

## Is it the shape of the cavity, or the shape of the water in the cavity?

Phillip W. Snyder<sup>1,b</sup>, Matthew R. Lockett<sup>1,b</sup>, Demetri T. Moustakas<sup>2</sup>,  
and George M. Whitesides<sup>1,3,a</sup>

<sup>1</sup> Department of Chemistry and Chemical Biology, Harvard University 12 Oxford Street, Cambridge, MA 02138, USA

<sup>2</sup> Infection Innovative Medicines Unit, Chemistry Department AstraZeneca Pharmaceuticals, R&D Boston 35 Gatehouse Drive, Waltham, MA 02451, USA

<sup>3</sup> Wyss Institute of Biologically Inspired Engineering, Harvard University 60 Oxford Street, Cambridge, MA 02138, USA

Received 15 September 2012 / Received in final form 30 March 2013

Published online 2 May 2013

**Abstract.** Historical interpretations of the thermodynamics characterizing biomolecular recognition have marginalized the role of water. An important (even, perhaps, dominant) contribution to molecular recognition in water comes from the “hydrophobic effect,” in which non-polar portions of a ligand interact preferentially with non-polar regions of a protein. Water surrounds the ligand, and water fills the binding pocket of the protein: when the protein-ligand complex forms, and hydrophobic surfaces of the binding pocket and the ligand approach one another, the molecules (and hydrogen-bonded networks of molecules) of water associated with both surfaces rearrange and, in part, entirely escape into the bulk solution. It is now clear that neither of the two most commonly cited rationalizations for the hydrophobic effect—an entropy-dominated hydrophobic effect, in which ordered waters at the surface of the ligand, and water at the surface of the protein, are released to the bulk upon binding, and a “lock-and-key” model, in which the surface of a ligand interacts directly with a surface of a protein having a complementary shape—can account for water-mediated interactions between the ligand and the protein, and neither is sufficient to account for the experimental observation of both entropy- and enthalpy-dominated hydrophobic effects. What *is* now clear is that there is no *single* hydrophobic effect, with a universally applicable, common, thermodynamic description: different processes (i.e., partitioning between phases of different hydrophobicity, aggregation in water, and binding) with different thermodynamics, depend on the molecular-level details of the structures of the molecules involved, and of the aggregates that form. A “water-centric” description of the hydrophobic effect in biomolecular recognition focuses on the structures of water surrounding

<sup>a</sup> e-mail: [gwhitesides@gmwhgroup.harvard.edu](mailto:gwhitesides@gmwhgroup.harvard.edu)

<sup>b</sup> These authors contributed equally to this work.

the ligand, and of water filling the binding pocket of the protein, both before and after binding. This view attributes the hydrophobic effect to changes in the free energy of the networks of hydrogen bonds that are formed, broken, or re-arranged when two hydrophobic surfaces approach (but do not necessarily contact) one another. The details of the molecular topography (and the polar character) of the molecular surfaces play an important role in determining the structure of these networks of hydrogen-bonded waters, and in the thermodynamic description of the hydrophobic effect(s). Theorists have led the formulation of this “water-centric view”, although experiments are now supplying support for it. It poses complex problems for would-be “designers” of protein-ligand interactions, and for so-called “rational drug design”.

## 1 The problem

### A. Hydrophobic effect or hydrophobic effects?

Water is the solvent in which “life” occurs. It is intimately involved—although often implicitly ignored—in many of the molecular processes that, together, make life what it is. A broad class of these processes (including, for example, protein-ligand binding, enzyme-catalyzed recognition and signaling, the formation of internal structure in biological macromolecules, and the aggregation of lipids and proteins into cell membranes) is that sheltered under the umbrella description of “biomolecular recognition”, and within this class, probably the most important single type of intermolecular interaction is the hydrophobic effect.

The “hydrophobic effect” (or more precisely the “hydrophobic effects” or the “varieties of the hydrophobic effect”) is a term describing the tendency of non-polar molecules or molecular surfaces to aggregate in an aqueous solution (or, again, more exactly, “to be expelled from water into an aggregate”). From the earliest discussions of hydrophobicity, an emphasis has been on the interaction of non-polar molecular surfaces with water, on the unique structure of liquid water, and on the differences in structure of water in the bulk and water close to non-polar interfaces [1–4]. The first experiments that examined the hydrophobic effect made the simplifying assumption that there is a *single* effect with a common structural, mechanistic, and thermodynamic description. This assumption is now evolving into an expanded and more complicated view, in which the “hydrophobic effect” appears to have different structural and thermodynamic origins in different molecular contexts: that is, a hydrophobic effect involving, for example, a convex non-polar surface may have a different thermodynamic basis than one involving a concave or planar surface.

Understanding hydrophobic effects (plural) is centrally important to understanding (i.e., predicting the strength and specificity of) biomolecular recognition—the noncovalent association of molecules in biological systems. Past explanations of molecular recognition, based on semi-quantitative experimental physical-organic studies in semi-polar organic solvents such as chloroform and methylene chloride [5–8], do not account for a (or perhaps, the) key component of all biomolecular recognition events: water. Understanding the restructuring of networks of hydrogen bonds among molecules of water is very difficult experimentally and requires high levels of theory (and, in fact, simulations have led experimentation in considering new ways of resolving the deficiencies of the simplest rationalizations of the hydrophobic effect).

There are important differences between historical views of the hydrophobic effect and the current, still evolving, view, both in its origin, and in its role in molecular

recognition. In brief, the older view was that hydrophobic effects reflected the release of entropically-unfavorable, “structured” water immediately adjacent to non-polar molecular surfaces (e.g., of protein and ligand), and required the contact of these non-polar surfaces. The emerging hypothesis takes a more complicated view of the role of water. It postulates that water near (but not necessarily in contact with) non-polar surfaces is, indeed, less favorable in free energy than water in bulk, but that a range of factors reflecting the complicated networks of hydrogen bonds that make water the remarkable fluid that it is contribute in several, not necessarily intuitively obvious, ways: these contributions might include the enthalpies of hydrogen bonds, the entropies of conformational and translational restrictions, the density filling (for declivities) or surrounding (for convexities) molecular surfaces, the interactions of water with proximate surfaces, entropic (osmotic) effects involving buffer ions, and other terms. Bringing two non-polar surfaces close together (but not necessarily into van der Waals contact) releases some of the near-surface waters into the bulk of the surrounding fluid, and perturbs the structure of the rest. The net change in the free energy is the basis of the “hydrophobic effect”. In this view, hydrophobic effects are due primarily to differences in the free energy of water in the vicinity of non-polar surfaces and water in bulk, and these free energies depend strongly on the details of the topography of these surfaces.

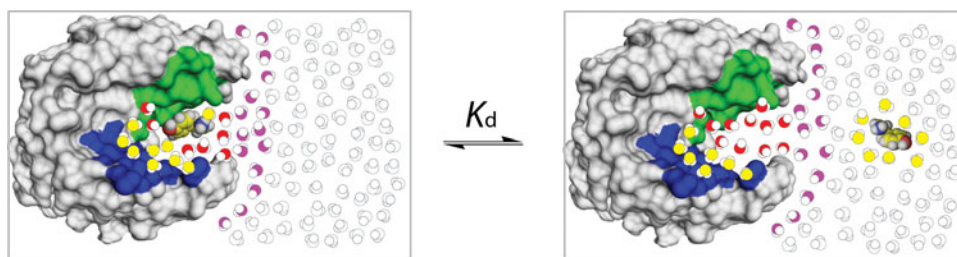
We suggest three core differences between “protein-ligand centric” and “water-centric” views of hydrophobic effects. These differences are in: i) *Thermodynamics*. The fundamental origin of hydrophobic effects, in the water-centric view, is that they do not represent a favorable interaction between non-polar molecular surfaces, but rather, an unfavorable free energy of water close to those surfaces. When hydrophobic aggregation occurs, it is due to the release of free-energetically unfavorable water into the lower free-energy bulk. This rationalization is related to that proposed by Kauzmann and Tanford, but differs in important thermodynamic details: in particular, near-surface water may be unfavorable in free energy for either enthalpic or entropic reasons, rather than being entirely entropic in origin [2,4]. ii) *The relative importance of two factors: complementarity in shape of protein and ligand, and the release of structured, free energetically unfavorable, near-surface water*. The deeply rooted idea that a ligand will fit into the binding pocket of a protein when the two are complementary in shape (“lock and key”) may have some truth. It is not, however, that the surface of the ligand interacts directly with the surface of the protein, but rather that a fit that is approximately complimentary in shape may be the one that releases the largest volume of free-energetically unfavorable water from the cavity of the protein and from the surface of the ligand. In this view, the hydrophobic effect in biomolecular recognition is influenced more by the shape (and free energy) of the water (more exactly, of the networks of water molecules) released from and rearranged in the binding pocket of the protein, and from the surface of the ligand, than by the interactions between the protein and the ligand. iii) *The importance of solvent properties in molecular recognition in organic solvents and water*. Extensive and excellent studies of molecular recognition in organic solvents (from crown ethers and metal ions to networks of hydrogen bonded molecules) have had, as one justification, the assertion that they teach important lessons for molecular recognition in water. In a water-centric view of molecular recognition, results obtained in organic solvents do not predict—and, in fact, are probably largely irrelevant to—molecular-level interactions in water. The hydrophobic effect, from the point of view of the release of unfavorable water molecules from a binding pocket, has no analogy with interactions occurring in organic solvents, because there are no networks of hydrogen bonds or other strong directional intermolecular interactions in  $\text{CH}_2\text{Cl}_2$  or  $\text{CHCl}_3$  to give these and other organic solvents the structures that uniquely characterize liquid water.

## B. Hydrophobic effects in biomolecular recognition

Molecular recognition, and especially the selective association of proteins with “ligands” (e.g., other proteins, substrates, transition states, drugs, etc.) is one of the most important molecular processes (and perhaps *the* most important) in life. “Hydrophobic effects” are central to molecular recognition, and to countless other processes in biology—the folding of proteins, the formation and structure of base-paired nucleic acids, the formation of lipid bilayers, the recognition of small-molecule ligands by proteins, and many others. Despite more than 50 years of research into the role of hydrophobic interactions in biology, and specifically in biomolecular recognition, we are still not able to predict the structure of a ligand that will bind tightly to a protein (even one whose active site is well-defined structurally), other than by empirical structural analogy. Instead, we observe that: i) hydrocarbons are poorly soluble in water, ii) proteins—alone or complexed with ligands—have large areas of apposed non-polar surface that are shielded from contact with water (i.e., buried), and iii) the interactions of low molecular weight ligands, substrates, or drugs with the active sites of proteins tend to involve the interaction of non-polar surfaces. These three observations have been unified under the umbrella of a single, common type of non-covalent interaction—called the “hydrophobic effect”.

In this view, the hydrophobic effect provides, perhaps, 75% of the free energy of most binding or association events in biomolecular recognition. This qualitative estimate derives from two observations: i) the distribution of non-polar and polar regions that accounts for the surface area of most ligands, and for the active sites of proteins, is approximately 75% non-polar and 25% polar; ii) the magnitude of the free energy of binding in molecular recognition, in water, is approximately linearly proportional to the amount of solvent-exposed surface area that is removed from contact with water upon binding [9]. This qualitative approximation does not provide an accurate prediction of the free energies of binding in any case, and it now seems increasingly likely that there are a number of different interactions that contribute to the free energy of hydrophobic effects, probably with different mechanistic and structural origins. This family of interactions, however, shares a common foundation in that the structure of networks of water molecules—especially of those molecules of water that are near surfaces—contributes to free energies; the components of this contribution (i.e., their enthalpy and entropy) depend on the structure of the binding pocket of the protein and the ligand.

The free energy of a hydrophobic interaction results from a difference between the free energy of bulk water, and the free energy of water near non-polar surfaces; different hydrophobic effects (or, at least, hydrophobic effects that differ thermodynamically) seem to be responsible for protein-ligand binding, and for the low solubility of hydrocarbons in water. Both theory and experiment are beginning to support the hypothesis that the topography of the binding pocket plays a crucial role in determining the free energy of protein-ligand binding—entirely aside from specific interactions of the surfaces of proteins and ligands—because this topography determines, or influences, the free energy of the network of hydrogen bonds between water molecules within the pocket, and thus the change in free energy when this unfavorably structured water is replaced by a ligand and escapes into the energetically more favorable bulk solution. In biomolecular recognition, in particular, the hydrophobic effect may be the combination of (at least) two effects: i) the network of water molecules in the binding pocket of a protein may have a structure that is less favorable in free energy than bulk water (Figure 1); ii) water in contact with small hydrophobic molecules may be less favorable in free energy than water in the bulk (but for a different reason, or at least with a different distribution of enthalpies and entropies).



**Fig. 1. Schematic Representation of the Role of Water in Biomolecular Recognition.** The protein appears as a surface representation colored by chemical character (green represents hydrophobic surface and blue represents polar surface). The ligand appears with spheres representing the van der Waals surfaces of its atoms. Molecules of water appear as sphere representations and are shaded by free energy: white represents molecules of water that have free energies near that of bulk water, yellow represents molecules of water that are less favorable in entropy than bulk water, and red represents molecules of water that are less favorable in enthalpy than bulk water. Water molecules close to polar groups in the active site, or on the extended surface of proteins, may be more stable than those in bulk water, if strongly stabilized by hydrogen bonds or other electrostatic interactions. This paper does not deal with these waters.

In the classical, “protein-centric” view, water near non-polar surfaces is unfavorable in free energy because it is ordered, and thus entropically unfavorable. In the “water-centric” view, water near non-polar surfaces, or in cavities, can be unfavorable in free energy (or indeed, favorable, although free-energetically favorable, near-surface water has not been much explored) for any combination of enthalpy and entropy, and this excess unfavorable free energy of water (relative to bulk water) depends on the topography and molecular-level details of the exposed surfaces of both the protein and the ligand.

### C. Solvent, topography, and the thermodynamics of binding

To emphasize the role of solvent and topography in protein-ligand binding, and to make a conceptual point, we write the general equilibrium expression with explicit molecules of water ( $\text{H}_2\text{O}$ ) and the symbol “ $\mathbb{C}$ ” to indicate water adjacent to a concave surface of a protein (P) and the symbols “ $\mathbb{D}$ ” and “ $\mathbb{L}$ ” to indicate water adjacent to convex surfaces of a protein or a ligand (L) (Eqs. (1–2)). Although these expressions are impractically cumbersome for everyday use, they emphasize how much is omitted from conventional formulations of the dissociation constant.



$$K_d = [\mathbb{P}\mathbb{C}\mathbb{H}_2\text{O}][\mathbb{L}\mathbb{D}\mathbb{H}_2\text{O}] / [\mathbb{P}^*\mathbb{L}\mathbb{H}_2\text{O}]. \quad (2)$$

It is possible that non-polar (hydrophobic) surfaces of common topography are similar, but the extent of this similarity is neither proved nor defined. The surface of a protein is, of course, a continuum of topography composed of concave and convex regions of surface that can be located anywhere on a continuous space between “hydrophobic” and “hydrophilic”, and the presence of charged and polar groups in the interacting and proximal surfaces may have a profound effect on the free energy of proximate water [10].

The free energy of protein-ligand association ( $\Delta G_{\text{bind}}^\circ$ , Eq. (3)) is estimated by measuring the dissociation constant ( $K_d$ , Eq. (4)).

$$\Delta G_{\text{bind}}^\circ = -RT \ln K_a \quad (3)$$

$$K_a = K_d^{-1}. \quad (4)$$

The free energy of binding is composed of enthalpic ( $\Delta H^\circ$ ) and entropic ( $-T\Delta S^\circ$ ) terms (Eq. (5)).

$$\Delta G_{\text{bind}}^\circ = \Delta H_{\text{bind}}^\circ - T\Delta S_{\text{bind}}^\circ. \quad (5)$$

*Decomposing Experimentally Measured Thermodynamic Parameters.* The thermodynamic parameters describing binding (i.e., those measured or estimated experimentally:  $J^\circ = G^\circ$ ,  $H^\circ$ ,  $S^\circ$ , or  $C_p^\circ$ , the heat capacity) can sometimes be decomposed into contributions from differences between bound and unbound states in their hydration, functional-group-specific interactions, conformations, and translational and rotational freedom.

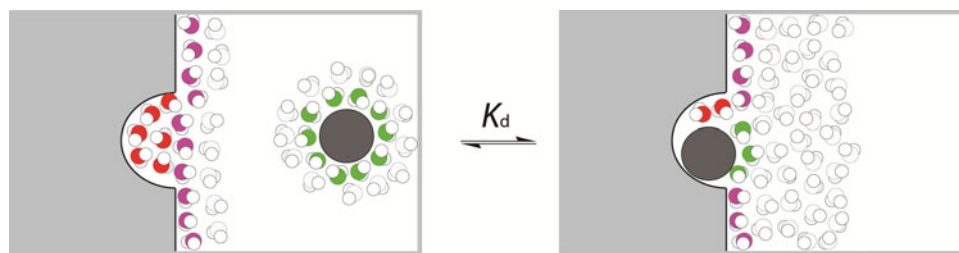
$$\Delta J_{\text{bind}}^\circ = \Delta J_{\text{hydration}}^\circ + \Delta J_{\text{interaction}}^\circ + \Delta J_{\text{conformation}}^\circ + \Delta J_{\text{trans,rot}}^\circ. \quad (6)$$

The magnitude of each of the terms on the right-hand side of Eq. (6) depends on the molecular details (i.e., the structures of the protein and the ligand, and the structure of water close to the protein and ligand) of each system. Predicting the values of these terms based on the available structural information (i.e., from crystallography or magnetic resonance spectroscopy) remains an exceptional challenge, and one that has not yet been solved, after five decades of thoughtful research [11–15].

As we discuss in some detail below, to understand hydrophobic effects—in other words, to determine the value of  $\Delta J_{\text{hydration}}^\circ$ —in the context of a protein-ligand interaction, it often makes sense to work with a model system in which changes in the structure of the ligand (and, ideally, the protein) have little effect on the values of  $\Delta J_{\text{interaction}}^\circ$ ,  $\Delta J_{\text{conformation}}^\circ$ , and  $\Delta J_{\text{trans,rot}}^\circ$ . In our own research, we have worked largely with carbonic anhydrase—an exceptionally rigid protein—and sketch some of the conclusions from this work later. Proteins can, however, also be mobile, plastic, and even completely disordered; proteins that are disordered in the absence of ligand [16–18], and develop tertiary structure only upon association with a ligand, provide particularly interesting and challenging systems to understand in the context of the hydrophobic effect.

*Solvent.* A key point to address is that differences between the free energy of water in the bulk, and the free energy of water near the hydrophobic surface of a small molecule or a protein, (reflected in the term,  $\Delta G_{\text{hydration}}^\circ$ ) depend on the system. The water-centric view of the free energies of molecules of water around the binding site of a protein, and around the ligand to which it binds, is that values of  $\Delta H_{\text{hydration}}^\circ$  and  $-T\Delta S_{\text{hydration}}^\circ$  depend on the details of the molecular structure of the protein and the ligand. The structures of the networks of water molecules surrounding these surfaces are different for different molecules, molecular topographies, and compositions (Figure 2). Based on limited theoretical and experimental evidence—which we discuss for several specific systems in detail below—two contributions seem to be important: i) the release of molecules of water from the surface of small (radii less than  $\sim 1$  nm), convex, hydrophobic molecules or groups to bulk water is entropically favorable (generally) at room temperature; ii) the release of molecules of water from concave, hydrophobic surfaces (like those often found in the active sites of proteins) to bulk water is, at least in some cases, enthalpically favorable.





**Fig. 2. Surface Topography and Xenophobic Water.** The concave surface represents a generic binding pocket of a protein and the gray circle represents a generic ligand. Water molecules are color coded to indicate their free energy, relative to that of bulk water (white). Waters with red oxygen atoms are enthalpically unstable relative to the bulk, and waters with yellow oxygen atoms are entropically unstable relative to bulk. We call these “unstable” waters “xenophobic”, in the sense (to be anthropomorphic) that they are “unhappy” (in free energy) to be close to strangers (i.e., non-water molecules). Note that this schematic picture does not require an exact (“lock-and-key”) fit between the pocket and the ligand, or even any interaction between the two, to result in a favorable change in free energy of association of protein and ligand.

## D. Entropy-enthalpy compensation

One puzzling phenomenon—which may also reflect changes in the networks of water molecules within a binding pocket—seems to limit the strength of association that can be achieved through the putative *design* of tight-binding, low-molecular-weight ligands for proteins: that is, a so-called “entropy-enthalpy compensation”. Despite the dismissal (on the grounds of correlated errors) of linear correlations between changes in  $\Delta H_{\text{bind}}^\circ$  and  $-T\Delta S_{\text{bind}}^\circ$  that have been claimed empirically for protein-ligand association (and numerous other chemical processes) [19–24], it is clear that, often, changes in the structure of a ligand leads to large changes in enthalpy and entropy of binding, but that these changes compensate in a way that results in small changes in  $\Delta G_{\text{bind}}^\circ$  [14, 25–29].

There is, however, no unequivocal, molecular-level explanation for entropy-enthalpy compensation, and its origin—even at a conceptual level—remains a controversial subject [30], despite the qualitative rationalizations for this phenomenon advanced by Dunitz, Williams, and others [25, 31, 32]. Although these suggestions “make intuitive sense,” [32, 33] at some level, there is so much in the hydrophobic effect that is non-intuitive (or perhaps intuitive at some level, but still very complicated), that we are currently suspicious of simple, intuition-based rationalizations of the even more difficult subject of entropy-enthalpy compensation.

## 2 The power of metaphor (for bad or good)

### A. “Ice-like water”

One of the oldest—and now most pervasive—rationales for a single hydrophobic effect is the formation of “structured” or “iceberg-like” water near non-polar solutes, as proposed by Frank and elaborated by Kauzmann, Tanford, and others [1–4]. This model rationalizes the transfer of small, simple hydrophobic molecules such as methane or ethane from a non-polar phase (i.e., the gas phase or a non-polar liquid phase) to an aqueous phase: the free energies of these transfers, at room temperature, are unfavorable and seem to be dominated by a large, unfavorable entropic term. The “iceberg”

model postulates the unfavorable entropy of transfer results from a network of structured waters that form around the hydrophobic molecules. Experimental programs employing neutron scattering, which is exquisitely sensitive to water structure, have repeatedly probed the structure of aqueous solutions containing non-polar molecules, and have not provided support for the notion that water in contact with these hydrophobic solutes is more “ordered” than water in the bulk [34–36].

## B. “Lock and key”

The “lock and key” metaphor was originally proposed by Emil Fischer to explain how an enzyme recognizes a substrate, catalyzes a covalent reaction, and releases its product. It has now, with the passage of time, and the lack of questioning, achieved the status of religious revelation. “Lock and key” has become an engrained principle in structure-guided ligand design, although it is increasingly questioned by sophisticated analysis [37].

## C. “Xenophobic water”

The network of hydrogen bonds between water molecules in contact with a non-polar surface is more “constrained” than the networks of water molecules in the bulk. This constraint, which arises from a “xenophobic aversion” of the waters in the proximity of non-polar surfaces, is both thermodynamic and kinetic in nature, and involves (at least) the shell of water in direct contact with the surface or solute; in the case of a protein, three to four shells of water surrounding the protein can be constrained [38,39]. Measurements of dielectric relaxation of water near the surfaces of proteins also conclude that these molecules of water have longer pre-exchange lifetimes [40] than the hydrogen bonds formed in bulk water.

Water molecules at the surface of a protein facilitate protein folding [41], and stabilize the structure of a native protein [42] as well as the complex formed between a ligand and a protein [27,43,44]. Grossman *et al.* found that the lifetimes of the hydrogen bonds between molecules of water in the active site of human membrane type-1 metalloproteinase, a zinc-containing metalloprotease, and those surrounding its peptide substrate increase (e.g., exchange within the hydrogen bond network slows) upon binding of the substrate [43]. In this case, the binding of the ligand seems to be coupled to the constrained motion of water molecules in the active site.

Studies of fluorescent probes attached to the surface of a protein also show that the pre-exchange lifetimes of hydrogen bonds among the molecules of water in the first few hydration layers of the protein are much longer than those in bulk water [38,45]. The fluorescence lifetime of the single tryptophan residue on the surface of subtilisin Carlsberg is significantly longer than the fluorescence lifetime of a free tryptophan in bulk water (38 psec vs. 1.1 psec), and is attributed the decreased frequencies of motion within the constrained network of hydrogen bonds at the surface of the protein.

Reduced rates of reaction are also observed in the cavity of cyclodextrins [46,47], as well as other molecular capsules [48]; the ability of molecules of water to reorient their dipole moments, and adopt a conformation that stabilizes a reaction intermediate, is two to four orders of magnitude slower within the cavity than in the bulk. A notable example is the deprotonation of 1-naphthol inside the cavity of a cyclodextrin; this elementary reaction is approximately 25 times slower in the cavity than in bulk water [49,50].



## D. “Water networks”

Both experimental information and interpretation of the thermodynamics and kinetics of xenophobic water at the surface of proteins or ligands is evolving. There also remain significant gaps in our understanding of the thermodynamics of networks of water in the bulk, and there is very little information on the structure and thermodynamics of water in buffer.

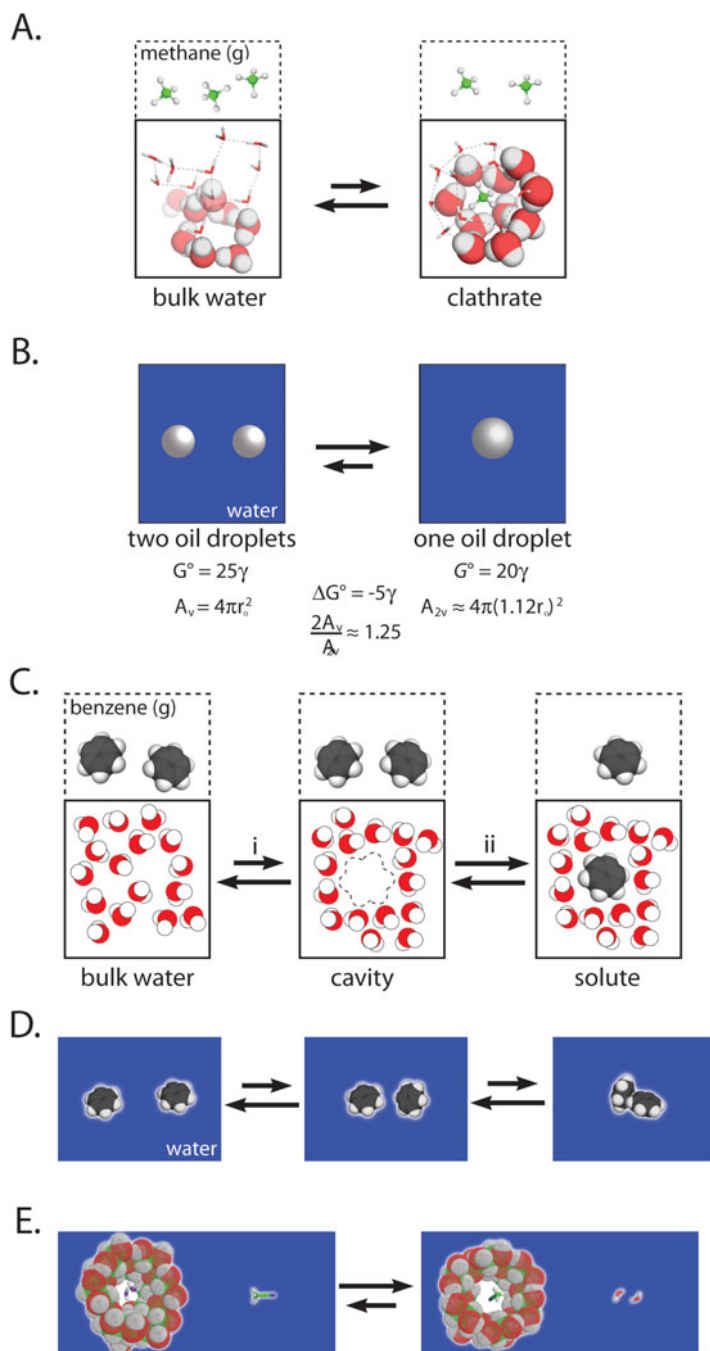
Molecules of water form hydrogen bonds that are directional [51], and the strength of a hydrogen bond between two molecules of water depends on the number of non-covalent interactions (i.e., other hydrogen bonds) in which each molecule of water participates [52]. Theoretical simulations suggest that the average hydrogen bond between two waters in a dimer is weaker than the average hydrogen bond between two waters in a trimer. The distribution of charge density of an individual molecule of water changes upon formation of a dimer, and this change results in the increased (cooperative) strength of the second hydrogen bond [53,54]. Cooperative interactions among molecules of water are observed in several systems in which hydrogen bonding is important, and include the intermolecular bonding of molecules of water, formamide, and urea [55,56], and water-mediated interactions between mono- and disaccharides [57,58].

## 3 Summary of models of the hydrophobic effect

### A. Iceberg model (Frank, Kauzmann, Tanford, *et al.*)

During the early 1940s, Frank and Evans analyzed the thermodynamics of mixing of liquids, and observed that water is anomalous among solvents: the entropy of mixing of water and non-polar liquids is unfavorable [1]. This unfavorable entropic term dominated the free energy of mixing, and was interpreted to mean that water, in aqueous solutions containing hydrocarbons, was more “ordered” than water alone. This interpretation was consistent, seemingly, with increases in the observed heat capacity of mixing: increased “order” in the water near non-polar solutes is intuitively consistent with increased heat capacity. To illustrate this ordering, in their seminal paper in 1945, Frank and Evans proposed the “iceberg” model to rationalize this experimental data. More than a decade later, Kauzmann drew on this conceptual iceberg model to rationalize the favorable entropy of the folding of proteins [2]. In this approximation, he suggested that the driving force for protein folding was the entropically favorable desolvation of non-polar groups, which are most often buried in the interior of the native structures of proteins (Figure 3A). Early support for Frank’s iceberg model of hydrophobic hydration appeared to come from the crystal structures of the gas hydrates [59]. These co-crystals of water and small organic molecules (e.g., methane, tetramethylammonium salts, etc.) contain molecules of water that are tetrahedrally coordinated to one another through a network of hydrogen bonds. The organic molecules fit into the intermolecular structures between the waters. Water around the organic molecules form clathrate structures, often with regular pentagonal faces, and, importantly, retained four hydrogen bonds per molecule of water—the same structure observed in ice.

There are a number of experimental programs—of particular note are the neutron diffraction studies of Soper and Finney—that characterize the structure of water near non-polar solutes [34,36]. Interestingly, these experiments provide no support for an “ice-like” water structure near non-polar solutes in aqueous solution. Despite decades of sophisticated experimental and theoretical studies of the structure of water near non-polar solutes, no rigorously complete model rationalizes the thermodynamics of solvation of small, hydrophobic molecules.



**Fig. 3. Models of Hydrophobic Effects.** (A) The clathrate or “iceberg” model of hydrophobic effects speculates that bulk water (left) is less ordered (more entropically favorable) than water near non-polar solutes (right; in this example a molecule of methane). Echoing the classical view, the structure of water that solvates the methane molecule is modeled on the crystal structure of methane hydrate. (B) The surface tension model suggests that the high surface tension of water minimizes the surface area of the interface of water with a non-polar substance; in this example two droplets of oil coalesce to minimize the

**Fig. 3.** (Continued). interacting surface area between oil and water. (C) The cavity formation model describes the solvation of non-polar molecules with two steps: the accumulation of a void volume that is large enough to accommodate the solute (indicated with the dashed outline), followed by the insertion of the solute into the cavity. Void volumes (indicated as white space) are distributed randomly throughout the bulk solvent, and the initial step of accumulating void volume is entropically unfavorable. (D) The van der Waals model points out that the noncovalent interactions between water and hydrocarbons—particularly aromatic groups like benzene (pictured)—should be less favorable in enthalpy than those among hydrocarbons. The model provides one explanation for enthalpically-dominated hydrophobic effects, and is rationalized by differences in the polarizabilities of water and hydrocarbons. Other states, such as solvent-separated pairs (indicated in the central pane) are also thermally accessible in the Boltzmann sense. (E) The confined water model describes the importance of the structure of water *in* binding cavities in influencing the free energy of binding, and incorporates theoretical predictions that water near concave surfaces is less favorable in free energy than bulk water. The hydrophobic effect, in this model, is determined by the drop of water (depicted here as two molecules of water in the cavity of  $\alpha$ -cyclodextrin) that hydrates the binding pocket: as the ligand (a molecule of acetonitrile) fills the binding pocket, it displaces molecules of water that are constrained in the cavity into bulk water, and is favorable in free energy.

## B. Surface tension model (Hildebrand)

The large cohesive energy density of water gives it a high surface tension. The surface tension of water forces droplets of oil, when suspended in water, to minimize the surface area of contact between oil and water (Figure 3B). This model allows one to calculate the free energy (in units of  $\text{cal mol}^{-1} \text{\AA}^{-2}$ ) of forming a *macroscopic* interface between water and oil. The extrapolation of the free energy of coalescence at the macroscopic scale leads to an overestimation of the free energy for hydrating a small hydrophobic molecule (i.e., methane, ethane, etc.).

This discrepancy between the macroscopic and molecular levels has been the subject a contentious discussion in the literature over the last four decades [3,60,61], and was the first indication that the mechanism of the hydrophobic effect differs depending on the size—and more importantly, from a water-centric point of view, the shape—of the hydrophobic solute.

## C. Cavity formation or “void volume” model (Stillinger, Chandler, Lee, Pratt, Hummer, *et al.*)

In contrast to the iceberg model, Stillinger applied scaled-particle theory to describe dissolution of non-polar molecules in water [62]. This idea, and its conceptual progeny, explain the entropically unfavorable solvation of small hydrocarbons by the accumulation of voids in bulk water to form “void volumes” that are large enough to accommodate the solute (Figure 3C) [62–65]. These models have been criticized because they do not predict changes in heat capacity that result from the solvation of hydrocarbons in water, although the most recent work by Chandler seems to address this limitation [66].

## D. Van der Waals Model (Saenger, Diederich, Homans, *et al.*)

Entropy-dominated models for the hydrophobic effect do not resolve an important disagreement between mechanistic theories and experimental fact: the origin of the hydrophobic effect(s) that dominates the free energy of most protein-ligand interactions is enthalpically favorable, whereas the origin of the hydrophobic effect in the

iceberg and the cavity-formation models is entropic [27,67]. Early rationalizations for this incompatibility suggested that noncovalent interactions between proteins and ligands were more favorable in enthalpy than interactions between water and either the surface of the protein or the face of the ligand (Figure 3D) [68]. Jencks, and several others, began to discuss, as far back as the 1970s, hydrophobic interactions that were driven by enthalpy, rather than entropy [69]. These so-called “non-classical hydrophobic effects” were observed for the denaturation of bovine serum albumin and ovalbumin [70]. More recent discussions—in particular by Diederich—have focused on a subset of so-called non-classical hydrophobic effects that are ostensibly important for the binding of substituted aromatics to cyclophanes in water [71].

## E. Mercedes Benz (Dill)

The Mercedes Benz model simplifies the structure of water by treating each molecule of water as a two-dimensional disk with three prongs (i.e., each molecule of water is the symbol of Mercedes Benz). These disks interact with one another through a Lennard-Jones interactions and the formation of hydrogen bonds; each prong represents a site at which a potential hydrogen bond can form. The formation of a hydrogen bond is dependent on the distance and the angle between two disks, and occurs when the prong of one disk overlaps with the prong of a second, and separate, disk [72,73].

Dill and colleagues suggest that many of the macroscopic properties of water are not due to its three-dimensional structure, nor to the detail of its atomic structure, but are, in fact, a reflection of the angles of the hydrogen bonds that form between the molecules of water. The decrease in dimensionality results in a model system that is less difficult to address computationally than molecular dynamics simulations, and predicts some of the properties of bulk water [74].

Model studies, by Dill and coworkers, of the hydrophobic effect show that the transfer of small, non-polar solute molecules into water [72,73] matches experimental evidence (i.e., large positive heat capacities and unfavorable entropy terms at room temperature). Increases in the radius of the non-polar solute show that the solvation process becomes enthalpically unfavorable as the hydrogen bonds of the molecules of water directly adjacent to the solute become geometrically strained, and this strain leads to the breaking of hydrogen bonds.

## F. Binding pocket model (Saenger, Lemieux, Diederich, Homans, Rossky, Friesner, Lazaridis, Berne, *et al.*)

A major rethinking of the mechanistic origins of hydrophobic interactions between a protein and its ligand occurred in the 1980s—based originally on qualitative speculation that rationalized experimental data, and later on theory and simulation—and suggested that the molecules of water in the binding pocket of a protein adopt a structure that is less favorable in free energy than that of bulk solvent (Figure 3E). The early speculation—primarily by Saenger, who studied cyclodextrin complexes of hydrocarbons, and based on the qualitative intuition of Lemieux, who studied the binding of carbohydrates to proteins—was that the release of weakly associated water in cavities rationalized the favorable enthalpy of hydrophobic interactions in molecular recognition [12,75]. Diederich and coworkers studied cyclophane-arene inclusion complexes [71], Toone and coworkers focused on the association of carbohydrates and lectins [76], and Ladbury analyzed the recognition of double-stranded nucleic acids by DNA-binding proteins with calorimetry [77]. Each of these studies implied that the structure of water—and in particular the difference between the free energy of water at the solvent/biomolecule interface and that of bulk water—seemed to play

an important, if not dominant, role in determining the free energy of biomolecular recognition.

In explicit-water simulations of the melittin tetramer [10], Rossky and coworkers determined that the overall topography (i.e., flat, concave, or convex) of the surface of the protein had a profound effect on the structure of networks of water hydrating the surface of melittin. In the case of a convex surface, molecules of water adopt a clathrate-like structure similar to those predicted for water near small hydrophobic surfaces, and these waters were  $\sim 1 \text{ kcal mol}^{-1}$  less favorable in enthalpy than waters in the bulk. The structure of water filling a concave surface is quite different, and interconverts between a clathrate-like structure and a geometry in which a hydrogen points directly toward the surface. The enthalpy of waters near a concave surface is much less favorable (near  $5 \text{ kcal mol}^{-1}$ ) than the enthalpy of waters in the bulk.

In a separate series of modeling studies with melittin, Berne and coworkers determined that the free energy of hydration of a hydrophobic pocket was determined by the shape of the pocket [78]. Similar studies with BphCdimer indicate, by comparison, that the concave nature of the melittin cavity determines the energetically unfavorable nature of its hydration [79]. Follow-up work by Rossky and coworkers—in which they compared the structure of water near the native structure of the melittin binding pocket to the structure of water near an idealized, flat surface with the same surface chemistry as melittin—corroborated the importance of the concavity of the pocket in determining its hydrophobicity [80]; they concluded that concave hydrophobic cavities are more hydrophobic than flat hydrophobic surfaces.

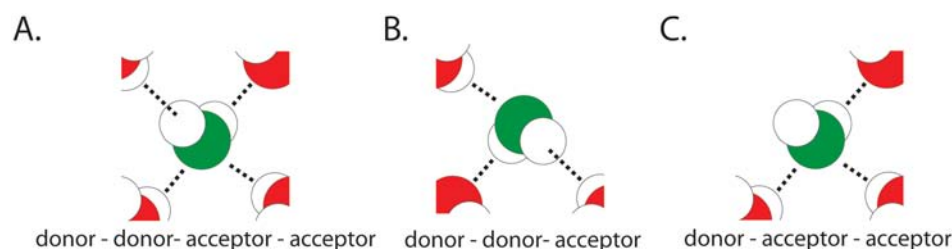
An approach to molecular recognition in water that attributes binding to the release of free-energetically unfavorable water from the binding cavity of the protein and from the surface of the ligand has become (in our opinion) one of the most attractive rationalizations for hydrophobic effects, and is compatible with a range of experimental data. Detailed studies of melittin support this idea, and suggest that the structure, and free energy, of networks of water at the surface of a protein is determined not only by the chemical groups present on the surface, but also by the topography of the surface. Below, we introduce some of the still outstanding but important questions concerning this approach, and describe some of its technical aspects to guide the reader.

To address the centrally important issue of water structure, we dedicate two sections of this review to the properties and structure of water. This subject is immense and complicated, and we provide only a summary of the most relevant information (in our view) to the hydrophobic effect. The following sections discuss, in detail, some of the important experimental, and theoretical, thermodynamic studies that lead to the conclusion that the free energy of the hydrophobic effect in biomolecular recognition is dependent on the “shape of the water”: that is, the shape—the structures and free energies of the networks of water molecules—of the water surrounding the ligand, and the analogous shape of the networks of water molecules within the binding pocket of the protein.

## **4 The structure of water in the bulk, and the structure of water at macroscopic hydrophobic interfaces**

### **A. What is the structure of water in the bulk, and how does it incorporate small molecules?**

A water-centric view of hydrophobic effects is most concerned with the changes in the network of hydrogen bonds that result when a hydrophobic surface is introduced



**Fig. 4. Patterns of Hydrogen Bonding of Molecules of Water.** The hydrogen bonds that form between molecules of water, from the perspective of a single molecule of water (highlighted in green) can be classified as a donation of a hydrogen (donor) or acceptance of a hydrogen (acceptor). Three examples of hydrogen bonding patterns are (A) the donor-donor-acceptor-acceptor pattern in which the molecule of water participates in four hydrogen bonds, (B) the donor-donor-acceptor, and (C) donor-acceptor-acceptor pattern. The donor-donor-acceptor-acceptor (A) pattern is common in bulk water, and each molecule of water assumes a tetrahedral arrangement that utilizes the maximum number of hydrogen bonds. Patterns of reduced hydrogen bonds (B and C) are found in water molecules at the interface of hydrophobic substances.

into bulk water. The plasticity of the networks of hydrogen bonds within bulk water allows the molecules to adopt configurations that can: i) incorporate an ion or small hydrophilic molecule into the network of hydrogen bonds; ii) surround a small hydrophobic molecule; iii) form a structured interface with large planar surfaces that are either hydrophilic or hydrophobic in nature; iv) surround and incorporate proteins and other larger molecules, whose surfaces are heterogeneous in composition and topographically complex; and v) fill cavities in proteins and other large molecules.

The structure of bulk water is a transient network of hydrogen bonds; each hydrogen bond in the network is strong ( $\sim 2.5 \text{ kcal mol}^{-1}$ ) but exchanges readily (the average lifetime for a hydrogen bond between two molecules of water in the bulk is 0.8–0.9 nsec) [81]. A hydrogen bond results from some combination of non-covalent dipole-dipole interactions and the direct orbital overlap of a lone pair of electrons on one molecule of water with the  $\sigma_{\text{OH}}^*$  orbital of a second molecule of water [81–83]. We describe a hydrogen bond from the vantage point of a single molecule of water: an “acceptor” interaction occurs when the lone pair of electrons interacts with a hydrogen from a neighboring molecule of water; a “donor” interaction occurs when the hydrogen interacts with the lone pair of a neighboring molecule of water. A water molecule with four hydrogen bonds participates in a donor-donor-acceptor-acceptor (DDAA) interaction (Figure 4A).

The theoretical and experimental methods used to probe the structure of water support a common view: bulk water is highly disordered, and comprises a network of hydrogen bonds that has a continuous distribution of bond lengths and bond angles [84]. Each molecule of water participates in three to four hydrogen bonds, and retains a local symmetry that is (more or less) tetrahedral [81,84]. Monte Carlo simulations of bulk water indicate that the number of hydrogen bonds in which each molecule of water participates, over a 10-nsec simulation, fluctuates between three and four; molecular dynamics simulations estimate that each molecule of water participates in approximately 3.2 hydrogen bonds [81], and that over 10 – 15% of the time, a given hydrogen is not participating in a hydrogen bond [85]. Vibrational spectroscopies [85,86], which provide an averaged view of the networks of hydrogen bonds within the bulk, and neutron scattering experiments [87–89], which provide information about the hydrogen bonds for each molecule of water in the bulk, agree with theoretical models, and support a structure in which the majority of waters in the bulk participate in a DDAA interaction.



The iceberg model predicts that molecules of water surrounding a small hydrophobic molecule of solute will be more ordered than molecules of water in the bulk. In actuality, the structure of bulk water appears not to be perturbed by the presence of small hydrophobic molecules such as argon, methane, or tetramethylurea [36,66,90]. Small hydrophobic molecules are not topographically complex, and can be viewed as a single convex surface that molecules of water must surround. The incorporation of methane or argon (both of which are less than 1 nm in diameter) into bulk water does not disturb the network of hydrogen bonds in bulk water, and a negligible change in enthalpy of hydration is observed (i.e., no hydrogen bonds are broken) [65,66]. There is an entropic cost because a small cavity must form to accommodate these small molecules [62,91], and because the orientation and translation of molecules of water near this cavity are more constrained than they are in bulk water.

The average strength of the hydrogen bonds (infrared spectroscopy) [92,93], and the average distance between each molecule of water (small-angle neutron scattering) [90] of bulk water are unchanged by the presence of tetramethylurea. Neutron scattering experiments of molecules of methane dissolved in water support the findings from tetramethylurea, and do not suggest that icebergs (i.e., regions of water with a density less than that of bulk water) form around the molecules of gas [34]. The molecules of water surrounding methanol participate, on average, in three or fewer hydrogen bonds (and are responsible for the an unfavorable enthalpy of solvation) [94], but retain a disordered structure similar to the bulk.

## What is the structure of water at a macroscopic, and planar, surface?

Although planar surfaces are not representative of the surface of a protein, they do provide a system that can be probed readily with spectroscopy. The structure of molecules of water at an interface with air [86,95], a non-polar liquid [96–98], or a solid surface presenting hydrophobic functional groups [99–101] share a commonality: the layer of water in direct contact with the non-water surface is xenophobic, and the water molecules it contains participate in fewer hydrogen bonds, on average, and is ~40% less dense [102] than water in the bulk. Molecules of water one layer away from the non-water surface have a structure similar to bulk water, and participate in the DDAA (Figure 4A) pattern of hydrogen bonding. The molecules of water in direct contact with the non-water surface participate in either a donor-donor-acceptor (DDA, Figure 4B) or a donor-acceptor-acceptor (DAA, Figure 4C) pattern of hydrogen bonding. A lone pair, on the oxygen atom, faces the interface in a DDA pattern. A DAA pattern results in a “dangling” O–H bond that points into the non-polar liquid, or toward the solid surface. Richmond *et al.*, showed that the vibrational frequency of the dangling O–H bond is dependent upon the dipole moment of the molecules in the non-polar liquid (or the absence of a dipole in the case of air) [103]; the frequency of the O–H stretch decreases with increasing dipole moment, and suggests that this bond interacts with, and is influenced by, the non-polar liquid.

While the structure of the hydrogen bonds of water in direct contact with a solid surface is similar to those at a non-polar liquid, the overall structure of water at a solid surface is distinct from that of water in direct contact with a non-polar liquid in two ways: i) the waters are more ordered (“ice like”) than the molecules of water in contact with a non-polar liquid (which are disordered and resemble the bulk) [86,104,105]; ii) the density of water is much less than that of the bulk. The origin of this decrease in density is debated [99–101]. In a recent review of the literature, Ball [12] concluded that a low-density region, approximately one molecule of water in thickness (i.e., a “molecular void”), exists at the surface of a hydrophobic solid. This molecular void is attributed to the “dewetting” of the surface. Dewetting refers to the formation

of low-density region between water molecules and a hydrophobic surface; dissolved gases within the solution are thought to partition selectively to this low-density region and adsorb onto the hydrophobic surface. The formation of a dewetted hydrophobic surface is more favorable in free energy than the free energy required to solvate it [63, 106, 107].

## What is the structure of water at the surface of a protein?

There are no experiments (of which we are aware) that directly probe the structure of water at the hydrophobic surface of a protein. We must, therefore, extrapolate that the structure of water at the surface, and in the active site, of proteins could be similar to the structure of water at planar surfaces, namely: i) the density of water in contact with hydrophobic regions is less than bulk water, and results from partial or complete dewetting of the surface; ii) the structure of water at a solid hydrophobic surface is more “ice like” than waters in the bulk.

Dewetting of a surface becomes more favorable in free energy when the surface is transformed from a planar interface to one that is concave or convex in shape [108]. Hummer *et al.* [109] proposed that the dewetting of a concave hydrophobic surface is favored in free energy because there are few hydrogen bonds formed with the surface, and the water is confined in volume (there is no restriction due to volume for water contacting a planar substrate). The free energy of the molecules of water at a hydrophobic surface depends upon its shape [110–112]. The surface of a protein is certainly not, however, completely hydrophobic, and molecules of water can form hydrogen bonds with polar residues on its surface as well as with exposed portions of the amide backbone.

Vibrational spectroscopic measurements provide a great deal of information about the structure of the water (water-like *vs.* ice-like) at the surface of a protein, and the networks of hydrogen bonds between these molecules of water. The structure of water at the surface of a silica substrate changes dramatically when BSA is adsorbed onto the surface. The adsorption of BSA onto the surface of silica reduced the intensity of the peaks that correspond to hydrogen bonds that are ice-like and water-like that are observed on bare silica [113]; Cremer *et al.* postulate that this reduction in intensity supports the formation of a “molecular void” at the surface of the adsorbed protein. The vibrational spectrum of water in the presence of BSA [113, 114], lysozyme [115], and poly-(*N*-isopropylacrylamide) [116] do not indicate the presence of dangling O–H bonds. It is unclear, however, if dangling O–H bonds are present at the surface of the protein, or if the bonds simply could not be resolved spectroscopically [117, 118].

## 5 Simulations of water in the vicinity of proteins

Crystals of proteins contain large numbers of molecules of waters (greater than 27% of the total volume of a typical protein crystal is water), and a small fraction of these waters at the surface, and within the active site, of a protein is resolvable with X-ray crystallography. Molecules of water that are “ordered” through the formation of hydrogen bonds with polar and charged groups on the surface of a protein can be resolved in a crystal structure; non-polar regions often do not seem to order waters. X-ray crystallography, therefore, does not resolve every molecule of water within the binding pocket of a protein. Even high-resolution X-ray crystal structures (1.0–1.2 Å) contain regions in the binding pocket that appear empty [109, 119].

In order to build a more comprehensive view of the structure of water molecules within a binding pocket, it is currently necessary to combine X-ray crystallography with computational approaches that explicitly model water molecules. Theoretical approaches tend to use two classes of methods: i) methods that use empirically derived potential functions to identify tightly-bound water molecules in the binding pocket; and ii) methods that map the hydration energy landscape in a binding pocket to predict which sites will be occupied with waters, and suggest the interactions that make these sites favorable.

In cases where both the structure of the protein and the positions of the waters are known, the HINT program [120,121], the CONSOLV program [122], and the WaterScore program [123] use empirically derived potential functions to estimate which crystallographic waters are tightly bound, and which are weakly bound (and thus readily displaced by a ligand). In a similar vein, the SuperStar [124] and AcquaAlta [125] programs use X-ray crystal structures from the Protein Data Bank (PDB) and Cambridge Structural Database (CSD) to predict the locations of water molecules within the binding pocket of a protein by identifying water molecules from crystal structures of proteins with chemically similar environments. The strength, and weakness, of these empirical methods is that their accuracy is limited by the data—most often X-ray crystal structures—used to develop the empirical models. When applied to binding sites with familiar structures, the empirical scoring functions tend to produce accurate results, and are fast to calculate. One would expect these methods to perform most poorly when applied to protein sites with novel structures and chemistries that might be poorly represented in the structural databases. Additionally, although these methods classify waters in binding sites as “stable” or “unstable”, they do not provide more quantitative estimates of the thermodynamics of solvation. The most frequent use of this class of methods is to understand which waters observed in a crystal structure are energetically significant, and should be considered for further modeling studies; including these non-bulk waters can significantly improve the accuracy of structure-based drug design (e.g., docking).

Approaches based on solvent mapping share a common strategy: they sample the overall free energies of different configurations of water at the surface of the protein, in order to predict the structural and energetic characteristics of the water molecules near the surface. These methods differ greatly in their implementations, however; they use different models of water, a wide variety of sampling techniques, and representations of the receptor, and have differences in computational expense, performance, and domains of applicability.

One of the most computationally efficient of the mapping methods is the 3D-RISM approach [126], which predicts the average solvent density in the binding pocket of a protein, using a model of liquid water based on the density functional theory developed by Chandler, McCoy, and Singer for non-uniform polyatomic liquids [127,128]. This approach predicts a density map of the water in a binding site, without explicitly sampling the energy of a solvent probe in the binding pocket. It is extremely efficient, as it does not require the sampling and energetic evaluations of many water configurations in a binding pocket, although it will only model the water in the binding pocket as well as the underlying theory allows. The model for water underlying 3D-RISM implicitly accounts for solvent-solvent interactions, so the 3D-RISM density maps should, in principle, recreate the networks of solvent often observed in the binding pockets of proteins. It uses a conformationally rigid model of the protein (solute), and is easily applied to the structure of the ligand, the structure of the unbound binding pocket, and the structure of the protein-ligand complex. 3D-RISM can therefore be used to identify positions on a ligand molecule that may be modified to improve both the ligand and receptor desolvation energies; this capability is significant because desolvation of the ligand is an important component of the overall

free energy of desolvation [27]. While modeling the receptor and ligand molecules with fixed conformations is certainly a major approximation, the ability to model the solvation energies of all major states in the binding reaction is also an important capability.

One of the first computational tools for predicting the binding of water molecules to proteins (when the water structure may not already be known) was the GRID program [129], which maps the interaction energy, obtained using molecular mechanics, of multiple isotropic probes with a protein structure, to identify sites with favorable chemical potentials for ligand binding. This approach was the first of several that use a probe molecule (or molecules) to model the free energy landscape of solvation in a binding site. Since the protein is treated as a rigid body, and water is modeled as an isotropic molecule, the method is highly computationally efficient, and has demonstrated good results in identifying water positions that are important for protein function or ligand binding [130,131], although quantitative estimates of the energetics of those water positions requires the use of more sophisticated approaches. In addition, since only one probe molecule is used to map the solvation energy landscape, the water positions identified by GRID do not take solvent-solvent interactions explicitly into account. GRID is therefore often used to build in “missing” water molecules to augment the receptor structure to account for the presence of highly stable waters that are tightly bound directly to the receptor. Wallnoefer *et al.* provide an example of this use of GRID, and describe its use to build a water network in the binding site of Factor Xa; knowledge of this network stabilizes subsequent molecular dynamics simulations [132].

The SZMAP program (OpenEye Scientific Software) is conceptually similar to the approach used in GRID, in that it maps the energy of a single probe water molecule through the volume of the binding pocket of a protein. Rather than using an isotropic model of water, SZMAP uses an all-atom water molecule, for which the interaction energy with the protein depends on the orientation of the probe; if hydrogen bond donors and acceptors are aligned, the electrostatic potential decreases favorably. SZMAP also uses Poisson-Boltzmann continuum electrostatics theory to model the remaining waters within and without the binding pocket, so that the energy of the probe water is not evaluated in vacuum. The probe water molecule is translated through the volume of the binding pocket, and at each point, the program samples many orientations of the water. The molecular mechanics energies of each orientation are evaluated, and a partition function is constructed to estimate the binding affinity of the water at that point in space. The SZMAP approach is a compromise between speed and sampling; a SZMAP calculation requires much less CPU time than more computationally intensive methods that sample the structures of many water molecules, or that allow the protein atoms to move. The protein is, however, treated as a rigid body, and the use of a single water probe prevents tools like SZMAP from elucidating water-water interactions that are important in the confined space of a binding pocket. SZMAP, like 3D-RISM, is also capable of computing solvation energy landscapes of the unbound receptor, of the free ligand, and of the complex, in order to model the solvation component of the binding reaction completely.

In order to describe the structure of water in a binding pocket more completely, several approaches use Monte Carlo (MC) or Molecular Dynamics (MD) methods to sample structures of a fully solvated binding pocket. These approaches are much more computationally expensive than those previously described because even a small number of water molecules can adopt a large number of potential hydrogen-bonding configurations. Barillari *et al.* [133] used a Replica-Exchange Thermodynamic Integration (RETI) Monte Carlo simulation to sample the configurations and energies of waters in the binding pocket of six different proteins including trypsin and factor Xa and correlated these energies with the degree to which crystallographic

waters in a binding pocket are conserved across multiple structures of the same receptor, in order to determine the predictive utility of the energies derived using this method.

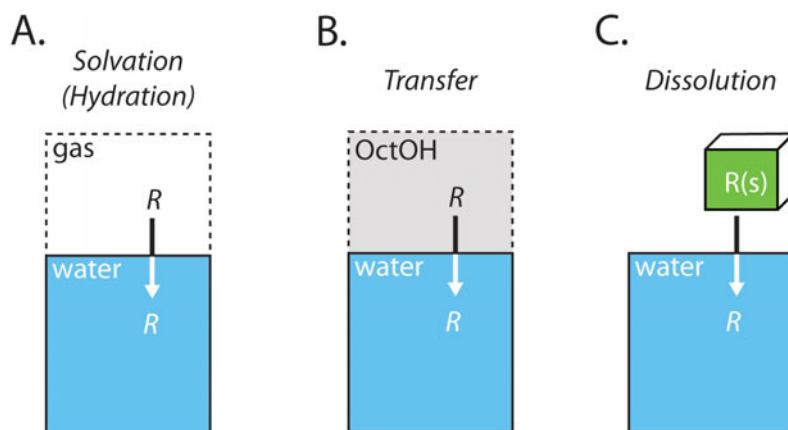
The major problem with this approach to sampling the configurations of molecules of water near the surface of a protein is that waters in highly confined regions of the protein exchange slowly with the bulk solvent; these slow exchange rates hinder the computation of their free energies of binding. To address this problem, methods such as the JAWS program [132] employ a Grand Canonical Monte Carlo (GCMC) sampling approach to accelerate the diffusion of water molecules in confined regions; as a result, these simulations converge more quickly, and produce more accurate estimates of energies for confined waters. Although they have the potential to model water-water interactions in a binding pocket accurately, these approaches are much more computationally expensive than the probe-mapping approaches of GRID and SZMAP. These methods also often treat the protein as a rigid (or semi-rigid) body, in order to reduce computational complexity and promote convergence of the simulations.

The last class of simulations uses molecular dynamics to sample the average positions of equilibrated waters in a binding pocket; the WaterMap program [134,135] uses MD to sample the configurations of waters within the binding pocket, and small motions of the side chains of proteins. Rather than directly computing the water binding energies from the simulation (a computation that would require much longer simulation times), WaterMap uses an endpoint-style approach, and post-processes the trajectory from an MD simulation to identify clusters of waters in the binding pocket. These clusters represent the preferred solvation sites described by inhomogeneous solvation theory, which postulates that water at the surface of a protein will vary widely in density, structure, and energetics. The thermodynamic binding parameters for waters at each of the solvation sites are computed using the ensemble of water orientations sampled in each cluster. WaterMap has been applied to a wide variety of biological systems, both to guide ligand design, and to understand the protein-ligand-solvent interactions underlying protein function [27,135–137]. WaterMap is much more computationally efficient than methods based on either MC or MD free energy simulations, since the MD simulation used by WaterMap do not need to sample the binding and unbinding of water molecules to each site of interest. They are, however, much more computationally expensive than single-probe approaches.

In many ways, these programs are relatively new, and more research is required to define the strengths, weaknesses, and utility of each. As the applications of computational water models have evolved in sophistication—from early usage to predict qualitative characteristics of binding site waters (e.g., position, conservation), to modeling the desolvation of binding in order to guide ligand design, to calculating the absolute binding free energies of binding site waters—it has become possible to model the solvation of many important biological processes, using modest computer resources, in computationally reasonable times.

## 6 Thermodynamic data for transfer of solutes from non-polar phases to aqueous phases

Most models of the hydrophobic effect described in Section 3 were devised to rationalize the unusual thermodynamics of transferring small ( $< 500$  Da) hydrophobic molecules from a non-polar liquid (or a vapor) to water. The iceberg model, proposed by Frank and Evans, suggests that when molecules of a non-polar gas dissolve in water, entropically-unfavorable networks of water form around them. When comparing



**Fig. 5. Schematic Definitions for Thermodynamic Processes.** (A) Solvation or hydration refers to the transfer of a molecule ( $R$ ) from the gas phase (system indicated with a dotted line) to the aqueous phase (light blue box). (B) Transfer is a general term that describes the transfer of a molecule from a non-polar phase (in this diagram octanol) to an aqueous phase. (C) Dissolution refers to the solution of a solid compound (depicted as a green box) in an aqueous phase.

two small molecules, three key principles arise from the data we summarize below: i) the molecule with a larger hydrophobic surface area will have a less favorable free energy of transfer from a hydrophobic phase to an aqueous phase than the molecule with smaller hydrophobic surface area; ii) at room temperature, entropy makes the dominant contribution to this unfavorable free energy of transfer; iii) the difference in heat capacity, at constant pressure, between the larger and smaller molecule will be negative in sign, and linearly proportional in magnitude to the difference in non-polar surface area of the two molecules.

## A. Definitions (transfer, dissolution, solvation, hydration)

Data that describe the free energy of water near small, non-polar molecules have been reviewed extensively. In particular, the painstaking calorimetric measurements of Wadsö, Gill, Murphy, Riebsehl, and others provides an excellent starting point for considering hydrophobic effects that pertain to small molecules [138–142].

Before discussing these data, we clarify several terms: i) *Solvation and Hydration*. Both words are general, and refer to the interaction of solvent with a molecule when it transfers from the gas phase to infinite dilution in that solvent. Solvation refers to that process, generally, whereas hydration refers specifically to solvation in water (Figure 5A). ii) *Transfer*. This general term describes the movement of a molecule from one liquid phase into a second liquid phase; the movement of a molecule at infinite dilution in a non-polar liquid (e.g., octanol) to an aqueous solution of that molecule at infinite dilution (Figure 5B), and to the differences in thermodynamic functions between these two liquid phases. iii) *Dissolution*. This general term refers to the conversion of a solid (crystalline or amorphous) state of a compound to a solution of that molecule at infinite dilution (Figure 5C), and to the differences in thermodynamic functions between these states.



## B. Thermodynamic data (values of $\Delta G^\circ$ , $\Delta H^\circ$ , $-T\Delta S^\circ$ and $\Delta C^\circ$ ) for alkanes and normal alcohols

### i) Transfer from gas phase to aqueous phase (hydration) of straight-chain alkyl groups

The free energy associated with the transfer of straight-chain alkanes and normal alcohols into water, from the gaseous state, ( $\Delta G_{\text{hydration}}^\circ$ , Eq. (7)) can be determined from the solubility of the gaseous molecule (i.e., the concentration at which a saturated solution of the molecule is formed), where  $G_{\text{solution}}^\circ$  is the free energy of a saturated solution of solute X,  $G_{\text{water}}^\circ$  is the free energy of the solution prior to the introduction of X,  $[X]_{\text{sat,solution}}$  is the concentration of a saturated solution of solute X at equilibrium, and  $[X]_{\text{vapor}}$  is the pressure of X at equilibrium.

$$\Delta G_{\text{hydration}}^\circ = G_{\text{solution}}^\circ - G_{\text{water}}^\circ = -RT \ln \frac{[X]_{\text{sat,solution}}}{[X]_{\text{vapor}}}. \quad (7)$$

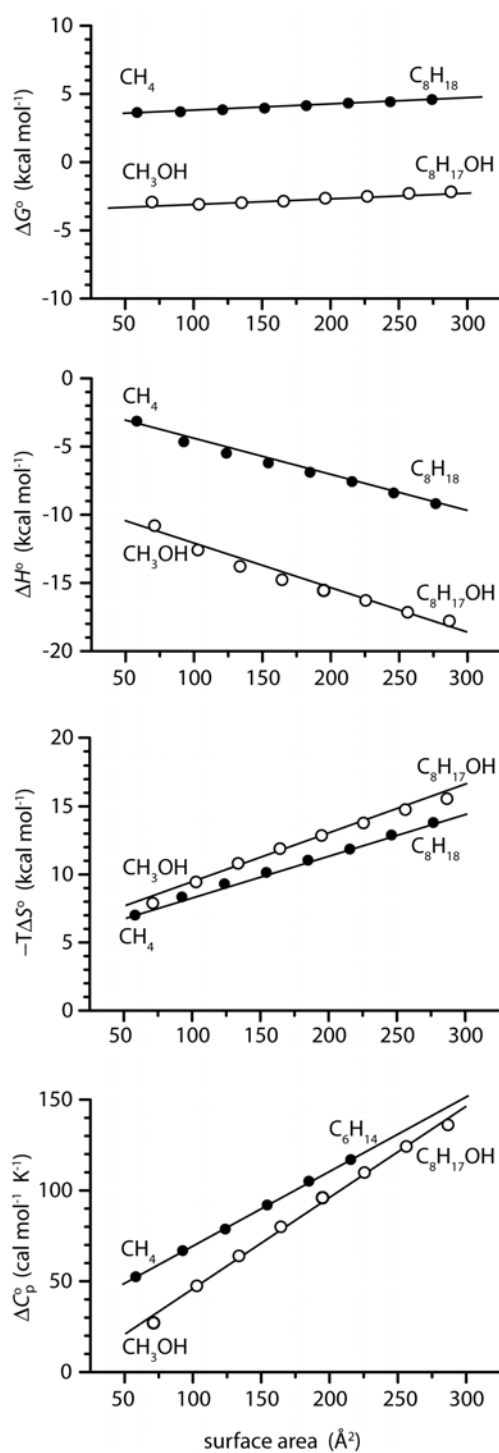
Calorimetry measures the enthalpy of hydration ( $\Delta H_{\text{hydration}}^\circ$ ) for these compounds. The  $\Delta H_{\text{hydration}}^\circ$  for gaseous compounds with high vapor pressures (e.g., straight-chained alkane gases such as ethane, propane, and butane) is estimated from the heat evolved when the gas dissolves into water, and the quantity of gas dissolved [143]. The value of  $\Delta H_{\text{hydration}}^\circ$  for liquid compounds is the difference between the molar enthalpy of dissolution (i.e., the heat to dissolve the pure liquid in water) and the molar enthalpy of vaporization (i.e., the heat of vaporization of the pure liquid) [144,145]. Measurements of  $\Delta H_{\text{hydration}}^\circ$  over a range of temperatures provide an estimate of  $\Delta C_{p\text{hydration}}^\circ$ , which is derived from the first derivative of  $\Delta H_{\text{hydration}}^\circ$  with respect to temperature. Figure 6 plots  $\Delta G_{\text{hydration}}^\circ$ ,  $\Delta H_{\text{hydration}}^\circ$ ,  $-T\Delta S_{\text{hydration}}^\circ$ , and  $\Delta C_{p\text{hydration}}^\circ$  for straight-chained alkanes and normal alcohols [146], ranging from  $C_1 - C_8$ , as a function of the molecular surface area of each compound.

The thermodynamic parameters of hydration vary linearly with increasing molecular surface area. Increasing the length of a straight-chain alkane, or a normal alcohol, by a single methylene unit, decreases the solubility of the molecule in water at room temperature, and  $\Delta G_{\text{hydration}}^\circ$  becomes increasingly unfavorable. Values of  $\Delta\Delta G_{\text{hydration}}^\circ$  – the incremental change in  $\Delta G_{\text{hydration}}^\circ$  as a function of increases in  $\text{\AA}^2$  – are calculated from the slope of the linear fit for the data for alkanes and alcohols: the values of  $\Delta\Delta G_{\text{hydration}}^\circ$  for the alkanes ( $\Delta\Delta G_{\text{hydration}}^\circ = 4.1 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ ) and the alcohols ( $\Delta\Delta G_{\text{hydration}}^\circ = 4.6 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ ) are effectively indistinguishable.

This hydrophobic effect, in the case of the hydration of alkanes and alcohols, results from an unfavorable entropic contribution to  $\Delta G_{\text{hydration}}^\circ$  that is only partially offset by a favorable enthalpy of hydration ( $\Delta\Delta H_{\text{hydration}}^\circ = -28.2 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ ). These data are compatible with the iceberg and void-volume models of hydrophobic hydration, and are consistent with the generalization that the dehydration of small hydrophobic molecules, with radii less than 1 nm, is favorable in entropy.

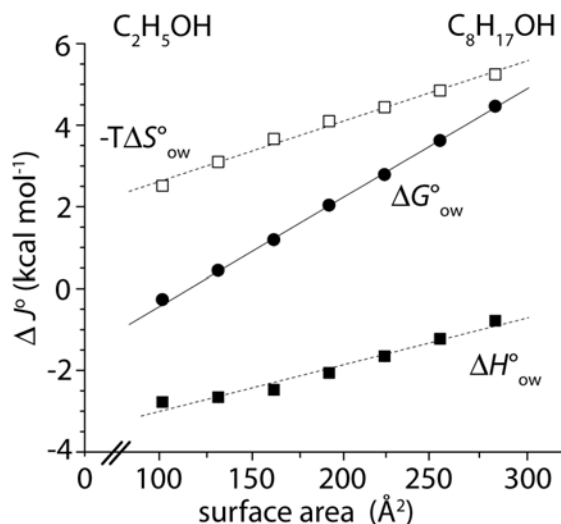
### ii) Transfer of liquid-phase, normal alcohols from octanol to aqueous phase

The relationships between each of the thermodynamic parameters associated with transferring a normal alcohol from an aqueous buffer to octanol, and increasing the surface area of the molecule, are also linear. Riebesehl and Tomlinson measured the enthalpy and free energy to transfer a normal alcohol (ranging in size from ethanol



**Fig. 6. Thermodynamic Data for the Hydration of Straight-Chained Alkanes and Normal Alcohols at Room Temperature.** Data from Plyasunov [146] are plotted against values of molecular surface area determined for the energy-minimized structures of the compounds in the gas phase (e.g., the area calculated at the van der Waals radii). The solid dots represent the data for straight-chain alkanes (we label methane and octane

**Fig. 6.** (Continued). for reference), and the open dots represent the data for normal alcohols (we label methanol and octanol for reference). The data for  $\Delta C_{p\text{hydration}}^\circ$  of each molecule are a compilation of values of  $\Delta H_{\text{hydration}}^\circ$  values from the literature, assembled by Plyasunov.



**Fig. 7.** Thermodynamics of Transfer for Normal Alcohols from Octanol to Water. The figure is adapted from the calorimetric data of Riebesehl and Tomlinson [147]. Solid circles represent  $\Delta G_{\text{ow}}^\circ$ , solid squares represent  $\Delta H_{\text{ow}}^\circ$ , and open squares represent  $-T\Delta S_{\text{ow}}^\circ$ .

to octanol) from an aqueous solution (pH = 7) to water-saturated octanol [147]. We adapted the data from these experiments in Figure 7 to represent the thermodynamics of transfer from octanol to water (e.g.,  $\Delta H_{\text{ow}}^\circ$  is the enthalpy of transfer from octanol to water).

The free energy of transfer for a normal alcohol from octanol into water, ( $\Delta G_{\text{ow}}^\circ$ ) is unfavorable for alcohols larger than propanol, and reflects an unfavorable entropic term. Like hydration, the entropy of transfer of normal alcohols from octanol to water increases (becomes more unfavorable) with increasing surface area—a trend consistent with iceberg and void-volume models. The value of  $\Delta\Delta G_{\text{ow}}^\circ$ —the value of the slope of the best-fit line through the values of  $\Delta G_{\text{ow}}^\circ$ —is unfavorable ( $\Delta\Delta G_{\text{ow}}^\circ = 27.1 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ ), but larger in magnitude than  $\Delta\Delta G_{\text{hydration}}^\circ (= 4.64 \text{ cal mol}^{-1} \text{ \AA}^{-2})$ . This difference is primarily due to enthalpy: the difference in the values of  $\Delta\Delta H_{\text{hydration}}^\circ$  and  $\Delta\Delta H_{\text{ow}}^\circ$  ( $\Delta\Delta H_{\text{ow}}^\circ - \Delta\Delta H_{\text{hydration}}^\circ$ ) is  $+15 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ , and corresponds to the transfer of a methylene group from octanol to the gas phase. This value is attributed to favorable dispersion interactions among alkyl groups in liquid alkanes.

### C. Anomalies in solubility with changing temperature

Remarkably, and unlike many polar solutes, which display increasing solubility in water with increasing temperature, the solubility of hydrocarbons and other non-polar molecules in water does not change significantly with increasing temperature. For example, the solubility of benzene in water is 23 mM at 25° C and 32 mM at 70° C, and the solubility of cyclohexane in water is 0.7 mM at 25° C and 1.1 mM at 70° C [148]. Although the free energy of hydration of cyclohexane does not display a

strong dependence on temperature, interestingly, both the enthalpy and entropy of hydration do.

Near room temperature, the free energetically unfavorable hydration of cyclohexane results from an unfavorable  $-T\Delta S_{\text{hydration}}^{\circ}$ , and the  $\Delta H_{\text{hydration}}^{\circ}$  is nearly zero. Near 70° C, however, the unfavorable  $\Delta G_{\text{hydration}}^{\circ}$  is caused by an unfavorable  $\Delta H_{\text{hydration}}^{\circ}$ . The iceberg model of the hydrophobic effect rationalizes this anomaly: at room temperature, the molecules of water surrounding the cyclohexane adopt an ice-like structure that is constrained entropically, and this unfavorable  $-T\Delta S_{\text{hydration}}^{\circ}$  is responsible for an unfavorable  $\Delta G_{\text{hydration}}^{\circ}$ ; at high temperatures, the molecules of water surrounding cyclohexane form fewer hydrogen bonds and are less entropically constrained (i.e., are less ice-like), and the unfavorable  $\Delta G_{\text{hydration}}^{\circ}$  results from an unfavorable  $\Delta H_{\text{hydration}}^{\circ}$ . There is no model available currently to rationalize this entropy-enthalpy compensation.

## 7 Molecular scale topography: Xenophobic water

### A. Does partitioning between water and hydrophobic liquids correlate with biomolecular recognition?

We believe that the short answer to this question is “no”. In view of numerous studies of protein-ligand interactions that combine structural and thermodynamic information, and data that characterize the thermodynamics of partitioning of small molecules from aqueous to hydrophobic phases, the weak correlation between the free energy of binding and free energy of partitioning is not replicated in terms of enthalpy or entropy. The examples we discuss below indicate that different hydrophobic effects determine the thermodynamics of binding and partitioning, although both classes of hydrophobic effect probably result from the differences between the structures and free energies of water near solutes and those of bulk water.

### B. Enthalpy-dominated hydrophobic effects

It is becoming increasingly apparent that the interactions between two non-polar surfaces—and in particular the formation of protein-ligand complexes—are not caused by the release of entropically unfavored waters alone, but rather by interactions in which the enthalpy is a favorable, and often dominant contributor, to the free energy of binding [149]. We can classify enthalpy-dominated interactions into three categories: i) enthalpic gains from solute-solute interactions in which water that weakly interacts with a hydrophobic surface of a protein (or synthetic host) is replaced by a more favorable interaction between the protein and a ligand (or a synthetic host and a guest molecule); ii) enthalpic gains associated with solute-solute interactions that are mediated by molecules of water (i.e., solute-water-solute interactions); and iii) enthalpic gains from the reorganization of water in a binding pocket that results from ligand binding. The complexation of aromatic molecules to synthetic hosts (e.g., cyclophanes [150], hemicarcerands [151], and cyclodextrins [8, 152]) is an enthalpically-dominated process in which weak interactions between the host and the water molecules within the host are replaced with host-guest dispersion interactions, and these dispersion interactions are stronger than those between the molecules of water and the host. The enthalpy-dominated interaction of *n*-alcohols, of increasing length (from pentanol to decanol) with major urinary protein (MUP) is analogous to these host-guest interactions with synthetic hosts, because water molecules interact

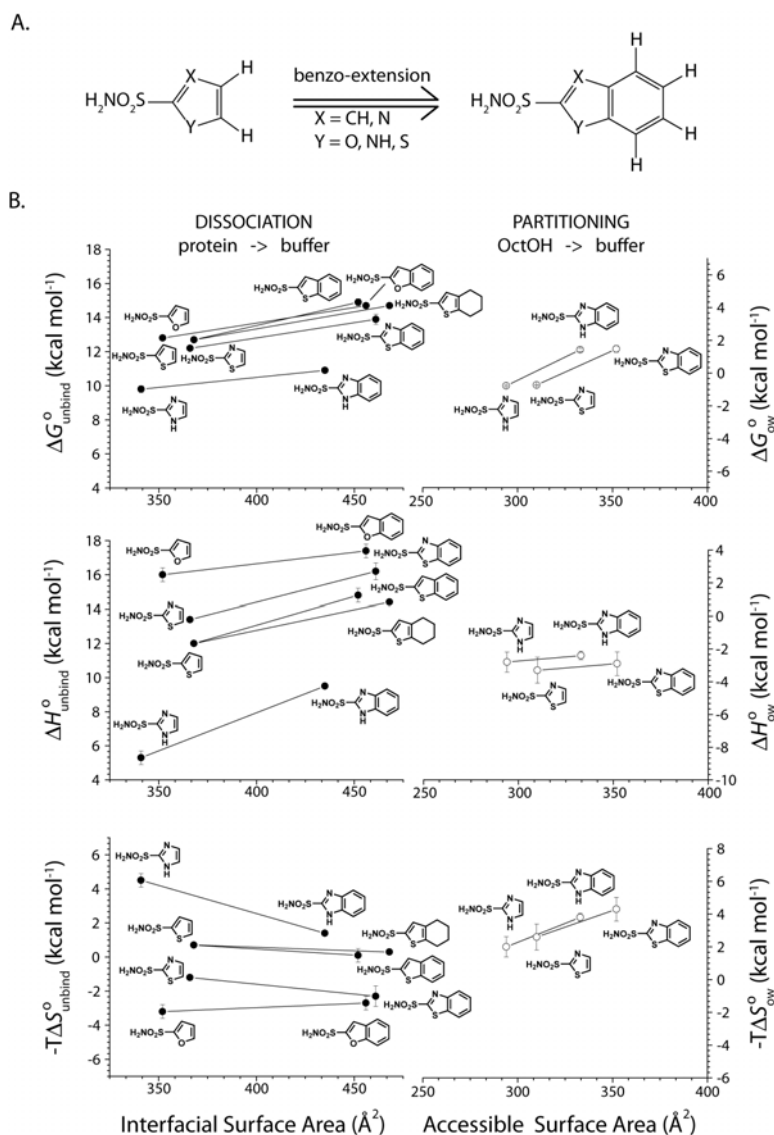
weakly with the hydrophobic binding pocket of MUP, and escape when replaced by alcohols [153–155]. The binding of carbohydrates to lectins results in a decrease in both enthalpy (more favorable) and entropy (less favorable) [14, 75, 76, 156], and this balance of effects has been attributed to: i) increased intramolecular hydrogen bonding, in which the hydroxyl groups of the carbohydrate hydrogen-bond to one another, and ii) increased intermolecular hydrogen bonding in which hydrogen bonds form between the carbohydrate and the lectin, either directly or via a molecule of water. The binding of arylsulfonamide ligands to human carbonic anhydrase (HCA) is an interaction in which the hydrophobic component seems to result primarily from water-mediated interactions between the protein and the ligand (See Figure 8; discussion presented in section D) [27].

### C. The prevalence of entropy-enthalpy compensation in protein-ligand interactions

A motivation for trying to understand the hydrophobic effects involved in protein-ligand interactions is that the understanding might make it more practical to design (rather than screen for) tight-binding ligands. A common frustration encountered in efforts of ligand design is, however, that small (“rational”) perturbations to the structure of a ligand such as increasing molecular weight, or hydrophobic surface area often do not increase binding affinity—by decreasing  $\Delta G_{\text{bind}}^{\circ}$ —but instead produce anti-correlated changes in  $\Delta H_{\text{bind}}^{\circ}$  and  $-T\Delta S_{\text{bind}}^{\circ}$ . This entropy-enthalpy compensation can result in a small or no change in affinity. One component of successful ligand design is, thus, the identification of chemical modifications that are independent of this type of entropy-enthalpy compensation. A water-centric view of the hydrophobic effect supports increasing both the volume and the hydrophobic surface area of a ligand as a strategy for increasing affinity because both, in principle, result in a larger volume of water being expelled from the binding pocket of a protein into the bulk. The increase in surface area or volume of the ligand does not have to correlate with an increased, apposed interfacial surface area between ligand and protein, but simply in the displacement of waters from the binding pocket.

Although controversy surrounds the statistical validity of many reported examples of entropy-enthalpy compensation [21, 157], there are nevertheless many systems of protein and ligand that clearly display statistically significant entropy-enthalpy compensation [158]. Olsson and colleagues [159] review two competing theories to explain the prevalence of entropy-enthalpy compensation: i) entropy-enthalpy compensation is a result of fundamental thermodynamic and statistical mechanical responses to small perturbations in the protein–ligand system, and ii) entropy-enthalpy compensation is a consequence of the shape and depth of the potential wells describing the protein, ligand, and solvent in the bound and unbound state. The statistical thermodynamic argument described by Sharp [21], proposes that entropy-enthalpy compensation results from the linear relationship between  $\Delta H_{\text{bind}}^{\circ}$  and  $-T\Delta S_{\text{bind}}^{\circ}$  for small perturbations to a statistical mechanical model system. This theory models the compensation as a consequence of the effect of small perturbations of the distribution of energy levels in a potential well, but it does not incorporate any aspects of protein, ligand, or solvent structure and bonding into the formulation of the model system.

The second theory, sketched by Williams [26] and Dunitz [31], is based on the intuitively plausible idea—within the context of a lock-and-key-like model—that a ligand that is more tightly bound will also be more entropically constrained. This conceptual model has been stimulating, but it is not obvious how to extend it to a water-centric view of binding. Ford made an effort to extend this theory of entropy-enthalpy



**Fig. 8.** (A) Ligands used in the benzo-extension study; (B) The thermodynamics of dissociation of the parent and benzo-extended ligands from HCA are plotted (in the left column) as a function of the surface area of the interface between the ligand and HCA. The thermodynamics of partitioning from octanol to water for the parent and benzo-extended ligands plotted (in the right column) as a function of surface area of the ligand.

compensation to include solvent [30] and other interactions. Olsson *et al.* conclude that this theory, while attractive, is also more illustrative rather than predictive.

NMR spectroscopy and computational simulations of protein-ligand and protein-protein binding reveal that binding results in a loss of the conformational and vibrational entropy of the side chains of the protein, and can contribute significantly to  $-\Delta S^\circ_{\text{bind}}$  [160–163]. These results reinforce the theoretical framework of Williams and Dunitz, as they demonstrate both experimentally and computationally that steric



interactions in the protein-ligand complex can reshape potential energy wells for atoms at the binding interface, and result in large losses or gains of vibrational entropy in both the protein and ligand. Ligand binding also can induce allosteric changes in protein dynamics and structure in regions of the protein that are distant from the site of ligand binding [16, 18, 164]; ligand binding therefore has the potential to influence many more protein motions than simply those at the interface between protein and ligand.

There are two implications of the Williams and Dunitz model. The first is that the number of factors that contribute to the thermodynamics of binding is sufficiently large that it is intrinsically difficult to design a “simple” system to understand ligand binding. Model systems, in which a physical-organic approach is applied to study the binding of structurally similar ligands to the model protein, may provide the most interpretable information about mechanisms of protein-ligand binding. The second implication is that the thermodynamics of binding is determined by a large number of weak interactions, rather than a single dominating interaction. Predicting compensation will thus require each interaction between protein and ligand to be computationally enumerated and characterized. Such a predictive implementation would require the construction of a Hamiltonian for all possible interactions, over the timescale relevant for ligand binding. Therefore, in order to predict how a single modification to the structure of ligand will affect the  $\Delta H_{\text{bind}}^\circ$  and  $-T\Delta S_{\text{bind}}^\circ$ , and ultimately  $\Delta G_{\text{bind}}^\circ$ , these calculations require a tremendous computational effort to sample the shapes of all the potential wells; this effect is presently theoretically and computationally impractical. Future studies combining experimental and theoretical/computational components hand-in-hand, may ultimately provide the needed capability, but accurate theoretical/computational estimations of the thermodynamics of protein-ligand interaction is currently impractical for all but the simplest and most rigid systems.

## D. Carbonic anhydrase as a model system for studying the hydrophobic effect

*A model system for hydrophobic protein-ligand interactions.* The nature of “models” in science is that the more that is known about them, the more useful they become. A protein model system, combined with a physical-organic approach to probe the complexities of the hydrophobic effects involved in protein-ligand interactions, provides information about the very complex problem of molecular recognition that can be interpreted more readily and with less ambiguity than most other experimental approaches. Carbonic anhydrase (CA) is an attractive model protein for biophysical studies, [29] and, in particular, for studies that focus on the thermodynamics of the hydrophobic effects in biomolecular recognition, for five reasons [29]: i) CA is exceptionally stable, structurally. Nearly 300 crystal structures of the native protein, its mutants, and its complexes indicate that the secondary and tertiary structures are indistinguishable by X-ray crystallography [29]. ii) The mechanism by which an arylsulfonamide (of the general structure R-Ar-SO<sub>2</sub>NH<sub>2</sub>, with some restriction on the structures of “R” and “Ar”) binds to CA is known in detail [29]. The sulfonamide anion (R-Ar-SO<sub>2</sub>NH<sup>−</sup>) coordinates to the Zn<sup>II</sup> cofactor by displacing a coordinated hydroxide ion. iii) The binding pocket of CA is a conical cleft (~15 Å deep) that is lined on one side with hydrophobic residues (Phe, Pro, Ile, and Val, referred to as the “hydrophobic wall”) and on the other side by hydrophilic residues (Asn, His) [165]. Crystal structures of CA complexed with ligands of the structure R-Ar-SO<sub>2</sub>-NH<sub>2</sub> show that the R and Ar groups are positioned over the hydrophobic wall of the binding pocket. The conserved binding geometry of these ligands offers a convenient system to study hydrophobic interactions between the hydrophobic wall of CA and

a series of ligands. iv) CA is readily available in quantities large enough to perform calorimetric experiments ( $\sim 1$  mg of protein per experiment) and protein crystallography. v) HCA is easily crystallized [166] in a polymorph that diffracts X-rays to  $\sim 1$  Å, and soaking experiments with crystals in this polymorph typically provide structures of the HCA-ligand complexes at resolutions in the range of 1 – 2 Å [167]. Crystals grown under these conditions are also stable in solutions of poly(ethylene glycol). These solutions provide better solubility for hydrophobic ligands than other solutions used for crystallization and, thus, allow soaking experiments to be conducted with ligands with large, pendant hydrophobic groups [28, 168, 169].

What is clear from studies of the binding of hydrophobic ligands to CA (of which we have highlighted three examples below) is that the hydrophobic effects within this model system, in which only the identities of the ligands binding to CA are changed, are situation-specific.

*Benzo-Extension.* A comparison of arylsulfonamide ligands composed of a five-membered heterocycle (parent), and their “benzo-extended” analogs (Figure 8A) revealed that the binding of a rigid ligand system that extends over, but is not in direct contact with, the hydrophobic wall of HCA is an enthalpy-dominated hydrophobic interaction [27]. The benzo-extended ligands have a more favorable  $\Delta G_{\text{bind}}^{\circ}$  (i.e., have a lower  $K_d$ , and thus bind more tightly) than the five-membered parent compounds. The change in heat capacity ( $\Delta \Delta C_p^{\circ}$ ) upon benzo-extension of the ligand suggests that the benzo moiety orders additional molecules of waters within the binding pocket of HCA; crystal structures of the protein-ligand complexes of the larger, benzo-extended ligands contain three to five more ordered (e.g., crystallographically defined) molecules of water in the binding pocket than the smaller, parent ligands.

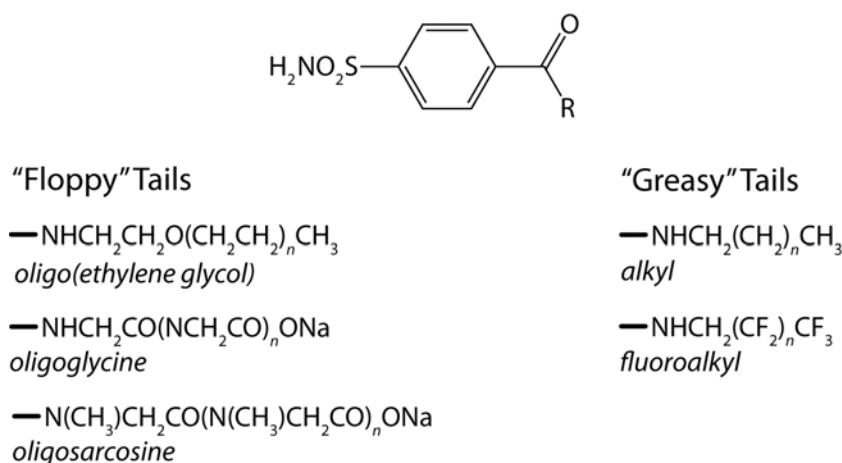
Interestingly, the hydrophobic effect responsible for the favorable transfer ( $\Delta G_{\text{ow}}^{\circ}$ ) of the ligands, upon benzo-extension, from water to octanol is not the same as the hydrophobic effect responsible for the favorable binding ( $\Delta G_{\text{bind}}^{\circ}$ ) of the ligands, upon benzo-extension (Figure 8B), to carbonic anhydrase.

The favorable enthalpy of transfer ( $\Delta H_{\text{ow}}^{\circ}$ ) and of binding ( $\Delta H_{\text{bind}}^{\circ}$ ), upon benzo-extension, can be rationalized in terms of dehydration: dehydration of the ligand in the partitioning between octanol and water, and dehydration of enthalpically unfavorable water from the binding pocket of HCA upon binding of the ligand. The unfavorable entropy of partitioning ( $\Delta S_{\text{ow}}^{\circ}$ ) can be rationalized in terms of re-ordering of bulk water to accommodate a molecule that is larger than 1 nm in diameter [106]; the entropy of binding ( $-T\Delta S_{\text{bind}}^{\circ}$ ) is approximately zero for most of the pairs of ligands in the series.

The benzo-extended system is conformationally rigid, and provides a strategy based on a well-defined, physical-organic approach to rationalize the role of water in protein-ligand binding in one, specific system; this study compliments our previous efforts to rationalize the binding of CA with sulfonamide ligands with hydrophobic tails, which are less rigid than the benzo-extension.

*Enthalpy-entropy compensation of “floppy tails” and “greasy tails”.* We studied the binding of two series of para-substituted benzene sulfonamide ligands (Figure 9) with alkyl chains (i.e., “tails”) of increasing length to CA: i) “floppy tails” of oligoglycine, oligosarcosine, and oligo(ethylene glycol) ranging in length from one to five units [28]; ii) “greasy tails” of alkyl and fluoroalkyl chains ranging in length from one to four methylene (or fluoromethylene) units [168].

The interactions between the two series of ligands and CA are quite different. The  $\Delta G_{\text{bind}}^{\circ}$  of the ligands with floppy tails is, astonishingly independent of the length of the tail, whereas  $\Delta G_{\text{bind}}^{\circ}$  of the ligands with greasy tails increases favorably (i.e., binds more tightly) with the length of the tail. A second, and noteworthy, distinction between the floppy tails and the greasy tails is the heat capacity of binding ( $\Delta C_{\text{pbind}}^{\circ}$ ), which is indicative of changes in the solvent-exposed surface area and a hallmark of a



**Fig. 9. Para-Substituted Benzene Sulfonamides with Floppy Tails and Greasy Tails.** The floppy tails (left column) consisted of a oligo(ethylene glycol), oligoglycine, and oligosarcosine tails that ranged in size from  $n =$  one to five units. The greasy tails (right column) consisted of alkyl chains of increasing methylene units and fluoroalkyl chains of increasing fluoromethylene units that ranged in size from  $n =$  zero to four units.

“hydrophobic effect”. The  $\Delta C_{\text{p bind}}^{\circ}$  for the floppy tails are independent of tail length, whereas the  $\Delta C_{\text{p bind}}^{\circ}$  becomes more favorable with increasing length of the greasy tails.

The enthalpy, entropy, and free energy of binding of the ligands with greasy tails become increasingly favorable with increasing tail length; we attributed this increase to an entropically favorable dehydration of the ligand, and an enthalpically favorable removal of waters from the active site of CA upon ligand binding. The general decrease in  $-T\Delta S_{\text{bind}}^{\circ}$  with increasing length is the same for the alkyl and fluoroalkyl tails after correction for their surface area ( $-7 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ ); dehydration of the tail, based on values for the entropy of transfer of normal alcohols (Figure 7) from octanol to water, is the major component of the changes in entropy of binding with increasing length of tail ( $-T\Delta\Delta S_{\text{bind}}^{\circ}$ ). The increase in the enthalpy of binding ( $\Delta\Delta H_{\text{bind}}^{\circ}$ ) with increasing length of the greasy tails suggests that longer tails displace a larger number of water molecules from the binding pocket of CA; the  $\Delta\Delta H_{\text{bind}}^{\circ}$  for fluoroalkyl tails is slightly larger than that of alkyl tails ( $-5 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  for fluoroalkyls, relative to  $-7 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  for alkyls), but within experimental error, and suggests that tails of equal length, surface area, and molecular volume replace the same number of waters.

While increasing the length of the floppy tails does not increase the  $\Delta G_{\text{bind}}^{\circ}$ , it does result in unfavorable changes in  $\Delta H_{\text{bind}}^{\circ}$  that are compensated by favorable changes in  $-T\Delta S_{\text{bind}}^{\circ}$ . This result is counterintuitive in light of the results of binding for the greasy tails series and for the benzo-extended ligands, but suggests that: i) like the greasy tails, increasing length of the tail results in an favorable increase in the entropy for desolvation of the ligand (which present both polar and non-polar surfaces), or perhaps desolvation of the surface of the protein; ii) like the greasy tails, increasing length of the tail or, more probably, hydrophobic volume displaces increasing amounts of water from the binding pocket of CA.

The binding of different arylsulfonamide ligands (in particular, the benzo-extended, floppy tail- and greasy tail-series of ligands) to CA emphasizes that the hydrophobic effects involved in protein-ligand binding depended on the structure of

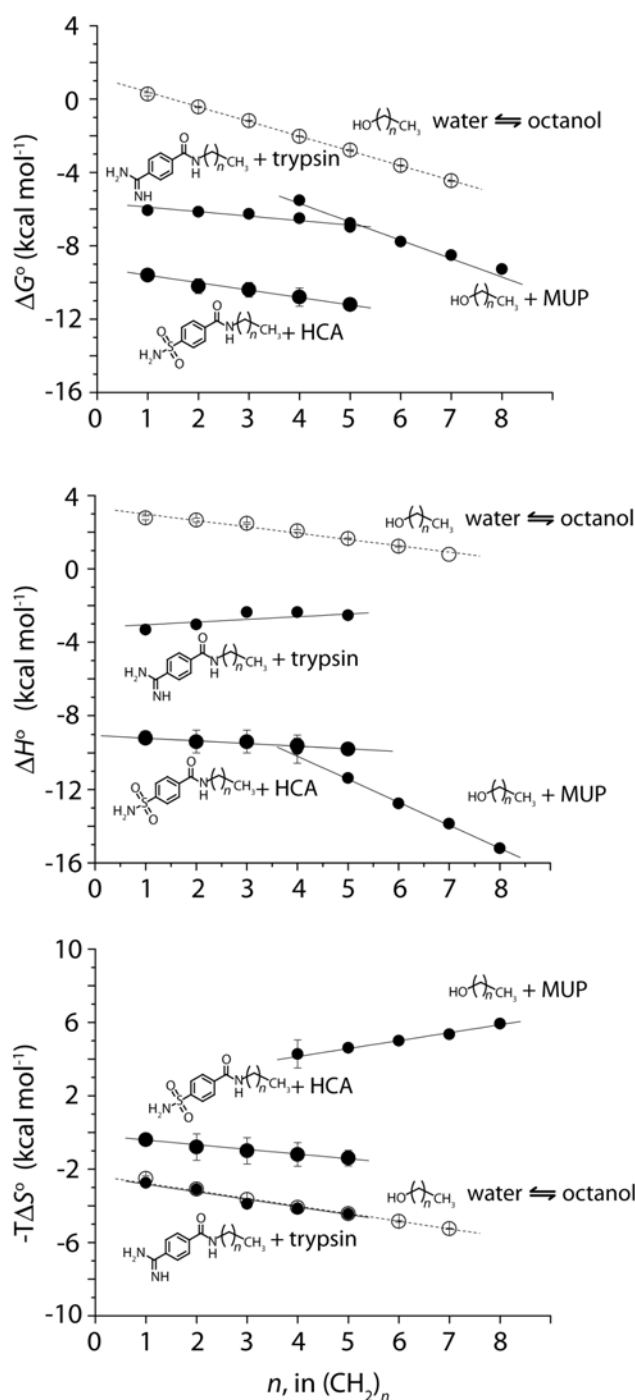
water around the ligand, and the structure of water within the binding pocket of the enzyme. In these three examples, the binding pocket of the enzyme does not change, and we can look specifically at the dehydration of the ligand and the dehydration of the binding pocket. The rigidity of the benzo-extended ligands shows that displacement of waters to the bulk (and not the direct interaction of the ligand and the binding pocket) results in an enthalpically-favored hydrophobic effect. The  $\Delta G_{\text{bind}}^{\circ}$  of the ligands with greasy tails is also enthalpy-dominated, and we rationalize it much as we do the benzo-extended ligands: enthalpically-unfavored waters are released from the binding pocket of CA into the bulk. The ligands with floppy tails require further scrutiny, and in particular studies of the enthalpy of hydration of the ligands, to determine the extent to which ligand dehydration is reflected in the enthalpy of binding, and to suggest the origin of the astonishing entropy-enthalpy compensation observed in this system.

## E. Hydrophobic effects in other systems of proteins and ligands

A physical-organic approach to understanding hydrophobic effects in protein-ligand association monitors the thermodynamic parameters of binding for a series of ligands whose structure is altered by a single, and predictable, perturbation. Figure 10 compares the thermodynamics of binding ( $\Delta J_{\text{bind}}^{\circ}$ ,  $J = G, H, S$ ) for three series of ligands whose hydrophobic alkyl chains (i.e., the “hydrophobic tail”) are increased in size by a single methylene group: i) modified arylsulfonamides to human carbonic anhydrase, HCA [168]; ii) normal alcohols to major urinary protein, MUP [154]; and iii) modified benzamidinium chlorides to trypsin [170,171]. We have also included the octanol-water partitioning data for the normal alcohols to illustrate relationships between trends of protein-ligand binding and octanol-water partitioning.

In each case, the thermodynamic parameters indicate that hydrophobic effects—in different molecular contexts—have thermodynamic origins that differ significantly. An increase in the length of the alkyl chain, for each protein, makes the value of  $\Delta G_{\text{bind}}^{\circ}$  more favorable; this result is compatible with a favorable contribution from dehydration of the alkyl group of the ligands on binding to the protein. The incremental terms (i.e., the slopes of the best-fit lines) for enthalpy and entropy of binding, however, are favorable, unfavorable, or near zero depending on the active site. In only one case does either term correlate with the partitioning of alkyl groups between water and octanol.

For MUP, Homans and coworkers rationalize the favorable changes in  $\Delta H_{\text{bind}}^{\circ}$  of normal alcohols, with increasing length, through increased, favorable, dispersion interactions between the hydrocarbon of the ligand and the aromatic- and alkyl-rich active site of the protein [154]. The authors considered the contributions from each term on the right-hand side of Eq. (6); the authors dismiss contributions from  $\Delta G_{\text{hydration}}^{\circ}$  from the release of water molecules from active site of the protein on the basis of a classical molecular interactions potential (CMIP) analysis of the active site of MUP [172]. More recently, Friesner and coworkers revisited the work of Homans, and analyzed the binding of normal alcohols to MUP with simulations utilizing an implicit and an explicit solvation model, and found that parts of the active site are almost completely devoid of water [173]. Interestingly, the values predicted for  $\Delta H_{\text{bind}}^{\circ}$  using the implicit solvation model did not correlate with the experimentally measured values; this result suggests that the increasingly favorable interaction between alcohols and MUP is not primarily the direct result of increasingly favorable dispersion interactions. The explicit solvation model does correlate with the experimentally measured values, and supports a water-centric view of the hydrophobic effect: even in MUP, in which portions of the active site are practically dry, the structure of water (or lack



**Fig. 10. Comparison of the Thermodynamics of Partitioning and Binding for  $-\text{CH}_2-$  groups.** Data collected from the binding of modified arylsulfonamides to HCA [168], modified modified benzamidinium chlorides to trypsin [170,171], and normal alcohols to MUP [154] are plotted against the number of methylene groups in the “tail” of each ligand. Data from the partitioning of normal alcohols [147], between octanol and water, are also plotted against the number of methylene groups.

thereof) in the binding pocket dominates most (or at least many) hydrophobic effects in biomolecular recognition.

In the case of trypsin, Talhout *et al.* observed that increasing the length of n-alkyl groups in the para-position of benzamidinium increased the strength of binding of a series of ligands to trypsin [170]. Increasing lengths of the alkyl chain resulted in unfavorable changes in  $\Delta\Delta H_{\text{bind}}^\circ$  and favorable changes in  $-\text{T}\Delta\Delta S_{\text{bind}}^\circ$ ; this trend is opposite to those observed by Homans and coworkers [154]. Although the authors attributed this result to “hydrophobic interactions,” they pointed out that classical models for the hydrophobic effect appeared to be “oversimplistic” [170]. Specifically, the free energy of transfer from water to octanol did not correlate with the free energy of binding in this series of ligands—a patent demonstration that, in this case, partitioning does not correlate with binding.

Each of the detailed thermodynamic, structural, and computational analyses described here deals with an exactly analogous perturbation (an increasing length of greasy tail) to a conserved ligand structure (p-carboxybenzenesulfonamide, hydroxyl, and benzamidinium) in three different active sites of structurally stable proteins (carbonic anhydrase, MUP-I, and trypsin). Within each system, the free energy, enthalpy, and entropy correlate linearly with the hydrophobic surface area of the ligand, but the values of the incremental terms, and the trends in these values, are not consistent across proteins, nor are they consistent with the thermodynamics of partitioning from octanol to water. There is every indication that the hydrophobic effect that determines the free energy of partitioning is unique, and different in the details of its origin from the hydrophobic effects observed in biomolecular recognition. In the latter context, the structures and energetics of the molecules of water *in* binding pockets may dominate the thermodynamics of binding. In any event, these thermodynamics are not captured (in detail) by water-octanol partitioning experiments.

## 8 Conclusions

### A. Partitioning and dissociation constants probably respond to different structures of networks of water molecules

In each of the examples this review describes, detailed comparisons of thermodynamic data for binding of ligand to protein, to data for partitioning of ligand from water to octanol, show different contributions from entropy and enthalpy (for identical, or closely related, ligands). We have, therefore, no reason to believe that partition constants describing the distribution of a hydrophobic ligand between a non-polar medium and water, and dissociation constants describing dissociation of that ligand from the non-polar cavity of a protein into water, involve the same structures of water.

Essentially all of the empirical and semi-empirical potential functions commonly used in computational approaches to estimate the solvation component of free energies of interaction in aqueous solution (e.g., PARSE, AMSOL, BIPSE etc.) employ terms that are derived from measurements of the solvation of small molecules in bulk water. These empirical potential functions thus model hydrophobic interactions for a process (oil-water partitioning) that may not necessarily correlate closely with the process of interest in molecular recognition (binding site-water partitioning). Our comparison of binding and oil/water partitioning support the analysis made by Dill and coworkers [13]: if the thermodynamic terms for solvation measured in bulk solution are different from those for solvation in the active sites of proteins (that is, if the molecular basis for the hydrophobic effect is different in the two cases), then we cannot expect these



empirical functions to provide accurate representations of the thermodynamics of solvation (partitioning between oil and water) and of binding to active sites.

## **B. Enthalpy and entropy are both important in hydrophobic binding (to different extents) depending on the topography/molecular details of the binding site and the ligand**

Entropy and enthalpy can *both* make important contributions to the free energy of hydrophobic interactions between proteins and ligands. The picture that is slowly emerging is that the magnitude of these contributions can be *very* different for related ligands binding to different active sites, or partitioning between environments of different hydrophobicity. Whether or not there are “rules of thumb”, or pictorial metaphors, that will aid (generally) in the design of ligands that bind tightly to proteins remains unclear. What is clear is that the old metaphors (i.e., “lock-and-key” and “ice-like water”) are at best incomplete pictures of protein-ligand binding, and at worst misleading in their simplicity, or simply wrong. The key points seem to be that water in the binding pocket, and around the ligand, is a (and perhaps, the) critical component of the problem, that every active site is unique in its structure and dynamics, and thus in the structure and dynamics of the water it contains. Looking for “rules of thumb” to guide the design of ligands that bind tightly to proteins may be difficult, or simply not possible. At present, ligand design must rather be informed by the most complete set of empirical data (from calorimetry, crystallography, and other biophysical techniques) and predictions (from molecular dynamics simulations that include water explicitly) possible. The problem of estimating the thermodynamics of protein-ligand binding seems to be one of adding large numbers of individually small terms; solving this type of problem requires quantitation.

## **C. What is the molecular basis for entropy-enthalpy compensation?**

The current answer to this question is, “We don’t know”. Our own work with carbonic anhydrase is leading us to look closely at the structure of the network of water molecules that hydrate the binding sites of proteins. In our example of benzo-extension, structural changes to the ligand lead to changes in free energy that are compatible with other observed hydrophobic effects, but suggest an unexpected (other than perhaps to theorists) origin of this hydrophobic effect—the displacement of *enthalpically* unfavorable waters by the benzo group [27]. In addition to predicting an enthalpically favorable hydrophobic effect, molecular dynamics simulations examining the enthalpy and entropy of the water molecules in the active site of CA also show compensating changes in the enthalpies and entropies of “some” of the molecules of water that are not displaced by the benzo group. It is difficult to generalize such observations to other active sites and to other ligands, but our observations are compatible with the hypothesis that enthalpy-entropy compensation arises—in some way—from interactions and organization of waters in cavities of proteins, rather than (as in the Dunitz model) from a tradeoff in entropy and enthalpy of interactions between ligand and protein. The key idea of the Dunitz proposal—that tight binding leads to enthalpic gain but entropic loss—still remains, however, the best available guiding principle in rationalizations of entropy-enthalpy compensation.

#### **D. The shape of the water droplet in the active site, rather than the shape of the active site, determines the hydrophobic effect**

Model systems of protein-ligand binding (e.g., the binding of normal alcohols to major urinary protein studied by Homans, the binding of arylsulfonamides to CA studied by our group, and others), in which a physical-organic approach to the hydrophobic effects responsible for binding can be rationalized, and for which there are complementary sets of data on the thermodynamics of ligand binding and structures of the protein-ligand complex, support a “water-centric” mechanism for the hydrophobic effect. In this mechanism, the enthalpy and entropy of individual molecules of water within the binding pocket determine the strength of binding because these molecules are displaced into the bulk upon ligand binding. Interactions directly between protein and ligand at least in some cases may be less important than the release of free-energetically unfavorable water.

The few proteins that have, so far, produced interpretable data argue strongly that hydrophobic effects result from differences in the structure of water in the binding pocket, around the ligand and in bulk water, and from the release of water in the binding pocket and around the ligand into the bulk on association of the protein and ligand. What is unclear is the role of water in proteins that undergo significant conformational changes upon ligand binding—an extreme example being intrinsically disordered proteins ( $\sim 25\%$  of the proteins within the cell contain an intrinsically disordered region [16]). These systems, while complicated by the entanglement of the “folding” and “binding” problems, offer a unique opportunity for the physical-organic approach to provide interpretable experimental results in systems operating (perhaps) by principles different than those characterizing simple, rigid proteins and ligands.

#### **E. What can studies of molecular recognition in typical non-aqueous solvents (e.g., MeOH, $\text{CH}_2\text{Cl}_2$ , etc.) teach us about molecular recognition in water?**

The properties of water, as a liquid, are very different than organic solvents. If—as we believe—the properties of water dominate many protein-ligand binding events, then studies of molecular recognition in organic solvents will hold few useful lessons for our understanding of molecular recognition in water.

#### **F. Assuming that hydrophobic effects are a substantial part of the free energy of association of proteins and ligands, what do we need to learn about them to be able to predict the structure of tight-binding systems?**

Detailed thermodynamic analysis will be an important part of the path forward in rational ligand design, but it is not sufficient. What is needed, we believe, is not simply more data. What is needed (at least in part) is more *interpretable* data. There are at least five considerations for obtaining interpretable sets of data: i) selection of good model systems that are minimally complicated by the structural dynamics of proteins and ligands, ii) characterization of the thermodynamics of protein-ligand binding by calorimetry, iii) rationalization of thermodynamics of binding with biostructural data from X-ray crystallography (and, ideally, from neutron diffraction) [166,167], and nuclear magnetic resonance spectroscopy, and iv) comparison of those data to the

estimates of binding free energies made by computational analyses that include water explicitly, and iv) modification of the theories applied to the computations to address the differences between computation and experiment. Bringing together these data, for most research groups, will require the close collaboration between physical-organic chemists, protein biochemists, structural biologists, biophysicists and computational chemists.

**Note added in proof:** Since this paper was submitted, a number of new, relevant papers have appeared in the literature. Among these, are the following: i) Two studies focused on the binding of a series of peptidomimetic ligands to thermolysin [1, 2] found that the interactions between binding pocket of the protein and the hydrophobic portion of the ligand are mediated by a network of crystallographically resolved waters that completely surround the ligand [1]. Increasing the hydrophobic portion of the ligand causes the network of waters surrounding the ligand to increase in size, but does not result in significant changes in the free energy of binding. Increasing the size of the ligands does result in large, compensating changes in the enthalpy and entropy of binding [2], and support the hypothesis that increasing the network of waters in the active site of a protein-ligand complex can be an important contributor to the thermodynamics of binding. ii) Biela *et al.* prepared a series of ligands to displace increasing numbers of water from the binding pocket of thrombin, and found that increases in the surface area or volume of the ligand did not result in gains of the free energy of binding, which was relatively unchanged because of enthalpy-entropy compensation [3]. This result suggests that displacing molecules of water from the binding site (with ligands of increased bulk) is not, by itself, a sufficient strategy to design tightly binding ligands, and that the interplay among protein, ligand, and water must be further dissected before we can “rationally” design these intermolecular interactions. iii) Shivakumar *et al.* calculated the absolute solvation energies of 239 small neutral molecules with explicit solvent molecular dynamics free energy perturbation simulations, and found high correlation between the coefficient of determination for the calculated and experimental absolute free energies of solvation ( $R^2 = 0.95$ ) [4]. Their approach, although computationally expensive, is the most accurate currently available to calculate solvation free energies.

1. A. Biela, N.N. Naseif, M. Betz, A. Heine, D. Hangauer, G. Klebe, *Angew. Chem. Int. Ed.* **52**, 1822 (2013)
2. N.N. Naseif, H. Tan, J. Kong, D. Hangauer, *J. Med. Chem.* **55**, 8283 (2013)
3. A. Biela, F. Sielaff, F. Terwesten, A. Heine, T. Steinmetzer, G. Klebe, *J. Med. Chem.* **55**, 6094 (2013)
4. D. Shivakumar, E. Harder, W. Damm, R.A. Friesner, W. Sherman, *J. Chem. Theory Comput.* **8**, 2553 (2012)

This work was partially funded by the National Science Foundation (NSF CHE-1152196), and the Wyss Institute for Biologically Inspired Engineering. The authors would like to acknowledge our many collaborators who have contributed to our evolving view of the role of water in protein-ligand interactions: Dr. Benjamin Breiten, Dr. Patrick R. Connelly (Vertex Pharmaceuticals), Dr. Annie Heroux (Brookhaven National Laboratory), Dr. Vijay Krishnamurthy, Dr. Heiko Lange, Dr. Eric T. Mack, Dr. Mathai Mammen, Dr. Woody Sherman (Schrödinger, Inc), and Prof. Eric Toone (Duke University).

## References

1. H.S. Frank, M.W. Evans, *J. Chem. Phys.* **33**, 507 (1945)
2. W. Kauzmann, *Adv. Prot. Chem.* **14**, 1 (1959)
3. C. Tanford, *Science* **200**, 1012 (1978)

4. C. Tanford, Proc. Natl. Acad. Sci., U.S.A. **76**, 4175 (1979)
5. G. M. Whitesides, E. Simanek, J. Mathias, C. Seto, D. Chin, M. Mammen, D. Gordon, Acc. Chem. Res. **28**, 37 (1995)
6. M. Barboiu, J. Lehn, Proc. Natl. Acad. Sci., U.S.A. **99**, 5201 (2002)
7. H. Schneider, Angew. Chem., Int. Ed. **30**, 1417 (1991)
8. J. Rebek, Acc. Chem. Res. **23**, 399 (1990)
9. K.N. Houk, A.G. Leach, S.P. Kim, X. Zhang, Angew. Chem., Int. Ed. **42**, 4872 (2003)
10. Y.K. Cheng, P.J. Rossky, Nature **392**, 696 (1998)
11. P. Ball, *Life's Matrix: a Biography of Water* (University of California Press, Los Angeles, 2001)
12. P. Ball, Chem. Rev. **108**, 74 (2008)
13. N.T. Southall, K.A. Dill, A.D.J. Haymet, J. Phys. Chem. B **106**, 521 (2002)
14. M.C. Chervenak, E.J. Toone, J. Am. Chem. Soc. **116**, 10533 (1994)
15. W. Blokzijl, J.B.F.N. Engberts, Angew. Chem., Int. Ed. **32**, 1545 (1993)
16. P. Tompa, K.H. Han, Phys. Today **65**, 64 (2012)
17. V. Uversky, Protein Sci. **11**, 739 (2002)
18. S. Lee, D. Kim, J. Han, E. Cha, J. Lim, Y. Cho, C. Lee, K.H. Han, Curr. Protein Pept. Sci. **13**, 34 (2012)
19. J.R. Beasley, D.F. Doyle, L. Chen, D.S. Cohen, B.R. Fine, G.J. Pielak, Proteins: Struct. Funct. Bioinf. **49**, 398 (2002)
20. A. Cornish-Bowden, J. Biosci. **27**, 121 (2002)
21. K. Sharp, Protein Sci. **10**, 661 (2001)
22. R.R. Krug, W.G. Hunter, R.A. Grieger, J. Phys. Chem. **80**, 2335 (1976)
23. R.R. Krug, W.G. Hunter, R.A. Grieger, Nature **261**, 566 (1976)
24. O. Exner, S. Bohm, New J. Chem. **32**, 1449 (2008)
25. C. Clarke, R.J. Woods, J. Gluska, A. Cooper, M.A. Nutley, G.J. Boons, J. Am. Chem. Soc. **123**, 12238 (2001)
26. M.S. Searle, M.S. Westwell, D.H. Williams, J. Chem. Soc., Perkin Trans. 2 **2**, 141 (1995)
27. P.W. Snyder, J. Mecinovic, D.T. Moustakas, S.W. Thomas, M. Harder, E.T. Mack, M.R. Lockett, A. Heroux, W. Sherman, G.M. Whitesides, Proc. Natl. Acad. Sci. U.S.A. **108**, 17889 (2012)
28. V.M. Krishnamurthy, B.R. Bohall, V. Semetey, G.M. Whitesides, J. Am. Chem. Soc. **128**, 5802 (2006)
29. V.M. Krishnamurthy, G.K. Kaufman, A.R. Urbach, I. Gitlin, K.L. Gudiksen, D.B. Weibel, G.M. Whitesides, Chem. Rev. **108**, 946 (2008)
30. D. Ford, J. Am. Chem. Soc. **127**, 16167 (2005)
31. J.D. Dunitz, Chem. Biol. **2**, 709 (1995)
32. D.H. Williams, E. Stephens, D.P. O'Brien, M. Zhou, Angew. Chem., Int. Ed. **43**, 6596 (2004)
33. J.D. Dunitz, A. Gavezzotti, Angew. Chem., Int. Ed. **44**, 1766 (2005)
34. P. Buchanan, N. Aldiwan, A.K. Soper, J.L. Creek, C.A. Koh, Chem. Phys. Lett. **415**, 89 (2005)
35. P. Buchanan, A. Soper, H. Thompson, R. Westacott, J. Creek, G. Hobson, C. Koh, J. Chem. Phys. **123**, 164507 (2005)
36. J. Turner, A.K. Soper, J. Finney, Mol. Phys. **70**, 679 (1990)
37. W. Jorgensen, Science **254**, 954 (1991)
38. S. Pal, J. Peon, A.H. Zewail, Proc. Natl. Acad. Sci., U.S.A. **99**, 1763 (2002)
39. T. Arikawa, M. Nagai, K. Tanaka, Chem. Phys. Lett. **457**, 12 (2008)
40. S. Balasubramanian, S. Pal, B. Bagchi, Phys. Rev. Lett. **89**, 115505 (2002)
41. Y. Levy, J. Onuchic, Ann. Rev. Biophys. Biomol. Struct. **35**, 389 (2006)
42. A. Der, L. Kelemen, L. Fabian, S. Taneva, E. Fodor, T. Pali, A. Cupane, M. Cacace, J. Ramsden, J. Phys. Chem. B **111**, 5344 (2007)
43. M. Grossman, B. Born, M. Heyden, D. Tworowski, G. Fields, I. Sagi, M. Havenith, Nat. Struct. Mol. Biol. **18**, 1102 (2011)

44. E. Eisenmesser, D. Bosco, M. Akke, D. Kern, *Science* **295**, 1520 (2002)
45. X. Jordanides, M. Lang, X. Song, G. Fleming, *J. Phys. Chem. B* **103**, 7995 (1999)
46. N. Nandi, K. Bhattacharyya, B. Bagchi, *Chem. Rev.* **100**, 2013 (2000)
47. K. Bhattacharyya, *Acc. Chem. Res.* **36**, 95 (2003)
48. L. Palmer, J. Rebek, *Org. Biomol. Chem.* **2**, 3051 (2004)
49. J. Hansen, E. Pines, G. Fleming, *J. Phys. Chem.* **96**, 6904 (1992)
50. A. Douhal, *Chem. Rev.* **104**, 1955 (2004)
51. T.M. Truskett, P.G. Debenedetti, S. Sastry, S. Torquato, *J. Chem. Phys.* **111**, 2647 (1999)
52. S. Matysiak, P.G. Debenedetti, P.J. Rossky, *J. Phys. Chem. B* **115**, 14859 (2011)
53. T.M. Truskett, K.A. Dill, *J. Chem. Phys.* **117**, 5101 (2002)
54. F. Weinhold, *J. Chem. Phys.* **109**, 373 (1998)
55. N. Kobko, L. Paraskevas, E. del Rio, J. Dannenberg, *J. J. Am. Chem. Soc.* **123**, 4248 (2001)
56. A. Masunov, J. Dannenberg, *J. Phys. Chem. B* **104**, 806 (2000)
57. A. Surolia, N. Sharon, F.P. Schwarz, *J. Biol. Chem.* **271**, 17679 (1996)
58. G. Bradbrook, J. Forshaw, S. Perez, *Eur. J. Biochem.* **267**, 4545 (2000)
59. G.A. Jeffrey, *Acc. Chem. Res.* **2**, 344 (1969)
60. J.H. Hildebrand, *Proc. Natl. Acad. Sci., U.S.A.* **76**, 194 (1979)
61. K.A. Sharp, J.M. Vanderkooi, *Acc. Chem. Res.* **43**, 231 (2010)
62. F.H. Stillinger, *J. Solut. Chem.* **2**, 141 (1973)
63. K. Lum, D. Chandler, J.D. Weeks, *J. Phys. Chem. B* **103**, 4570 (1999)
64. L.R. Pratt, A. Pohorille, *Chem. Rev.* **102**, 2671 (2002)
65. D. Chandler, *Nature* **437**, 640 (2005)
66. D. Chandler, *Nature* **445**, 831 (2007)
67. L.M. Salonen, M. Ellermann, F. Diederich, *Angew. Chem., Int. Ed.* **50**, 4808 (2011)
68. P.D. Ross, S. Subramanian, *Biochemistry* **20**, 3096 (1981)
69. W.P. Jencks, *Catalysis in Chemistry and Enzymology*, 1st edn. (McGraw-Hill, New York, 1969)
70. J. Gordon, W.P. Jencks, *Biochemistry* **2**, 47 (1963)
71. D.B. Smithrud, T.B. Wyman, F. Diederich, *J. Am. Chem. Soc.* **113**, 5420 (1991)
72. K. Silverstein, A. Haymet, K.A. Dill, *J. Am. Chem. Soc.* **120**, 3166 (1998)
73. K.A. Dill, T.M. Truskett, V. Vlachy, B. Hribar-Lee, *Ann. Rev. Biophys. Biomol. Struct.* **34**, 173 (2005)
74. T. Urbic, V. Viachy, K.A. Dill, *J. Phys. Chem. B* **110**, 4963 (2006)
75. R.U. Lemieux, *Acc. Chem. Res.* **29**, 373 (1996)
76. B.A. Williams, M.C. Chervenak, E.J. Toone, *J. Biol. Chem.* **267**, 22907 (1992)
77. J.E. Ladbury, J.G. Wright, J.M. Sturtevant, P.B. Sigler, *J. Mol. Biol.* **238**, 669 (1994)
78. P. Liu, X. Huang, R. Zhou, B.J. Berne, *Nature* **437**, 159 (2005)
79. R. Zhou, X. Huang, C.J. Margulis, B.J. Berne, *Science* **305**, 1605 (2004)
80. N. Giovambattista, C.F. Lopez, P.J. Rossky, P.G. Debenedetti, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2274 (2008)
81. R. Kumar, J. Schmidt, J. Skinner, *J. Chem. Phys.* **126**, 204107 (2007)
82. F. Weinhold, C.R. Landis, *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective* (Cambridge University Press, New York, 2005)
83. A. Reed, L. Curtiss, F. Weinhold, *Chem. Rev.* **88**, 899 (1988)
84. G.N.I. Clark, C.D. Cappa, J.D. Smith, R.J. Saykally, T. Head-Gordon, *Mol. Phys.* **108**, 1415 (2010)
85. H. Bakker, J. Skinner, *Chem. Rev.* **110**, 1498 (2010)
86. G.L. Richmond, *Annu. Rev. Phys. Chem.* **52**, 257 (2001)
87. J. Sorenson, G. Hura, R. Glaeser, T. Head-Gordon, *J. Chem. Phys.* **113**, 9149 (2000)
88. G. Hura, J. Sorenson, R. Glaeser, T. Head-Gordon, *J. Chem. Phys.* **113**, 9140 (2000)
89. A. Soper, *Chem. Phys.* **258**, 121 (2000)
90. V. Bezzabotnov, L. Cser, T. Grosz, G. Jancso, Y. Ostanevich, *J. Phys. Chem.* **96**, 976 (1992)

91. S. Rajamani, T. Truskett, S. Garde, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9475 (2005)
92. Y. Rezus, H. Bakker, *Phys. Rev. Lett.* **99**, 148301 (2007)
93. A. Bakulin, C. Liang, T. Jansen, D. Wiersma, H. Bakker, M. Pshenichnikov, *Acc. Chem. Res.* **42**, 1229 (2009)
94. S. Dixit, J. Crain, W. Poon, J. Finney, A. Soper, *Nature* **416**, 829 (2002)
95. Q. Du, R. Superfine, E. Freysz, Y.R. Shen, *Phys. Rev. Lett.* **70**, 2313 (1993)
96. E. Raymond, T. Tarbuck, M. Brown, G. Richmond, *J. Phys. Chem. B* **107**, 546 (2003)
97. C. McFearn, D. Beaman, F. Moore, G. Richmond, *J. Phys. Chem. C* **113**, 1171 (2009)
98. L.F. Scatena, M.G. Brown, G.L. Richmond, *Science* **292**, 908 (2001)
99. J. Tyrrell, P. Attard, *Phys. Rev. Lett.* **87**, 176104 (2001)
100. N. Ishida, T. Inoue, M. Miyahara, K. Higashitani, *Langmuir* **16**, 6377 (2000)
101. R. Steitz, T. Gutberlet, T. Hauss, B. Klösgen, R. Krastev, S. Schemmel, A.C. Simonsen, G.H. Findenegg, *Langmuir* **19**, 2409 (2003)
102. A. Poynor, L. Hong, I. Robinson, S. Granick, Z. Zhang, P. Fenter, *Phys. Rev. Lett.* **97**, 266101 (2006)
103. C. McFearn, G.L. Richmond, *J. Mol. Liq.* **136**, 221 (2007)
104. Q. Du, E. Freysz, Y. Shen, *Phys. Rev. Lett.* **72**, 238 (1994)
105. Y. Shen, V. Ostroverkhov, *Chem. Rev.* **106**, 1140 (2006)
106. B.J. Berne, J.D. Weeks, R. Zhou, *Annu. Rev. Phys. Chem.* **60**, 85 (2009)
107. X. Huang, R. Zhou, B. Berne, *J. Phys. Chem. B* **109**, 3546 (2005)
108. S. Huang, B. Sjöblom, A.E. Sauer-Eriksson, B.H. Jonsson, *Biochemistry* **41**, 7628 (2002)
109. J. Rasaiah, S. Garde, G. Hummer, *Annu. Rev. Phys. Chem.* **59**, 713 (2008)
110. A. Wallqvist, B. Berne, *J. Phys. Chem.* **99**, 2885 (1995)
111. T. Lazaridis, *J. Phys. Chem. B* **102**, 3542 (1998)
112. T. Lazaridis, *J. Phys. Chem. B* **102**, 3531 (1998)
113. J. Kim, P.S. Cremer, *Phys. Chem. Phys.* **2**, 543 (2001)
114. X. Chen, S.C. Flores, S.M. Lim, Y. Zhang, T. Yang, J. Kherb, P.S. Cremer, *Langmuir* **26**, 16447 (2010)
115. G. Kim, M.C. Gurau, J. Kim, P.S. Cremer, *Langmuir* **18**, 2807 (2002)
116. X. Chen, T. Yang, S. Kataoka, P.S. Cremer, *J. Am. Chem. Soc.* **129**, 12272 (2007)
117. A.J. Hopkins, C.L. McFearn, G.L. Richmond, *Curr. Opin. Solid State Mater. Sci.* **9**, 19 (2005)
118. D. Bearman, E. Robertson, G.L. Richmond, *J. Phys. Chem. C* **115**, 12508 (2011)
119. M. Quillin, P. Wingfield, B. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19749 (2006)
120. A. Amadasi, J. Surface, F. Spyralis, P. Cozzini, A. Mozzarelli, G. Kellogg, *J. Med. Chem.* **51**, 1063 (2008)
121. G. Kellogg, D. Abraham, *Eur. J. Med. Chem.* **35**, 651 (2000)
122. M. Raymer, P. Sanschagrin, W. Punch, S. Venkataraman, E. Goodman, L. Kuhn, *J. Mol. Biol.* **265**, 445 (1997)
123. A. Garcia-Sosa, R. Mancera, P. Dean, *J. Mol. Model.* **9**, 172 (2003)
124. M. Verdonk, J. Cole, R. Taylor, *J. Mol. Biol.* **289**, 1093 (1999)
125. G. Rossato, B. Ernst, A. Vedani, M. Smiesko, *J. Chem. Inf. Model.* **51**, 1867 (2011)
126. D. Beglov, B. Rioux, *J. Phys. Chem. B* **101**, 7821 (1997)
127. D. Chandler, J. McCoy, S. Singer, *J. Chem. Phys.* **85**, 5971 (1986)
128. D. Chandler, J. McCoy, S. Singer, *J. Chem. Phys.* **85**, 5977 (1986)
129. P. Goodford, *J. Med. Chem.* **28**, 849 (1985)
130. K. Appelt, R. Bacquet, C. Bartlett, C. Booth, S. Freer, M. Fuhry, M. Gehring, S. Herrmann, E. Howland, E. Janson, T. Jones, C. Mohr, E. Moomaw, S. Oatley, R. Ogden, M. Reddy, S. Reich, W. Schoettlin, W. Smith, M. Varney, J. Villafranca, R. Ward, S. Webber, K. Welsh, J. White, *J. Med. Chem.* **34**, 1925 (1991)
131. M.J. Jedrzejewski, S. Singh, W.J. Brouillette, G.M. Air, M. Luo, *Proteins: Struct., Funct., Genet.* **23**, 264 (1995)
132. H. Wallnoefer, K. Liedl, T. Fox, *J. Chem. Inf. Model.* **51**, 2860 (2011)
133. C. Barillari, J. Taylor, R. Viner, J. Essex, *J. Am. Chem. Soc.* **129**, 2577 (2007)



134. A. Abel, T. Young, T. Farid, B. Berne, R. Friesner, *J. Am. Chem. Soc.* **130**, 2817 (2008)
135. R. Abel, N.K. Salam, J. Shelley, R. Farid, R.A. Friesner, W. Sherman, *Chem. Med. Chem.* **6**, 1049 (2011)
136. T. Beuming, R. Farid, W. Sherman, *Protein Sci.* **18**, 1609 (2009)
137. C. Higgs, T. Beuming, W. Sherman, *ACS Med. Chem. Lett.* **1**, 160 (2010)
138. K. Kusano, J. Suurkuusk, I. Wadso, *J. Chem. Thermodyn* **5**, 757 (1973)
139. P. Privalov, S. Gill, *Pure Appl. Chem.* **61**, 1097 (1989)
140. V. Payne, N. Matubayasi, L. Murphy, R. Levy, *J. Phys. Chem. B* **101**, 2054 (1997)
141. S. Gill, N. Nichols, I. Wadso, *J. Chem. Thermodyn.* **7**, 175 (1975)
142. S. Gill, N. Nichols, I. Wadso, *J. Chem. Thermodyn.* **8**, 445 (1976)
143. P. Gilli, L. Pretto, V. Bertolasi, G. Gilli, *Acc. Chem. Res.* **42**, 33 (2009)
144. A. Ben-Naim, J. Wilf, *J. Phys. Chem.* **84**, 583 (1980)
145. A. Ben-Naim, Y. Marcus, *J. Chem. Phys.* **81**, 2016 (1984)
146. A.V. Plyasunov, E.L. Shock, *Geochim. Cosmochim. Acta* **64**, 439 (2000)
147. W. Riebesehl, E. Tomlinson, *J. Solution Chem.* **15**, 141 (1986)
148. *CRC Handbook of Chemistry and Physics*, 91 edn. (CRC Press, Cleveland, 2010)
149. T.S.G. Olsson, M.A. Williams, W.R. Pitt, J.E. Ladbury, *J. Mol. Biol.* **384**, 1002 (2008)
150. E.A. Meyer, R.K. Castellano, F. Diederich, *Angew. Chem., Int. Ed.* **42**, 1210 (2003)
151. Piatnitski, E., Flowers, R., Deshayes, K. *Chem. Eur. J.* **6**, 999 (2000)
152. M. Rekharsky, Y. Inoue, *Chem. Rev.* **98**, 1875 (1998)
153. N. Shimokhina, A. Bronowska, S.W. Homans, *Angew. Chem., Int. Ed.* **118**, 6522 (2006)
154. R. Malham, S. Johnstone, R.J. Bingham, E. Barratt, S.E.V Phillips, C.A. Laughton, S.W. Homans, *J. Am. Chem. Soc.* **127**, 17061 (2005)
155. S.W. Homans, *Drug Discovery Today* **12**, 534 (2007)
156. T.G. Oas, E.J. Toone, *Adv. Biophys. Chem.* **6**, 1 (1997)
157. L. Liu, Q. Guo, *Chem. Rev.* **101**, 673 (2001)
158. E. Gallicchio, M. Kubo, R. Levy, *J. Am. Chem. Soc.* **120**, 4526 (1998)
159. T.S.G. Olsson, J.E. Ladbury, W.R. Pitt, *Protein Sci.* **20**, 1607 (2011)
160. A. Lee, S. Kinneer, A.J. Wand, *Nat. Struct. Mol. Biol.* **7**, 72 (2000)
161. M.S. Marlow, J. Dogan, K.K. Frederick, K.G. Valentine, A.J. Wand, *Nat. Chem. Biol.* **6**, 352 (2010)
162. B. Carrington, R. Mancera, *J. Mol. Graphics Model.* **23**, 167 (2004)
163. C. Chang, W. Chen, M. Gilson, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1534 (2007)
164. P.K. Agarwal, S.R. Billeter, P.T.R. Rajagopalan, S.J. Benkovic, S.J.S. Hammes-Schiffer, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2794 (2002)
165. S.K. Nair, T.L. Calderone, D.W. Christianson, C.A. Fierke, *J. Biol. Chem.* **266**, 17320 (1991)
166. S.Z. Fisher, A.Y. Kovalevsky, J.F. Domsic, M. Mustyakimov, R. McKenna, D.N. Silverman, P.A. Langan, *Biochemistry* **49**, 415 (2010)
167. S.Z. Fisher, C.M. Maupin, M. Budayova-Spano, L. Govindasamy, C. Tu, M. Agbandje-McKenna, D.N. Silverman, G.A. Voth, R. McKenna, *Biochemistry* **46**, 2930 (2007)
168. J. Mecinovic, P.W. Snyder, K.A. Mirica, S. Bai, E.T. Mack, R.L. Kwant, D.T. Moustakas, A. Heroux, G.M. Whitesides, *G J. Am. Chem. Soc.* **133**, 14017 (2011)
169. V.M. Krishnamurthy, V. Semetey, P.J. Bracher, N. Shen, G.M. Whitesides, *J. Am. Chem. Soc.* **129**, 1313 (2007)
170. R. Talhout, J.B.F.N. Engberts, *Eur. J. Biochem.* **268**, 1554 (2001)
171. R. Talhout, A. Villa, A.E. Mark, J.B.F.N. Engberts, *J. Am. Chem. Soc.* **125**, 10570 (2003)
172. J. Gelp, S. Kaiko, X. Barril, J. Cirera, X. Cruz, F. Luque, M. Orozco, *Proteins: Struct., Funct., Bioinformat.* **45**, 428 (2001)
173. L. Wang, B.J. Berne, R.A. Friesner, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1326 (2011)