummer, G., and D. M. Soumpasis. 1994b. A new approach to calculate the hydration of DNA molecules. *In* Structural Biology: The State of the Art; Proceedings of the Eighth Conversations in the Discipline Biomolecular Stereodynamics. Vol. 2. R. H. Sarma and M. H. Sarma, editors. Adenine Press, Schenectady, NY. 273–278.

Hummer, G., and D. M. Soumpasis. 1994c. Statistical mechanical treatment of the structural hydration of biological macromolecules: Results for B-DNA. *Phys. Rev.* E50:5085–5095.

Hummer, G., D. M. Soumpasis, and A. E. García. 1994. The hydration of A-, B-, and Z-DNA studied with the potentials-of-mean-force approach. *Biophys. J.* 66:A25.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Kim, K. S., and E. Clementi. 1985a. Energetics and pattern analysis of crystals of proflavine deoxydinucleoside phosphate complex. *J. Am. Chem. Soc.* 107:227–234.

Kim, K. S., and E. Clementi. 1985b. Hydration analysis of the intercalated complex of deoxydinucleoside phosphate and proflavine: computer simulations. *J. Phys. Chem.* 89:3655–3663.

Kim, K. S., G. Corongiu, and E. Clementi. 1983. Networks of water molecules in a proflavine deoxydinucleoside phosphate complex. *J. Biomol. Struct. Dyn.* 1:263–285.

Kirkwood, J. G. 1935. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3:300–313.

Klement, R., D. M. Soumpasis, and T. M. Jovin. 1991. Computation of ionic distributions around charged biomolecular structures: Results for right-handed and left-handed DNA. *Proc. Natl. Acad. Sci. USA.* 88:4631–4635.

Kubinec, M. G., and D. E. Wemmer. 1992. NMR evidence for DNA bound

water in solution. *J. Am. Chem. Soc.* 114:8739–8740.

Levitt, M., and B. H. Park. 1993. Water: now you see it, now you don't. *Structure.* 1:223–226.

Liepinsh, E., G. Otting, and K. Wüthrich. 1992. NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. *Nucleic Acids Res.* 20:6549–6553.

Mezei, M., D. L. Beveridge, H. M. Berman, J. M. Goodfellow, J. L. Finney, and S. Neidle. 1983. Monte Carlo studies on water in the dCpG/proflavin crystal hydrate. *J. Biomol. Struct. Dyn.* 1:287–297.

Münster, A. 1969. Statistical Thermodynamics, Vol. 1. Springer, Berlin. p. 338.

Neidle, S., H. M. Berman, and H. S. Shieh. 1980. Highly structured water network in crystals of a deoxydinucleoside-drug complex. *Nature.* 288:129–133.

Niesar, U., G. Corongiu, E. Clementi, G. R. Kneller, and D. K. Bhattacharya. 1990. Molecular dynamics simulations of liquid water using the NCC ab initio potential. *J. Phys. Chem.* 94:7949–7956.

Otting, G., E. Liepinsh, and K. Wüthrich. 1991. Protein hydration in aqueous solution. *Science.* 254:974–980.

Pitt, W. R., J. Murray-Rust, and J. Goodfellow. 1993. Aquarius 2: Knowledge-based modeling of solvent sites around proteins. *J. Comp. Chem.* 14:1007–1018.

Ramakrishnan, B., and M. A. Viswamitra. 1988. Crystal and molecular structure of the ammonium salt of the dinucleoside monophosphate d(CpG). *J. Biomol. Struct. Dyn.* 6:511–523.

Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer, Berlin.

Saenger, W. 1987. Structure and dynamics of water surrounding biomolecules. *Annu. Rev. Biophys. Biophys. Chem.* 16:93–114.

Savage, H., and A. Wlodawer. 1986. Determination of water structure around biomolecules using x-ray and neutron diffraction methods. *Meth. Enzymol.* 127:162–183.

Schneider, B., D. Cohen, and H. M. Berman. 1992a. Hydration of DNA bases: analysis of crystallographic data. *Biopolymers.* 32:725–750.

Schneider, B., S. L. Ginell, and H. M. Berman. 1992b. Low temperature structures of dCpG-proflavine: conformational and hydration effects. *Biophys. J.* 63:1572–1578.

Schneider, B., D. M. Cohen, L. Schleifer, A. R. Srinivasan, W. K. Olson, and H. M. Berman. 1993. A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J.* 65:2291–2303.

Seeman, N. C., J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, and A. Rich. 1976. RNA double-helical fragments at atomic resolution: I. The crystal and molecular structure of sodium adenylyl-3',5'-uridine hexahydrate. *J. Mol. Biol.* 104:109–144.

Shieh, H.-S., H. M. Berman, M. Dabrow, and S. Neidle. 1980. The structure of drug-deoxydinucleoside phosphate complex; generalized conformational behavior of intercalation complexes with RNA and DNA fragments. *Nucleic Acids Res.* 8:85–97.

Soumpasis, D. M. 1993. Formal aspects of the potentials of mean force approach. *In* Computation of Biomolecular Structures. D. M. Soumpasis and T. M. Jovin, editors. Springer, Berlin. 223–239.

Soumpasis, D. M., P. Procacci, and G. Corongiu. 1991. A: On the computation of N-particle correlations in classical fluids via computer simulations. B: Triplet correlations in NCC water. IBM DSD report, October 1991.

Swaminathan, S., D. L. Beveridge, and H. M. Berman. 1990. Molecular dynamics simulation of a deoxydinucleoside-drug intercalation complex: dCpG/proflavin. *J. Phys. Chem.* 94:4660–4665.

Texter, J. 1978. Nucleic acid–water interactions. *Prog. Biophys. Mol. Biol.* 33:83–97.

Westhof, E., and D. L. Beveridge. 1990. Hydration of nucleic acids. *In* Water Science Reviews. Vol. 5. F. Franks, editor. Cambridge University Press, Cambridge. 24–136.

Zhang, X.-J., and B. W. Matthews. 1994. Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. *Protein Sci.* 3:1031–1039.