Faraday Discuss., 1996, **103**, 175–189

A statistical mechanical description of biomolecular hydration

Gerhard Hummer,*† Angel E. García and D. Mario Soumpasis‡
Theoretical Biology and Biophysics Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

An efficient and accurate theoretical description of the structural hydration of biological macromolecules is presented. The hydration of molecules of almost arbitrary size (tRNA, antibody–antigen complexes, photosynthetic reaction centre) can be studied in solution and in the crystalline environment. The biomolecular structure obtained from X-ray crystallography, NMR or modelling is required as input information. The structural arrangement of water molecules near a biomolecular surface is represented by the local water density, analogous to the corresponding electron density in an X-ray diffraction experiment. The water-density distribution is approximated in terms of two- and three-particle correlation functions of solute atoms with water using a potentials-of-mean-force expansion.

1 Introduction

As the natural solution environment of biological macromolecules, water influences many aspects of biological function.[1] Spectroscopic methods are the major source of experimental information about the interactions of water with biomolecules. X-ray, neutron[2] and NMR spectroscopy[3] make it possible to study the structural hydration of biomolecules at atomic resolution. Single-crystal diffraction methods can provide a detailed, three-dimensional picture of the localisation of water molecules at the surface and in the interior of biomolecules; but clearly, the application of single-crystal diffraction requires that the biomolecule can be crystallised, and that the crystal diffracts to short enough wavelengths to get the spatial resolution required for unambiguously assigning water molecules. Moreover, re-refinements and studies of crystals with different space groups reveal methodological difficulties in crystallographic hydration studies.[4,5] Surface regions of interest must not be affected by crystal packing but should be exposed to sufficiently large solvent channels in the crystal. X-Ray and neutron diffraction crystallography also face problems when the water phase surrounding the biomolecule is disordered. The crystal ordering is much less stringent for small molecules occupying solvent channels, with the consequence that the water structure can vary considerably between different unit cells of the crystal. Overall, this disorder reflects an equilibrium water density distribution in the crystal unit cell. The electron density, as determined in an X-ray diffraction experiment, contains, in principle, information about this position-dependent one-particle density of water. However, the practical hydration analysis is limited to regions showing a high localisation of water molecules.

Theoretical methods are used to overcome some of these obstacles in studying the structural hydration of biomolecules. The current theoretical modelling of biomolecules in solution and in the crystal environment relies almost exclusively on computer simulations. Molecular dynamics (MD) and Monte Carlo (MC) simulations of biomolecules

† Also at: Center for Nonlinear Studies. E-mail: hummer@t10.lanl.gov
‡ Biocomputation Group, Max Planck Institute for Biophysical Chemistry, P.O. Box 2841, D-37018 Göttingen, Germany.

and water provide versatile tools to study solute–solvent interactions.[6–8] Aside from the difficulties of specifying sufficiently accurate interaction potentials, the major problem associated with computer simulation studies is the enormous computer time requirement. This matters less if a study is devoted to only a single and relatively small molecule. However, the conventional simulation approach rapidly reaches its limitations if a large number of mutations,[5] large biomolecules such as photosynthetic reaction centres[9] or antibody–antigen complexes,[10] or a large number of ligands such as in a typical small-drug docking study are analysed.

For large molecules and large sample sizes, fast and accurate methods other than computer simulations are needed. One direction of research is to use so-called 'knowledge-based' approaches.[11–14] Rules are 'learned' from a training set of known X-ray crystal structures. These rules are assumed to be transferable to 'predict' the hydration of a molecule not in the training set. Despite the interesting results of these approaches, problems remain, such as insufficient statistics because of, for instance, the enormous number of side-chain conformations[15] and side-chain interactions in proteins that prohibit exhaustive sampling. Another problem is that crystallographic hydration sites are often caused by interactions with two or more symmetry-related molecules in the crystal and are thus not conserved between different crystals.[4,5]

A theory based on molecular principles avoids these difficulties. We shall discuss such a method to describe the equilibrium structural hydration of biomolecules. This statistical-mechanical method is the most recent offspring of the potentials-of-mean-force (PMF) strategy[16,17] to handle the problems of biomolecule–solvent interactions in a computationally efficient, reasonably accurate and physically sound way. The method is based on a truncated expansion of the local density of water in terms of two- and three-particle correlation functions. A similar method was used previously to study ionic densities near nucleic acids at a simpler level using ionic pair correlations and a crude dielectric continuum model for water.[18] An accurate analysis of the structural hydration even of molecular systems with 10000 or more heavy atoms becomes feasible within CPU minutes, rather than weeks as in comparable computer simulation studies. Unlike 'knowledge-based' approaches, the method is soundly based on statistical mechanics. It can be applied to study the hydration of any biomolecular complex without previous knowledge of similar structures.

Here, we first develop the theoretical foundation. Then we show how this general theory can be applied to biomolecular systems, emphasising approximations and limitations. Water triplet correlations will be presented. To illustrate the applicability and accuracy of the method, results will be discussed for a broad class of systems (ice/water interface, small peptides, nucleic acid fragments, B-DNA oligonucleotides, transfer RNA, antibody–antigen complexes, photosynthetic reaction centres). Directions for future applications to the calculation of binding entropies will be outlined.

2 Theoretical development

2.1 Local densities in inhomogeneous systems

Central to hydration calculations is the local water-density distribution, which characterises the structural equilibrium properties of water near a biomolecule. The density distribution is expected to approach the bulk-water density of 1 g cm^{-3} within a few Å from a biomolecule in solution. Close to the molecular surface, however, the density distribution will show sharp oscillations, reflecting the structural ordering of water molecules.

The solvated macromolecule is characterised by atomic coordinates, s_{i_α}, of N_α atoms of type α and M different types of atoms, $\alpha = 1 \cdots M$. This set of atomic coordinates, $\{s_{i_\alpha}\}$, can be obtained from X-ray crystallography, NMR or modelling. A rigid equi-

librium structure will be assumed for the solute. We express the local solvent density as a configuration-space integral. To simplify the presentation, results for an atomic solvent will be given. The generalisation to molecular solvents is straightforward. For a canonical ensemble of N solvent particles with coordinates $\{r_i\}$, we obtain for the conditional solvent density at a position r_1

$$\rho(r_1 \mid \{s_{i_\alpha}\}) = N \frac{\int \mathrm{d}r_2 \cdots \mathrm{d}r_N \exp[-\beta U(\{r_i\}, \{s_{i_\alpha}\})]}{\int \mathrm{d}\{r_i\} \exp[-\beta U(\{r_i\}, \{s_{i_\alpha}\})]} \tag{1}$$

where $\beta = 1/k_B T$ and U is the total potential energy in a classical mechanical description. In computer simulations, eqn. (1) is used directly to obtain local water densities as ensemble or, equivalently, time averages. Here, we further simplify eqn. (1). The first step is to express the local density in terms of particle correlation functions,[17,19,20]

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

where $\rho_0 = N/V$ is the solvent density in the volume V. Eqn. (2) follows directly from the definition of multiparticle correlation functions, $g^{(n)}$. However, eqn. (2) also shows a close analogy with probability theory relating the conditional probability of observing event A given event B to the ratio of the probability of event A and B to that of event B, $P(\mathrm{A} \mid \mathrm{B}) = P(\mathrm{A} \cap \mathrm{B})/P(\mathrm{B})$. The high dimensionality of the correlations in eqn. (2) prohibits a direct application. We therefore apply a PMF expansion.[19,21,22] This yields an expression for the density in terms of lower-order correlations,[20,23]

$$\rho(r \mid \{s_{i_\alpha}\}) = \rho_0 \left[\prod_{\alpha=1}^{M} \prod_{i_\alpha=1}^{N_\alpha} g^{(1;\alpha)}(r, s_{i_\alpha}) \right] \times \left[\prod_{\beta=1}^{M} \prod_{j_\beta=1}^{N_\beta} \prod_{\gamma=\beta}^{M} \prod_{j_\gamma=1+\delta_{\beta\gamma} j_\beta}^{N_\gamma} \frac{g^{(1;\beta,\gamma)}(r, s_{j_\beta}, r_{j_\gamma})}{g^{(1;\beta)}(r, s_{j_\beta}) g^{(\beta,\gamma)}(s_{j_\beta}, s_{j_\gamma}) g^{(\gamma;1)}(s_{j_\gamma}, r)} \right] \tag{3}$$

where $\delta_{\beta\gamma}$ is the Kronecker symbol. To lowest order, the expansion gives the bulk density of water, ρ_0. The second- and third-order corrections involve products over all single solute atoms and all distinct pairs, respectively. Truncation at the two- and three-particle correlation level corresponds to the Kirkwood[21] and Fisher–Kopeliovich[24] superposition approximations, KSA and FKSA, respectively.

2.2 Towards a description of the structural hydration of biomolecules

Eqn. (3) involves a prohibitively large number of correlation functions when applied to a macromolecule in aqueous solution. However, based on energetic and geometric arguments it is possible to group atoms into similar classes. This reduces drastically the number of different atom types required in the calculations. Hydrogen bonds provide the strongest interactions between water and a biomolecule. In proteins and nucleic acids, electronegative oxygen, nitrogen and sulfur atoms are possible interaction partners for water to form hydrogen bonds. Geometrically, these hydrogen bonds are similar (in bond length and bond angle) to those formed between water molecules. In addition, currently used empirical force fields for biomolecular simulations do not warrant too fine-grained descriptions which are visible only at the quantum level.

These observations allow us to obtain a description of the structural hydration that should capture the essential features by equating all nitrogen, oxygen and sulfur atoms to water oxygen with respect to their effect on ordering water. Non-polar carbon atoms

are expected to have only a local, weakly structuring effect on water.[25,26] Carbons will be represented as excluded volume by setting the water-oxygen density to zero inside a sphere with radius $r_c = 0.3$ nm, where r_c corresponds approximately to the distance of closest approach of a water oxygen to a methane in aqueous solution. The carbon–water pair correlations in eqn. (3) are thus represented by a unit step function, $\Theta(r - r_c)$. Three-particle correlations involving carbon atoms are neglected. If large hydrophobic regions are exposed to water, the treatment needs to be refined as discussed below.[26] The solvent-exposed surface of proteins and nucleic acids is mostly hydrophilic. Hydration calculations based on polar atoms alone should therefore prove to be sufficiently accurate.

Another possible complication arises from the treatment of charged groups. One could generate a more extensive database of correlation functions for different charged groups with water. However, in nucleic acids and proteins, the charges on charged groups reside essentially on oxygens and nitrogens as well as on the polar hydrogens attached to them. The structural hydration of charged groups is therefore expected to be represented well by considering only oxygen, nitrogen and, possibly, polar hydrogen atoms.

Polar hydrogens on non-rotating groups such as NH and NH_2 can be treated explicitly to represent the directionality of hydrogen bonds better. A correlation function, $\chi(r \mid s_D, s_H) = \rho(r \mid s_D, s_H)/\rho_0 g^{(2)}(r, s_D)$, describes the conditional water-oxygen density caused by an OH bond with O at s_D and H at s_H. We thus obtain a working formula,

$$\rho(r \mid s_1, \ldots, s_n) \approx \rho_0 \prod_{i=1}^{n} g^{(2)}(r, s_i) \prod_{j=1}^{n-1} \prod_{k=j+1}^{n} \frac{g^{(3)}(r, s_j, s_k)}{g^{(2)}(r, s_j) g^{(2)}(s_j, s_k) g^{(2)}(s_k, r)} \times \prod_{\text{carbons } s_C} \Theta(|r - s_C| - r_c) \prod_{\substack{\text{polar hydrogens } s_H \\ \text{(with donors } s_D)}} \chi(r \mid s_D, s_H) \tag{4}$$

$g^{(2)}$ and $g^{(3)}$ are the two- and three-particle correlation functions of water oxygens in bulk water, respectively; $s_1 \cdots s_n$ are the positions of electronegative atoms. Unless otherwise stated, the correction for polar hydrogens, χ, was not applied in the following calculations.

2.3 Expected limitations

Several approximations enter PMF hydration calculations. The first is the truncation of the expansion, eqn. (3), at the three-particle correlation level. For a molecular fluid such as water with a highly anisotropic interaction potential owing to the directional hydrogen-bond interactions, one expects the KSA to give poor structural results at distances corresponding to the first hydration layer. This is best understood if one looks at the structure of hexagonal ice, where a water molecule has only four nearest neighbours in a tetrahedral arrangement. Water-oxygens in contact form isosceles triangles with edges of about 0.28, 0.28 and 0.45 nm length. Locally ice-like structures are also observed in liquid water. Equilateral triangles with edges of 0.28 nm length would result in strongly unfavourable interactions. Since the hydrogen-bond interactions dominate under standard conditions of temperature and density, water is not close packed and the KSA gives poor results for three-particle correlations. Therefore, three-particle correlations are indispensable.

It is less clear how well the FKSA approximates four-particle correlations. Computer-simulation calculations of four-particle correlations are not trivial because of insufficient statistical sampling. We can instead identify the smallest quadruplet (as measured by its longest edge) in an ice lattice, which would be given a finite FKSA probability, but which does not appear in the lattice. In hexagonal ice, the smallest three triangles formed by oxygen atoms on lattice sites have edges A = (1,1, $\sqrt{(24)}/3$), B = (1,

$\sqrt{(24)}/3$, 5/3) and C = ($\sqrt{(24)}/3$, $\sqrt{(24)}/3$, $\sqrt{(24)}/3$) in units of the nearest-neighbour distance nm). Among all possible combinations of triangles A, B and C in Euclidian space, a combination of four triangles B can be formed into a tetrahedron that is not found in the lattice. Therefore, the smallest quadruplet in hexagonal ice incorrectly predicted by the FKSA is rather large having four of its six edges longer than 0.45 nm. Its contributions should therefore be relatively small. More convincing, however, than these lattice-based arguments, are the results of the following section comparing calculations based on the FKSA for density profiles at an idealised ice/water interface with explicit computer simulations.

Another problem associated with the PMF expansion for correlation functions is its multiplicative form. As a consequence, the expansion expressions for local densities, eqn. (3) and (4), are highly non-linear in the correlation functions. This problem will surface when one attempts to use eqn. (3) with correlation functions for all atoms. For any given point, a large number of factors will then contribute to the calculated density. The use of only polar atoms (with non-polar atoms modelled as excluded volume) has the advantage that statistical errors in the correlation functions and systematic errors in the truncated expansion are better controlled because the number of contributing terms is relatively small.

In applications to biomolecular hydration calculations, the first problem arises from the assumption of a rigid equilibrium structure. As a consequence, correlated fluctuations of water and surface residues of the hydrated biomolecule are neglected. This problem is expected to be less significant if the biomolecule has a well defined structure. If, however, large motions occur at the biomolecular surface, a truly dynamic treatment is important.

Regarding the effect of hydrophobic residues on the local water density in their vicinity, recent calculations[26] showed that a PMF expansion is satisfactory if individual non-polar groups do not overlap. For covalently bonded hydrocarbons, however, serious deficiencies of a PMF expansion for non-polar molecules become apparent. The strongest interaction of non-polar groups with water is volume exclusion. If individual sites overlap, each describing a spherical excluded volume, the truncated product expansion, eqn. (3), guarantees zero density in the interior, but for the exterior the predicted density will be wrong. The worst case is the limit of complete overlap of n hard spheres, where the FKSA, eqn. (3), results in a density profile of $\rho_0\, g(r)^{n-n(n-1)/2}$, *cf.* the correct result of $\rho_0\, g(r)$, where $g(r)$ is the hard-sphere–water pair correlation function.[26] However, the effect of non-polar solutes in water can be described well by the simplest approximation based on proximity, *i.e.* taking only a pair correlation with the nearest solute site.[25,26]

As a consequence of the discussed approximations and limitations, one can expect good qualitative results for the PMF expansion of the density. This means that the PMF expansion should primarily be able to locate regions with high local water density at a molecular surface. However, the actual density values are less reliable. The analysis of the structural hydration using PMF calculations should therefore primarily focus on the identification of regions with high water density.

3 Results and Discussion

3.1 Water correlation functions

Pair correlations of water were studied extensively by X-ray and neutron diffraction experiments[27] and computer simulations using various water models.[28] Much less is known about water triplet correlations. Experimentally, water three-particle correlations have been accessible only through indirect, maximum-entropy type approaches.[29] Using computer simulations, water three-particle correlations were first calculated by Soumpasis *et al.*[30] for the *ab initio* NCC[31] water model. In Fig. 1, we show results for the

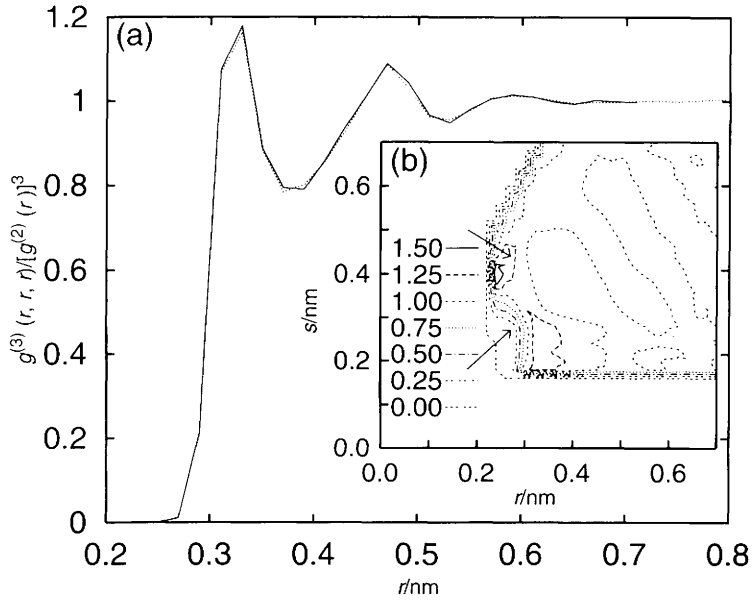

Fig. 1 Water-oxygen triplet correction, $\Gamma(r, r, s) = g^{(3)}(r, r, s)/[g^{(2)}(r)]^2 g^{(2)}(s)$, for (a) equilateral ($s = r$) and (b) isosceles triangles (inset). Arrows in (b) indicate equilateral and isosceles triangles with hydrogen-bond distance. Results in (a) and (b) are from an MC simulation of 256 SPC water molecules using generalised reaction-field (GRF) electrostatics. Also shown in (a) are results of a 512-particle simulation using Ewald-summation electrostatics (dotted line).

three-particle correction, $\Gamma(r, r, s) = g^{(3)}(r, r, s)/[g^{(2)}(r)]^2 g^{(2)}(s)$, of the SPC model of water[32] for isosceles (and equilateral) triangles with edges r, r and s formed by three water oxygens. The simulation calculations were described previously.[19,20] For oxygen pair distances between 0.19 and 0.25 nm, a constrained method was used to determine $g^{(3)}$.[20] Unless otherwise stated, SPC correlations will be used in the PMF calculations.

As expected from our previous discussion based on the structure of hexagonal ice, the locally tetrahedral organisation of water results in Γ being close to zero for three oxygens at hydrogen-bond distances of 0.28 nm. For isosceles triangles with distances close to the ice-values for nearest neighbours, 0.28, 0.28 and 0.45 nm, Γ reaches its maximum for isosceles triangles. In addition, Fig. 1 shows that the triplet correction is of short range, decaying rapidly to one. For equilateral triangles with $r > 0.3$ nm, Γ deviates less than 20% from unity. The consequence for the PMF density expansion is that triplet corrections are indispensable at short range, but contribute little at larger distances.

3.2 A test case: The interface of ice and water

To illustrate the applicability of the expansion formula, eqn. (3), for inhomogeneous aqueous systems, we study a model ice/water interface. The interface is described by fixing SPC water oxygens at ideal hexagonal-ice I_h lattice positions without specifying molecular orientations. The thermal disorder in the solid phase is limited to reorientation of water molecules. The basal plane of the ice layer is covered by liquid water. More realistic models of the interface were studied previously by computer simulations.[33–35]

Fig. 2 shows results of explicit MC computer simulations and PMF calculations for the density profile as a function of the distance d from the plane defined by the last layer

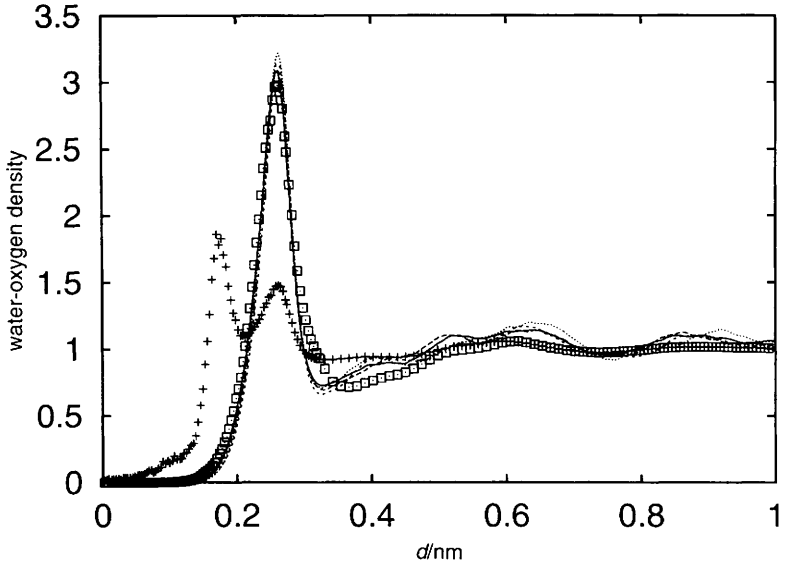

Fig. 2 Water-oxygen density at the ice/water interface in units of the bulk-water density, ρ_0. d is the distance from the closest plane of oxygen atoms in the ice phase. (+) PMF expansion including only pair correlations; (□) PMF expansion including pair and triplet correlations. Solid lines: explicit MC simulations using different system sizes and methods for electrostatic interactions.

of ice-oxygen atoms.[19] In the PMF expansion, the discrete two- and three-particle correlation functions of SPC water are used with linear and trilinear interpolation for pair distances. The PMF expansion truncated at the two-particle level clearly fails to reproduce the simulation densities. The first peak is too close because of the high KSA probability of equilateral triangles. The second peak at the KSA level is too small. If, however, three-particle correlations are included, the picture improves drastically. Position and height of the first peak are well reproduced by the FKSA–PMF density. Beyond the first peak, qualitative agreement is observed. The FKSA–PMF result is less structured, but shows extrema at the right positions. In biomolecular applications, the emphasis is on identifying high-density regions. Therefore, accuracy of the FKSA–PMF expansion in the interfacial region is important, where the strongest variations in the density occur.

3.3 Local density of water around rigid peptides

To test the accuracy of the PMF method for biomolecular hydration, we performed MD simulations of an α-helix and a β-hairpin in aqueous solution. Constant (N, V, T = 300 K) simulations have been extended for 1.0 ns with a time step of 0.002 ps, after an equilibration period of 125 ps. The united-atom OPLS force field and the rigid TIP3P model of water[36] were used for all simulations. N- and C-terminal residues were uncharged. The PMF calculations used eqn. (4), including the hydrogen-bond correction, χ. The correlation functions $g^{(2)}$, $g^{(3)}$ and χ were those of TIP3P water.

A polypeptide, $(Ala)_{18}$, was constructed in an ideal α-helical conformation with $\phi = -48°$, $\psi = -57°$. As a system that better represents the composition of amino acids in a protein, we have chosen a β-hairpin from BPTI with sequence[21] YFYNAKAGLCQT.[32] The conformation of this peptide was taken from the crystal structure of Wlodawer *et al.*[37] (PDB[38] code 5pti). This peptide contains one charged amino acid (K26), polar residues (Y21, Y23, N24, Q31 and T32), a cysteine residue (C30), five non-

polar residues (F22, A25, A27, G28 and L29) and three residues with aromatic rings (F22, Y21 and Y23). In BPTI, C30 forms a disulfide bridge with C51. Here, we take C30 in the reduced form.

The peptides were immersed in a previously equilibrated cubic box of water. All water molecules in this box that were within a distance of 0.27 nm of a non-hydrogen atom of the peptide were deleted from the system. The resulting α-helix (β-hairpin) system contains 112 (125) peptide atoms (including united atoms and hydrogen atoms forming part of polar groups) and 266 (255) water molecules. The β-hairpin carries a net charge of +1. The peptides were kept rigid during the simulations. The Coulomb interactions were modelled by a generalised reaction field (GRF)[19] with a cutoff of 0.95 nm. The local water density was calculated in a cubic grid of *ca.* 0.1 nm width and averaged over 10000 configurations. The local density of water gives values as high as $15\rho_0$(α-helix) and $16\rho_0$ (β-hairpin). Local densities averaged over 125 ps blocks show a uniform variation of $\pm 2\rho_0$ for $\rho > 3\rho_0$.

Plate 1(a) shows the local water density around the ideal α-helix. The local water densities calculated by MD and PMF calculations are in good agreement. A region where the density does not compare well is the C-terminus, with PMF overestimating the extent of the high-density region. Plate 1(b) shows several views of the hydration pattern of the β-hairpin. There is excellent agreement between the MD and PMF densities at the NH_3^+ of Lys, the SH group of Cys (with S coloured yellow) and various NH (with N coloured blue), OH and CO (with O coloured red) groups in the peptide. We find that the SH group in Cys is well represented by the water correlation functions and, therefore, SH should be treated as a polar group. There are two regions where the PMF and MD calculated densities do not agree. The MD simulation shows a large density of water (5.0) at a site surrounded by a carbonyl oxygen (O of C30 at hydrogen-bond distance of 0.263 nm) and five carbon atoms (CB, CG, CD1, CD2 of L29 and CG of N31) at distances close to contact (0.375, 0.423, 0.403, 0.442 and 0.413 nm,

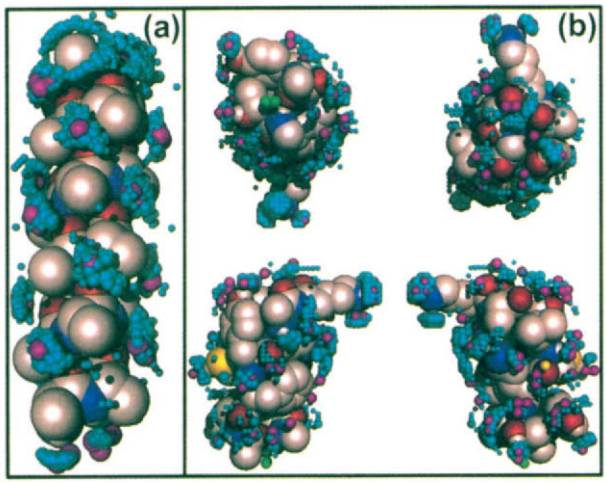

Plate 1 Hydration of rigid model peptides from MD simulations and PMF calculations. Results for an α-helix and a β-hairpin from BPTI are shown in panels (a) and (b), respectively. A magenta sphere is located at each grid point (0.1 nm grid width) with MD water density $\geq 5\rho_0$. PMF grid points (0.025 nm grid width) of high density are labelled with cyan spheres of various sizes. The local densities are divided into classes of low density ($3 \leq \rho/\rho_0 < 5$), medium density ($5 \leq \rho/\rho_0 < 7$) and high density ($\rho/\rho_0 > 7$). These grid points are illustrated by spheres of radii 0.03, 0.04 and 0.05 nm, respectively. In panel (b), MD points of high density with no corresponding PMF density are shown in green (top left; near N31) and yellow (bottom right; near T32).

respectively). The MD simulation shows another region of high density (4.0) at a site surrounded by four polar groups beyond hydrogen-bond distance (0.38 to 0.45 nm) and near contact with CB of T32 (0.368 nm) and close to the *N*-terminal acetyl methyl (4.10 nm), CG2 and T32 and CA of G28 (at 0.423 and 0.427 nm, respectively). In summary, PMF and simulation results for the positions of high density agree well. Significant discrepancies are observed only when water molecules are localised by interactions with several non-polar groups and only one or two polar groups. For these cases, a refined treatment of combined polar and non-polar groups is required, as outlined above.

3.4 Crystals of small nucleic acid molecules

Crystals of small nucleic acid fragments were studied in ref. 23. The structures of these small-molecule crystals were determined with high X-ray crystallographic resolution, such that the positions of ordered water molecules could be determined accurately. With one exception, the solvent channels in the crystals were small such that most of the water molecules were well ordered. The PMF hydration calculations showed spatial regions of high water density ($\geq 5\rho_0$) in excellent agreement with crystal-water positions. PMF high-density regions often connect different crystal-water sites, indicative of flexibility in the hydration network. This also agrees with the crystallographic observations: Seeman *et al.*[39] proposed to view the hydration of these crystals on the basis of a solid/liquid interface with thermal and statistical disorder in the water phase, to account better for the diffraction data than a conventional model using Gaussian electron-density distributions for water sites.

The radial dependence of the calculated water density around crystal-water sites peaked at or near $r = 0$ for most of the crystal-water positions. This demonstrates good positional agreement of calculated and measured water positions. However, the agreement was less satisfactory for integrated densities. When the integration extended over the volume of *ca.* one water molecule, sites with highly peaked densities contained less than one water molecule (typically 0.3 to 0.8), whereas less ordered regions often showed integrated densities closer to one. This suggests that in the PMF calculations, water-density peaks might actually be underestimated, although positioned correctly.

3.5 B-DNA oligomers

The hydration of B-DNA oligomers has been studied extensively using a variety of experimental techniques.[40,41] In X-ray crystallographic and, more recently, NMR studies,[42] the minor groove showed the most conserved hydration patterns. Two characteristic features emerged: In A·T base-pair regions, a highly localised so-called spine of hydration[43] occupies the minor groove. Two less localised (and less conserved) side-by-side ribbons of high water density[44] are found in wider G·C base-pair regions of the minor groove.

PMF calculations of B-DNA oligomer hydration did reproduce these experimental findings.[20,45,46] An interesting finding of those PMF studies was that hydration in the minor groove shows cooperativity. To account for the influence of neighbouring bases, the structural hydration in the minor groove is best described as a function of base-pair steps. In particular, marked differences were observed for the hydration of 5′-TA-3′ steps compared with that of 5′-AT-3′ steps, with TA steps showing less water localisation and disruption of the spine of hydration.[20,45,46] This is in agreement with earlier findings based on simple energetic arguments for the interaction of a single water molecule with DNA[47] and later NMR hydration studies comparing AATT and TTAA sequences.[48]

3.6 Transfer RNA molecules

Crystallisation and X-ray structure determination of transfer RNA (tRNA) molecules[49] initiated tRNA hydration studies.[50] Westhof *et al.*[4,50] described results for a small

region in $tRNA^{Asp}$ involving bases A9, G10 and G45. Water molecules in that region appear to be conserved between $tRNA^{Asp}$ and $tRNA^{Phe}$. However, the assignment of water is difficult as the crystallographic resolution of these structures is only 0.27 nm ($tRNA^{Phe}$) and 0.3 nm ($tRNA^{Asp}$).

Results of a PMF calculation for that region are shown in Plate 2. We compare calculations for two structures A and B of $tRNA^{Asp}$ with PDB codes 2tra and 3tra,[51] as well as a $tRNA^{Phe}$ structure (6tna[49]), which was later re-refined by Westhof *et al.* (4tra[51]). The calculations for $tRNA^{Asp}$ were performed both for isolated molecules and in the crystal environment to identify effects of crystal contacts.

The PMF hydration calculations indeed reveal some degree of conservation between different crystals and different tRNA molecules in the region of bases 9, 10 and 25 forming the interface between the dihydro-uridine loop and the anti-codon stem.[50] The conserved water molecules identified by Westhof *et al.* appear at or close to regions with high calculated water density. In particular, although none of the water molecules could be assigned in 2tra, the calculated high-density regions agree well with those of 3tra. PMF calculations indicate that one of the water molecules in 2tra and 3tra (marked with a green arrow in Plate 2) can, in part, be attributed to crystal packing. The PMF calculations show only few grid points with densities between 3 and $4\rho_0$ near that water site for the isolated molecules, but densities of $7\rho_0$ in the crystal.

PMF calculations can therefore provide direct information about the effects of crystal packing on the hydration structure. In addition, PMF calculations allow the hydration to be analysed, even for structures with intermediate to low resolutions, where the crystallographic assignment of water is difficult.

3.7 Antibody–antigen complexes

The X-ray structure of a complex of the Fv fragment of the anti-hen egg white lysozyme (HEL) antibody D1.3 with its antigen bound was solved at 0.18 nm resolution (PDB

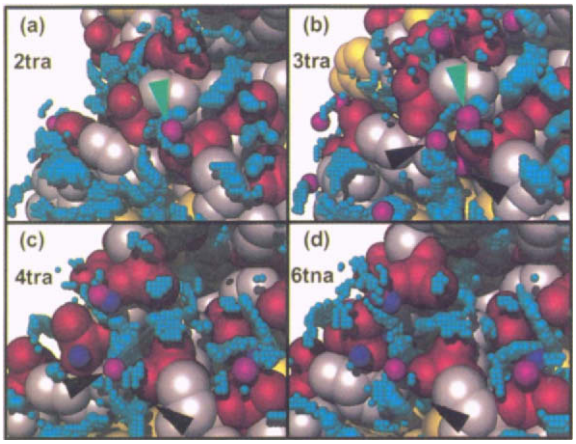

Plate 2 Hydration of tRNA from X-ray crystallography and PMF calculations. Results are shown for $tRNA^{Asp}$ [A and B form 2tra and 3tra in panels (a) and (b), respectively] and $tRNA^{Phe}$ [4tra and 6tna in panels (c) and (d)]. Small magenta spheres show water molecules in the crystal. The smallest cyan spheres indicate grid points (on a 0.03 nm grid) with calculated PMF densities between $3\rho_0$ and $5\rho_0$; larger cyan spheres indicate densities greater than $5\rho_0$. Atoms of bases, ribose and phosphate groups are shown in yellow, white and red, respectively. In panels (c) and (d), water molecules of the original refinement[49] are shown in blue. Crystal water of the re-refinement[51] is shown in magenta. Black arrows indicate conserved water molecules, as identified by Westhof *et al.*[50]

code 1vfb[10]). The larger Fab fragment of the D1.3 antibody complexed to HEL was solved at a lower resolution of 0.25 nm (PDB code 1fdl[52]). In Plate 3, results of PMF calculations are shown for the antibody–antigen contact regions of 1vfb and 1fdl. Only the higher-resolution 1vfb structure (as deposited at PDB) contains water positions, and only in the contact interface. Remarkably, the interface is strongly hydrated. In 1vfb, water molecules were found in partly connected clusters in the interface, thus correcting for a lack of direct surface complementarity of antigen and antibody.

The crystal-water sites are, with few exceptions, within or near high water-density regions of the PMF calculations (Plate 3). The reverse holds as well: In or near regions of high calculated density, one also finds crystal water (no water molecules were reported outside the interface, however). For the lower-resolution 1fdl structure, the predicted hydration agrees almost perfectly with that of 1vfb.

This illustrates that PMF calculations are rather insensitive to small structural changes, such as those between two crystal structures with different resolution. The agreement with the crystallographic hydration structure shows that PMF calculations can indeed be used in cases where the assignment of water is difficult, as for the Fab-D1.3 complex, 1fdl. Furthermore, PMF calculations can complement or substitute crystallographic hydration studies when only a low-resolution structure is available and water cannot be assigned unambiguously.

3.8 Photosynthetic reaction centre

Deisenhofer *et al.*[9] presented an X-ray structure of a photosynthetic reaction centre at a high resolution of 0.23 nm resolution (PDB code 1prc). Strongly hydrated interior regions were identified near the special pair and near the ubiquinone, Q_B. In Plate 4, we analyse those regions using PMF calculations. Again, we find agreement between the crystal-water positions and regions of high calculated density. Interestingly, water molecules hydrating the special pair and its nearest haem overlap with localised PMF density peaks. The hydration structure near the ubiquinone on the other hand shows wires of high density connecting individual water sites. Deisenhofer *et al.*[9] suggested that these water molecules near the ubiquinone, Q_B, could play a role in its protonation. The network-type structure observed in the PMF calculations indicates flexibility of the water structure in that region. Mobility is important for water to perform its function as a polar molecule screening charge interactions and lowering electrostatic barriers. Thus,

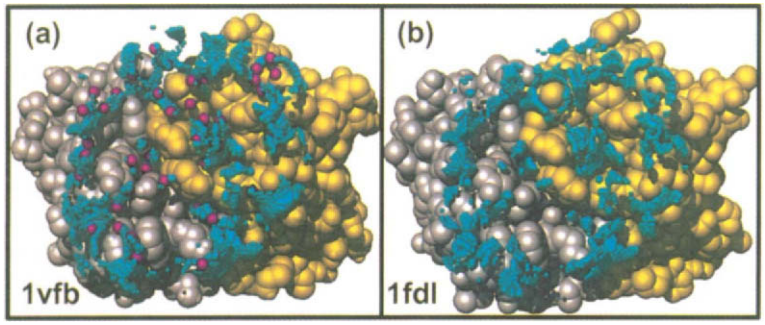

Plate 3 Hydration of the antibody–antigen contact interface of the Fv fragment [panel (a); 1vfb[10]] and Fab fragment [panel (b); 1fdl[52]] of antibody D1.3 complexed to HEL. The HEL molecule is removed for visualisation purposes, but was considered in the calculations. Yellow and white are used for the heavy and light chains. Crystal-water molecules are shown in magenta. PMF grid points with high water density are shown in cyan (see also Plate 2). No crystal water was reported in the PDB file 1fdl.

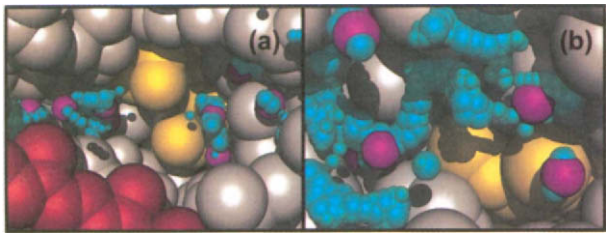

Plate 4 Hydration of a photosynthetic reaction centre.[9] Panel (a) shows the hydration between the special pair (yellow) and the nearest haem (red). Panel (b) shows the water channel near the ubiquinone, Q_B. Protein atoms are shown in grey, with parts of the protein removed to visualise the hydration of the protein interior. Crystal water is shown in magenta; PMF high-density grid points are shown in cyan (see also Plate 2).

the results of the PMF hydration calculation are further indications of a possible functional role of those water molecules near Q_B.

3.9 Future directions: Energetics of antibody–antigen binding

The crystallographic and PMF analysis of the antibody D1.3 complexed to its antigen HEL revealed a strongly hydrated interface. Thermodynamic data[10] show that the binding is accompanied by a significant negative entropy at room temperature. For a series of anti-lysozyme antibodies, we tried to correlate the degree of water delocalisation upon binding with measured binding entropies. As measure for the delocalisation of water, we used the lowest-order term in an entropy expansion in terms of correlation functions,[53] $\sigma = -k_B \int dr\, \rho(r) \ln[\rho(r)/\rho_0]$. The change of σ upon binding, $A + B \rightarrow AB$, is $\Delta\sigma = \sigma_{A+B} - \sigma_A - \sigma_B$. The PDB structure of the complex AB is used. The isolated molecules are those of the complex with B or A removed. Clearly, this treatment is very crude. It assumes that entropies associated with orientational degrees of freedom of water, side-chain orientations and higher-order correlations cancel each other in $\Delta\sigma$ or are constant for a class of antibodies. Therefore, we can at the best expect to get the trends right in a comparison with actual measurements. However, $\Delta\sigma$ should describe approximately the entropic effects of changes in polar hydration upon binding.

Fig. 3 shows results for six complexes of antibodies with lysozyme: D1.3 (1vfb[10] and 1fdl[52]), D11.15 (1jhl[54]), D44.1 (1mlc[55]), F9.13.7 (1fbi[56]), HyHel-5 (3hfl[57]) and HyHel-10 (3hfm[58]), with two complexes in the asymmetric unit of 1mlc and 1fbi. We indeed find a positive correlation between measured entropy and calculated $\Delta\sigma$. Also, calculations always show positive entropies of complex formation. The major neglected contribution is loss of side-chain orientational degrees of freedom upon binding, which would make entropy estimates more negative. The calculated absolute values are of the right order of magnitude, but the variation of experimental values between different antibodies is about three times that of $\Delta\sigma$.

Comparison of the results for antibodies with more than one crystal structure (D1.3) or more than one complex in the asymmetric unit (D44.1 and F9.13.7) shows quite a wide spread of $\Delta\sigma$. In the case of D1.3, this spread is mostly caused by differences in the hydration of the individual molecules and not the complex. An improved modelling could therefore use the crystal structures of individual molecules, if available.

We also used a different measure for the water delocalisation, $\omega(\rho_c) = \rho_0 \int dr \Theta[\rho(r) - \rho_c]$, measuring the extent of regions with density above ρ_c. For a density threshold $\rho_c = 3\rho_0$, we observed a somewhat better correlation between $\Delta\omega$ and measured entropies, compared with that of $\Delta\sigma$.

It is clear that this kind of entropic analysis is only at a preliminary stage. Antibody–antigen complexes are sufficiently complicated that no final conclusions can be drawn.

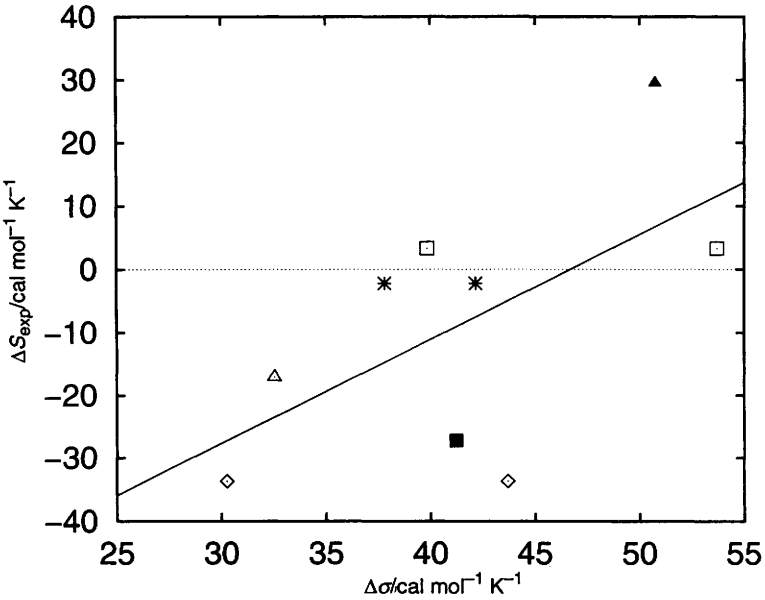

Fig. 3 Measured antibody–antigen binding entropies *vs.* estimated entropy, $\Delta\sigma$, of dehydration of polar groups. Thermodynamic results are shown for antibodies D1.3[60] (◇), F9.13.7[60] (□ HEL data), D44.1[60] (*), D11.15[61] (△), HyHel-5[62] (■) and HyHel-10[63] (▲). The solid line is a linear fit.

Conformational searches in ligand-docking should prove to be more accessible to this kind of analysis because there the changes are more local and trends should be easier to see. A simple example is a small molecule replacing one or a few very localised water molecules. The enthalpic contributions of this binding reaction can be estimated using conventional force fields. $\Delta\sigma$ should give a rough estimate of the entropic contributions owing to dehydration which is otherwise difficult to calculate.

4 Conclusions

We have presented a method to study theoretically the structural hydration of biomolecules. The local density of water near a macromolecule in solution or in a crystal is expressed in terms of particle correlation functions. These correlations have to be calculated only once and can then be used for any number of molecules. This results in a computationally very efficient method to study the structural hydration of large biomolecules. Illustrating this point, the PMF calculation for an antibody–antigen complex (volume 4.2 × 4.2 × 2.1 nm^3; 1.4 × 10^6 grid points) took less than 25 CPU minutes on a Silicon Graphics Indigo 2 workstation. Computer simulations of comparable systems[59] require significantly larger resources. In addition, a study of the hydration of the interfacial region is difficult using computer simulations, as the diffusion of water molecules into and out of interfacial cavities will occur on much longer timescales than are currently accessible to MD. Therefore, in that most interesting region, equilibrium averages cannot be calculated easily.

We have shown that PMF calculations show good agreement with computer simulations and crystallographic studies. However, it is important to notice that none of the methods to study structural hydration is without limitations. Crystallographic studies require structures at high resolution with no crystal contacts in the region of interest. In

simulations, model potentials have to be specified and even moderate statistical sampling of water densities on coarse grids requires long simulation times. PMF calculations rely on several approximations, but avoid the problem of insufficient statistics. In addition, as individual contributions can be identified and structural variations can be considered easily, PMF calculations can provide real insight.

PMF calculations can be used to study the hydration in the interior and at the surface of biological macromolecules in solution and in the crystal environment. Potential water sites can be identified on the basis of the calculated water-density distributions. Effects of crystal packing on the hydration structure can be estimated. Hydration studies of NMR and intermediate-to-low resolution X-ray structures are possible, where experimental information on hydration is limited. Beyond assigning water positions, the equilibrium density distribution contains information about the mobility of water. For instance, the water molecules hydrating the ubiquinone, Q_B, in the reaction centre 1prc[9] are connected through wires of high density, indicating flexibility in the hydration structure. Additional applications are the interactive modelling of structurally changed or mutated molecules, where no experimental data are available, or the screening of large numbers of candidate structures in docking studies.

This work has been funded by the Department of Energy (US). D.M.S. acknowledges support by the BMFT (FRG), the Max-Planck Society (FRG) and the Department of Energy (US).

References

1 J. A. Rupley and G. Careri, *Adv. Protein Chem.*, 1991, **41**, 37.
2 H. Savage and A. Wlodawer, *Meth. Enzymol.*, 1986, **127**, 162.
3 G. Otting, E. Liepinsh and K. Wüthrich, *Science*, 1991, **254**, 974.
4 E. Westhof, *Annu. Rev. Biophys. Biophys. Chem.*, 1988, **17**, 125.
5 X-J. Zhang and B. W. Matthews, *Protein Sci.*, 1994, **3**, 1031.
6 C. L. Brooks III and M. Karplus, *Meth. Enzymol.*, 1986, **127**, 369.
7 M. Gerstein and R. M. Lynden-Bell, *J. Phys. Chem.*, 1993, **97**, 2982.
8 R. M. Brunne, E. Liepinsh, G. Otting, K. Wüthrich and W. F. van Gunsteren, *J. Mol. Biol.*, 1993, **231**, 1040.
9 J. Deisenhofer, O. Epp, I. Sinning and H. Michel, *J. Mol. Biol.*, 1995, **246**, 429.
10 T. N. Bhat, G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza and R. J. Poljak, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 1089.
11 N. Thanki, Y. Umrania, J. M. Thornton and J. M. Goodfellow, *J. Mol. Biol.*, 1991, **221**, 669.
12 W. R. Pitt, J. Murray-Rust and J. M. Goodfellow, *J. Comput. Chem.*, 1993, **14**, 1007.
13 B. Schneider, D. M. Cohen, L. Schleifer, A. R. Srinivasan, W. K. Olson and H. M. Berman, *Biophys. J.*, 1993, **65**, 2291.
14 S. M. Roe and M. M. Teeter, *J. Mol. Biol.*, 1993, **229**, 419.
15 R. L. Dunbrack Jr. and M. Karplus, *J. Mol. Biol.*, 1993, **230**, 543.
16 D. M. Soumpasis, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 5116.
17 D. M. Soumpasis, *Computation of Biomolecular Structures*, ed. D. M. Soumpasis and T. M. Jovin, Springer, Berlin, 1993, pp. 223–239.
18 R. Klement, D. M. Soumpasis and T. M. Jovin, *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 4631.
19 G. Hummer and D. M. Soumpasis, *Phys. Rev. E*, 1994, **49**, 591.
20 G. Hummer and D. M. Soumpasis, *Phys. Rev. E*, 1994, **50**, 5085.
21 J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300.
22 A. Münster, *Statistical Thermodynamics*, Springer, Berlin, 1969, vol. 1, p. 338.
23 G. Hummer, A. E. García and D. M. Soumpasis, *Biophys. J.*, 1995, **68**, 1639.
24 I. Z. Fisher and B. L. Kopeliovich, *Dokl. Akad. Nauk SSSR*, 1960, **133**, 81.
25 H. S. Ashbaugh and M. E. Paulaitis, *J. Phys. Chem.*, 1996, in the press.
26 S. Garde, G. Hummer, A. E. García, L. R. Pratt and M. E. Paulaitis, *Phys. Rev. E*, 1995, **53**, 4310.
27 A. K. Soper and M. G. Phillips, *Chem. Phys.*, 1986, **107**, 47.
28 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
29 A. K. Soper, *Nucl. Instrum. Meth. Phys. Res. A*, 1995, **354**, 87.
30 D. M. Soumpasis, P. Procacci and G. Corongiu, IBM DSD report, October 1991, IBM.

31 U. Niesar, G. Corongiu, E. Clementi, G. R. Kneller and D. K. Bhattacharya, *J. Phys. Chem.*, 1990, **94**, 7949.
32 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Intermolecular Forces: Proceedings of the 14th Jerusalem Symposium on Quantum Chemistry and Biochemistry*, ed. B. Pullman, Reidel, Dordrecht, 1981, pp. 331–342.
33 O. A. Karim and A. D. J. Haymet, *Chem. Phys. Lett.*, 1987, **138**, 531.
34 O. A. Karim and A. D. J. Haymet, *J. Chem. Phys.*, 1988, **89**, 6889.
35 O. A. Karim, P. A. Kay and A. D. J. Haymet, *J. Chem. Phys.*, 1990, **92**, 4634.
36 W. L. Jorgensen and J. Tirado-Rives, *J. Am Chem. Soc.*, 1988, **110**, 1657.
37 A. Wlodawer, J. Deisenhofer and R. Huber, *J. Mol. Biol.*, 1987, **193**, 145.
38 F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.*, 1977, **112**, 535.
39 N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim and A. Rich, *J. Mol. Biol.*, 1976, **104**, 109.
40 W. Saenger, *Principles of Nucleic Acid Structure*, Springer, Berlin, 1984, pp. 368–384.
41 E. Westhof and D. L. Beveridge, *Water Science Reviews*, ed. F. Franks, Cambridge University Press, Cambridge, 1990, vol. 5, pp. 24–136.
42 E. Liepinsh, G. Otting and K. Wüthrich, *Nucleic Acids Res.*, 1992, **20**, 6549.
43 M. L. Kopka, A. V. Fratini, H. R. Drew and R. E. Dickerson, *J. Mol. Biol.*, 1983, **163**, 129.
44 G. G. Privé, K. Yanagi and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 177.
45 G. Hummer and D. M. Soumpasis, *Structural Biology: The State of the Art; Proceedings of the Eighth Conversations in the Discipline Biomolecular Stereodynamics*, ed. R. H. Sarma and M. H. Sarma, Adenine Press, Schenectady, NY, 1994, vol. 2, pp. 273–278.
46 G. Hummer, D. M. Soumpasis and A. E. Garcia, *Nonlinear Excitations in Biomolecules, Les Editions de Physique*, ed. M. Peyrard, Springer, Berlin, 1995, pp. 83–99.
47 V. P. Chuprina, *Nucleic Acids Res.*, 1987, **15**, 293.
48 E. Liepinsh, W. Leupin and G. Otting, *Nucleic Acids Res.*, 1994, **22**, 2249.
49 J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church and S-H. Kim, *J. Mol. Biol.*, 1978, **123**, 607.
50 E. Westhof, P. Dumas and D. Moras, *Biochimie*, 1988, **70**, 145.
51 E. Westhof, P. Dumas and D. Moras, *Acta Crystallogr., Sect. A*, 1988, **44**, 112.
52 T. O. Fischmann, G. A. Bentley, T. N. Bhat, G. Boulot, R. A. Mariuzza, S. E. V. Phillips, D. Tello and R. J. Poljak, *J. Biol. Chem.*, 1991, **266**, 12915.
53 H. S. Green, *The Molecular Theory of Fluids*, North-Holland, Amsterdam, 1952, p. 73.
54 V. Chitarra, P. M. Alzari, G. A. Bentley, T. N. Bhat, J-L. Eiselé, A. Houdusse, J. Lescar, H. Souchon and R. J. Poljak, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 7711.
55 B. C. Braden, H. Souchon, J-L. Eiselé, G. A. Bentley, T. N. Bhat, J. Navaza and R. J. Poljak, *J. Mol. Biol.*, 1994, **243**, 767.
56 J. Lescar, M. Pellegrini, H. Souchon, D. Tello, R. J. Poljak, N. Peterson, M. Greene and P. M. Alzari, *J. Biol. Chem.*, 1995, **270**, 18067.
57 G. H. Cohen, S. Sheriff and D. R. Davies, *Acta Crystallogr., Sect. D*, 1995, **52**, 315.
58 E. A. Padlan, E. W. Silverton, S. Sheriff, G. H. Cohen, S. J. Smith-Gill and D. R. Davies, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 5938.
59 F. Alary, J. Durup and Y-H. Sanejouand, *J. Phys. Chem.*, 1993, **97**, 13864.
60 F. P. Schwarz, D. Tello, F. A. Goldbaum, R. A. Mariuzza and R. J. Poljak, *Eur. J. Biochem.*, 1995, **228**, 388.
61 D. Tello, F. A. Goldbaum, R. A. Mariuzza, X. Ysern, F. P. Schwarz and R. J. Poljak, *Biochem. Soc. Trans.*, 1993, **21**, 943.
62 K. A. Hibbits, D. S. Gill and R. C. Willson, *Biochemistry*, 1994, **33**, 3584.
63 K. Tsumoto, Y. Ueda, K. Maenaka, K. Watanabe, K. Ogasahara, K. Yutani and I. Kumagai, *J. Biol. Chem.*, 1994, **269**, 28777.

Paper 6/00065J; *Received* 3rd *January*, 1996