1
Biomolecular Solvation in Theory and Experiment
Michael Feig

1.1
Introduction

The modern understanding of biological macromolecules is unthinkable without a consideration of the solvent environment [1]. Solvent is crucial for the functioning of biological systems both directly, by actively participating in biological processes, and indirectly, by stabilizing biologically active conformations of proteins and nucleic acids. The aqueous milieu of virtually all living environments singles out water as the most important biomolecular solvent. Water is an extremely versatile solvent owing to its small size, and its inherently polar and polarizable nature. Its ability to act as both a hydrogen bond acceptor and donor allows it to form a variety of structures that can easily adapt to changes in environmental conditions and in particular to the presence of biomolecules and other co-solutes.

The interaction of water with biomolecules is a complex subject that has been studied for many decades. While the study of biomolecular solvation remains the focus of current research efforts, there are a number of major themes that have crystallized from past studies:

1) Solute–water interactions profoundly affect the conformational sampling of peptides [1]. In particular, protein folding to a unique native structure that can carry out biological function relies on the presence of water and in particular the hydrophobic effect [2–4] and, in peptides, the interaction with solvent determines secondary structure preferences [5]. Sufficient hydration is also necessary to maintain the biologically active B-form of DNA [6, 7].

2) Water is a critical participant in many biochemical reactions, where it can act as either an acid or a base and where it can rapidly transfer protons through the Grotthuss mechanism [8].

3) Solvent plays a key role in complex formation [9–11], ligand recognition [12–15], and DNA sequence recognition by DNA binding proteins [16].

4) The dynamic properties of water generally smooth solute energy landscapes [17] and, in particular, water has been found to act as a lubricant during protein folding [18].
5) Solvent interactions in biological environments are modulated further by the presence of co-solvents such as ions and a variety of small molecules [19], while entirely non-aqueous environments do not necessarily abrogate biological function for natively water-soluble proteins [20, 21]. Such environments are interesting in the context of industrial process biochemistry, where they may confer certain practical advantages [22]. Furthermore, non-aqueous biological environments exist in the form of lipid bilayers, where hydrophobic biomolecules can be sequestered from the aqueous solvent environment.

While a comprehensive review of biomolecular solvation in a single chapter is not possible, this chapter does attempt to provide a cursory overview of only those aspects that are most important to the subject of this book, the development of computational solvent models. For more in-depth discussions, the reader is referred to the relevant literature. This chapter is divided as follows: First, solvation is discussed from a thermodynamic and statistical mechanics perspective, followed by an overview of computational and experimental methods that are commonly used to study solvation around biomolecules. Next, the current knowledge about water–protein and water–nucleic acid interactions is summarized. Finally, the effects of non-aqueous solvents are described briefly.

1.2 Theoretical Views of Solvation

1.2.1 Equilibrium Thermodynamics of Solvation

The energetics of a given biomolecule is commonly separated into bonded and non-bonded interactions [23]. Bonded interactions stem from covalent bonding and effectively restrain local molecular geometries according to molecular bonding orbitals. The remaining non-bonded interactions add long-range charge–charge interaction, van der Waals type dispersion, and strong hard-sphere-like repulsion to avoid electron density overlap of non-covalently interacting atoms. The presence of solvent introduces solvent–solute and solvent–solvent terms in addition to the solute–solute non-bonded terms that are present in vacuum as well. The total free energy of a given biomolecule is therefore given as follows:

\[
\Delta G_{\text{total}} = \Delta H_{\text{banded solute–solute}} + \Delta H_{\text{bonded solute–solvent}} + \Delta H_{\text{banded non-bonded solute–solvent}} + \Delta H_{\text{bonded non-bonded solute–solvent}} + \Delta H_{\text{non-bonded solute–solvent}} + \Delta H_{\text{non-bonded solute–solvent}} - T\Delta S_{\text{solute}} - T\Delta S_{\text{solvent}}
\]  

(1.1)

where \(\Delta H_{\text{bonded solute–solvent}}\), \(\Delta H_{\text{non-bonded solute–solvent}}\), and \(\Delta S_{\text{solvent}}\) denote the changes in solvent–solute enthalpies and solvent entropy from bulk values.

As a good first approximation, the bonded terms in solute and solvent are largely unaffected by solute–solvent interactions, so that it is the balance of the non-
bonded enthalpic terms and entropic terms that determines the conformational preference of biomolecules in a given solvent environment. With the solvation free energy

\[ \Delta G_{\text{solvation}} = \Delta H_{\text{solute-solvent}}^{\text{non-bonded}} + \Delta \Delta H_{\text{solute-solvent}}^{\text{non-bonded}} - T\Delta S_{\text{solvent}} \]  

the biomolecular free energy then becomes

\[ \Delta G_{\text{total}} = \Delta H_{\text{solute-solute}} - T\Delta S_{\text{solute}} + \Delta G_{\text{solvation}} \]  

For proteins in aqueous solvent, the formation of well-defined, compact native structures results in favorable solute–solute enthalpies but decreased solute entropy. At the same time, native protein structures generally have unfavorable free energies of solvation compared to fully extended structures because favorable direct interactions between water and the peptide backbone are traded for intramolecular hydrogen bonding in the formation of secondary structure elements. Successful folding in aqueous solvent then requires that the folded structure retains sufficiently favorable interactions between the solvent and amino acid side chains. This results in the well-known hydrophobic effect, where hydrophobic amino acids that do not contribute significantly to \( \Delta H_{\text{solute-solvent}}^{\text{non-bonded}} \) are sequestered into the core of the folded structure while a large part of the protein surface consists of polar and charged residues that interact favorably with water. Hydrophobic surface residues are further discouraged energetically because the lack of hydrogen bond donors or acceptors forces the surrounding water to adopt a more ordered, clathrate-like structure that is both enthalpically and entropically less favorable than bulk water.

1.2.2 Radial Distribution Functions

The thermodynamic view of solvation can be connected to the molecular level with the aid of statistical mechanics by using correlation functions that capture the molecular organization of the liquid solvent. Typically, pairwise radial distribution functions (RDFs) \( g_{ij}(r) \) are used to describe the density distribution of an atomic or molecular species \( i \) as the function of distance from another species \( j \). Using RDFs it is then possible to calculate thermodynamic quantities. For example, the total interaction energy between \( N \) species \( i \) in a monatomic system that is only subject to pairwise interactions can be calculated according to

\[ U_{ii} = \frac{N \rho_i}{2} \int_0^\infty 4\pi r^2 g_{ii}(r) U(r) \, dr \]  

where \( \rho_i \) is the density of species \( i \) and \( U(r) \) is the pairwise interaction energy as the function of distance. This approach can be extended to systems with non-pairwise interactions by including higher-order correlation functions. Pairwise (or higher-order) distribution functions can be extracted directly from computer simulations (see Section 1.3.4) or from experiments.
As an example, Figure 1.1 shows RDFs between the oxygen atoms of water and the N, O, and C$\beta$ atoms of blocked alanine dipeptide from a molecular dynamics simulation in explicit solvent. For distances up to about 10 Å, the water oxygen density varies significantly from that of bulk solvent and also between different alanine dipeptide reference sites, indicative of the specific solvent structure around this solute.

1.2.3 Integral Equation Formalisms

The pairwise distribution function can be obtained alternatively by solving the Ornstein–Zernike integral equation, which decomposes the pairwise distribution function $h(r_{12}) = g(r_{12}) - 1$ between molecules 1 and 2 into a direct correlation function $c(r_{12})$ and an indirect contribution due to the presence of a third molecule:

$$h(r_{12}) = c(r_{12}) + p\int dr_{13} c(r_{13}) h(r_{13})$$

Solution of the Ornstein–Zernike integral equation requires an expression for the direct correlation function $c(r_{12})$, which can be approximated under certain assumptions. In the reference interaction-site model (RISM), this approach is extended to heterogeneous systems, and in particular solute–solvent interactions, by introducing site–site correlation functions between solute and solvent interaction sites.
1.2 Theoretical Views of Solvation

[24–27] as in Figure 1.1. Using the pairwise site–site correlation functions \( c(r) \) and \( h(r) \), free energies, enthalpies, and entropies of solvation can then be calculated directly [28, 29]. This approach has been used to understand the solvation of peptides [30–32] and co-solvent effects [33], and more recently to predict ligand binding sites [34]. It is described in more detail in Chapter 2.

1.2.4 Kirkwood–Buff Theory

Another route for connecting thermodynamics with the molecular nature of solutions is through Kirkwood–Buff (KB) theory [35, 36] (see Chapter 3). In this theory, thermodynamic quantities are related to the so-called KB integrals,

\[
G_{ij} = \int_0^\infty [g_{ij}(r) - 1] 4\pi r^2 \, dr
\]

(1.6)

In KB theory, RDFs are required as input to calculate \( G_{ij} \) and from those to derive thermodynamic quantities. The inversion of KB theory also allows \( G_{ij} \) values to be determined from thermodynamic measurements [37], but RDFs and the corresponding information about liquid structure cannot be extracted from KB theory. In contrast, the integral equation approach based on the Ornstein–Zernike equation allows the determination of pair correlation functions based on a given interaction potential.

Like integral equation theory, KB theory has also been applied to analyze the solvation of peptides [38] and in particular to investigate co-solvent effects [39–42]. Recently, KB theory has been used in the parameterization of force fields [43, 44].

1.2.5 Kinetic Effects of Solvation

Solute–solvent interactions also play a critical role in kinetic processes of the solute, such as chemical reactions and conformational transitions. The solvent environment may alter the free energy of the transition states through specific solute–solvent interactions. Examples of such interactions are found in enzyme active sites, where the presence of water molecules has been proposed to lower the transition state to accelerate the kinetics of enzyme reactions [45–47]. In contrast, less-specific stochastic collisions between solute and solvent atoms provide both frictional drag forces and the activation energy necessary to overcome kinetic barriers. A decrease in solvent viscosity has been found to accelerate protein folding kinetic rates [48], and even with friction, the mere presence of stochastic collisions between solute and solvent leads to an enhancement of barrier crossings [49]. As a result, transition state barriers are effectively lowered and, more generally, the energy landscape is smoothed. Over short time scales, such collisions are non-equilibrium processes, with the rate of collisions linked to the dynamics of solvent molecules at the biomolecular surface. The stochastic effect of solvent on kinetics is captured by Grote–Hynes theory [50] and has been applied successfully
in the context of variational transition-state theory [51] to explain the effect of solvent on reaction rates [52].

1.3
Computer Simulation Methods in the Study of Solvation

1.3.1
Molecular Dynamics and Monte Carlo Simulations

While the main subject of this book is the development of solvent models that can describe the effect of solvent on a given biological macromolecule in computer simulations, this chapter reviews how computer simulations can be used to investigate biomolecular solvation properties.

Typically, such simulations employ either molecular dynamics or Monte Carlo techniques, and rely on a classical model for the solute with an explicit representation of the solvent molecules. Both molecular dynamics and Monte Carlo simulations require an intra- and intermolecular interaction potential $V(r)$ commonly known as the “force field” [53]. In Monte Carlo simulations, the potential is used directly to accept or reject moves according to the Metropolis probability [54]

$$P = \min(1, e^{-\frac{1}{kT}(V(r')-V(r))})$$

(1.7)

In molecular dynamics simulations, the gradient of the potential is used to describe the dynamics of a given system according to Newton’s equation of motion,

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}) = -\nabla V(\mathbf{r})$$

(1.8)

This equation is then integrated in a stepwise fashion to obtain a trajectory.

The force field $V(\mathbf{r})$ commonly consists of bonded terms for covalent bonds, angles, and torsions along with non-bonded electrostatic and Lennard-Jones terms [53]:

$$V(\mathbf{r}) = \sum_{i=1}^{N_{\text{bonds}}} k_{\text{bond},i} (d - d_0)^2 + \sum_{i=1}^{N_{\text{angles}}} k_{\text{angle},i} (\theta - \theta_0)^2$$

$$+ \sum_{i=1}^{N_{\text{tensions}}} k_{\text{tension},i} [1 + \cos (n\phi - \phi_0)]$$

(1.9)

$$+ \sum_{i=1}^{N_{\text{atoms}}-1} \sum_{j=i+1}^{N_{\text{atoms}}} \left[ 4\varepsilon_{\text{LJ},ij} \left( \frac{\sigma_{\text{LJ},ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{\text{LJ},ij}}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi \varepsilon_0} \frac{q_i q_j}{\left| \mathbf{r}_{ij} \right|}$$

The Lennard-Jones potential captures both attractive van der Waals type dispersion and hard-sphere repulsion upon atomic contact formation [55]. Parameters for bond and angle terms are typically obtained from spectroscopic and crystallographic data, while the partial charges $q_i$, the Lennard-Jones coefficients $\sigma_i$ and $\varepsilon_i$, and the torsion parameters are usually fit to quantum-mechanical calculations. The classical formalism can be extended by including atomic polarizabilities.
through either dynamically fluctuating charges [56, 57], inducible dipoles and higher-order multipoles [58–60], or Drude oscillators [61, 62].

A major practical issue with the application of $V(r)$ is the calculation of the pairwise non-bonded terms. The Lennard-Jones term contributes significantly up to pairwise distances of about 8 Å [63]. The electrostatic term remains non-negligible up to at least 18 Å in a polar medium with high dielectric screening [64, 65] and much further, up to 30 Å, in non-polar environments [66], which becomes computationally intractable for large systems. In the case of an infinitely periodic system, the total Coulomb interaction can be calculated without a cutoff at moderate cost using the Ewald summation technique [67, 68]. A drawback of this scheme is that a sufficiently large simulation system is required to avoid artifacts when repeated periodically [69]. Nevertheless, simulations with periodic boundaries that use Ewald summation and in particular the efficient particle-mesh Ewald implementation [70, 71] have become the de facto standard for explicit solvent simulations.

1.3.2 Water Models

Realistic water models are crucial for the success of explicit aqueous solvent simulations and have been the subject of intensive research over many decades [72, 73]. The most straightforward and most widely used model of water consists of three sites, one at the center of the oxygen atom, and one each at the centers of the hydrogen atoms. Such a three-site model has only eight parameters: partial charges $q_o$ and $q_h$, Lennard-Jones parameters $\varepsilon_o$, $\sigma_o$, $\varepsilon_h$, and $\sigma_h$, and the O–H bond distance and H–O–H angle. With such a model, hydrogen bonds are formed as a result of a balance between electrostatic attraction and electronic repulsion through the $r^{12}$ part of the Lennard-Jones potential. The two most popular models, TIP3P [74] and SPC/E [75], are shown in Figure 1.2.

Water models are parameterized and tested by their ability to reproduce experimentally measured water structure (in form of radial distribution functions), density, enthalpy of vaporization, self-diffusion rate, and dielectric constant in bulk solvent. Three-site models can reproduce many of those quantities reasonably well at 298 K and 1 bar pressure, but different models trade good reproduction of one property with poor reproduction of another. For example, TIP3P matches the enthalpy of vaporization and dielectric constant well, but slightly underestimates the density and significantly overestimates the self-diffusion rate [72]. In contrast, SPC/E overestimates the enthalpy of vaporization and underestimates the dielectric constant, but accurately reproduces the density and self-diffusion rates [72]. Three-site models generally have difficulties in correctly describing the temperature dependence of liquid water properties. In particular, the density maximum at 282 K is often poorly reproduced, if present at all [72, 76]. These problems can be partially addressed by including additional off-nucleus electrostatic interaction sites to better represent the electronic structure of water and in particular to model the effect of the oxygen lone pairs. Some variants of these models, such as TIP4P
Biomolecular Solvation in Theory and Experiment

and TIP5P (see Figure 1.2), have resulted in significantly improved, although still not perfect, agreement with the experimental properties of bulk water. However, these models are often not used in practice because of the increased computational expense due to the additional interaction sites.

The electron distribution of water can significantly redistribute in response to external electric fields. The polarizability of water is reflected in a change of the dipole moment from 1.85 D in the gas phase to about 3 D in the bulk solvent. Fixed-charge models that do not allow the dipole moment to change and are parameterized to reproduce bulk properties are therefore problematic when used in combination with solutes that may alter the polarization of water, such as ions or hydrophobic solutes. This has motivated the development of polarizable water models, with the most successful models relying on the fluctuating charge approach, such as the TIP4P/FQ model. Polarizable water models are even more expensive than multi-site water models, but may be necessary to correctly reflect water interactions in cases where the environment of water molecules differs significantly from the bulk solvent.

1.3.3 Solvent Structure and Dynamics from Simulations

The main advantage of computer simulations is that, given sufficient sampling, solvent structure and dynamics can be extracted directly from the resulting trajectories. Solvent structure is characterized in the form of either radial distribution functions (see Figure 1.1) or volumetric density distributions (see Figure 1.3).
Radial distribution functions are commonly calculated by collecting pairwise distances in a histogram and normalizing the resulting distribution for the bin $[r, r + \Delta r]$ by the number of solvent molecules in a spherical shell of width $\Delta r$:

$$4\pi r^2 \Delta r \cdot \rho_{\text{solvent}} \cdot \frac{N_A}{M_{\text{solvent}}}$$

where $\rho_{\text{solvent}}$ is the solvent density, $M_{\text{solvent}}$ is the molar mass of the solvent, and $N_A$ is Avogadro’s number. When radial distribution functions are calculated with respect to solute sites, they reflect the structure of the solvation layer(s).

Solvent density distributions are easily counted with a three-dimensional histogram, but they require either that the solute remains fairly rigid or that the solute is reoriented appropriately for each configuration before the solvent density is accumulated. The resulting volumetric map shows preferential solvent interaction sites as high-density regions and identifies the solvent-excluded volume. An example of the three-dimensional solvent distribution around alanine dipeptide is shown in Figure 1.3. The high-density regions often correlate with crystallographic water sites, but the apparent order does not necessarily imply altered dynamic properties, since a site with a single long-lived solvent molecule is not distinguished from a site where solvent molecules rapidly exchange with the environment.

Dynamic properties of the solvent can also be readily calculated from simulations, typically through time correlation functions. Because solvent dynamics might be substantially altered in the vicinity of a solute, such analyses often consider conditional correlation functions that take solute proximity into account.

**Figure 1.3** Water oxygen density around alanine dipeptide from molecular dynamics simulation. The gray surface encloses the solvent-excluded volume. High-density regions are shown in purple, with the highest densities represented by solid surfaces.
property that is commonly calculated in this manner is the distribution of solvent
residence times in a specific region according to the correlation function [80–82]

\[
c_\alpha(t) = \frac{1}{N_{\text{solv}}} \sum_{i=1}^{N_{\text{solv}}} \frac{1}{N_{\alpha,i}} \sum_{t=0}^{t} p_{\alpha,i}(t', t' + t)
\]

(1.10)

where \( p_{\alpha,i}(t', t' + t) \) is a binary function with a value of 1 if solvent molecule \( i \)
remains within the confined area \( \alpha \) from time \( t' \) to \( t' + t \) and a value 0 otherwise.
\( N_{\alpha,i} \) is the number of times that a solvent molecule is present in \( \alpha \) and the outer
sum is over all solvent molecules \( i \). Other correlation functions can be calculated
to determine rotational correlation times, the lifetime of specific interactions such
as hydrogen bonds, and so on. In all cases, the resulting correlation functions can
be fit to a single, double, or stretched exponential function [82, 83] to extract characteristic
time scales that can be compared with experimental measurements. The
comparison with experiments, however, relies on a definition of the region \( \alpha \) in
Equation 1.10 that matches the experimental measurements best, which is not
always straightforward.

The self-diffusion coefficient measures the rate of mean square displacement
in the long-time limit. It is calculated according to the Einstein relation [84] from

\[
D = \frac{1}{6} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle |\vec{r}_i(t_0 + t) - \vec{r}_i(t_0)|^2 \rangle
\]

(1.11)

where the average is taken over all solvent molecules \( i \) and all time origins \( t_0 \). It
is possible to calculate a local diffusion coefficient from a short-time (1–2 ps) finite-
difference approximation to Equation 1.11 and to accumulate the resulting values
of \( D \) on a grid according to the location of a given solvent molecule to obtain
spatially resolved local diffusion coefficients [85, 86]. Furthermore, the three-
dimensional diffusion coefficient can be decomposed into components that are
parallel and perpendicular to a given biomolecular surface in order to fully capture
the variation of solvent dynamics near biomolecular solutes [85].

1.3.4 Free Energy Simulations

Free-energy simulations [87] using thermodynamic integration, free-energy perturbation,
or more advanced methods [88] can calculate solvation free energies
[89–92] or its components, for example, the solute–solvant component or the
charging free energy that results from turning on solute charges in an already
solvated system [93]. Compared to the calculation of solvation free energies from
integral equation theory, the use of computer simulations is very expensive since
extensive sampling is required to obtain converged results [94]. However, in the
simulations, no further assumptions are made beyond the classical interaction
model, and all multi-body effects are included implicitly. As a result, solvation free
energies calculated from free-energy simulations can provide results in close
agreement with experimental measurements for small molecules [91]. While
experimental solvation free energies of biomolecules are generally not available,
calculated free energies of solvation take their place as important reference quantities in the development of solvent models [95].

1.4 Experimental Methods in the Study of Solvation

A wide array of experimental methods have been applied in the study of solvation around biomolecules. The main techniques are described in this section.

1.4.1 X-Ray/Neutron Diffraction and Scattering

The most direct evidence for solvent molecules interacting with biological macromolecules comes from X-ray crystallography. X-ray crystallography averages electron densities over time and over a large number of aligned molecules. Solvent molecules are visible if they are ordered with respect to the crystal cell. Most protein and nucleic acid structures contain water molecules at various sites on the biomolecular surface, and much has been learned about the interaction of solvent with biomolecules from X-ray crystallography [96–98].

Crystal water sites are generally understood to reflect the presence of ordered water molecules, but a static interpretation as a single water molecule occupying a specific site according to the crystallographic coordinates can be misleading. Because of the dynamic nature of water, the solvent-associated electron densities are generally less well defined than for the solute. Therefore, a single water molecule oscillating in place can often not be distinguished from dynamically moving and exchanging solvent atoms that frequently occupy a certain site [99]. In other cases, water sites have been reported to lie at average positions between two actual water sites, with the reported water sites actually rarely being visited according to simulations [100]. Furthermore, in lower-resolution crystal structures, water molecules may be added at seemingly random positions simply to reduce the crystallographic R-factor during the refinement stage [97, 101]. In contrast, the very highest-resolution crystal structures often reveal water sites very clearly and may provide additional information about fractional occupancy and alternate hydration patterns [102, 103].

Most of today’s crystal structures are obtained at cryogenic temperatures. While the lowered thermal fluctuations of the solute improve the resolution, the effect on the solvent is less clear. It has been suggested that not just solvent dynamics but also solvent structure are altered significantly [104]. One apparent effect seems to be the formation of ordered, presumably enthalpically stabilized, hexagonal and pentagonal water networks [102, 103] that may otherwise not be present at room temperature. Additional solvent ordering is also likely due to packing forces in the imposed crystal lattice environment.

Neutron diffraction is similar to X-ray crystallography, but scattering occurs at the nuclei rather than the electrons. In contrast to X-ray crystallography,
neutron diffraction is sensitive to hydrogen atoms and it is therefore well suited for the study of water structure. Neutron diffraction has been especially useful for the experimental determination of pairwise radial distribution functions [105].

Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) provide low-resolution structural information, in particular about the distribution of intramolecular distances, from which the size of a given biomolecule can be estimated. The presence of a hydration layer with increased solvent density results in an apparent increase in size in SAXS measurements but a decrease in size in SANS measurements owing to different relative scattering lengths between water and biomolecules with X-rays and neutron radiation. If the structure of a given molecule is known, for example, from X-ray crystallography, SAXS and SANS can therefore provide information about the extent of the hydration layer [106].

Quasi-elastic and other inelastic neutron scattering methods measure the spatially resolved loss of kinetic energy upon scattering in the form of a so-called dynamic structure factor [107, 108]. The dynamic structure factor contains information about solvent dynamics that can be extracted either directly [109] or by interpreting the experimental data with the help of molecular dynamics simulations [110, 111]. Quasi-elastic neutron scattering typically results in estimates of the translational and rotational diffusion rates of interfacial solvent molecules interacting with biomolecules.

1.4.2 Nuclear Magnetic Relaxation

Nuclear magnetic relaxation (NMR) techniques have been used for a long time to study the properties of biomolecular solvation [112–114]. They can provide both structural and dynamic information about water near biological solutes. There are essentially two main NMR-related techniques.

The first relies on the nuclear Overhauser effect (NOE) between solvent and solute atoms based on the relaxation-induced transfer of magnetization. This transfer process is sensitive to short distance separations and therefore allows a localized analysis of solvent molecules interacting directly with a given biomolecular surface. However, NOE measurements are limited in temporal resolution and arguments have been made that NOE measurements cannot fully distinguish between interfacial and bulk solvent, so that the resulting solvent residence times may be significantly overestimated [115].

The other NMR technique, magnetic relaxation dispersion (MRD), measures the longitudinal magnetic relaxation rate of water deuteron or quadrupolar $^{17}$O nuclei as a function of resonance frequency. This technique has a shorter time resolution but lacks spatial resolution [116, 117]. Nevertheless, the application of MRD has led to detailed insight into rotational and translational relaxation of water in the vicinity of biomolecules [117].
1.4.3 Optical Spectroscopy

Femtosecond-resolution fluorescence spectroscopy is the third major technique for studying the dynamics of biomolecular solvation [108, 118–120]. It involves the time-resolved analysis of the fluorescence of a laser-excited fluorophore probe located at a biomolecular surface. The location of the fluorescence peak is sensitive to the probe dipole, which in turn depends on the solvent polarization. As the solvent relaxes in response to the excited probe, the fluorescence peak exhibits a (Stokes) shift to longer wavelengths. The dynamics of the fluorescence shift is therefore assumed to be a direct reflection of the solvent relaxation dynamics [119]. Dynamic fluorescence spectroscopy is in principle a very powerful technique, since it combines very high temporal resolution with spatial resolution based on the known location of the fluorescence probe. However, in practice, the results require further interpretation with the help of molecular dynamics simulations [121, 122], since the solvent dynamics extracted from the fluorescence shift measurements depends also on the degree of surface exposure of the probe [122].

Infrared spectroscopy has been used to probe the change in vibrational spectra of both water and solute atoms as a result of solvation. Vibrational spectroscopy is particularly sensitive to the formation of hydrogen bonding and allows the study of specific solvent–solute interactions through site-specific labeling [123, 124]. Such experiments provide mostly qualitative information about the degree of solvation of different parts of a given solute and have been able to suggest two-dimensional networks of hydrogen-bonded water molecules on the surface of solvated proteins [125]. However, more quantitative information is often difficult to extract from such studies [1].

1.4.4 Dielectric Dispersion

Dielectric dispersion measurements are one of the oldest techniques for studying the dynamics of solvated biomolecules [108]. They provide information about characteristic time scales from the frequency dependence of the dielectric constant. However, this method does not provide spatial resolution, the time resolution is limited, and the features observed in the dielectric spectra are often difficult to interpret in terms of specific aspects of solvent dynamics. Therefore, this technique is not widely applied today.

1.5 Hydration of Proteins

Numerous theoretical and experimental studies of protein–water interactions have resulted in a complex picture of protein and peptide hydration that transcends a
range of spatial and temporal scales. In the following, the effect on protein folding and peptide conformational sampling is described first before the properties of water near protein surfaces are reviewed.

1.5.1

**Protein Folding and Peptide Conformations in Aqueous Solvent**

The overall effect of aqueous solvent on proteins in terms of the hydrophobic effect and dielectric screening of electrostatic interactions between polar and charged groups is well understood. For longer peptides and proteins, solvent promotes the formation of regular secondary structure elements and folding into unique native structures [126]. In short peptides, where hydrophobic residues cannot be fully sequestered from the aqueous solvent environment, the electrostatic screening effect is dominant and results in competition for intramolecular salt bridges and hydrogen bonds [127, 128]. In fact, the formation of a single backbone hydrogen bond versus individual solvation of the carbonyl and amide groups in a short peptide is energetically nearly net neutral, with only a slight enthalpic advantage of about 1 kcal mol\(^{-1}\) per residue, which is countered by an entropic cost [129]. This gives rise to a delicate energetic balance that allows short peptides to adopt a variety of different conformations with similar relative free energies depending on the amino acid sequence [1, 130]. Recent experimental and computational research suggests a general preference of short peptides for polyproline II (PPII) conformations with backbone dihedral torsion angles in the vicinity of \(\phi = -85^\circ\) and \(\psi = 140^\circ\) that is apparent in short alanine-based peptides [131–133] but is also found for longer peptides in the unfolded state [134]. It is clear that solvent plays a significant role in stabilizing the PPII conformation, especially in polyalanine peptides, but the fact that the exact mechanism remains a subject of debate [135–137] is testament to the complexity of peptide–water interactions even for relatively simple biomolecular solutes. Longer peptides often show a preference for the formation of \(\alpha\)-helices [5, 138, 139] or extended (\(\beta\)) structures [140–143]. \(\alpha\)-Helices are frequently observed for alanine-rich peptides [144] and are believed to be stabilized in part as a result of helical dipole formation due to the alignment of multiple backbone hydrogen bonds [145, 146] as well as secondary effects due to side chain–solvent interactions [147]. Extended structures are favored entropically, but the more extensive backbone hydrogen bonding interactions with the solvent incur a penalty in solvent entropy due to ordering of solvent molecules [133]. Therefore, especially the formation of \(\beta\)-hairpins often requires additional stabilizing interactions between amino acid side chains [141, 148–150].

1.5.2

**Molecular Properties of Water Near Protein Surfaces**

On a molecular level, the interactions between proteins and the surrounding solvent are equally complex. Water forms a distinct hydration layer around proteins that consists of one or more solvent shells containing water molecules with
physical characteristics that may differ from those of the bulk solvent. The first hydration layer extends about 3 Å from the protein surface and has an average density that is increased by about 10–15% over the value for bulk solvent [106, 151]. Furthermore, the properties of water molecules within the first hydration layer may vary significantly depending on the type of interactions with the protein. A comparatively small fraction of water molecules in the first hydration shell, on the order of 5–30% depending on the property that is being measured [117, 119, 152], exhibit significantly reduced rotational and translational diffusion times typically on the order of tens of picoseconds [108, 117] and may occupy localized interaction sites relative to the solute [101]. The remaining water molecules retain nearly bulk-like properties, with diffusion times reduced by no more than a factor of 2–3 over bulk solvent despite close interaction with the solute [117, 152, 153].

A related measure is the so-called thermal volume of solvent surrounding proteins, which captures the fraction of water molecules that are thermodynamically behaving as if they were part of the protein. The interpretation of experiments in the context of scaled-particle theory has resulted in a thickness of about 0.6 Å for the thermal volume [154], which again translates into a fraction of 20% of the water molecules in the first hydration layer interacting strongly with a given protein. Finally, some studies have determined the minimum level of hydration necessary to achieve native-like protein dynamics and function as about 0.4 g of water per 1 g of protein, which translates into less than a single layer of hydration [155]. The combination of those observations suggests the following somewhat simplified view of protein hydration. A relatively small, but crucial, number of water molecules strongly associate with part of the protein. Those water molecules exhibit reduced dynamics and occupy well-defined hydration sites that may coincide with crystallographic water sites. The resulting protein–water complex interacts with a near bulk-like hydration layer with only slightly increased density and reduced dynamics, which in turn is surrounded by bulk solvent. The modern view of hydration does not assume an ice-like water layer but allows for fluid-like individual water molecules that visit certain preferred interaction sites where they exhibit temporarily retarded dynamics while continuing to exchange with the surrounding bulk on sub-nanosecond time scales. Furthermore, the notion of reduced solvent dynamics near the protein surface needs to be differentiated further based on observations that water molecules in the first hydration shell largely retain bulk-like rotational dynamics [156] and lateral diffusional dynamics [85] while the diffusional component perpendicular to the protein surface is slowed down most dramatically [85].

In addition to the strongly associated but still exchanging water molecules, there is also a small number of water molecules that do in fact have much longer residence times. These water molecules may be completely buried [157] or bound in deep cavities [158] and often play specific roles that may directly complement protein function—for example, in the form of water molecules that are essential partners in an enzyme active site. Especially the latter roles of water as an essential functional part of proteins have resulted in water being considered the “21st amino acid” [1].
1.5.3 Water Molecules at Protein–Ligand and Protein–Protein Interfaces

The presence of strongly associated water molecules on the protein surface has implications for ligand binding and protein complex formation. Often, water molecules are retained at ligand–protein [15, 158, 159] and protein–protein interfaces [11, 160–162]. A pre-organization of water molecules in the absence of a binding partner reduces the entropic penalty during binding, but the presence of ordered water molecules also essentially changes the molecular shape during molecular recognition [9, 163].

1.6 Hydration of Nucleic acids

Nucleic acids, DNA and RNA, differ significantly from proteins in two major ways. (1) Each nucleotide carries a negative charge associated with the backbone phosphate group. As a result, even a short piece of nucleic acid carries a high net charge compared to proteins and peptides, which are often nearly net neutral. (2) DNA molecules do not form globular structures but rather have flexible rod-like shapes. RNA molecules may fold to more compact shapes, but the RNA folding mechanism and resulting role of solvent are very different from protein folding. RNA folding is not driven by hydrophobic collapse—because of the absence of significant hydrophobic groups—but instead relies on counterion condensation and promiscuous base pairing and stacking interactions.

As a result of the high charge of nucleic acids, interactions with solvent molecules (both ions and water) are stronger than for proteins. While proteins essentially have a single hydration layer where water molecules are perturbed to a significant extent from bulk solvent, solvent around nucleic acids may be perturbed in terms of density and dynamics over two or even three layers [85, 99]. Furthermore, ions play a critical role in nucleic acid solvation as they may interact either directly with the nucleic structure [164] or through a shell of coordinated water molecules [80].

Water molecules in the first hydration layer often occupy very well-defined interaction sites [99, 165, 166]. The most famous hallmark of DNA hydration is the so-called “spine of hydration” in the minor groove that is readily visible in crystal structures [7]. Well-defined hydration sites are also present in the major groove, most prominently in the form of pentagonal water clusters [167], and around the phosphate oxygens in the form of so-called “cones of hydration” [168]. In both grooves, the hydration site patterns vary as a function of nucleotide sequence [99].

The interaction with solvent is also closely related to the equilibrium of the polymorphic A and B right-handed nucleic acid helices. While RNA is essentially locked into A-form helices as a result of steric and electrostatic interactions with the extra 2′-hydroxyl at its ribose unit, DNA is found in both A- and B-forms
Hydration of Nucleic acids depending on the environment. It has long been known that A-form DNA is prevalent at low water activities while B-form DNA is preferred at high water activities [6, 169], where “low water activity” means either partial dehydration [6] or the presence of co-solvents such as alcohols and certain ions [170–172]. Consistent with this effect would be a smaller number of strongly interacting water molecules per nucleotide in A-DNA compared to B-DNA according to experimental studies [173]. A different number of water molecules per nucleotide could be rationalized based on the more compact form of A-DNA and increased solvent exposure of the partially hydrophobic ribose ring [174], which would become relatively more favorable under conditions of partial dehydration and/or in the presence of co-solvents. However, simulation studies of A- and B-DNA in aqueous solvent do not necessarily find significantly different numbers of water molecules interacting with A- and B-DNA through hydrogen bonding [175, 176], and A-RNA appears to be hydrated more than B-DNA, presumably due to the presence of the 2’-hydroxyl group [177]. Secondary effects such as increased water bridge formation and increased counterion condensation might therefore be the main driving force for stabilizing A-DNA [176]. The stabilization of A-DNA in the presence of ions would result from additional screening of the repulsive electrostatic interaction between neighboring phosphate groups, since the phosphate–phosphate distance in A-DNA is reduced to 5.9 Å from about 7 Å in B-DNA.

There is also a difference in hydration of A:T versus C:G base pairs [99] that can be correlated to differential propensities to undergo a B- to A-DNA transition. Compared to A:T base pairs, C:G base pairs have an extra hydrogen bond donor in the minor groove from the exocyclic amino group in guanine. Because the total number of water molecules around a base pair is roughly the same, this means that there are more water molecules around A:T base pairs with less favorable, non-hydrogen-bond-forming interactions than around C:G base pairs. This may explain why C:G base pairs more readily assume the A-form upon dehydration and/or addition of co-solvents where polar waters are lost at the expense of hydrophobic waters interacting with the more exposed ribose ring [99].

As in protein hydration, some fraction of the water molecules interacting with nucleic acids exhibit reduced diffusion rates, and, as in the case of protein hydration, reduced mobility and occupancy of well-defined hydration sites are not necessarily correlated [178, 179]. Nevertheless, it has become clear that water molecules located in the grooves have relatively long residence times between tens of picoseconds up to nanoseconds [180–182]. Such relaxation times are about one to two orders of magnitude longer than for water molecules interacting with proteins. The combination of more extensively ordered water molecules and longer residence times around nucleic acids has prompted the notion that “water is an integral part of nucleic acid structures” [183] but also bears implications for the interaction of nucleic acids with proteins and other ligands. As a consequence, proteins scanning nucleic acids, such as transcription factors, may be guided by interactions with the hydration layer exhibiting sequence-specific patterns rather than direct interactions with the nucleic acid.
1.7 Non-Aqueous Solvation

The interaction of biomolecules with non-aqueous solvent is relevant in the biological context (for example, protective solvents that allow cells to survive under freezing conditions or membrane environments) or in biotechnology applications (for example, denaturants such as urea and guanidium hydrochloride) [19]. Many studies have investigated solvation in non-aqueous solvent or co-solvent, but the detailed mechanisms of how certain non-aqueous solvents interact with biomolecules is not always fully understood. One aspect that is often investigated for solvent mixtures is the preferential solvation at the biomolecule–solvent interface. Depending on the type of co-solvent, water may be displaced partially or completely, or water may continue to interact directly with the biomolecule but exhibit altered properties due to secondary interactions with the co-solvent. The specific effects of different types of non-aqueous solvents are reviewed briefly in the following:

1.7.1 Alcohols

Alcohol (methanol, ethanol, 1,1,1,3,3,3-hexafluoropropan-2-ol (HFIP), and 2,2,2-trifluoroethanol (TFE)) mixtures lower the dielectric constant of the solvent and thereby provide reduced screening of electrostatic interactions, which leads to an enhanced propensity to form intramolecular hydrogen bonds and salt bridges within proteins but also reduces the penalty for solvent exposure of hydrophobic side chains. As a result, α-helix-forming peptides display enhanced α-helix propensities in mixtures containing methanol or ethanol [184–188] and secondary structure formation is generally enhanced in the presence of TFE and HFIP [189]. TFE and HFIP overall interact preferentially with peptides [190, 191] while ethanol preferentially interacts with peptide backbone atoms [192], thereby enhancing their effect in creating an environment with reduced polarity.

1.7.2 Urea

Urea is commonly used as a denaturant for proteins. Urea has an amphiphilic character that allows it to interact with both polar and hydrophobic groups. Its effect on protein solvation is actually fairly subtle. Based on simulation studies, it appears that urea affects the stability of proteins both directly and indirectly. The direct effect may involve preferential interaction with hydrophobic residues, thereby destabilizing the hydrophobic core, or interaction with polar constituents and in particular the peptide backbone. The indirect effect alters water structure in such a way that water becomes free to compete more successfully for backbone hydrogen bonds and other intra-protein interactions [193–197].
1.7.3 Glycerol

Glycerol, along with other similar compounds such as trehalose, is a natural compound that has a stabilizing effect on protein structures to allow cells to withstand freezing temperatures or dehydration. Glycerol increases the viscosity of the solvent and thereby dampens the dynamic fluctuations of a given protein [198, 199]. As a result, the protein may be trapped in a native-like conformational state in a glass-like environment that can be maintained at low temperatures. Glycerol also substitutes surface-bound water molecules and thereby reduces the consequences of freezing. A secondary effect of the presence of glycerol appears to be entrapment of water molecules in the first solvation shell due to the increased viscosity that would result in prolonged water residence times [197, 199].

1.8 Summary

This chapter has surveyed the current knowledge about the interplay between solvent and biomolecular structure and dynamics. As an introduction to the remainder of this book, it introduces the theoretical framework, simulation methodologies, and experimental techniques that are commonly applied in the study of biomolecular solvation. With the development of solvent models in mind, the structural and dynamic characteristics of solvent around proteins and nucleic acids are sketched out. In both cases, there are well-defined hydration sites that are preferentially occupied by solvent. Water molecules at those sites may have significantly reduced diffusion rates, but, contrary to the seemingly static pictures portrayed by crystallographic water sites, actual water molecules often continue to exchange with the surrounding solvent and display only moderately reduced rotational diffusion rates. The observation that many water molecules in the first hydration shell or proteins retain near-bulk-like dynamic properties justifies the development of mean-field solvent models that are based on a separation into explicit solute and surrounding bulk-like implicit solvent. However, such an approach essentially ignores the relatively small but significant fraction of water molecules in the first hydration shell that possess significantly perturbed properties and hence do not behave like bulk solvent. In the case of nucleic acids, water molecules are even more ordered and a larger fraction of water molecules is dynamically perturbed. It appears that the most realistic solvent models short of a fully explicit solvent representation would need to take into account the differentiated character of water molecules in the first hydration shell.
References

References


171 Cheatham, T.E. and Kollman, P.A. (1997) Insight into the stabilization of
A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d(ACCCGCGGGT)₂ in the presence of hexaamminecobalt(III). *Structure*, 15, 1297–1311.


