Acc. Chem. Res. **2002,** *35,* 376−384

Solvation and Hydration of Proteins and Nucleic Acids: A Theoretical View of Simulation and Experiment

VLADIMIR MAKAROV AND
B. MONTGOMERY PETTITT*
Department of Chemistry, University of Houston, Houston, Texas 77204-5641

MICHAEL FEIG
Department of Molecular Biology, TPC-6, The Scripps Research Institute, La Jolla, California 92037

Received October 3, 2001

ABSTRACT

Many theoretical, computational, and experimental techniques recently have been successfully used for description of the solvent distribution around macromolecules. In this Account, we consider recent developments in the areas of protein and nucleic acid solvation and hydration as seen by experiment, theory, and simulations. We find that in most cases not only the general phenomena of solvation but even local hydration patterns are more accurately discussed in the context of water distributions rather than individual molecules of water. While a few localized or high-residency waters are often associated with macromolecules in solution (or crystals from aqueous liquors), these are readily and accurately included in this more general description. The goal of this Account is to review the theoretical models used for the description of the interfacial solvent structure on the border near DNA and protein molecules. In particular, we hope to highlight the progress in this field over the past five years with a focus on comparison of simulation and experimental results.

Introduction

Water plays a central role in the thermodynamics and structure of macromolecules. In particular, the stability and functionality of proteins and nucleic acids are dictated by specific as well as nonspecific solvent effects. The biological activity of these molecules generally occurs within a relatively narrow range of temperature, solvent chemical potential, and ionic concentration. Most cellular functions are driven not by temperature gradients but by changes in solvent environments including varying pH and ionic activities as well as different solute concentrations between cellular and subcellular compartments. It is thus of both practical and fundamental interest to understand the relation of the aqueous solvent to these important macromolecules.

In this Account, we will consider the central role water plays in biochemistry from a structural perspective. We give a perspective on the hydration patterns of both proteins and nucleic acids so that we may compare them. It is the understanding of context-sensitive hydration patterns which ultimately yields our most detailed understanding of molecular recognition. Such recognition, whether intermolecular (binding) or intramolecular (folding), occurs because of specific hydration patterns which mediate all outer-sphere interactions and contacts.

In our own work we have found the language of probability distributions to describe most hydration and solvation phenomena for these macromolecules. We find that in most cases not only the general phenomena of solvation but even local hydration patterns are more accurately discussed in the context of water distributions rather than individual molecules of water. While a few localized or high-residency waters are often associated with macromolecules in solution (or crystals from aqueous liquors), these are readily and accurately included in this more general description.[1,2]

We wish to consider the distribution of water given a nearby macromolecule. While much of the theoretical literature has used pair distribution functions in one form or another,[3] we and others have found it more convenient to use conditional single-molecule densities.[4] These have been termed perpendicular or proximal distribution functions,[4−6] and they essentially count water atoms with respect to the closest atom on the macromolecule or perpendicular to the surface. These distributions give a direct measure of the local density of solvent in the context of the macromolecular structure. Rather than consider individual, distinguishable molecules of hydration, one can consider the probability of finding a solvent molecule near the macromolecular solute of interest. It is natural to consider X-ray data in this probabilistic context rather than attempting to fit whole waters into density which may be only partially occupied.[1,7] It is useful to distinguish between water correlations (and thus interactions) with differing solute atoms since they exhibit different equilibrium distances.[8,9] In Figure 1, we show a typical simulated density of water near a protein. If we consider the water distribution radially out from the surface near an atom, we find the distributions given in Figure 2 for C, N, and O.[9] This model has had a striking experimental confirmation by careful analysis of X-ray data.[10] This method gives a quantitative measure to some familiar concepts for small molecules which can be transferred to macromolecules. The positions of the first minima in the solvent proximal distributions can be used to define the solvation shell or hydration layer around a given macromolecule. We will use the language of these distributions throughout the rest of this Account.

Proteins

Over the past decade there has been an extensive accumulation of evidence pointing to the critical role that solvation plays in the function and structural stability of proteins.[11,12] Simulation and theoretical studies of solva-

Vladimir Makarov was a scientist at Paracel when this article was written and currently is with BionomiX.

B. M. Pettitt is the Cullen Professor of Chemistry and Director of the Institute for Molecular Design. His research area is theoretical chemistry.

Michael Feig is currently a research scientist at the Scripps Institute.

10.1021/ar0100273 CCC: $22.00
Published on Web 03/08/2002

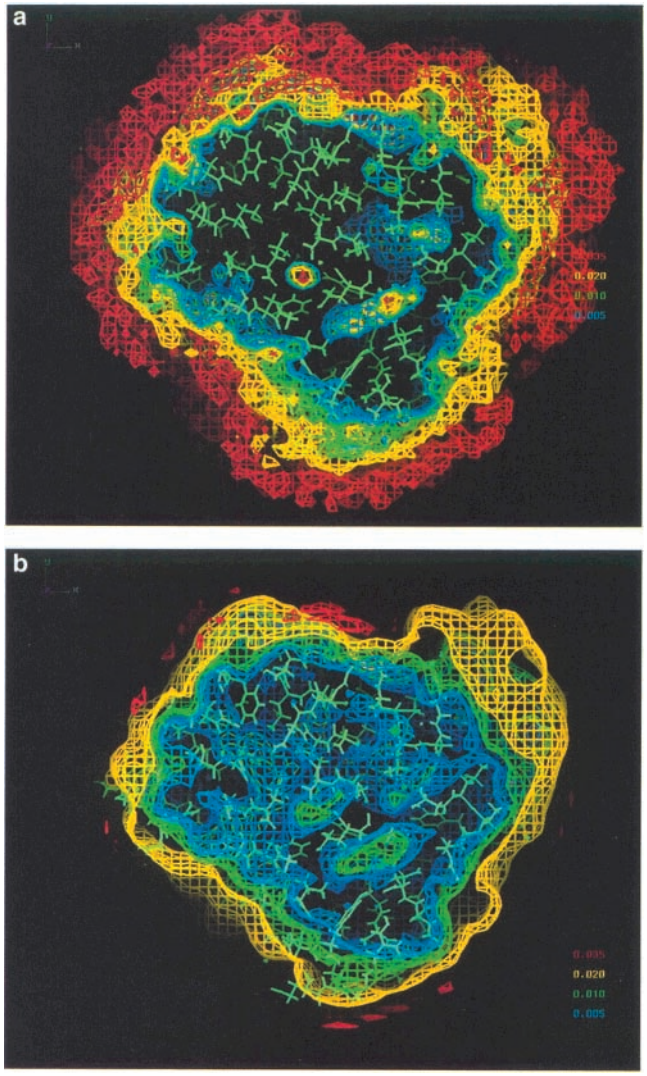

FIGURE 1. Three-dimensional density. Three-dimensional solvent number density distribution around myoglobin is shown as a slice computed from a molecular dynamics trajectory. The solvent density is overlaid with an average structure of myoglobin and contoured at 0.005 $Å^{-3}$ (blue), 0.01 $Å^{-3}$ (green), 0.02 $Å^{-3}$ (yellow), and 0.035 $Å^{-3}$ (red). The bulk solvent density is 0.033 $Å^{-3}$. (a) Density from simulation, (b) density from prediction. Reproduced with permission of the authors from ref 24.

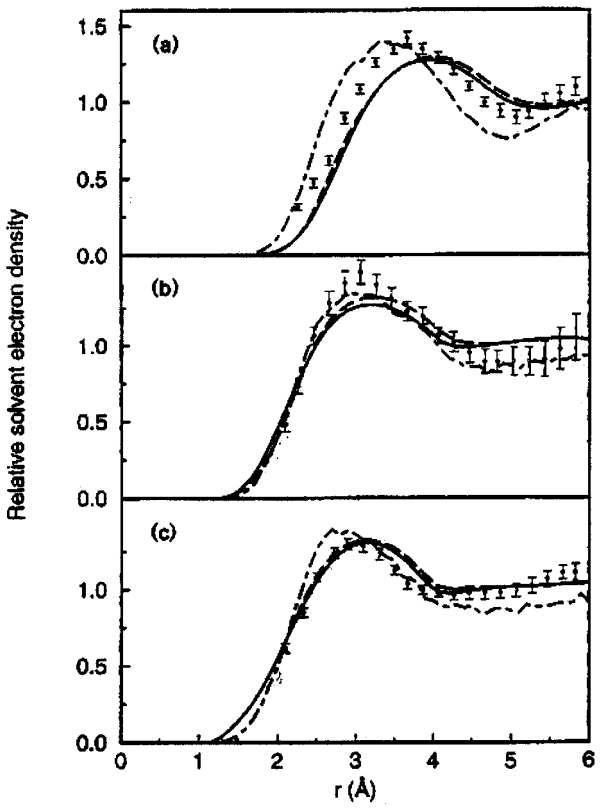

FIGURE 2. Radial distributions density. A comparison is made of the averaged radial distributions of solvent electron density around the surface atoms of various proteins: (a) oxygens, (b) nitrogens, and (c) carbons. Circles, rat mannose binding protein A, experimental data from Burling et al. (1996); solid line, sperm whale myoglobin, simulation; dashed line, azurin, simulation; dot−dashed line, BPTI, simulation. Note a remarkable agreement between the simulated density distributions and the experimental data, which holds even despite the differences in the nonbonded parameters and the water models used in the simulations. Reproduced with permission of the authors from ref 9.

tion have utility, much as in the case for DNA below for interpreting and refining macromolecular crystallographic data, for attempting ab initio protein structure prediction, and for drug design/screening.[13]

Density Distribution of Solvent Structure around Proteins. While a variety of experimental techniques have been successfully used for solvent structural studies around proteins,[12] the bulk of accurate spatial information results from diffraction experiments. At the same time, a consistent interpretation of the diffraction density maps presents its own difficulties.

Conventional treatments of solvent spatial distribution used in diffraction studies prior to 1996 relied upon refinement of a limited number of discrete, well-ordered solvent sites in an overall continuous solvent density.[4,12,14] Coordinates of these solvent sites are typically taken as a representation of an actual solvent molecule. The disordered solvent region is treated by various solvent flattening procedures,[15] and information contained in it is effectively discarded, even though it is not completely featureless.[10] Limitations of this approach were first recognized in the mid-1970s.[1,16] These limitations become particularly evident when one considers families of independently refined structures for the same protein. Comparisons yield a nearly perfect agreement of the protein portion of the structure, while the solvent sites do not, in general, agree with each other (Figure 3).[11,16−19] This discrepancy has been a source of some confusion in the past. Two factors have been pointed out as potential sources. The first one is the influence of crystal packing. The number of experimental observations that support it is indeed very large, and they have been extensively reviewed earlier.[11,12] Loris et al.[17] provide one of the more recent examples in this area. The second factor is the general error associated with interpretation of the continuum solvent density in terms of discrete water molecules.[12,14,19] In a 1986 review, Savage and Wlodawer[14] illustrate a number of cases when the "continuous" solvent sites cannot, and should not, be described this way. In subsequent work it has been shown

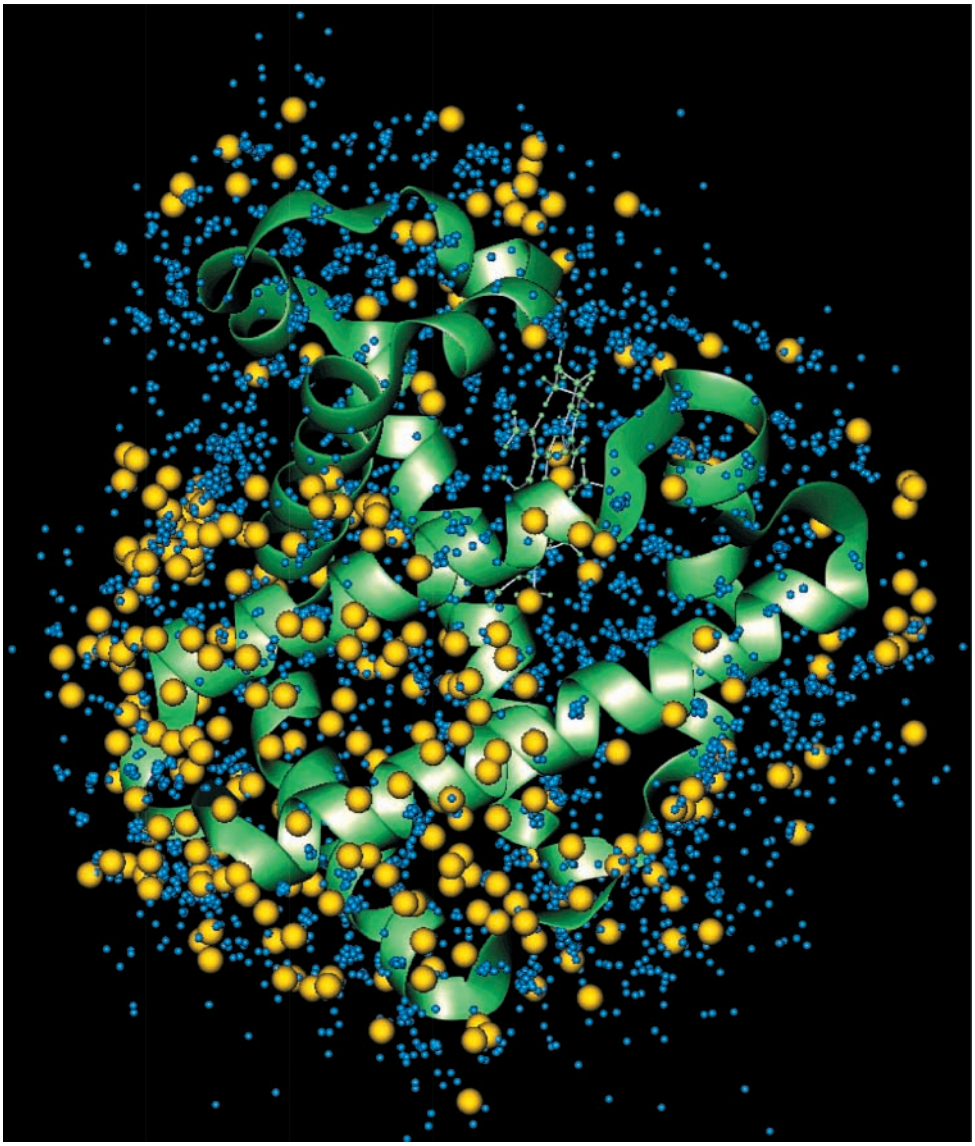

FIGURE 3. Comparison of simulation and experiment. A view of the hydration sites contained in the 39 independent Protein Data Bank entries for sperm whale myoglobin (blue dots) compared to the hydration number density maxima generated in the molecular dynamics simulation (yellow circles). Note the generally poor agreement between many discrete representations of hydration structure around the protein. Reproduced with permission of the authors from ref 22.

that this interpretation may introduce model bias into the crystal structure[4,10,20] and may serve as an additional source of error at a higher resolution.[4,7]

One more critical limitation of the discrete solvent site interpretation is realized in molecular simulation studies.[21,22] Molecular dynamics and three-dimensional free energy mapping point to the existence of a large number of local free energy minima on the protein surface. In a molecular dynamics simulation,[22] we observed that of the 294 high-density hydration sites around myoglobin, only at most 164 are occupied at any given time. No appreciable difference exists between the estimated free energies of conserved and nonconserved solvent sites. Thus, for a typical globular protein, any given set of ordered solvent molecules represents one of many possible configurations. Consequently, description of solvation based upon such one—or even many such descriptions—is bound to remain incomplete.

An alternative to the discrete approach would be to use a continuous solvation density distribution model. The need for a continuous representation was recognized early,[14] but the technical realization of it lagged.[10,20] The goal of such a method is to treat the solvent region as a single three-dimensional density distribution function. Probability distributions of this kind have been used in the past as a convenient way of describing and displaying the average solvent structure.[4−8,21,23−25]

One important advantage of the three-dimensional probability density distribution is that it is an experimentally observable quantity from X-ray diffraction in the form of the electron density map.[10,26] Using multiwavelength anomalous dispersion (MAD) phasing, the need for the initial fit to an atomic model is essentially eliminated, and the resulting map may be treated as a direct experimental result free from human bias. At the same time, it may be calculated from a molecular dynamics trajectory of a

protein with explicit solvent as described in refs 4, 7, 8, 21, 24, and 25 (Figure 1). Some authors have even subjected the simulated density to a mock refinement.[7,25] Finally, the three-dimensional density may be decomposed into quasi-component radial distributions. Historically this special type of distribution function, known as a quasi-component distribution function, was introduced by Beveridge and co-workers in the early 1980s[26] and applied to biological macromolecules.[7,8,24] This decomposition typically relates the property of interest (i.e., electron or number density) at a point to the closest chemical group or moiety on the solute. For groups of the same chemical type, the data are averaged, taking the corresponding normalization volume into account (which is a nontrivial computational task[7,8,24,26]). The resulting proximal or perpendicular radial distribution functions depend only on the reference chemical group and the distance. They are computed in groups, one function for every chemically distinct moiety of the solute. For proteins, one of the simplest possible decompositions would be into all oxygen, all nitrogen, and all carbon atoms.[7,9,10,24] The value of the density is then determined by the distance of the water with respect to the reference solute atom. Given this definition, we find that the proximal radial distribution functions are relatively insensitive to the overall shape of the protein and to the parameters of the model used in simulations, and therefore readily transferable between proteins irrespectively of their structure. This is indeed what is observed in practice (Figure 2).[9,10,24] These structural considerations are indirectly supported by the notion that the fractional composition of nonpolar, polar, and charged surface regions is very similar among diverse proteins, as observed in a protein volumetric study.[27] Good agreement with experiment suggests an immediate utility of this continuum solvent model for structural and theoretical studies.[7,9,39]

We have used a very simple 3-d density reconstruction procedure that consists simply of finding the proximal reference solute group followed by a subsequent assignment of the local density from the respective proximal RDF.[7,8,24] While this is obviously a pairwise approximation that ignores the contributions of all other, nonproximal, solute groups[24,26] and neglects the general many-body character of hydration,[29] it has been shown to perform remarkably well for solvent density reconstructions.[7,8,24] This proximal density distribution function decomposition has been strikingly confirmed by experiment.[10]

Solvent Structure and Mobility at the Interface with Proteins. The presence of a large macromolecular solute produces significant effects on both structural organization and mobility of the surrounding solvent. It is convenient to consider an average distribution of solvent in the form of a density map such as the one shown in Figure 4. Simulations show that there is a region inside the protein where the solvent does not penetrate easily or where the probability of finding it is very low (region A in Figure 4).[4,7] Outside of the "hydrophobic core" there is a transitional region (region B) that can be characterized by a significant degree of interpenetration of protein and

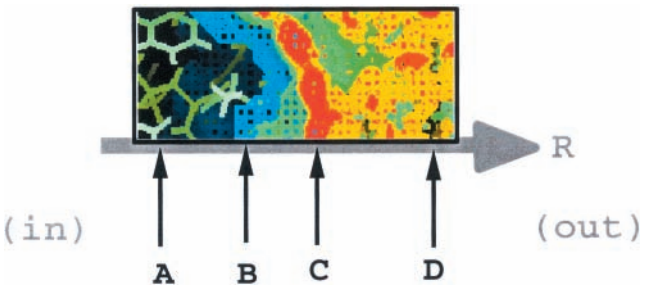

FIGURE 4. Global picture of protein hydration. This figure shows a one-dimensional slice through the protein-water interface. Markings "in" and "out" correspond to the interior and exterior of the protein. There are four distinct regions in the interface: (A) solvent-free region; (B) region of interpenetration of protein and solvent; (C) local maxima in the solvent density, "hydration sites"; and (D) remote region, where no distinct density maxima are found, but the diffusion rate is perturbed.

solvent. Observations in the field of protein crystallography and molecular dynamics simulations[7,24,30,31] appear to confirm this effect. The value of solvent density in this region is smoothly increasing as we are moving away from the protein and into the solution (Figure 4). It has been proposed[30] that solvent penetration provides a significant contribution to the high apparent polarizability inside proteins, although more work remains to establish this.

There are also isolated spots of high solvent density inside the otherwise hydrophobic core of the protein that correspond to regions where the water is encapsulated inside the protein.[22,24,32,33] It is well established that solvent penetration into the core influences protein dynamics on the global scale[32] and that an excessive water penetration, forced in, for example, by elevated pressure, may denature proteins.[34,62]

A large number of local density maxima are found in the 2.5−4.0 Å distance range from the protein (region C).[1,4,16,21,24,25] However, the density value averaged over the first hydration shell of the protein is higher than that of the bulk solvent by a modest value of at most 10%, as evidenced by simulations[4,7,22,24] and experimental observations.[36] This is in contrast to the effect for a polyion such as DNA (see below). The local peaks in the density are often referred to as "hydration sites".[4,21,22] These regions are also evident as minima in the three-dimensional free energy maps.[18] While sites are often too close to each other to be interpreted as individual water molecules, many authors have compared the positions of these statistical sites to those of the hydration crystal waters.[4,18,22] In general, these comparisons require other criteria to determine agreement. Nonetheless, the distributions observed in different studies are very similar. In reasonable cases, most crystal waters fit within 1.5 Å, on the average, from the theoretical sites.[18,22]

The mobility of buried and interfacial water is often severely restricted. Attempts to calculate the diffusion constant in the 2.5−4 Å range from the protein yield a value which is usually 25−50% lower than that in the bulk,[37,38] accompanied by a significant anisotropy.[22,38] In particular, the diffusion rate in the direction perpendicular

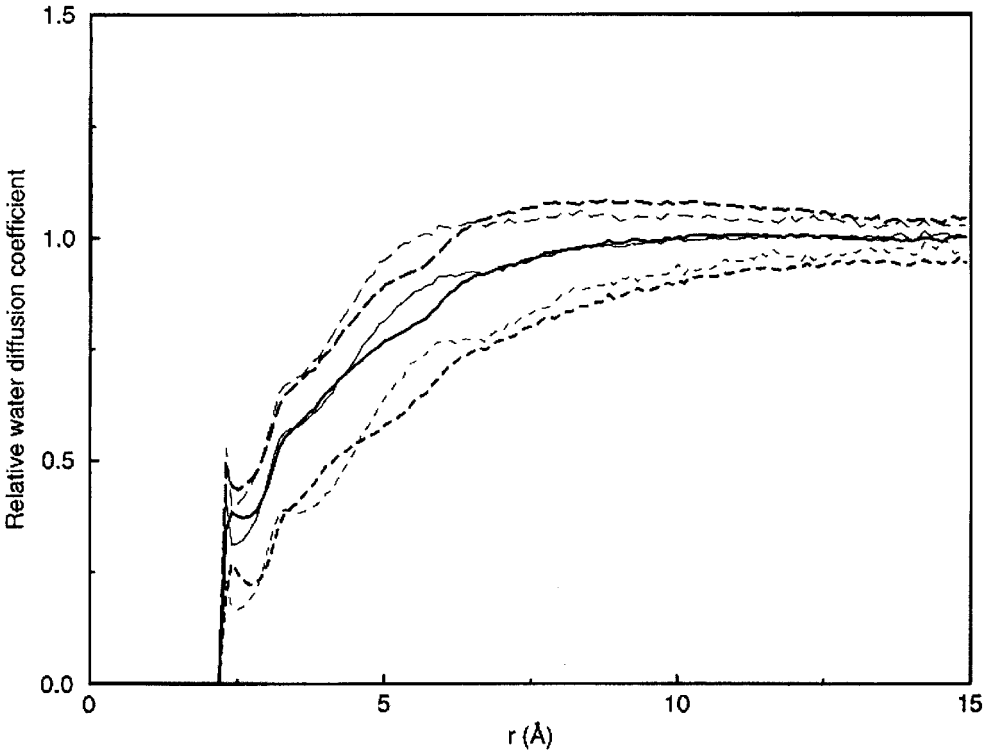

FIGURE 5. Radial diffusion profiles. Computed radial profiles of the water diffusion coefficient around myoglobin (thin lines) and DNA (thick lines). The overall diffusion coefficient is shown in the middle (solid), and perpendicular and parallel components are shown with short and long dashed lines, respectively. The curves are scaled so that the bulk value, achieved at high *R*, corresponds to 1. This anisotropy persists as far as 15 Å from the solute surface. Three relative depressions in the diffusion coefficient profile (at 2.7, 3.5, and roughly 5 Å) correlate with the locations of the peaks in the radial distribution functions. Reproduced with permission of the authors from ref 38.

to the solute surface is slower compared to the overall diffusion rate, while diffusion parallel to the solute surface is around 20% faster[22] (Figure 5). Diffusion rates are also lowered in the various hydration shells around the solute. This results in characteristic patterns of fine depressions in the radial profiles of the diffusion coefficient that roughly correspond to the major peaks in the radial distribution functions.[22,38]

For the buried and tightly bound water, the diffusion model does not apply well.[39] Such water typically has a residence time within its own volume on the order of hundreds of picoseconds, and occasionally much longer.[22,31,33] On the other hand, those water molecules that are more exposed to bulk have much shorter residence times, typically 5−50 ps.[21,22,41,43,44] The characterization of protein hydration water mobility by a residence time is an approximation in itself,[22,38,43] as residence times are almost universally obtained by an exponential fit to some "survival probability" time correlation function[38,41,43] that does not follow in general single-exponential kinetics. More accurate descriptions are achieved utilizing either polyexponential or the Kohlrausch−Williams−Watts stretched exponential models typically used for phenomena governed by multiple relaxation rates.[22]

This points to the existence of complicated scales in the time domain manifested in the form of a distribution of solvent residence times, which is specific for every location at the interface.[22] It appears that the primary factor determining the magnitude of the residence time is the surface geometry of the site.[22] Some of this is demonstrated in Figure 6. Water in the sites that are exposed to bulk typically has a shorter residence time than that in the grooves, clefts, and buried inside the protein.[22,45] Other factors that may potentially influence residence times appear to be secondary to the surface-geometry. A number of attempts have been made to correlate water residence times with the chemical nature of the protein residues closest to the site or the shell where the residence time is measured.[43] "Ranking relations" of residence times of water in proximity to (typically) charged, polar, and nonpolar groups appear to be conflicting or inconclusive.[19,42] Luise et al.[40] found that chemical differences between hydration sites, such as polarity of the surrounding residues or the ability of hydration water to hydrogen bond to the protein at the given site, only affect the residence times of water in the less accessible locations. Water residence times on the exposed surfaces of the protein apparently do not depend in a simple way on the surface polarity.[19,40] On the other hand, water in the nonpolar atomic sites has a short residence time irrespective of the solvent accessibility of the site.[22,40]

There is no direct correlation between the spatial (number density) and temporal (residence time) order of solvent in the protein hydration shell; i.e., simulated hydration sites with the highest local density do not on average contain water with the longest residence time.[19,22,40] A majority of waters appearing as ordered in a crystal cannot be resolved by NMR,[42] yet at the same time, the presence of positionally disordered water is indicated in

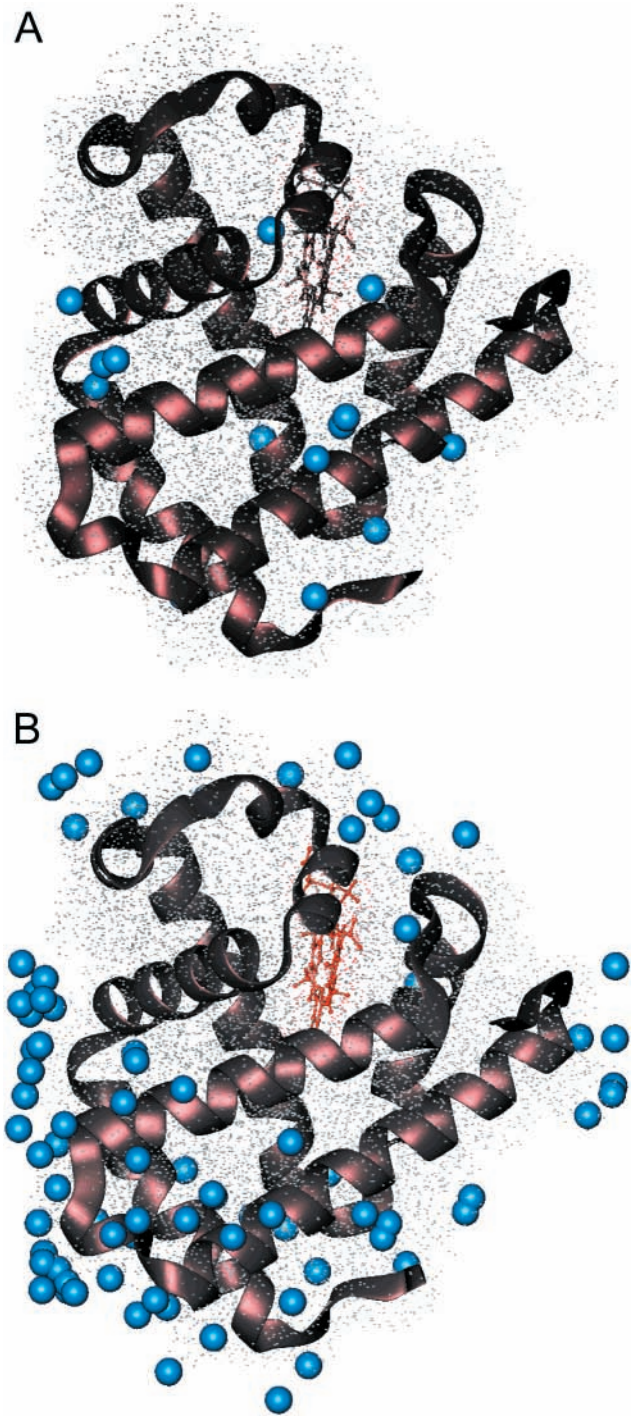

FIGURE 6. Long-life versus short-life waters. Spatial distributions of the hydration sites around myoglobin: sites with long residence times (longer than 80 ps) are buried inside (A) and in the grooves of the protein, while sites with short residence times (less than 10 ps) are exposed (B). The dotted area outlines the surface of the protein. Reproduced with permission of the authors from ref 22.

regions of the protein that appear not to contain water in the crystal structure.

Further into the bulk of solvent (at a distance of roughly 5−7 Å from the protein), one finds only weak maxima corresponding to the second hydration shell. While the overall value of the water diffusion coefficient in this region approaches the bulk value,[22] there is a big deviation between the rates of diffusion in directions parallel and perpendicular to the solute surface. This deviation reaches a maximum at about $R = 6$ Å. From 7 Å on, this deviation diminishes, but it is still clearly visible as far as 15 Å away from the protein, where there is no solvent structure present in either three-dimensional density maps or the radial distributions (region D in Figure 4).[22]

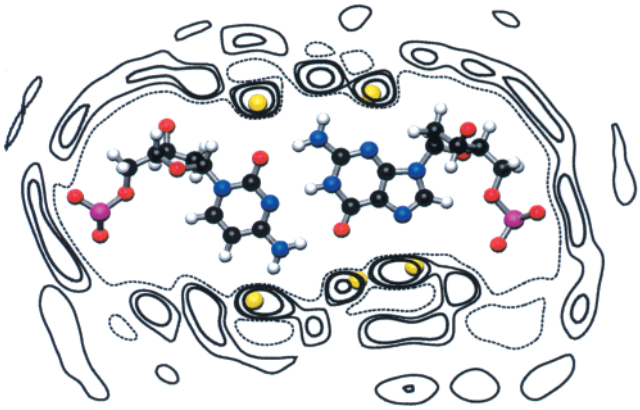

FIGURE 7. Solvation density around DNA. Density maxima (solid) and minima (dashed) are shown around a base plane computed from an MD simulation. The gold spheres are the consensus site from X-ray analyses (Berman et al.). The agreement is striking. From Feig et al. (1998) with permission of the authors.

It is clear that the solvation environment of proteins in part determines folding and is important in ligand design.[13] In our view, while the macromolecule is the focal point, its properties are inextricably linked to the solution conditions. Next we consider DNA.

Nucleic Acids

Nucleic acids and DNA in particular present a picture that is different from but complementary to that given above for proteins. The stability and conformational flexibility of nucleic acids are due largely to interactions with aqueous surroundings.[47,48] However, the strength of interactions with the surroundings is much greater than that for proteins with a similar molecular weight due to their highly ionic character. The binding of ligands to nucleic acids is also well known to induce aspects of the conformational polymorphism which is so important for the diverse nature of DNA's bioactivity.[49,50] High-resolution hydration patterns around the entire DNA molecule are clearly correlated with the predominant biological conformations.[2] The role of ions associating with DNA has been discussed with respect to changes in the hydration structure which correlate with DNA conformation.[51]

The hydration layers around DNA from computer simulations are easily visualized. Figure 7 shows the distribution of water averaged over like base planes. The proximal distribution functions for the system are quite similar to those shown above from proteins. The volume-normalized distributions show that the water density is increased up to 6 times the bulk density in the first hydration layer near oxygen and nitrogen atoms and around twice the bulk density near carbon atoms. Second and even third solvation shells can also be seen to

contribute to a locally enhanced density of water out to 0.8−1.0 nm from the molecular surface of DNA.

A to B and Back. Considerable attention has focused on the hydration patterns as they relate to equilibrium of the predominant biological forms of A and B. The early crystallographic evidence from DNA fibers demonstrated a preference for the A-form of DNA at low water activity and the B-form at high water activities.[52] The role of base pair composition is better understood as C/G base pairs generally tend less toward B-like conformations than A/T base pairs.[53] Solvent conditions which lower the activity of water by cosolvents or salt tend to cause a sequence-dependent B-to-A transition.[54]

A good deal of experimental data exist on the hydration patterns around A- and B-DNA.[47,56] A number of studies have contributed to the X-ray diffraction analysis literature on proximal waters.[14,56−59] In addition, complementary neutron diffraction studies which are more sensitive to hydrogen positions have been performed.[60] While NMR has some well-known sensitivity limitations, such experiments provide a view which is not biased by crystal packing, an issue more relevant for DNA than for proteins, as discussed above.[61] This laboratory and others have used theory and simulation to better understand DNA conformations mechanistically and their dependence on solution composition and structure.[2,62,63]

DNA molecules have considerably more favorable solvent interactions than proteins. While the fundamental forces which cause both to fold to unique structures are the same (i.e., the less soluble, "hydrophobic" groups are on the interior and the more soluble are on the exterior), the energetic contributions from free energies of solvation for DNA are stronger. The free energy of transfer from gas to water of the interior base guanine is similar to that of the most hydrophilic groups (the acids and bases) of proteins. The highly favorable free energy of solvation for DNA is due to the assortment of polar groups on DNA's surface with hydrogen bond donor and acceptor groups in both grooves and along the polyionic backbone interacting with the proximal waters.

The classic view distilled from many past studies shows that proximal waters of hydration near select sites on the surface DNA are much less mobile than bulk water, with residence times for some of a few hundred picoseconds.[64,65] These waters can often be associated with consensus waters from comparison of multiple experiments.[66] Such localized hydration sites correspond to well-defined density maxima. Given the number of strongly solvated sites, the density of water near DNA is increased over bulk water.[67] This is not unlike that seen in ionic hydration where the local density of water near even small ions is greatly increased, as evidenced by the solvent-to-ion radial distribution functions seen in both simulation and experiment.[68]

An array of nonstructural experiments on DNA fibers at various water activities, usually controlled by humidity, have demonstrated that roughly 18−30 water molecules per nucleotide are associated with B-DNA while fewer, 10−15, water molecules are found for A-DNA.[69] The more exposed sugar ring[70] and the more compact A-DNA conformation at the backbone are thought to be the major determinants of the reduced number of waters of hydration on A-DNA.[2] Due to the different probes used, some past controversies have existed.[67,71] Logically, given the strong free energy of hydration for G/C pairs[72] and G in particular, one might expect those bases to be better hydrated. However, as is well appreciated, under favorable conditions G/C base pairs are more likely to adopt the A conformation, while A/T base pairs are more stable as the B-form.[2] The apparent changes in differential hydration would then be more a reflection of the reduction in water activity around A-DNA compared to B-DNA than a direct effect of sequence.

Structures show that G/C base pairs have an additional hydrogen bond donor through the exocyclic guanine amino group in the minor groove compared to A/T base pairs. Computer simulations allow counting and averaging for the different canonical structures. By using the sequence of a block co-homo oligomer, the hydration preference for base type was also modeled.

From our simulations, we found the A-DNA hydration of the furanose ring is increased by nearly two water molecules, while the phosphate group and major groove each lost around a water when compared to B-DNA.[2,63] If we use a hydrogen-bonding criterion, then 18−21 water molecules were found to be "bound" to B-DNA. Using this technique gives results in good agreement with those of previous dielectric relaxation experiments which reported 18−19 (13−15) water molecules strongly interacting with B-DNA (A-DNA).[69] This emphasizes that many mixes of thermodynamic and structural criteria are reasonable and can yield distinctly different results.[2,63]

Thus, we see that the idea that G/C base pairs are less hydrated than A/T base pairs found in the early literature was due to the experimental technique used and reflects conformationally dependent hydration.[63,72] In addition, simulation then offers a mechanistic explanation for why G/C-rich sequences have a higher propensity toward A-DNA at modest water activity levels, while T/A-rich sequences remain in B-form until practically dehydrated: a lower number of bound waters is found around T/A base pairs in the A conformation. Concomitantly, ∼11 water molecules are found in hydrophobic contact in the first hydration layer of the A-form compared to ∼10 in the B-form and ∼9 around C/G base pairs.[2] The interplay between hydrogen-bonded (generally quite favorable) water and water exposed to less soluble groups ("hydrophobic" waters) thus reconciles the experimental phenomena and the detailed counting studies.

Solvent Structure versus Mobility at the Interface with DNA. As mentioned above, some water near DNA exhibits reduced mobility. We know that mobility and occupancy are not simply inversely correlated.[21] They are measures of different quantities, one kinetic, one thermodynamic, which we wish to understand in relationship to each other. At the extremes of long residence time and high occupancy, the concepts are certainly correlated though not identical. Thus, some water density is strong enough to

be observed by X-ray or neutron diffraction,[14] while highly localized water molecules are also found with NMR techniques.[61]

A variety of recurring themes in the patterns of the waters of hydration exist, including pentagonal water arrangements in A-DNA major grooves,[58] "cones of hydration" with three water molecules around each phosphate oxygen in B-DNA,[48] and the well-known spines of hydration seen in minor grooves of duplex DNA.[57,59,74] Ions are also seen in both simulation[51] and diffraction[59] to occupy the grooves with a reasonable probability. Different hydrogen-bonding patterns with DNA and between neighboring water molecules in the primary hydration shell around adenine and guanine have been seen.[51] This has clear implications for sequence-specific recognition of DNA through its base-specific solvation shell.

Ions have both kinetic and thermodynamic effects. A general reduction of water diffusion and therefore mobility near DNA has been seen in simulations of the Dickerson dodecamer or $d(CGCGAATTCGCG)_2$ with Mg^{2+} ions.[75] Monovalent ions in and around the major groove near guanine bases are thought to stabilize bridging water molecules between adjacent phosphates along the backbone that are characteristic of A-DNA hydration.

Thus, we have a clearer picture of how the presence of ions causes a structural rearrangement and a reduction in mobility of water near DNA. In this new picture, water molecules are redistributed in the presence of ions (monovalent or divalent) but not lost in a dehydration-like event.[55] They have a clear effect on water activity or chemical potential around DNA, and they reduce water mobility but not the number of waters.[38,75] This implies that reduced humidity A-form structures and the stabilization of A-DNA conformations of C/G-rich sequences in concentrated salt solutions are related but separate phenomena.

Simulated water distributions around the entire DNA molecule yield a consistent picture with experimental data. Both specific and general effects of ions on the water structure are seen in simulations and demonstrate that water in the first solvation shell is reordered and reduced in mobility by the presence of an ion, which has consequences for the populations of the specific water structures required to stabilize specific sequences.

Perspective

The combination of both continuum effects and specific solvent interactions is available from examining hydration in terms of water densities as described in this Account. The resulting view therefore gives the most complete picture of solvation effects in biological macromolecules available at this time.

Future directions for research in this area include studies concerning the important role of continuum protein hydration models in folding studies and in drug design. The implications of the results presented herein are broad and reach much farther than the fields of biomolecular simulation and protein/DNA X-ray crystallography from which they originate. In general, the current hydration description does not provide us sufficient information about the solvent structure around the protein in its natural aqueous environment. Hence, such descriptions can rarely be relied upon in accurate studies of molecular docking and ligand design[13] and folding.[46] In view of the above, it is becoming clear that macromolecular structure should not be considered without the environment. The entire notion of biopolymer structure is dependent upon the environmental conditions in which the protein exists.

The authors acknowledge partial support from the NIH, Texas Higher Education Coordinating Board, and the Robert A. Welch foundation. V.M. is grateful to Paracel, A Celera Business, for access to the FDF-4 TextFinder used for scanning of the library databanks for this Account. In addition, V.M. wishes to thank Dr. Yi Mo, Dr. Craig Hyde, David Beydler, and Andrew Basile for assistance and many inspiring discussions.

References

(1) Scanlon, W. J.; Eisenberg, D. Solvation of crystalline proteins. *J. Mol. Biol.* **1975**, *98*, 485−502.
(2) Feig, M.; Pettitt, B. M. Modeling Consistent with Experiment. *J. Mol. Biol.* **1999**, *286*, 1075−1095.
(3) Brooks, C. L.; Karplus, M.; Pettitt B. M. *Proteins: A Theoretical Perspective*; Rice, S., Prigogene, I., Eds.; Advances in Chemical Physics 71; Wiley: New York, 1988.
(4) Lounnas, V.; Pettitt, B. M. A connected-cluster of hydration around myoglobin. *Proteins* **1994**, *18*, 133−147.
(5) Mehrotra, P. K.; Marchese, F. T.; Beveridge, D. L. Statistical state solvation sites. *J. Am. Chem. Soc.* **1981**, *103*, 672−680.
(6) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. Monte-Carlo computer simulation of the aqueous hydration of the glycine zwitterion at 25 °C. *J. Biomol. Struct. Dyn.* **1984**, *2*, 1−27.
(7) Lounnas, V.; Pettitt, B. M.; Phillips, G. N. A global model of the protein−solvent interface. *Biophys. J.* **1994**, *66*, 601−614.
(8) Rudnicki, W. R.; Pettitt, B. M. Modeling the DNA−solvent interface. *Biopolymers* **1997**, *41*, 107−119.
(9) Pettitt, B. M.; Makarov, V. A.; Andrews, B. K. Protein hydration density: theory, simulations and crystallography. *Curr. Opin. Struct. Biol.* **1998**, *8*, 218−221.
(10) Burling, F. T.; Weis, W. I.; Flaherty, K. M.; Brunger, A. T. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **1996**, *271*, 72−77.
(11) Karplus, P. A.; Faerman, C. H. Ordered water in macromolecular structure. *Curr. Opin. Struct. Biol.* **1994**, *4*, 770−776.
(12) Phillips, G. N.; Pettitt, B. M. Structure and dynamics of the water around myoglobin. *Protein Sci.* **1995**, *4*, 149−158.
(13) Gane, P. J.; Dean, P. M. Recent advances in structure-based rational drug design. *Curr. Opin. Struct. Biol.* **2000**, *10*, 401−404.
(14) Savage, H.; Wlodawer, A. Determination of water structure around biomolecules using x-ray and neutron diffraction methods. *Methods Enzymol.* **1986**, *127*, 162−183.
(15) Deisenhofer, J.; Steigman, W. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor. *Acta Crystallogr. B* **1975**, *29*, 238−250.
(16) Scanlon, W. J.; Eisenberg, D. Solvation of crystalline proteins. *J. Phys. Chem.* **1981**, *85*, 3251−3256.
(17) Loris, R.; Langhorst, U.; De Vos, S.; Decanniere, K.; Bouckaert, J.; Maes, D.; Transue, T. R.; Steyaert, J., Conserved water molecules in a large family of microbial ribonucleases. *Proteins* **1999**, *36*, 117−134.
(18) Dennis, S.; Camacho, C. J.; Vajda, S. Continuum electrostatic analysis of preferred solvation sites around proteins in solution. *Proteins* **2000**, *38*, 176−188.
(19) Smith, P.; Pettitt, B. M. Modelling Solvation Effects in Biomolecules. *J. Phys. Chem.* **1994**, *39*, 9700−9711.
(20) Jiang, J.-S.; Brunger, A. T. Protein hydration observed by x-ray diffraction. *J. Mol. Biol.* **1994**, *243*, 100−115.
(21) Lounnas, V.; Pettitt, B. M. Distribution function implied dynamics versus residence times and correlations. *Proteins* **1994**, *18*, 148−160.
(22) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **2000**, *76*, 2966−2974.

(23) Beveridge, D. L.; Mezei, M.; Mehrotra, P. K.; Marchese, F. T.; Thirumalai, V.; Ravi-Shanker, G. Liquid state computer simulations of biomolecular solvation problems. *Ann. N. Y. Acad. Sci.* **1981**, *368*, 108−131.
(24) Makarov, V. A.; Andrews, B. K.; Pettitt, B. M., Reconstructing the protein−water interface. *Biopolymers* **1998**, *45*, 469−478.
(25) Schiffer, C. A.; van Gunsteren, W. F. Accessibility and order of water sites in and around proteins. *Proteins* **1999**, *36*, 501−511.
(26) Mezei, M.; Beveridge, D. L. Structural chemistry of biomolecular hydration via computer simulation. *Methods Enzymol.* **1986**, *127*, 21−47.
(27) Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M. Protein hydration. *Fold. Des.* **1998**, *3*, 105−118.
(28) Hummer, G.; Garcia, A. E.; Soumpasis, D. M. A statistical mechanical description of biomolecular hydration. *Faraday Discuss.* **1997**, *103*, 175−189.
(29) Perkyns J.;Pettitt, B. M. Peptide Conformations Are Restricted by Solution Stability. *J. Phys. Chem*. **1995**, *99*, 1−2.
(30) Dwyer, J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; Garcia-Moreno, B. E. High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* **2000**, *79*, 1610−1620.
(31) Likic, V. A.; Prendergast, F. G. Dynamics of internal water in fatty acid binding protein. *Proteins* **2001**, *43*, 65−72.
(32) Lounnas, V.; Pettitt, B. M.; Findsen, L.; Subramanian, S. A Microscopic View of Protein Solvation. *J. Phys. Chem.* **1992**, *96*, 7157−7158.
(33) Langhorst, U.; Loris, R.; Denisov, V. P.; Doumen, J.; Roose, P.; Maes, D.; Halle, B.; Steyaert, J. Dissection of the structural and functional role of a conserved hydration site in RNase T1. *Protein Sci.* **1999**, *8*, 722−730.
(34) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1552−1555.
(35) Denisov, V. P.; Jonsson, B. H.; Halle, B. Hydration of denatured and molten globule proteins. *Nature Struct. Biol.* **1999**, *6*, 253−260.
(36) Svergun, D. I.; Richard, S.; Koch, M. H.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2267−2272.
(37) Wong, C. F.; McCammon, J. A. Computer simulation and the design of new biological molecules. *Isr. J. Chem.* **1987**, *27*, 211−215.
(38) Makarov, V.; Feig, M.; Andrews, K. A.; Pettitt, B. M. Diffusion of Solvent around Biomolecular Solutes. *Biophys. J.* **1998**, *75*, 1−9.
(39) Bizzarri, A. R.; Cannistraro, S. Anomalous and anisotropic diffusion of plastocyanin hydration water. *Europhys. Lett.* **1997**, *37*, 201−206.
(40) Luise, A.; Falconi, M.; Desideri, A. Molecular dynamics simulation of solvated azurin. *Proteins* **2000**, *39*, 56−67.
(41) Brunne, R. M.; Liepinsh, E.; Otting, G.; Wuthrich, K.; van Gunsteren, W. F. Hydration of proteins. *J. Mol. Biol.* **1993**, *231*, 1040−1048.
(42) Denisov, V. P.; Halle, B. Protein hydration dynamics in aqueous solution. *Faraday Discuss.* **1996**, *103*, 227−244.
(43) Rocchi, C.; Bizzarri, A. R.; Cannistraro, S. Water residence times around copper plastocyanin. *Chem. Phys.* **1997**, *214*, 261−276.
(44) Carey, C.; Cheng, Y. K.; Rossky, P. J. Hydration Structure of the a-Chymotrypsin Substrate Binding Pocket. *Chem. Phys.* **2000**, *258*, 415−425.
(45) Tarek, M.; Tobias, D. J. The dynamics of protein hydration water. *Biophys. J.* **2000**, *79*, 3244−3257.
(46) Lazaridis, T.; Karplus, M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.* **1999**, *288*, 477−487.
(47) Westhof, E. Water: An Integral Part of Nucleic Acid Structure. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 125−144.
(48) Saenger, W. Structure and Dynamics of Water Surrounding Biomolecules. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 93−114.
(49) Robinson, C. R.; Sligar, S. G. Molecular Recognition Mediated by Bound Water. *J. Mol. Biol.* **1993**, *234*, 302−306.
(50) Chalikian, T. V.; Plum, G. E.; Breslauer, K. J. Influence of Drug Binding on DNA Hydration. *Biochemistry* **1994**, *33*, 8629−8640.
(51) Feig, M.; Pettitt, B. M. Sodium and Chlorine Ions as Part of the DNA Solvation Shell. *Biophys. J.* **1999**, *77*, 1769−1781.
(52) Franklin, R. F.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibers. *Acta Crystallogr.* **1953**, *6*, 673−677.
(53) Arnott, S.; Selsing, E. The Structure of Polydeoxyguanylic Acid Polydeoxycytidylic Acid. *J. Mol. Biol.* **1974**, *88*, 551−553.
(54) Pohl, F. M. Polymorphism of a Synthetic DNA in Solution. *Nature* **1976**, *260*, 365−366.
(55) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(56) Berman, H. M. Hydration of DNA: take 2. *Curr. Opin. Struct. Biol.* **1994**, *4*, 345−350.
(57) Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. Ordered Water Structure Around a B-DNA Dodecamer. *J. Mol. Biol.* **1983**, *163*, 129−146.
(58) Kennard, O.; Cruse, W. B. T.; Nachman, J.; Prange, T.; Shakked, Z.; Rabinovich, D. Ordered Water Structure in an A-DNA Octamer at 1.7 A Resolution. *J. Biomol. Struct. Dyn.* **1986**, *3*, 623−647.
(59) Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* **1998**, *37*, 8341−8355.
(60) Shotton, M. W.; Pope, L. H.; Forsyth, V. T.; Langan, P.; Grimm, H.; Rupprecht, A.; Denny, R. C.; Fuller, W. A High-Angle Neutron Fibre Diffraction Study of the Hydration of B-DNA. *Physica B* **1997**, *241/243*, 1166−1168.
(61) Liepinsh, E.; Otting, G.; Wuthrich, K. NMR Observation of Individual Molecules of Hydration Water Bound to DNA Duplexes. *Nucleic Acids Res.* **1992**, *20*, 6549−6553.
(62) Feig, M.; Pettitt, B. M. Structural Equilibrium of DNA Represented with Different Force Fields. *Biophys. J.* **1998**, *75*, 134−149.
(63) Feig, M.; Pettitt, B. M. Crystallographic Water Sites from a Theoretical Perspective. *Structure* **1998**, *6*, 1351−1354.
(64) Denisov, V. P.; Carlstrom, G.; Halle, B. Kinetics of DNA Hydration. *J. Mol. Biol.* **1997**, *268*, 118−136.
(65) Phan, A. T.; Leroy, J.; Gueron, M. Determination of the Residence Time of Water Molecules Hydration B′-DNA; B-DNA. *J. Mol. Biol.* **1999**, *286*, 505−519.
(66) Berman, H. M. Crystal Studies of B-DNA: The Answers and the Questions, Biopolymers. *Nucleic Acid Sci.* **1997**, *44*, 23−44.
(67) Chalikian, T. V.; Sarvazyan, A. P.; Plum, E. G. Influence of Base Composition, Base Sequence, and Duplex Structure on DNA Hydration. *Biochemistry* **1994**, *33*, 2394−2401.
(68) Enderby, J. E. *Annu. Rev. Phys. Chem*. **1983**, *34*, 155−185.
(69) Harmouchi, M.; Albiser, G.; Premilat, S. Changes of Hydration during Conformational Transitions of DNA. *Eur. Biophys. J.* **1990**, *19*, 87−92.
(70) Alden, C. J.; Kim, S. H. Solvent-Accessible Surfaces of Nucleic Acids. *J. Mol. Biol.* **1979**, *132*, 411−434.
(71) Tunis, M.-J. B.; Hearst, J. E. On the Hydration of DNA. II. *Biopolymers* **1968**, *6*, 1345−1353.
(72) Elcock, A. H.; McCammon, J. A. Sequence Dependent Hydration of DNA. *J. Am. Chem. Soc.* **1995**, *117*, 10161−10162.
(73) Tunis, M.-J. B.; Hearst, J. E. On the Hydration of DNA. I. *Biopolymers* **1968**, *6*, 1325−1344.
(74) Quintana, J. R.; Grzeskowiak, K.; Yanagi, K.; Dickerson, R. E. The Structure of a B-DNA Decamer with a Central T−A Step. *J. Mol. Biol.* **1992**, *225*, 379−395.
(75) MacKerell, A. D. Influence of Magnesium Ions on Duplex DNA. *J. Phys. Chem. B* **1997**, *101*, 646−650.

AR0100273