The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery

The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery

Francesca Spyrakis[†], Mostafa H. Ahmed[‡], Alexander S. Bayden[§◊], Pietro Cozzini[¶], Andrea Mozzarelli[*#^], and Glen E. Kellogg[*‡]

[†] Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria 9, 10125 Torino, Italy

[‡] Department of Medicinal Chemistry & Institute for Structural Biology, Drug Discovery and Development, Virginia Commonwealth University, Richmond, Virginia 23298-0540, United States

[§] CMD Bioscience, 5 Science Park, New Haven, Connecticut 06511, United States

[¶] Dipartimento di Scienze degli Alimenti e del Farmaco, Laboratorio di Modellistica Molecolare, Università degli Studi di Parma, Parco Area delle Scienze, 59/A, 43121 Parma, Italy

[#] Dipartimento di di Scienze degli Alimenti e del Farmaco, Laboratorio di Biochimica, Università degli Studi di Parma, Parco Area delle Scienze, 23/A, 43121 Parma, Italy

^ Istituto di Biofisica, Consiglio Nazionale delle Ricerche, Via Moruzzi 1, 56124 Pisa, Italy

Abstract. The value of thoroughly understanding the thermodynamics specific to a drug discovery/design study is well known. Over the past decade, the crucial roles of water molecules in protein structure, function and dynamics have also become increasingly appreciated. This Perspective explores water in the biological environment by adopting its point-of-view in such phenomena. The prevailing thermodynamic models of the past, where water was seen largely in terms of an entropic gain after its displacement by a ligand, are now known to be much too simplistic. We adopt terminology that describes water molecules as being "hot" and "cold", which we have defined as easy and difficult to displace, respectively. The basis of these designations, which involve both enthalpic and entropic water contributions, are explored in several classes of biomolecules and structural motifs. The hallmarks for characterizing water molecules are examined and computational tools for evaluating water-centric thermodynamics are reviewed. The Perspective's summary features guidelines for exploiting water molecules in drug discovery.

1. INTRODUCTION

Structure-based drug discovery/design (SBDD) has been shown to be a reasonably successful strategy for identifying and optimizing lead compounds of therapeutic importance. SBDD has not proven, however, to be the panacea that had been envisaged at its inception[1,2] nor in the successes[3-5] and hype[6-10] at the turn of the millennium. It has turned out to be much more difficult than anyone expected at the beginning to exploit an X-ray crystal structure for a substrate- or ligand-bound or native protein for designing, synthesizing and implementing a high efficacy small molecule. The problems inherent in turning an inhibitor into a drug are manifold[11] and beyond the scope of this Perspective.

There are a large number of contributing factors to the "failures" in SBDD, not the least of which was unrealistic expectations from the technology – especially at the time. Only a very small fraction of the potential targets had been structurally characterized and solving X-ray crystal structures were difficult, often multi-year, endeavors. In that light, Kuhn *et al*.[2] described a number of cutting edge innovations: 1) molecular biology tools such as parallel expression to alleviate the bottleneck of obtaining suitable protein samples, which enabled the exploration of multiple constructs, homologs and variants for specific protein targets; 2) new developments in robotics and associated apparatus to process crystallization trials more efficiently, i.e., in a high-throughput (HT) manner; and 3) diffraction data collection at synchrotron beamlines using flash-cooling, robotics and automated and better structure solution methods for structure determination, e.g., single/multiple-wavelength anomalous dispersion (SAD/MAD) techniques. This article proved to be prophetic as the technologies mentioned have come to the forefront and the number and quality of protein crystal structures has soared in the last fifteen years, reaching about 110,000 in 2016. The contributions of combinatorial and parallel syntheses must also be acknowledged: having much larger sources of putative ligands at our disposal is also clearly a boon to drug discovery,[12,13] although their drug-likeness has been questioned.[14] With those really big problems behind us, it would seem to be a simple matter to effectively apply physical principles of shape and electrostatic complementarity to SBDD in many biological problems, right? Not so fast! There are numerous other issues at play beyond matching shape and character

between a small molecule and binding site. Nearly all of these issues can be summarized in one word: **Thermodynamics!**

There are many moving parts in a protein-small molecule binding event – there can be, beyond the main actors, cofactors and solvent molecules (perhaps hundreds) that must be considered. The protein and small molecule are never totally rigid bodies, and often possess significant flexibility. Also, as stated by Dill,[15] the formation of protein-ligand, protein-protein and protein-nucleic acid complexes are not simply easy to dissect two body interactions. **Everything** is in motion, which is an inseparable and key part of thermo*dynamics*. Thus, each member of the biomolecular milieu may contribute statically and/or dynamically to the thermodynamics of the binding event (*vide infra*). Numerous articles and reviews have been published over the past few decades to illuminate the energetics of protein-small molecule binding. The reader is directed to a few.[16-18] A large subset of such studies have commented on and often focused on the energetic roles played by water molecules in such associations.[18-20] Klebe[21] and Wang[22] demonstrated that at least one water molecule is involved in the binding in about two-thirds of protein-ligand associations.

Our understanding of water in the biological environment *has* evolved over the past decade or so, but the terminology, even in key articles, is not 100% consistent and the conclusions can be seemingly contradictory. In our view, one large problem is that these articles are **not** generally written for medicinal chemists interested in drug discovery. The goal of this Perspective is to fill this need with useful advice on one very significant topic: when and how to design molecules that effectively interact with or displace water molecules in an active site or at a protein-protein interface.

1.1. A water-centric approach. This Perspective is largely written from the point-of-view of a water molecule. Almost nothing occurs in a biological event that doesn't involve water molecules playing one or (usually) multiple important roles in the process. Clearly, within a protein, a water molecule is better than an empty space.[23] While the protein and ligand(s) are the star players in a drug binding association, the actions of water molecules, whether as individuals or in bulk, require detailed attention to really understand the process. Moreover, when considered collectively, their influence multiplies and cannot be ignored or trivialized. Water molecules usually have a significant, if not dominant, contribution to the free energy of ligand-macromolecule binding usually associated with their regaining the entropy lost by the ligand

when it bound. This interpretation is seen in the hydrophobic effect and the release of protein-bound water molecules, but is an oversimplification of the mechanism and its energetics.

Perhaps counterintuitively, the water issue that has garnered the most attention in medicinal chemistry is actually what happens when they are *removed*.[24] There are really two water displacement problems in medicinal chemistry: i) displacing waters from empty sites by inserting ligands and ii) displacing waters from sites already ligand-bound by growing chemical substituent groups onto those ligands. The first problem is somewhat easier to solve because most empty sites are located in hydrophobic pockets, and the waters within are likely labile. They are easy to displace and will generally provide an entropic contribution. Water molecules found in a ligand-bound pocket will likely be in hydrophilic loci and predicting their displacement is not obvious. This is typically the problem encountered in lead optimization, of which the development of the cyclic inhibitors of human immunodeficiency virus 1 (HIV-1) protease from about two decades ago is a classic example of water as a player in drug discovery.

Acquired immune deficiency syndrome (AIDS) is caused by the HIV-1. By the late 1980s, AIDS was considered an epidemic, and while nucleoside reverse transcriptase inhibitor treatments such as zidovudine (ZDV), also known as azidothymidine (AZT) (approved 1986), were efficacious with largely manageable side effects, it was widely perceived that another target protein was needed.[25] To that end, HIV-1's retroviral aspartyl protease was crystallized and solved by the Wlodawer group in 1989,[26] and this ushered in an era of intense interest in SBDD, especially with respect to this protein.[27] Many inhibitory compounds were designed based on this structure and an impressive number of co-crystal structures were solved and reported in just a few years. One particularly interesting series of compounds for this story was the cyclic urea class developed by the DuPont Merck Pharmaceutical Co.[28-30] In addition to an open conformation, the unbound protease's (pdb: 1g6l[31]) active site (Figure 1A) contains several water molecules that are nominally displaced by the incoming ligands, a few that adapt to support ligand binding,[33,34] and another that is tightly bound and generally retained. Water "**300**", found between the two catalytic aspartyls,[32] is an easy target for displacement, as are a handful of other water molecules in the open pocket of the unbound HIV-1 protease. Most interesting, however, is water "**301**", which bridges between the isoleucines 50 and 50' and non-cyclic (peptidomimetic) inhibitors such as CGP 53820 (pdb: 1hih[29,30,35])(Figure 1B). Water **301** was not displaced until a concerted effort in drug design was undertaken to do so.[28] Thus, one part of the interesting story of the cyclic inhibitors such as AHA001 is the sulfamide oxygens or urea carbonyls that very effectively mimic

the hydrogen bond acceptor ability of the tightly bound and structurally key water **301** (pdb: 1ajx[36]) (Figure 1C). While that particular water molecule and the inspired drug design that displaced it have places in history, what were the real thermodynamic consequences? Water **301** was only one of several water molecules in the unbound protein's active site; what energetic roles did the others play? With the computational tools and understanding of today, in contrast to 20+ years ago, is rational water targeting in drug discovery/design better developed? To discuss these issues is the goal of this Perspective. In short, is water a largely untapped resource for drug discovery?

2. BIOMACROMOLECULAR STRUCTURE FROM WATER'S POINT OF VIEW

Water has a seemingly very limited set of tools to work with: just two hydrogen bond donors, its protons, and two hydrogen bond acceptors, its lone pairs. Clearly, all water molecules are identical and we will not learn much by trying to study the water molecules themselves. Instead, we have to query their environments in order to extract useful information. In this section of the Perspective, we are examining protein structure and chemistry from the water molecules' point of view. Our thesis is that a deeper understanding of water-involved phenomena from this viewpoint will illuminate the much more complex events of ligand binding and other biological macromolecule associations, which we begin to focus on in Section 3. First, we will explore the roles that water plays in the structure and (bio)chemistry of macromolecular systems. Next, we will explore how to evaluate the relationship between each water molecule and its environment, and discuss how such an understanding is the first step to exploiting water molecules for drug design.

2.1. Roles of Water in Structure. It is well known that water is fundamental to life at every scale of biology, from *subnano* to *macro*. In particular, many reviews have described protein structure and highlighted the contributions of water. Here, we briefly review the structural roles of water, but as mentioned above, are presenting the myopic view that the water molecules have of their structural roles: 1) water molecules are often found "standing in" for other functional groups (for a few examples, see Figure 2A). These functional groups can be "missing" from the protein(s), ligand(s) or other cofactors; 2) water molecules have a simple role as extenders –

adding a water to an acidic or basic residue functional group can project that same property up to about 3 Å away (see Figure 2B). They are often found associated with aspartate residues arranged in such a way that together they are "isosteric" to glutamate residues. Such a role is particular important in protein-polynucleotide recognition (see 4.2.5.); 3) similarly, a water has the ability to invert a residue or any functional group from an "acid" to a "base" or vice versa (see Figure 2C). We have previously proposed that water molecules act as "nanobuffers" because of their ability to adapt to pH changes by simply rotating;[ref 167] 4) lastly, water molecules fill in voids. Water molecules are found abundantly in unoccupied binding or active sites, but also can be trapped as a consequence of protein folding and assembly (Figure 2D), at protein-protein interfaces (Figure 2E) or in other scenarios. Nature generally fills most such voids with water molecules, even when their presence seems counterintuitive on energetic grounds (*vide infra*). Another, more dynamic, role for water is its influence on translational diffusion dynamics around protein surfaces [D. Russo, G. Hura, T. Head-Gordon. Hydration Dynamics Near a Model Protein Surface Biophysical J. 2004, 86, 1852–1862], which has a potentially large influence on protein function and protein-protein recognition.

2.2. Roles of Water in (Bio)chemistry. Water is, of course, both a media with unique properties and a potent chemical reactant. A number of those roles are discussed here.

2.2.1. Water as Solvent. While water is not really the "universal solvent", it does dissolve more substances than any other liquid. It is highly polar and thus able to charge-separate many ion pairs and solubilize inorganic salts (see Figure 2F). Only medium to highly hydrophobic substances do not have some water solubility. Also, because of its very high dielectric (relative permittivity, ε_r) of around 80, water modulates charge-charge interactions, including hydrogen bonding, to energetically accessible – and potentially reversible – ranges. It is highly unlikely that anything remotely similar to our notion of "life" would have developed in vacuum, where every interaction would be eighty times stronger.

2.2.2. Water as Reactant. Another of water's remarkable properties is it is both a weak acid and a weak base. Water can be both a proton donor and a proton acceptor to strong or weak inorganic acids and bases, to a host of other organic functional groups that are either acidic or basic, and even to other water molecules (amphoterism). Water is thus both a weak nucleophile and a weak electrophile. In essence, it is ready for just about anything! Reactions that involve water under biological conditions are termed hydrolysis and the enzymes that perform such

reactions are hydrolases. Prominent among this enzyme class are: esterases such as nucleases, phosphodiesterases, lipase and phosphatase; DNA glycosylases, glycoside hydrolase, proteases and peptidases. Lastly, water is consumed and formed, respectively, in perhaps the most important reactions in biology:

$$ATP + H_2O \rightarrow ADP + H_2PO_4^- + \textbf{ENERGY}$$
$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$$

The ATP hydrolysis (see Figure 2G) is *so* exothermic that it can be coupled with thermodynamically unfavorable reactions to give an overall favorable (negative) free energy for many biologically important reaction sequences.

2.2.3. Water as Catalyst. As a special case, water may be considered to be a catalyst as well as nucleophile and solvent in the same reaction. Such spontaneous reactions are pH-independent hydrolytic processes that involve transfer of ≥ 1 proton(s) to or from water molecules in the transition state of the rate-determining step. Here, one water molecule is a nucleophile while others are generally performing as bases. Well-studied examples include the hydrolysis of alkyl halides, alkyl and aryl sulphonates, esters and amides where water is a nucleophile.

2.2.4. Water as Lubricant. As water is by definition **not** hydrophobic, it can act as a lubricant to interfere with hydrophobic interactions in hydrophobic environments (see Figure 2H), which is kind of a stunning role reversal for polar and hydrophobic species! Also, water molecules, because of their ability to facilely form, break and reform hydrogen bonds: 1) promote fast conformational fluctuations as residues in unstructured proteins "flicker" between structured regions in Ramachandran space at picosecond rates,[37] a time scale that is similar to rearrangements in the bulk water hydrogen bond network. Solvent water molecules thus hydrate residue-residue interactions at various conformational transitions (see Figure 2I); and 2) new concepts for friction and lubrication are being applied to molecular machinery possessing hydrogen-bonded components. Adding small amounts of water accelerates the relative motion of components, but the addition of other protic liquids has much weaker or even opposite effects.[38] Key is water's ability to participate in evolving three-dimensional hydrogen-bond networks between critical parts of the machine.

2.2.5. Diffusion and Water Wires. Even in bulk, the structure of water is constantly changing as its hydrogen bonds (up to four per water) rapidly break and reform with different

partners at fairly low energy cost; this continues even when other polar (hydrogen bond donor or acceptor) moieties are involved. This reduces barriers for molecular motion. Water has been suggested by Kier, Cheng and Testa to structure itself around protein surfaces in a way that facilitates ligand diffusion to binding loci.[39] Similarly, proton hopping (the Grotthuss mechanism) through a water network (Figure 2J) is why the conductivity of water is relatively high. It has been proposed[40] that proton hopping is key in the functioning of membrane channels and axon nerve conduction, and may even offer an explanation for the initiation of nerve impulses following an effector-ligand encounter.

2.2.6. The Hydrophobic Effect. This alone is the subject of many books, but to keep it simple here, hydrophobic entities are nonpolar groups, which means they generally do not possess lone pairs or protic hydrogens and are not amenable to interactions with water. The hydrophobic effect arises because water molecules are driven to self-associate (Figure 3A) as well as associate with other chemical groups possessing protic hydrogens or lone pairs (Figure 3B). The driving force for this association is quite high, and nonpolar groups or species are excluded from, or may be encapsulated by, an evolving extended network of water (Figure 3C) and/or polar functional groups. This, in turn, forces hydrophobic moieties in the milieu together (Figure 3D). For example, the hydrophobic cores of most soluble proteins are the consequence of the more polar and water soluble sidechains being solvated by bulk water on the proteins' exteriors, resulting in the apolar sidechains being forced inside the protein to avoid interaction with the water. Much debate has ensued over the years concerning terminology to describe this phenomenon: while clearly there is no "hydrophobic bonding", in this scenario is there actually a "force"?

It is simple to argue that the van der Waals' (London Force) attraction (Figure 3E) supports the notion of a hydrophobic force because the ideal atom-atom distance for a hydrophobic interaction as seen in X-ray crystal structures exhibiting this structural feature is essentially the same as the sum of atomic van der Waals' radii. On the other hand, there is generally a small positive partial atomic charge (relative to the generally electronegative heteroatoms) on hydrophobic functional groups like methyls, etc. that is normally manifested with the proton charges > 0 and carbon charges < 0, with a net positive sum. This seems to indicate that, from a Coulombic pairwise atom-atom view (Figure 3F), hydrophobic interactions are repulsive! Thus, application of "first principles" physics with atom-atom pairwise potentials cannot account for the hydrophobic *effect*. The hydrophobic interaction is clearly an *emergent property*. To date, no physics-based forcefield calculation, i.e., with all-atom molecular dynamics, has simulated the

effect. Instead, more phenomenological approaches are usually used to represent and quantitate these effects (*vide infra*). Despite interactions between water molecules being the root cause of the hydrophobic effect, its observed effect on structure may not explicitly include them.

2.2.7. Desolvation from Water. Related to the hydrophobic effect is what might be thought of as its flipside – desolvation. Discrete molecules in water solution have many specific interactions with individual water molecules, are encaged by water molecules, or are subjected to some combination of these two extremes (Figure 4A). Encagement serves the best interests of the network of water and polar moieties because the resulting "complex" would be polar and very water-like. The role and point-of-view of the water molecules is simple: in solution each water molecule's role is to hydrogen bond with a polar atom of a molecule or with another water in the most energetically favorable way. Analogously, water molecules "fill" hydrophobic pockets, tunnels, etc. of proteins, and are largely stabilized by inter-water interactions,[refs] which would be particularly strong in low dielectric hydrophobic loci. Other water molecules make direct hydrogen bonding interactions with surface or pocket residues (Figure 4B).

However, upon the interaction of two solvated species, such as a ligand binding to a protein (Figures 4C and 4D), the cage-like portions of the networks and the unanchored waters in hydrophobic regions of proteins are generally stripped away with ease. These actual interactions were mostly *between* water molecules. All other "sterically redundant" water molecules are released to other roles, some energetically favorable for the resulting complex and others considerably less so. These phenomena define desolvation. A few water molecules, in either the protein site or associated with the incoming ligand, have interaction roles that must be reprised in the complex or require retention of that water molecule. This results in a potentially large number of resulting interactions between the species (Figure 4E) with great diversity.

All of the above are key points for later discussion, and are, in fact, the purpose of this Perspective. One construct that we believe is very useful for considering and discussing the energetics of water molecules is to classify them as *hot* or *cold* depending on their environments. A *hot* water is one that possesses relatively high internal energy, is mobile and not energetically stable in its environment. A *cold* water is then very much the opposite. Before we discuss hot and cold water further, we briefly present some important concepts and principles of thermodynamics.

3. THERMODYNAMICS OF WATER IN BIOLOGICAL SYSTEMS

3.1. Thermodynamic Laws. In Albert Einstein's words, "*A theory is the more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended is its area of applicability. Therefore the deep impression which classical thermodynamics made upon me. It is the only physical theory of universal content concerning which I am convinced that within the framework of the applicability of its basic concepts, it will never be overthrown.*"[41] The terminology of thermodynamics as it applies to water in the biological environment should be briefly reviewed.

3.1.1. Basic Thermodynamics in 60 Seconds. The Law of Thermodynamic Equivalence states that if two systems are in thermal equilibrium with a third, then they must be in thermal equilibrium with each other, which defines thermodynamic systems as possessing reflexive properties. The Zeroth Law is the basis for the scientific definition of temperature.

The Law of Conservation of Energy states that energy cannot be created or destroyed in an isolated system, i.e., $\Delta U = 0$, where ΔU is the total energy of the system. However, in a non-isolated (open) system, where energy and matter can flow from the system to the surroundings and vice versa, then the internal energy, U, can change. As we are interested in biological systems, it is convenient to define the additional state function, ***enthalpy***, which is a measurement of energy in a thermodynamic system equivalent to the total heat content of the system. In particular, enthalpy includes the internal energy, U, which is the energy required to create a system from scratch. It is defined as a state function because it depends only on the prevailing equilibrium state, i.e., its internal energy, pressure, and volume. The total enthalpy, H, of a system cannot be measured directly, but ΔH expresses system energy *changes* at constant pressure in many chemical, biological and physical experiments.

The Law of Irreversibility states that the entropy (S) of an isolated system is constant ($\Delta S = 0$) when a change is reversible and rises when a change is irreversible, i.e., spontaneous: $\Delta S > 0$. S is another state function and its increase is a measure of the degradation and dispersion of energy from the system to the universe. Energy always flows to achieve spatial homogeneity of matter, energy and temperature. This Law serves to define the concept of ***entropy***, which by definition measures how much thermal energy *cannot* be used to do work. Like enthalpy, it is usually more correct to describe changes in entropy (ΔS) rather than absolute entropy S. While entropy is commonly thought of as a measure of "disorder", this is just another way of defining irreversibility.

The Law of Absolute Zero states that the entropy of a system is zero at 0 K. In other words, the entropy of a system at absolute zero is zero, but it is impossible for any process, even ideal, to reach this absolute zero entropy in a finite number of steps. Reaching absolute zero temperature or entropy is thus impossible. While this is not practically important for most purposes of everyday thermodynamics, it does provide a reference point such that absolute entropy can be calculated.

The Gibbs energy is another state function also known as *free enthalpy*. It is the maximum work that may be obtained from a process. At the standard (T = 298 K) constant (i.e., isothermal) temperature and standard (1 atm) constant (i.e., isobaric) pressure, Gibbs energy is most often expressed as

$$\Delta G = \Delta H - T\Delta S.$$

It is evident that the value and the sign of ΔG depends of the algebraic signs of ΔH and ΔS and of the absolute temperature T. When $\Delta G > 0$ for a process (reaction), it is termed endergonic and external (non-pressure/volume) energy would have to be added to the system to make the reaction energetically possible; e.g., it may be enabled by coupling it with another reaction such that the *total* $\Delta G < 0$. In cases of a negative ΔG, the reaction is called exergonic and should be spontaneous until equilibrium is reached.

Clearly, the Gibbs energy is a gold standard metric for evaluating the merits of a particular reaction, and it can be measured directly because of the simple relation between the equilibrium constant (K) of a reaction and ΔG:

$$\Delta G = \Delta G^\circ + RT \ln \{ [\text{products}]/[\text{reactants}] \};$$
$$\text{at equilibrium: } \Delta G^\circ = -RT \ln K,$$

where R is the Boltzmann constant. Enthalpy can be measured by calorimetry and entropy is most often calculated from independently measured ΔG and ΔH, especially for systems – like protein-ligand complexes – that are difficult to obtain in large quantities. Thus, it is difficult to *predict* ΔG because to do so one would have to simultaneously predict ΔH and ΔS. But, more importantly, for a plethora of reasons, enthalpy and entropy are linked, and assigning the origin of an observation to be either enthalpic or entropic is not at all straightforward. One reason is enthalpy–entropy

compensation (see section 3.4), which has been suggested to result from an intrinsic property of hydrogen bonds[42] or is mechanically induced by solvent.[43] However, the compensation effect has been suggested to be an artifact of data from a limited temperature range or from a limited range for the free energies.[44,45] Importantly, the basis of compensation is in the isokinetic plot (i.e., ΔH vs. ΔS), which is unfit in principle for substantiation because ΔH vs. ΔS are, as mentioned above, measured in the same experiment.[46]

3.1.2. Chemical Potential. The chemical potential (or partial molar free energy) provides an alternative way of describing molecular energetics [Job, G.; Herrmann, F. Chemical potential–a quantity in search of recognition. European Journal of Physics. 2006, 27 (2): 353–371; Simonson, T. The Physical Basis of Ligand Binding, in In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications, Cavasotto, C. N., Ed., CRC Press, Boca Raton, Florida, 2016, pp. 3-44.]. It is potential energy that can be absorbed or released during a chemical reaction, e.g., as in ligand binding. The chemical potential of a particular species (like water molecules) in a mixture is the partial derivative of the free energy with respect to the amount of that species, while all other concentrations remain constant. Thus, at equilibrium, the total sum of chemical potentials is zero because the system's free energy is at a minimum value. Also at equilibrium, the chemical potentials of all identical participants – such as water molecules – should be identical. However, the process of ligand binding, and the behavior of the involved molecules, especially the waters, is a perturbation of the system and changes chemical potentials as we describe in Sections 4.1 and 4.2.

3.1.3. Applying Structural Data. Our knowledge of chemical reactions like ligand binding, which mostly involve non-bonded interactions, is most often just one or two snapshots (usually X-ray crystal structures) – before and after. (And the "before" data, when available, can be very misleading due to the likely-to-come structural rearrangements.) Unwinding the reaction can be like a detective arriving at a murder scene and sifting through the various clues. In this Perspective we explore the role of water molecules in protein-ligand complexes, and especially, in using them as tools for designing ligands with improved efficacy and properties. The difficulties of studying the thermodynamics of ligand binding are multiplied many-fold when we attempt to focus on one or a handful of water molecules of relevance to the reaction amongst the many thousands associated with each protein and/or ligand in the pre-interaction milieu. Also unfortunate is that we can't experimentally label a single water so that we can trace its behavior amongst all of the others. Nevertheless, there is a sizeable body of literature on water in biological

systems. We will first report what is known experimentally about the thermodynamics of such water molecules, and then will survey some of the vast collection of theoretical work from the past several decades.

3.2. What we *really* know. Despite all that is known about thermodynamics, it is quite difficult to carry out definitive experiments on the role of water molecules in biomolecular systems because water is also the solvent and in overwhelming abundance. In addition, very few experiments operate on a *single* molecule rather than a very large collection of them. Even the smallest crystals contain around 10^{12} complete protein molecules that are time and space averaged in their X-ray diffraction patterns. Cryo-electron microscopy samples molecules one at a time, and are approaching atomic resolution, but the final structure models are constructed from data of thousands to millions of identical frozen molecules that have been selected, aligned, and averaged to create high-resolution maps. It is thus almost impossible to study isolated water molecules in binding sites. Changing the medium to another solvent is also futile since this almost always affects macromolecule shape.

Experiments from the late 1970s highlighted the role of water in molecular recognition with organic host-guest complexes.[47-49] A variety of host compounds with multiple centers capable of interacting in complementary ways with numerous guests were synthesized, and evaluated in multiple solvents. Host-guest complexation was found to be particularly favorable in aqueous solutions,[49] thus showing for the first time that water plays a major role in molecular recognition. This was attributed to "hydrophobic effects", caused by a release of water molecules and a concomitant gain in entropy, ΔS°. This view of the role of water in molecular recognition was not challenged until 1988, when Ferguson *et al*. showed that complexation of benzene in a cyclophane host molecule is *enthalpy* driven, and accompanied by a slightly *negative* entropic change.[50] Studies by Harata *et al*.[51] of α-cyclodextrins showed similar results. Also important result from solution studies was that Raman scattering[52] confirmed the presence of bifurcated hydrogen bonds to water, i.e., a single water molecule may interact with five neighboring waters.[53] Such penta-coordinated water molecules cause defects in the tetrahedral hydrogen bond network preferred by liquid water, and lowers the energy barrier between different arrangements of the network. Introducing an inert hydrophobic solute, however, prevents water molecules in its hydration shell from interacting with a fifth neighbor and slows the involved water molecules' molecular motions, which ultimately impacts the system's thermodynamic characteristics.

3.2.1 Structural insight. By the mid-to-late 1990s, the emphasis in thermodynamic studies shifted to exploiting structural data. The rapidly increasing availability of X-ray (and neutron) diffraction-solved protein structures, many at atomic or near-atomic resolution, multidimensional NMR data,[54] which can with considerable effort yield information on water molecules in protein structures (and, in solution rather than solidified in crystals), and data from increasingly accurate isothermal titration calorimetry (ITC) experiments,[55] yielded an more detailed understanding of the roles of water in molecular recognition. Ringe showed with protein crystals soaked in organic solvents that active site water molecules of elastase are easily displaced, even compared to those on the protein surface.[56] These active site water molecules were also more disordered, and thus abide by balancing their entropies of liberation and enthalpies of removal. One set of protein X-ray crystal structures that highlighted the multiple roles of water molecules in protein-ligand binding is that of the bacterial oligopeptide binding protein (OppA),[57,58] which transports small peptides into gram-negative cells. This protein has to bind tightly to very chemically diverse ligands in order to fulfill its biological role, i.e., it requires binding plasticity in order to accommodate a large variety of ligand shapes and chemical types. The combination of X-ray crystallography and ITC showed that water molecules play a central role in peptide ligand binding by OppA.[59] Each peptide requires the presence of a specific number of water molecules in the active site; their behavior changes the pocket volume, shields electric charge due to their high dielectric, and by simply reorienting changes the site's hydrogen-bonding pattern. Thus, even a small set of water molecules can produce a major impact by weakening unfavorable charge-charge interactions and by preserving hydrogen bonding. Also, *water burial* near the ligand can greatly *favor* binding; in contrast to the then prevailing dogma that the entropy gain from water displacement was the primary energetic contribution to improved binding free energy.

While one cannot label a single water in order to quantitate its contributions, nor forcefully include or remove a water from a binding site except through chemically modifying the protein or ligand, studies that combined structural and calorimetric analyses have shed some light on hydration energetics. Ladbury's 1996 review[23] first brought into focus the importance of water in structure-based drug discovery, as well as understanding the associated thermodynamics. In the mid-1990s data were not yet available concerning the thermodynamic impact of incorporating water molecules into protein-ligand interactions, but related data suggested that such an inclusion could increase the affinity of the ligand for the protein. Upper limits for enthalpies (ΔH° = -3.8 kcal mol^{-1}) and entropies ($T\Delta S^\circ$ = -2.1 kcal mol^{-1}) for transfer of a single water molecule from bulk

solvent to an interface in crystalline salt hydrates,[60,61] yields an estimated free energy change (ΔG°) of -1.7 kcal mol^{-1}. Cooper's 2005 analysis of the growing body of direct experimental calorimetric data for protein folding, protein–protein, protein–ligand interactions, etc.[62] showed that, depending on the number of formed hydrogen bonds, each water site that is made unavailable yields a ΔH° contribution of -1.5 to -3.0 kcal mol^{-1} and a $T\Delta S^\circ$ contribution of about -0.8 kcal mol^{-1}, i.e., with ΔG° -0.7 to -2.2 kcal mol^{-1}. In aldose reductase, sorbitol was found to bind with nearly equal affinities to both the wild-type (-9.0 kcal mol^{-1}) and Leu300Pro (-9.2 kcal mol^{-1}) protein variants.[63] One water molecule bridging between sorbitol and the wild-type enzyme is not found in the proline mutant – probably because it was associated with the leucine's amide proton, which is no longer available. Sorbitol binding to the wild-type more favorable enthalpically by 1.2 kcal mol^{-1}, but more expensive entropically by 1.4 kcal mol^{-1} compared to binding to the mutant enzyme. Thus, retaining the water molecule in the wild-type protein's binding site has an enthalpic advantage that compensates for the entropic cost of constraining it, while the mutant gains in entropy by releasing the water but loses the enthalpy attributable to the hydrogen bond.

The caveat that quantitation of thermodynamic effects is case specific should be looked at in more depth. First, the conditions (temperature, solvent, etc.) under which structural and calorimetric studies are performed differ significantly. Second, completely attributing measured thermodynamic parameters for binding as a consequence of water is only possible if the state of *all* involved molecules does not change, which is not the case if the basis of the experiment is a chemical modification of either the protein or the ligand. Third, inclusion of water molecules in structural models is usually in late-stage crystallographic refinement and is generally based on data resolution, whether they are in the first hydration layer of the protein (i.e., directly interacting), and/or whether their inclusion improves the apparent fit of density data. Thus, care must be exercised in drawing conclusions about water in crystallographic models,[23] especially if the structural resolution is suboptimal.

3.3. What we *think* we know. Because of the importance of understanding water roles in structure and thermodynamics, and given the issues inherent in experimental studies, this is an ideal problem for computational work. The realization that water molecules had substantive roles in protein structure, and could be computationally studied, was emerging in the late 1960s and early 1970s. In his 1972 paper, *Protein Folding*,[64] Tack Kuntz noted: "Since unfilled proton-donor or proton-acceptor sites are energetically unfavorable, we presume that these sites are actually

occupied by water molecules, suggesting that solvent participation is an important influence in turn generation or stabilization." Later, Kuntz and Kauzmann offered a few proposals for predicting protein hydration based on the amino acid composition of the protein.[65] Monte Carlo simulations by Hagler and Moult in 1978[66] of the water structure around biological molecules revealed that the "ordering" of water molecules is greatest near the protein and decreases with distance from it. Also significant was very early work by Levitt.[67,68] Cho, Singh and Robinson (1996)[69] published *Liquid Water and Biological Systems: The Most Important Problem in Science That Hardly Anyone Wants to See Solved*, and provided a water model that gave the first correct temperature-dependent density behavior for liquid water. When replica exchange molecular dynamics and explicit all-atom water were used in the 2003 simulation of the 10–55 helical fragment B of protein A (*Staphylococcus aureus*), water molecules were seen to play dynamic roles in protein folding.[70] Further enhancements in algorithms and hardware over the past 10-15 years have made explicit solvent simulations the standard practice. Anton, the massively parallel supercomputer built by D. E. Shaw Research specifically for molecular dynamics is notable.

3.3.1. (Mostly) rigorous physics: FEP, TI and LIE. The most rigorous methods of predicting relative free energy of protein-ligand binding are free energy perturbation (FEP) or thermodynamic integration (TI) that explicitly consider solvent molecules and flexibility.[71,72] FEP methods can very accurately predict differences in binding affinity for structurally similar ligands, as in lead optimization. Deconvoluting free energy into enthalpy and entropy terms, and for each ligand functional group, unravels the various contributions to the binding process. Solvation free energy calculations evaluate desolvation energetics, i.e., relating structural changes to these effects. McCammon[73] and Kollman[74,75] showed that FEP calculations captured the influence of desolvation with respect to ligand binding to trypsin and thermolysin. The semi-empirical linear interaction energy method (LIE) also requires explicit water to estimate absolute binding free energies as the difference between two simulations: the ligand in water (free state) and the ligand bound to the protein and surrounded by water (bound state).[76]

3.3.2. Computational focus on water. These computational analyses can also be applied to water molecules themselves. A wide and surprising range of results are found, from water molecules tightly bound and possessing favorable binding free energy to water molecules with clearly unfavorable binding free energy and perhaps mysterious structural roles. The very early molecular dynamics (MD) study on protein-ligand interactions in cytochrome P450cam by Helms and Wade,[77] following up Wade's earlier GRID analysis of the same protein,[78] showed that a

hydrated cavity in the protein accepts a water with $\Delta G = -2.8$ kcal mol^{-1}, while an unhydrated cavity has an unfavorable solvation ΔG of +3.8 kcal mol^{-1}. This work concluded that "… an empty interfacial cavity can be tolerated and suggest that such cavities are likely to be unhydrated unless the surrounding protein and ligand atoms have the capacity to make more than one hydrogen bond to a water molecule in the cavity". Our understanding of the vast and diverse roles that water molecules play in structure and energetics has significantly changed since then! The reader should be aware, however, that free energies calculated for the same water in the same crystal structure could vary greatly depending on the method used and its parameterization: Hamelberg and McCammon calculated the binding free energy of water 301 in the HIV-1 protease/KNI272 complex as -3.3 kcal mol^{-1},[79] while Essex's group calculated it at -10.0 kcal mol^{-1}.[80]

The binding free energies of water molecules depend, foremost, on their environment. In fact, the free energy of binding for the same crystallographic water can vary greatly in response to different ligands. The calculated free energies of one of the waters in the various crystal structures of OppA were between -3.9 and -10.4 kcal mol^{-1}.[80] Strongly bound waters are mostly found in polar cavities, where they form 3+ hydrogen bonds with the protein, while waters that are loosely bound are mostly found in environments that are at least partially apolar and form < 3 hydrogen bonds. For the latter scenario, there are now many known cases where the free energies of binding for individual crystallographic water molecules have been calculated to be positive, e.g., in the structure of neuraminidase-inhibitor complexes,[80] in apolar cavities of the subtilisin Carlsberg-eglin C complex,[81] and in a mutant of barnase.[82] Parsing free energy into enthalpic and entropic terms reveals that both can vary over a wide range, possibly from favorable to unfavorable, depending on context. Huggins predicted with the statistical mechanical inhomogeneous fluid solvation theory (IFST) method (see section 5.3) that the entropy contributions ($T\Delta S$) to hydration free energy are in the range of -0.46 to -2.67 kcal mol^{-1}.[83] To sort some of this out, Yu and Rick[84] examined the effects of cavity size and properties on the free energy, entropy, and enthalpy change for transferring a water molecule from bulk to the cavity with TI calculations. Both the size and the number of available hydrogen bonding partners – to simulate hydrophobic and polar environments – for a water molecule inside the cavity were varied. Increasing the number of hydrogen bonds results in free energy of hydration decreases (from +4.57 kcal mol^{-1} with zero hydrogen bonds to -2.66 kcal mol^{-1} with four) because the enthalpy gained from each additional hydrogen bond ($\Delta H \sim$ -3.8 kcal mol^{-1}) exceeds its lost entropy ($-T\Delta S \sim$ +2.0 kcal mol^{-1}). Cavity volume changes were seen by Yu and Rick to produce a

much less significant effect (see section 4.3.8). In Section 4, below, we will describe how protein structures exemplify this range of water energetics.

3.4. Enthalpy-entropy compensation. Enthalpy–entropy compensation (H/SC) is a specific compensation effect where a series of closely related chemical reactions exhibit a linear relationship between their enthalpy (ΔH_i) and entropy ($T\Delta S_i$) changes, i.e., ΔH scales proportionately with ΔS, and enthalpy and entropy compensate for each other because they possess opposite algebraic signs in the Gibbs equation. The compensation is a consequence of the temperature dependence possessed by both ΔH and $T\Delta S$ in weakly interacting systems.[45,61,85] H/SC is a wide-ranging phenomenon and is particularly relevant to drug design efforts such as lead optimization, where relatively small substituent changes are made.[86-88] A common theme is that binding affinity, related to ΔG, cannot in general be correlated to either ΔH or $-T\Delta S$.[88] In recent studies, water molecules participating in active site interactions have demonstrated H/SC in terms of their contributions to the overall thermodynamics. Two types of water H/SC have been identified. In the first, tighter binding of a ligand (or water) into a site, yielding a larger favorable enthalpy, imposes a more severe constraint on that entity, yielding a correspondingly less favorable (and compensating) entropic change. In one example, KNI-10033, a potent pM affinity HIV-1 protease inhibitor was further optimized by replacement of its thioether with a sulfonyl group (KNI-10075) that produces an additional hydrogen bond with Asp 30' of the protease.[87] This changes the binding enthalpy by -3.9 kcal mol^{-1}, but is completely compensated by a loss of entropy by the more constrained sulfonyl ligand KNI-10075, and there is no net change in affinity due to a combination of conformational[89] and solvation effects.

Whitesides and collaborators discovered the second type of water H/SC in detailed analyses of human carbonic anhydrase.[90] We want to reserve a deeper exploration of this case for later in the Perspective (section 6.2), but in brief: a series of closely related ligands bind to human carbonic anhydrase (HCA) with similar affinity but quite different enthalpies and entropies. Systematic fluorination of the sulfonamide ligands changed both the ligands' orientations and the structure of the water network, with concomitant binding enthalpy changes that are more or less mirrored by binding entropy changes of the opposite sign. It was proposed that the changes in water network structures – resulting *from* the ligand binding – are key determinants in ΔH_{bind} and $T\Delta S_{bind}$.[90] In a more general sense, H/SC can be the result of any ligand changes that subtly weakens the direct ligand-protein interactions, but simultaneously allows more freedom of

motion, i.e., entropy, either in the ligand or in the solvent. (See Figure 4F.) Enthalpy-entropy compensation is believed to exist over a relatively narrow energy range;[61,91] thus, if ligand changes severely impact the protein-ligand interactions, H/SC will not be able to recover the lost free energy.

4. THERMODYNAMICS FROM WATER'S POINT OF VIEW: HOT AND COLD WATER

Proteins, excepting those that are membrane-bound, like to be in water and water molecules like to interact with these proteins. Indeed, protein folding is dictated by the desolvation of apolar amino acid residues, forcing them to form the protein core, as well as by solvation of polar residues on the protein surface. Other water molecules play specific structural roles as described above contributing to protein flexibility, which is both inherent and functionally essential. As we have noted above, the strengths of water-protein interactions vary significantly. Some water molecules are loosely bound and easily displaced by either incoming ligands or protein conformational changes, while other water molecules are tightly bound, shaping active sites and protein structure, and mediating protein-ligand and protein-protein interactions. Therefore, the energetics of water-protein interaction exists on a continuum and is quite difficult to evaluate, either experimentally or computationally (see Computational Approaches, Section 5, below). In this section, we present numerous scenarios describing how and where weakly and tightly bound water molecules, which we are referring to as "hot" and "cold" water, respectively, can be found and how they can be characterized.

4.1. Hot waters. "*Hot*" water molecules are generally defined as those that are: have significant internal energy and/or are flexible, those that are not energetically stable in protein binding sites or at protein surfaces, and those mostly in contact with hydrophobic regions of their site. From the water-centric view, we might think of these waters as being "unhappy" molecules that are not very comfortable in their environment, and are thus waiting for opportunities to be displaced and/or moved to the water network in bulk. In bulk, such water molecules should be more easily able to meet what is their core obligation: maximizing their own hydrogen bonding. This model assumes a *favorable* change in binding free energy ($\Delta G < 0$) for these waters in moving

from the bound protein phase to the bulk and thus, correspondingly, their prior "freezing" in the protein matrix must have been accompanied by an *unfavorable* ΔG. While the enthalpic and entropic contributions to Gibbs energy largely compensate for each other, where they do not is where thermodynamics is most interesting. Clearly, ΔH and ΔS (and ΔG) are strongly dependent on the specifics of the protein environment, that is, the protein physical, chemical and geometrical properties. In the alternative formalism of chemical potential, some "hot" water molecules will have *significantly* increased chemical potential caused by the perturbation of ligand binding and are thus forced to move from their positions in order to decrease their chemical potential in the system's new equilibrium.

One view is that such hot waters are *only* filling voids, or at the extreme, that they do not even really exist and might be only crystallographic artifacts or misassigned electron density.[92,93] The latter hypothesis is simple to disprove: many hot waters are found in high-resolution X-ray structures and even in neutron diffraction structures, although the possibility of an occasional crystallographic error is real. As to whether such water molecules are simply filling voids, our arguments to follow should make it clear that hot waters play crucial roles in protein macromolecular structure. Their high entropic character makes them freer to move and, consequently, enhances the flexibility of the protein matrix that they occupy. Such structural roles call for hot waters to also have fundamental dynamic and mechanistic roles in the function and regulation of proteins.

Any water molecule in contact with a non-water and partially nonpolar surface can be considered a hot water, because water molecules have xenophobic character (see Figure 2). A layer of water in such contact is uncomfortable and is generally 40% less dense than the other water layers or than water in bulk, and is involved in the formation of fewer hydrogen bonds.[94-97] This phenomenon is called "dewetting", which at a macroscopic scale is the rupture of a thin liquid film on a substrate and the concomitant formation of droplets. Dewetting in biological systems occurs when the free energy of dewetting is more favorable than the free energy of solvation.[98-100] Dewetting free energy is a function of the surface's shape. For instance, concave surfaces have more favorable dewetting energies because water molecules can form fewer hydrogen bonds compared to water molecules on planar surfaces.[101-104] One of the features that makes the thermodynamics of water in biological systems so interesting – and difficult – is that no two situations are exactly the same. Each protein active site or interface has specific and unique characteristics of shape, hydropathy and electrostatics. The generation of hot waters thus comes

from many places. The remainder of this section describes several illuminating examples where we focus on the functions and mechanisms of action for hot waters.

4.1.1. Poorly solvated binding sites. Although it is a very attractive and commonly applied simplification of water thermodynamics, it is currently accepted that hydrophobic effects are **not** always the result of an entropic gain produced by the displacement of structured waters to bulk but, instead, by an *enthalpic* contribution favorable to the free energy of binding.[105] This is especially the case for proteins having hydrophobic binding sites that are poorly solvated by just a few water molecules. In these pockets, it is difficult to attach a significant entropic contribution to the reorganization of the few displaced waters upon ligand binding, so the overall thermodynamics must be enthalpy driven – a result of the displaced water molecules joining and gaining stabilization from the water network (Figure 5A). A very well studied case of this scenario is the mouse major urinary protein (MUP), a pheromone binding protein.

The energies of binding for a series of pyrazine-derived ligands have been thoroughly investigated by means of ITC, nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography and extensive MD simulations.[106] Upon binding of these ligands, all the waters solvating the binding cavity are displaced. While desolvation of the hydrophobic binding site and the ligand provides a favorable, albeit limited, entropic contribution, the "freezing out" of the ligand, i.e., its loss of conformational freedom, strongly counteracts this gain, producing in the end, a final negative (and unfavorable) $T\Delta S$. The major enthalpic contributions to the binding energy arise from protein desolvation (probably negligible, since there are few interactions between the pocket and water molecules), desolvation of hydrophobic small ligands (unfavorable, about +12 kcal mol^{-1}), and solute-solute interactions (favorable). However, as described above, the waters in the MUP binding site are few and disordered. Thus, the pocket desolvation allows these waters to form significantly more stable and fruitful interactions with the water bulk, which produces a favorable *enthalpic* contribution of about -6 to -12 kcal mol^{-1}. The unfavorable ligand desolvation enthalpy mostly compensates for this solvent-solvent effect. Altogether, the driving force for the binding is then due to solute-solute, that is, protein-ligand interactions.

In the usual cases, van der Waals', hydrophobic and hydrogen bond interactions formed within the binding site between protein and ligand are able to balance the lost solute-solvent interactions existing prior to the association. In mouse MUP, however, given the site's poor solvation, solute-solute dispersion interactions are not compensated in this way and they provide a notably favorable enthalpic contribution. To generalize, hydrophobic interactions in poorly

solvated hydrophobic sites are mainly driven by *enthalpy* rather than by the more expected *entropy*.[62] This may be referred to as a *non-classical* hydrophobic effect, but is mostly a manifestation of London forces. It was also previously observed for bovine serum albumin and ovoalbumine denaturation.[20] Note that the *hot* waters of this class are *losing* entropy and gaining enthalpy as they move from the somewhat non-restricting binding pocket into more locked positions in bulk.

Overall, we can differentiate between three different categories of an enthalpic-driven hydrophobic effect. The first, as above in the mouse major urinary protein, is driven by the formation of more favorable solute-solute interactions, in which poorly bound waters are displaced by solute, thus forming more stable contacts with the surrounding environment. This can be tracked for the mouse MUP when interacting with alcohols of increasing length,[19,107] and for carbohydrates binding to lectin.[108,109] The second category considers an enthalpic contribution arising from water-mediated solute-solute interactions (Figure 5B), as in the case of arylsulfonamide ligands binding to carbonic anhydrase (see section 6.2),[110] while the third is a gain associated with water reorganization within a binding site upon ligand binding (Figure 5C).

4.1.2. Non-solvated binding sites. Bovine β-lactoglobulin (BLG) represents a case of a *completely* dehydrated free binding pocket. BLG is part of the calycin superfamily, whose members all bind nonpolar molecules, much as does MUP. While in other members a variable number of waters is displaced upon ligand binding,[111,112] BLG apparently presents an empty pocket even when unliganded.[113] This cavity, also referred to as the calyx, shows a largely cylindrical shape, with a narrow entrance gauge of about 1.4 Å and a volume of 315 Å^3. It is an intriguing question to ask: what is the biological function or purpose in preserving totally empty cavities in folded proteins? In this specific case, because of the narrow entrance and the elongated channel shape, ligands would have to wait for the exit of all water molecules before being able to enter the cavity, which would have a very significant effect on the association rate and the resulting binding affinity. Thus, the *absence* of a stable hydration network within the cavity, which also could be described as a "void effect", facilitates the binding of non-polar ligands (Figure 5D).

Other well-studied protein cavities have been reported as empty and not hydrated. A classic example is the fully nonpolar (volume = 48 Å^3) cavity in T4 lysozyme, where no enclosed water molecules were located even after extensive crystallographic refinement.[114] Similarly, interleukin 1β presents five cavities, four containing one or two waters and a central pocket of about 40 Å^3 that is apparently empty and lined by non-polar side-chains.[115] And, while the hen

egg-white lysozyme contains three apolar cavities, with volumes ranging from 10 to 40 Å^3, none are occupied by a water molecule.[116]

In fact, the issue of protein cavities really being desolvated has been debated for decades in the literature.[117] The debate about *horror vacui* (Nature abhors a vacuum) started with Aristotle and was commented on by Sir Isaac Newton.[118] However, it is clear that moving a single water molecule from the bulk into a protein matrix should be highly unfavorable if most, if not all, of the water's hydrogen bonds are not restored.[119] Unsurprisingly, evidence of hydrophobic pockets being populated by single water molecules is fairly rare.[120,121] On the other hand, although placing a single water molecule in a small apolar cavity apparently yields only an unlikely and energetically unfavorable scenario for that water, in larger cavities – that are likely populated by small water clusters – the formation of intermolecular (inter-water) interactions can be energetically favorable and promote hydration. Still, no minimum cavity size, nor minimum number of clustered waters for self-stabilizing hydration, has been determined or reported. Altogether, in such cases of dehydrated cavities, the absence of water does not make the thermodynamics of ligand binding noticeably simpler to understand. Even when absent, the influence of water is inescapable!

4.1.3. Concave hydrophobic cavities. The relevance of site architecture in determining the solvation energy is well known. Narrow tubular hydrophobic sites have a much higher probability of undergoing dewetting events compared to larger but globular pockets, even when very hydrophobic.[98,122-126] Commonly, dewetting cavities are found to be confined and mostly hydrophobic. This is because in such spaces, water molecules have more difficulty in clustering and tightly networking with each other (Figure 6A-C). Moreover, in a water-centric view, the protein's surface topography is believed to strongly influence the hydrophobic effect, i.e., the free energy difference between water in bulk and water close to non-polar surfaces.[20] Looking at it in this way, higher shape and hydropathic complementarity between proteins and ligands does not only allow the formation of more hydrogen bonds or van der Waals' interactions, but also induces the release of a larger number of free-energetically unfavorable waters from the protein cavity and from the ligand surface.[20] Viewed in still another way, the structures of water networks (the "water shape") that are lying close to non-polar surfaces, and thus contributing to the free energy of binding in biomolecular interactions, are strongly influenced by the structure (shape) of the binding pocket and the ligand. The topography of the protein binding site, in fact, affects the nature and number of hydrogen bonds formed within the network of waters (Figure 6A and 6B)

and, thus, the change in free energy resulting from the binding of a ligand and the associated release of the waters into the bulk. For instance, waters close to concave surfaces, with respect to waters in the bulk, have quite unfavorable enthalpies of about +5 kcal mol^{-1}.[20] Because of this, concave hydrophobic surfaces apparently have "more hydrophobic" character than flat hydrophobic regions.[127]

Dewetting most often occurs as waters move from a protein binding site to the bulk.[128,129] However, in some cases, waters in channels of the protein matrix that are near the active site can be displaced, either into the pocket or transiting to bulk.[130-132] The substrate association reaction of thrombin, for example, has been shown to be enhanced by waters moving back and forth from the binding site to such an internal water channel.[133] This appears to be a new mechanism of dewetting, where a channel is acting as a reservoir for dewetted water molecules instead of those waters being dumped to bulk (Figure 6C). MD simulations[133] suggest that the water molecules within the channel are quite labile and regularly and rapidly exchanging with those in the pocket.

4.1.4. Hot water in GPCRs. The revolutionary crystallographic efforts that led to the resolution of about 160 G-coupled protein receptor (GPCR) structures in the last decade identified the constant presence of a significant but highly variable number of waters in the binding site and in the entire protein matrix.[134-141] Water molecules in this set play both "hot" and "cold" structural and biochemical roles. On the cold ledger, the importance of water molecules for proton transport in membrane-bound proteins such as bacteriorhodopsin has been revealed over the past two decades to be as important as that of the amino acids for proton transport and biological function.[142] This and other cold water molecule roles will be explored in section 4.2.6.

However, with their large crystallographic B-factors and/or poorly resolved electron density, water molecules near the activation switches and allosteric sites are very labile, i.e., "hot", and seem to be more involved in the mechanism of receptor activation than in ligand binding and stabilization. Their facile rearrangements appear to be associated with disruptions of GPCR internal hydrogen bonds and to subsequent structural changes between the inactive and active forms of these receptors. Simulations have suggested that, in the inactive state, the internal water networks are confined within regions separated by hydrophobic layers, while upon transition to the active form, e.g., by binding of agonists, a water channel connecting these regions is formed within each receptor (see Figure 7A). Yuan *et al*. recently investigated this mechanism by extensive molecular dynamics simulations performed on the adenosine A_{2A} receptor, the β_2-adrenergic receptor and rhodopsin.[140] The X-ray structure of the A_{2A} receptor, when complexed

with the antagonist ZM241385 (ZMA, 4-{2-[(7-amino-2-furan-2-yl[1,2,4]triazolo[1,5-a][1,3,5]triazin-5-yl)amino]ethyl}phenol),[137] showed two layers of hydrophobic residues between the orthosteric and allosteric sites and close to the conserved and key NPxxY motif, which has been implicated in activation switching. MD simulations supported the presence of hydrophobic layers separating these two regions as well as confirming the waters' mobility. While receptor metastates were obtained after dynamics on complexes of the A_{2A} and an agonist, simulations in the presence of both an agonist and the G-protein indicated a significant change in the water network, leading to the disappearance of the hydrophobic layers and to the concomitant formation of a long and continuous channel of waters connecting the orthosteric site with the cytoplasm. Also, the volume increases more than two-fold at the G-protein-binding site and the number of waters inside the receptor dramatically increases. The activation of this receptor class can be associated with the reorganization of internal hydrogen bonds involving the ligand, the receptor and especially the water molecules. A closer look at adenosine A_{2A} with respect to exploiting hot waters to identify druggable sites is presented in section 6.3.

Other recent studies have shown similar results. Comparisons between a new crystallographic structure of activated bovine rhodopsin and earlier models of inactive rhodopsin[143] found waters confined in discontinuous local areas in the transmembrane region of the inactive state but in a continuous water channel connecting the chromophore binding site to the G-protein site when activated by light stimulus, which triggers the *cis-trans* isomerization of 11-*cis* retinal to all-*trans* retinal.[144,145] Three regions in GPCRs host the majority of water molecules: the CWxP motif at the bottom of the ligand-chromophore binding pocket, the NPxxY motif at the intracellular end of TM7, and the (D/E)RY motif at the intracellular end of TM3. In this state, these three water groups allow communication between the "switch" components and act to couple the ligand binding site to the G-protein. The same solvent reorganization was also observed for the antagonist-and agonist-bound μ-opioid receptor (MOR).[146,147,148]

The highly mobile "hot" waters found in GPCRs appear to support receptor structure reorganization in several ways: 1) by interacting with crucial residues on the transmembrane helices such as tyrosine 7.53 (of the NPxxY), which undergoes well-documented conformational transitions between the inactive and the active receptor forms;[140] 2) once the transition has occurred, the newly-formed continuous water channel provides necessary receptor flexibility; 3) regulating the distance between reactive functional groups, such as Asp 181 and the Schiff base in rhodopsin;[142] and 3) water from the bulk can enter the receptor either from extracellular locations

or from near intracellular loops. Many of these latter water molecules (which may be "cold") will be able to improve their hydrogen bonding compared to outside the receptor. While it is clear that the mobile waters found in and around GPCRs support the protein's structure and function by facilitating the transition among its various conformational states, these are *very* complex processes that will require many more high-resolution structures, many more experiments and many more simulations before complete understanding can be obtained. Further, there are cold water roles playing out simultaneously in GPCRs (*vide infra*).

4.1.5. Hot water in ion channels. Like GPCRs, ion channels exploit the presence of crucial water molecules to properly function. Although the mobility of such waters, as indicated by diffusion coefficients, is up to an order of magnitude less than in bulk,[149,150] they move with ions and assist in passing them through the channels and cell membranes. In particular, the KcsA potassium ion channel has been well studied and is representative of the relevance of hot waters in ion channels. Hodgkin and Keynes (1955) hypothesized the "knock-on" model to explain potassium transfer in polarisable membranes. The model in this pioneering work considered ions as being constrained to move in a single file queue across narrow tubes or channels. Also, ions were described as moving together in the same direction and assisting each other in this directed movement.[151] It is important to point out that virtually no structural information was yet available to support this hypothesis. In 1998 the structure of the potassium ion channel from *Streptomices lividans* was solved by Doyle *et al*.[152] It showed a narrow selectivity filter 12 Å long, lined by carbonyls from the protein backbone and contained three K^+ ions with a water molecule lying between two of them. This configuration appeared to promote ion conduction by exploiting electrostatic repulsive forces between the K^+ ions to overcome the selectivity filter-ion attractive forces.[152]

Further crystallographic studies confirmed that water molecules often filled sites left empty by ions. Thus, both K^+-H_2O-K^+-H_2O and H_2O-K^+-H_2O-K^+ configurations for the ordering in the single file queue within the filter were equally probable. The K^+ occupancy at each of the four sites (S1-S4, Figure 7B) was estimated at 0.5. This maintained a charge balance in the filter: two ions separated by water are stable in the filter, while three would represent a transition state.[153] Further experimental and computational analyses supported this model of alternating water and potassium ions, which was first termed "knock-on" (Figure 7B),[154] but following electrophysiological measurements of the streaming potential of ions and waters through the KcsA channel by Iwamoto and Oiki, was revised and renamed the "permeation model".[155] Their

experiments suggested a flow-based model where the permeation is driven by the water flux across the channel and the number of ions in the channel at any one time is concentration dependent.

However, alternative models that were less reliant on water have continued to appear;[156,157] in particular, Kopfer *et al*. performed extensive MD-based calculations and highlighted Coulombic repulsion between adjacent ions as the main driving force for high-efficiency K^+ conduction,[157] which was termed the "hard-knock" model (see Figure 7B). The hard-knock model has no explicit role for water within the channel. To resolve this issue, a multidisciplinary team led by Perozo, Roux, Valiyaveetil and Zanni very recently combined 2D infrared (IR) spectroscopy and molecular dynamics simulations.[158] The ultrafast time resolution of 2D IR spectroscopy (1-2 ps) can probe short-lived configurations on a time scale compatible with MD simulations, and the resulting spectra report on the near-instantaneous positions and conformations of molecules, ions and solvent. Their key observation in this case was that, when waters were removed from the models, e.g., as in the hard-knock model, little agreement between simulated spectra and 2D IR data was seen. However, agreement between simulation and experiment was excellent for the knock-on model.

Waters seem to also play a crucial role also in regulating the recovery of ion channels from inactivation.[159] Reactivation of a channel is generally a quite slow ion-dependent mechanism, requiring 5-20 seconds after inactivation.[160,161] Three waters molecules buried behind the selectivity filter of each subunit have been proposed as key elements of the inactivation/activation mechanism. In the "pinched" inactive conformation these waters are stable and immobile, with the role of preventing spontaneous transition towards the conductive state. Only after a K^+ ion binds to the filter and the waters are released does the recovery process commence. In this case, the same water molecules are converted from cold to hot as needed by the hosting protein.

4.1.6. Hot waters at protein-protein interfaces. Protein-protein interfaces (PPIs) are an emerging area of interest in drug discovery because many biological processes involve at some point at least one protein-protein association that could, in principle, be mediated by a small molecule or peptidic agent. There are numerous roles that water molecules play in such associations: from the obvious solvation of the interacting proteins, and all that entails, to specific roles in facilitating complex formation, e.g., sidechain extenders, acid-base inverters, or filling in voids. (See Figure 6D.) Presumably, prior to the association, both proteins were well solvated, even at the putative interface surfaces, yielding a potentially large number of displaced water

molecules when the proteins come together. Depending on the specific pre-association engagement of these waters with their respective proteins or each other, the association from this displacement of waters likely provides a favorable entropic contribution to the overall energetics.

With respect to our definitions, the various roles of the water molecules found in the final complex range between hot and cold. The cold ones might be classified as those with multiple strong and favorable interactions to one or both protein actors, but that is a static view suggesting that the path to inhibiting the PPI is to *break apart* the already-formed complex. A more dynamic view of inhibition, more in terms of *preventing* occurrence of the association, might yield another set of definitions for hot and cold. However, the latter form of inhibition is much more difficult to execute, and neither appropriately configured and accessible computational technology nor adequate and extensible experimental structural data yet exist to routinely simulate such events. Regardless, the static approach and careful examination of protein-protein complex structures can yield a wealth of information. We discuss hot waters at protein-protein interfaces here, and cold waters in Section 4.2.4.

Numerous research teams have studied PPIs over the past few decades[162-164] and several have zeroed in on the water molecules found therein.[165,166] We looked at a large data set (179 high-resolution X-ray structures, 4741 water molecules) with a simple question: what fractions of the water molecules are truly bridging between proteins, associated with only one protein, or associated with neither.[167] To make these assignments, we used a metric, Relevance (section 4.3.7),[168] that assesses geometry and interaction quality for each water. Interestingly, of those water molecules at the interface, 21% were found to be bridging, 53% were found to be associated with only one protein and 26% were found to be associated with neither. This last category is perhaps the most intriguing and these waters would seem to fit our definition of "hot". Further characterization of their solvent-accessible surface areas, etc., revealed that, overall, 7.4% of interface water molecules are *well buried* at the PPI and yet make no favorable interactions. We termed these waters as being trapped in hydrophobic bubbles[167] and suggest that these bubble motifs may have a functional role in regulating PPIs, e.g., such hot-water instability is required to ensure that protein-protein associations are dynamic. Similar conclusions were previously reported,[169,170] i.e., suggesting a lubricant role for waters at PPIs. It was also suggested[171] that interfacial water networks can decrease the interaction strength of a protein-protein interaction when necessary, that is, when proteins need to rapidly assemble and disassemble to mediate

cellular events, and that water networks close to hydrophobic surfaces have enhanced fluctuation over water in bulk or solvating hydrophilic regions.[172,173]

We should point out – to avoid confusion – that a common strategy in targeting PPIs for inhibitor intervention is to first identify the so-called "hot spots", which are the strongest and most favorable interaction points, e.g., salt bridges or hydrogen bonds, between the two proteins. These interactions are then used to design potential inhibitors that would serve to disrupt one or more of these key protein-protein interactions. Our definition of "hot" waters refers to water molecules located in what would presumably be "cold spots", i.e., loci where the protein-protein association is, at best, weakly favorable.

Altogether, hot waters play a dynamic role in regulating protein function and functional protein-protein interactions. Their intrinsic site instability and their high entropic character are necessary for protein movement and flexibility. More generally, solvent exposure ensures lability, as evidenced by the higher evolutionary rate of solvated residues.[174]

4.2. Cold Waters. Since virtually the onset of structural biology, it was recognized that water molecules are conserved features of native and functionally active protein tertiary structure, playing critical roles in protein dynamics, protein stability, conformational transitions, oligomer formation, active site shaping, ligand specificity and as reaction substrates or cofactors. Many of these structural roles were described above and illustrated in Figure 1. Furthermore, numerous reviews have been published to date that have focused on the many different and key functions of water molecules in proteins.[175-179] Here, we will briefly summarize these roles, with a specific emphasis on "***cold***" water molecules, i.e., those water molecules that are a constitutive part of protein structure and protein machinery. Using the water-centric view terminology above, these water molecules **are** happy in their roles and loci in biomacromolecular structures. Their displacement to bulk should be, at best, nearly energy neutral ($\Delta G \sim 0$) or, at worst, incur a cost ($\Delta G > 0$). Clearly, in most cases the thermodynamics of other changes in structure must be considered in order to compensate for the displacement of cold waters. The chemical potentials of "cold" water molecules are largely unchanged by the ligand binding event, and these waters have no impetus to relocate within the structure or to bulk. In fact, the chemical potentials of some of them may even decrease until a new equilibrium is established. In the text to follow we will classify cold waters on the basis of the role that they play: i) in determining protein structure, ii) in allowing protein dynamics, and iii) in achieving protein function and regulation.

4.2.1. Cold waters in protein binding sites. Highly conserved water molecules present in proteins of known three-dimensional structure are likely to be "cold" because they usually have multiple and significant *direct* interactions with the protein. There are numerous studies in the literature where such waters were classified on the basis of B-factors and contact/hydrogen bond counts.[80,179,180] We reported a study of this nature where we calculated: 1) a metric termed Rank[181] (Section 4.3.7) that scores the count, strength and geometry of potential hydrogen bonds for each water molecule in a structure and 2) the Hydropathic INTeractions (HINT) score[182,183] (Section 4.3.7) for each water molecule in its "pseudo-receptor" (environment).[184] These parameters were correlated with structural roles played by water in protein and protein-ligand complexes. In particular, second-shell (0), first-shell (1), active site (2), buried (3) and small cavity (4) water molecules (Figure 8A) were manually indexed and curated in twelve non-liganded protein X-ray structures. Fifteen ligand-bound structures, including nine in common with the unbound set, were examined for water molecules bridging between protein and ligand. Comparisons between bound and ligand-bound structures revealed some general guidelines for displacement based on Rank and HINT score. A later report[168] combined Rank and HINT score into the more quantitative Relevance metric that was about 90% accurate in predicting, from the unbound structure, the disposition of waters in its ligand-bound structure. The prediction for this data set using the B-factor alone was notably inferior and, when combined with Rank and the HINT score, did not improve the prediction accuracy.

The most interesting data are those that track water molecules between unbound and ligand-bound structures. We proposed[184] seven descriptions for such waters: conserved bridging, conserved active site, conserved cavity, conserved (largely surface) external, functionally displaced, sterically displaced and, lastly, just missing. Members of the first three categories should be considered as "cold" water molecules; the conserved external water molecules on the surface *may* also be cold, but there is no way to know if the "same" water was conserved or if another one has coincidentally landed in that same location. Some of the functionally displaced waters may be cold while others are hot depending on several structural factors and their environments. To clarify, we need to "take their temperatures" (*vide infra*). It is likely that the sterically displaced waters were hot, and, excluding other factors, the waters that "disappeared" between the unbound and ligand-bound structures were also hot. Thus, cold waters may be distributed in many different sites of proteins, some within sites that are involved in protein function, such as enzyme active sites and ligand binding sites, or in protein structure, such as

potential hot spots in protein-protein interactions.

Our opening example of HIV-1 protease (Figures 1A and 1B) clearly indicates that water 301 was well conserved and locked in-place by the isoleucines in the unbound structure, and easily adapted to a bridging mode in the peptidomimetic ligand-bound structure of Figure 1B. In fact, in early development of HIV-1 protease inhibitors, molecular design took into account that the very cold water **301** molds the shape of the active site. Accordingly, the resulting compounds, hydroxyethylenic ligands and peptidomimetic diol derivatives, displaced most of water molecules in the site, except for 301. Also often found in X-ray structures of HIV-1/ligand complexes are another pair of water molecules (labeled **313** and **313'**) that provides binding support by bridging between the ligand (at either end) and aspartates 29 and 29'. The second retained pair (**313bis** and **313bis'**) can be characterized as conserved active site water molecules as they occupy a more peripheral active site region than **313/313'** and are closer to the protein than to ligands.[33] NMR studies confirmed the relevance of retained binding site water molecules.[185] On the hot water side, the "catalytic" water **300**, which was coordinated to the two catalytic residues aspartates 25 and 25' in the unbound form of the enzyme, is obviously and easily displaced. Several other water molecules present in the unbound structure (**a**-**g**) have clearly been displaced by the ligand, either functionally or sterically.

Water **301**, however, *seemed* like an obvious target for displacement by a ligand that could replicate its structural and functional role. Such a ligand could gain the additional hydrogen bonding from its interactions to the pair of isoleucines, and displacement of that water could also yield an entropic contribution as it moves to bulk. Thus, series of ligands (e.g., Figure 1C) containing a ureidic moiety[33] or a sulfamide moiety[186] were designed to substitute the water's oxygen atom with a carbonyl group and sulfoxide, respectively. More than a decade ago, we explored the energetics of six such inhibitors as well as seventeen of the acyclics by calculating their HINT binding scores with respect to HIV-1 protease.[34] The overall scaled[187] protein-ligand HINT scores for the acyclic inhibitors averaged to -6.6 ± 1.1 kcal mol^{-1}. And, indeed, the averaged HINT scores for the six members of the cyclic set averaged to -8.1 ± 1.5 kcal mol^{-1}, which seems to confirm the rationale for water displacement. However, it must be considered that the former compounds also have a **significant** interaction with water **301**, which averaged to -2.1 ± 0.5 kcal mol^{-1}. *In toto*, these compounds are actually **more** tightly bound (-8.7 kcal mol^{-1}) than the uridyls and sulfamides. For reference, waters **313** and **313'** also support binding through protein-ligand bridging to both cyclic and acyclic inhibitors by about -0.7 ± 0.4 kcal mol^{-1} and waters **313bis** and

313bis' by another -0.1 ± 0.1 kcal mol^{-1},[34] which implies that the latter pair is more employed in maintaining the binding cavity wall than in supporting binding. Looking at this another way, this scenario is an example of entropy-enthalpy compensation: the displaced water **301** might indeed gain significant entropy upon release from its tight binding site, but the acyclic ligands' indirect interactions to protein through that water molecule likely provide more reliable hydrogen bonds in the complex than that water experiences in bulk. There are certainly quite a large number of other considerations in terms of inventing drugs ignored by these simple calculations, but displacing a cold water such as HIV-1 protease's **301** clearly may not produce the desired energetic effect.[188]

4.2.2. Cold waters and catalysis. Water networks that link the active site with the external medium have been proposed as a common feature for enzymes that, in addition to the incoming substrate or leaving product, need a supply or sink for protons, hydroxyls and/or water molecules during their catalytic cycle.[189] A recent study of the *Staphylococcus aureus* enoyl-acyl carrier protein reductase (saFabI), a NADP+ dependent enzyme involved in bacterial fatty acid biosynthesis and potential drug target,[190] found such a network linking the active site with a water cluster located within the homo-tetramer matrix. It is believed that this network serves to both close a flexible substrate-binding loop and transfer hydrides and protons during the saFabI catalytic cycle.[190]

The pattern of water molecules present in the ATP binding site of 19 kinome-diverse kinases derived from 171 crystal structures was studied by Barillari et al.[191] The striking differences amongst this set allow a satisfying explanation of inhibitor specificity for distinct kinases. Moreover, different patterns of conserved water molecule loci were observed between inactive and active kinase conformations. Both observations suggest that these cold water molecules are skilled players in shaping and defining the character of the kinase ATP binding sites and thus affecting specificity and affinity for molecules targeting them. In turn, this evidence strongly argues that exploiting water distribution in such binding sites for the design of specific ligands should be a key strategy of drug design for kinases, which are notoriously difficult to target with specificity, and likely many other enzymes (see Figure 8B). Another study[192] superposed the structures of 13 diverse active-conformation kinases and subjected each to molecular dynamics simulations. This revealed another role for conserved water molecules: six were found to be highly relevant for protein stability and for correctly positioning the catalytic residues within the active site. The Pfizer cancer drug bosutinib, a tyrosine kinase inhibitor, uses two conserved cold

water molecules protected by the DFG loop in an inner cavity of the ATP binding site to dictate its recognition and specificity to the Src kinase.[193] One of the waters forms a hydrogen bond with the nitrile moiety of bosutinib, which is engaged in only one other hydrogen bond to the enzyme. We will look at this case in more detail later in the Perspective (section 6.1).

In another enzyme class, two conserved water molecules that play a structural role have been observed in the catalytic domain of Pin1, an enzyme that catalyzes the cis/trans isomerization of peptidyl prolyl bonds. A careful investigation using both computational and spectroscopic methods highlighted a unique structural motif that traps water molecules, similar to that observed for the EF-hand domain for calcium.[194] In another example, DNA methyltransferases – key enzymes in the epigenetic pattern of DNA – have attracted great interest because their role in cell replication suggests a potential strategy for anticancer drug development. Structural insight into this interaction from the crystal structure of the Pro-Trp-Trp-Pro domain of DNMT3B (DNA (cytosine-5-)-methyltransferase 3β) in complex with the epigenetic mark H3K36me3, a methylated peptide histone, showed that the trimethylated Lys36 forms a strong hydrogen bond with a conserved water molecule that is locked in position by the enzyme's Ser270.[195] This finding explains the lack of affinity of the Ser270Pro mutant form of DNMT3B, which is implicated in the ICF (Immunodeficiency, Centromeric instability and Facial abnormalities) syndrome, for the H3K36me3 epigenetic mark. The mutation is accompanied by the loss of the conserved water and its hydrogen bonding properties. Another recent and interesting case of cold water molecules in enzyme active sites was reported by Salie *et al*.,[196] who solved the structure of HIV reverse transcriptase in the presence of 4-ethynyl-2-fluoro-2-deoxyadenosine. This compound, currently in phase I clinical trials, is the most potent nucleoside analogue enzyme inhibitor known and possesses an unusually long half-life. A key observation, possibly contributing to activity and half-life, was that a network of ordered water molecules at the polymerase active site acts to stabilize the enzyme's interaction with both the nucleotide and DNA substrates.

4.2.3. Cold waters in protein folding, misfolding and dynamics. The indirect roles played by water molecules in protein folding via hydrophobic interactions are well known,[197] but to our knowledge, only a few studies have pinpointed the direct actions of cold water molecules in folding and misfolding. It is not easy to assess whether such water molecules are participating in folding pathways since these are very fast events and there is a (current) lack of illuminating intermediate structures. However, the scenarios found to date involving water are intriguing. The

role played by a single water molecule (**Wat3**) in controlling the mobility of the 50S loop in the binding site of FK506 binding protein-12 was investigated by X-ray crystallography and MD simulations.[198] Mutations of Glu60, the residue hydrogen bonded to **Wat3**, to either Ala or Gln disrupt this interaction and cause a rearrangement of loop position. This finding supports the notion that water molecules are able to shape protein mobility as well as protein structure. Similarly, it was shown by polarization-resolved femtosecond infrared (fs-IR) spectroscopy combined with MD that water dynamics around bovine α-lactalbumin changes between the native and misfolded protein. Specifically, the protein misfolding leads to an increase of hydrophobic residue exposure to solvent, and, consequently, to an increased mobility of the water molecules that were previously locked in close contact with those hydrophobic residues.[199] In another type of scenario, the transition of prions from their native to aggregated forms has been demonstrated to be favored by dehydration.[200] The conformational transition from the normal prion protein to the pathological form involves water molecule reorganization. Time-resolved fluorescence anisotropy showed that the mobility of some water molecules in prion oligomers is reduced, possibly because they are trapped at specific sites.[201]

With respect to drug discovery and design, a significant effort is being devoted towards identifying compounds that, by binding to allosteric sites, alter the conformation of misfolded proteins and stabilize native-like conformations.[200,202,203] Whether cold water molecules are involved in this rescue effect is still open to investigation.

4.2.4. Cold waters in Protein-Protein Interactions (PPIs). Less than twenty years ago, proposing a project to develop compounds (small molecules or drugs) that interfere with a PPI would be considered unrealistic due to the relatively shallow and rather large sites that were thought to define the interaction. This view has significantly changed in recent years because our understanding of the determinants that control PPIs has deepened from the increasing availability of structural data. An emerging feature, generated by a careful analysis of high-resolution three-dimensional X-ray structures, is the presence of a variable number of water molecules at these interfaces. This feature has been investigated over recent years by exploiting the broadening database of available PPI structures and their improving resolution. This, in turn, has led to a much-improved understanding of water involvement in protein-protein interactions.

In a 1999 review, Janin described PPI interfaces as “dry” or “wet”.[204] Like a protein interior, where water is almost completely excluded with mostly empty small cavities, a dry interface has only a few small cavities. However, a wet interface has many cavities where nearly

all are filled with water. In both cases, the driving force is to maintain a close packing of atoms. A later, detailed examination of protein-protein interactions was carried out by Janin on 115 dimeric proteins and 46 protein-protein complexes.[205] A key result was that about 10 water molecules/1000 Å^2 contact surface area are present at protein-protein interfaces. Also reported were the fractions of these water molecules involved in hydrogen bonding with the carbonyl (~33%) and NH (~12%) moieties of the backbone chain, with charged, such as Glu, Asp and Arg, amino acid residue sidechains (~29%) and neutral, such as Ser, Thr, and Tyr, sidechains (~25%). The structural roles of water molecules at protein-protein interfaces are diverse: from, as discussed above, those that are trapped within the interface and bridging (likely cold) or otherwise interacting (hot) with the proteins, thus contributing to the association energy for complex formation, to those capping the interface to shield a mainly hydrophobic-based interaction and which are likely cold. This study further reported[205] that the water molecules present at protein-protein interface represent only about 10–15% of the those originally solvating the protein partner subunits. Thus, 85-90% of the solvating water molecules are released to the solvent during complex formation, which is, of course, a manifestation of the hydrophobic effect, and a potentially significant contributor to the energetics of protein–protein interaction.

A more recent and extensive study of water molecules at PPI analyzed 3295 homodimeric structures, classifying water molecules based on their distances to protein: i) waters at interface, ii) crystallographic waters and iii) not involved in the protein-protein interaction and only involved in water-protein interactions.[206] Distinct hydrogen bonding patterns were observed for the different classes, with water molecules involved in PPIs exhibiting a higher degree of hydrogen bonding. Furthermore, and contrary to expectations, it was reported that these PPI water molecules displayed geometries tending more towards planar rather than tetrahedral,[206] i.e., more liquid water-like than ice-like. In another recent study that analyzed five proteins in both their free and protein-bound states with molecular dynamic simulations, it was suggested that water molecules around both polar and charged residues, at sites presumed to be non-interacting, might actually be important for *preventing* functionally-irrelevant PPIs.[207]

The interesting biochemistry and biophysics associated with protein-protein interactions will hopefully lead to drug discovery strategies that exploit it. We are particularly enthusiastic about the opportunities offered by interfacial water molecules. Following are three intriguing scenarios of water-mediated PPIs. The first is the interaction between monomeric units of the dimeric tRNA-modifying enzyme tRNA-guanine transglycosylase (Tgt) from *Zymomonas*

mobilis.[208] Interestingly, since higher eukaryotes possess a heterodimeric enzyme, subunit interfaces of bacterial orthologues can be targeted by compounds that are potential drugs against Shigellosis, which is also known as bacillary dysentery or Marlow syndrome. A cluster of four aromatic amino acids is at the core of the subunit interface and crucial for stability of the dimer, as is a highly flexible extended loop-helix motif. Mutational analysis of interface residues led to a variable degree of dimer disruption – up to 99% – that can be traced to two causes: the mutated loop-helix motif adopting unfavorable conformations within the interface and, more dramatically, a series of water molecules entering into the interface and destabilizing it. This finding suggests a strategy for the development of potential interface breakers.

The second example involves protein aggregation, which is a special case of protein-protein interaction. Among the hypotheses for the etiology of Alzheimer's disease is that mutations in the amyloid precursor protein and presenilins 1 and 2 cause an increase in the production of the small protein Aβ42, which, when aggregated, is the main component of senile plaques.[209] While ordered polymers are formed by recruiting multiple proteins and/or peptides, these processes in water environments can lead to misfolding and unwelcome aggregation. Fluctuating thermodynamic analyses of Aβ42 misfolding and dimer formation from unfolded peptides released into water from membrane models indicates that the water-protein interaction free energy dictates the transition of the peptide from unfolded to misfolded, the interaction between two misfolded peptides to form a dimer, and the succeeding peptide interaction with the growing aggregate. Furthermore, the aggregation propensities of Aβ42 mutants and mutants of the N-terminal domain of the HypF protein and of acylphosphatase were found to correlate with protein solvation free energy.[210] Another proposal is that dehydrons, non-conserved residues with unsatisfied H-bonds, might lead to unwelcome protein aggregation.[211]

The third example highlights how dense interfacial water is an essential component in transient protein-ligand interfaces involving electron transfer. The number of water molecules at the PPI of the complex between cytochrome P450cam with putidaredoxin was found to be dependent on the redox state of the system [Furukawa Y, et al. J. BIOL. CHEM. 2001, 276, 12983-12990]. There are about ten more water molecules at the interface when putidaredoxin is in the reduced state. These waters increase the affinity of the complex by mediating hydrogen bonding. Furthermore, it was found in multiple protein complexes, either by experimental measurements or computational simulations, that structured water molecules at the PPI can enhance the tunneling of electrons from one protein to the other [Pelletier and Kraut, Science, 1992, 258, 1748-

1755; Tezcan et al, PNAS 2001 98, 5002–5006; Lin et al. Science, 2005, 310, 1311-1313; Bizzarri et al., J. MOL. RECOGNIT. 2007, 20, 122-131]. This role played by water molecules is another that might be exploited for the development of ligands that displace them, thus preventing or slowing down electron transfer, which is a key biological process. We are not aware of any current attempt of developing drugs along this direction, but it is an interesting prospect.

4.2.5. Cold Waters in Protein-Polynucleotide Recognition. Although we are focusing in this Perspective on protein-ligand systems, water plays as large a role, if not larger, in interactions involving polynucleotides.[212-216] Particularly interesting are the multiple roles played by water molecules in the recognition sites between DNA or RNA and proteins, i.e. in mediating specific interactions between amino acid side chains and bases [Reddy et al, Do water molecules mediate protein-DNA recognition? J. Mol. Biol. 2001, 314:619–632; Janin and Bahadur, Cellular and Molecular Bioengineering, 2008, 1, 327–338]. It was found that about 6% of the water molecules identified in X-ray structures of DNA-protein complexes contact both the protein and the DNA simultaneously and thus directly mediate recognition [Reddy et al, Do water molecules mediate protein-DNA recognition? J. Mol. Biol. 2001, 314:619–632]. About one-third of these water molecules act as an extension of a sidechain's polar (donor) group in order to contact a base (acceptor). A well-documented example is the bacterial trp repressor protein. This protein is a dimer that binds to specific recognition sequences of DNA and acts as regulator of the transcriptional machinery that leads to the formation of enzymes involved in the L-tryptophan biosynthetic pathway. Its affinity to DNA depends on the binding of L-tryptophan to the repressor binding site. This binding causes a conformational change of the protein that allows its recognition helices on two symmetric subunits to be at the ideal distance from the DNA recognition sequences. The recognition determinants between amino acid residues and DNA have been the subject of numerous structural, mutational and spectroscopic studies [Otwinowski et al. Crystal structure of trp Repressor-Operator Interactions 553 sor/operator complex at atomic resolution. Nature, 335, 321-329; Lawson, C. L. & Carey, J. (1993). Tandem binding in crystals of a trp repressor/operator half-site complex. Nature, 366, 178-182; Grillo et al. J. Mol. Biol. (1999) 287, 539-554]. While there are a few direct amino acid-base interactions, recognition seems to depend on the phosphates, and just two water molecules that mediate protein-DNA interactions, one of which acts by extending amino acid chains so as to contact the base. Specific recognition mediated by hydrogen bonding of water molecules is still not completely understood.

Similarly, analyses of 145 RNA-protein complexes showed, similar to PPI and protein-DNA associations, three classes of water molecules: 1) those present both in the bound and unbound state, 2) those that are present only when the protein-RNA complex is formed, and 3) those that are only present in the unbound state.[217] The aggregate of these data on water roles in protein-polynucleotide complexes, of course, provide numerous starting points for designing disruptive compounds,[ref. 214; ref. 245] with potential to target numerous diseases.

4.2.6. Cold water in GPCRs. Some of the hot and mobile waters found in GPCRs were discussed above; these are largely associated with and/or directly influence structural reorganization of the receptor. There are also a number of cold waters in such structures, and this label is assigned to highly conserved water molecules whose effects are dependent on them staying in place. Many integral membrane proteins, including GPCRs, contain ordered water molecules that are functionally like prosthetic groups and quite unlike bulk water. Such water molecules can mediate proton transfer in the active state,[142,144,218,219] can function as local energy storage while reducing the pK_a of the central proton in the binding site, in supporting the role of Na^+ as a cofactor in signal transduction,[220] and in other specific roles. Well-ordered and conserved water molecules appear to be responsible for the functional plasticity for transmitting activation signals from their binding pockets to the cytoplasmic GPCR side.[134] Water molecules were shown to be involved in bacteriorhodopsin's mechanism of proton pumping,[142] but it is not clear whether this involves proton hopping such as by the Grotthuss mechanism. Radiolytic footprinting of bovine rhodopsin confirmed that only small structural variations occur in the vicinity of the bound chromophore. Also, these experiments show – by rapid mixing with ^{18}O labeled water – that the water molecules indicated by X-ray diffraction to be conserved similarly do not appear to exchange with bulk solvent.[221,222] Thus, the set of conserved water molecules in rhodopsin, and their ability to transport protons, provide an attractive explanation for how rhodopsin transmits visual signals without exhibiting large conformational changes.

4.2.7. Other roles of cold water molecules. Cold water molecules can be found in the structures at a variety of other critical locations in proteins. In one recent example, the combination of neutron and X-ray diffraction crystallography, both at room temperature, was used to characterize the complex between heart fatty acid binding protein (H-FABP) and oleic acid,[223] and in particular the roles of water molecules within the complex. The all-atom coordinates for over 200 D_2O molecules were reported. First, the detailed location of the fatty acid within the active site was unveiled. Its hydrophobic tail associates with a wall of hydrophobic

residues whereas its other end faces a cluster of 14 ordered water molecules. These water molecules are conserved in the active site with a quite low average B-factor (15.6 $Å^2$). Most of the other water molecules present in the unbound structure are displaced by the incoming fatty acids with their long apolar chains while some of the remaining few interact with the substrate's carboxylate moiety. Most interesting, however, is that the role of the conserved water cluster does not seem to be associated with ligand stabilization, but instead, to a significant reduction of volume in the cavity, thus forcing the fatty acid into a U shape arrangement and governing discrimination amongst different fatty acids, as suggested by Matsuoka *et al*.[224] Figure 8C illustrates this scenario schematically. Only because the structural study was performed with neutron diffraction, and the hydrogen (deuterium) positions of the water molecules were directly observed, it could also be determined that the dipoles of the waters in this cluster were aligned with the protein's electrostatic field, which is additional rationale for their low mobility. These observations imply that hydration layers on protein surfaces are endowed with specific features, e.g., electrical, which can be especially important when these water molecules become trapped within a protein-protein interface. As a corollary, note that the occurrence of a decreased dielectric constant, e.g., through a rearrangement of water dipoles, leads to a reinforcement of the interaction strength between polar residues, thus further favoring complex formation. However, the thorough understanding of the water roles and electrostatics offered by these studies suggests possible strategies for inhibitor design, perhaps by displacing one of the cluster waters while respecting its contribution to the overall electric field.

In another example, hydrogen bond networks involving water molecules mediate hydrolysis of adenosine triphosphate (ATP) at the nucleotide binding site of the bacterial protein SecA. This protein exchanges ADP for ATP and thus provides energy to membrane translocation of secretory pre-proteins. Energy is obtained by exploiting the binding and hydrolysis of ATP, which is then used to produce large conformational changes at domains far away from the ATP binding site. Despite the availability of crystal structure snapshots of SecA along its reaction path, the coupling mechanism between the nucleotide binding site and the other domains of SecA has been unclear. However, molecular simulations of wild type and mutant SecA proteins revealed the presence of extensive networks of hydrogen bonds connecting the nucleotide binding site to the protein's remote regions.[225] A number of water molecules were integral to these networks and thus enabled the reaction.

4.3. How Do You Take the Temperature of a Water Molecule? What we are actually asking is: what are the characteristics of hot and cold water molecules? As described above, the roles that water molecules play are very diverse, with significant differences between those that are hot or cold. It is not a simple matter of measuring rotational and/or translational entropy for each water molecule – if that was even possible. As discussed above, the energetics of water molecules are a case-dependent and complex mix of entropic and enthalpic contributions.[226] As in most endeavors in science, it is best to start with what can be extracted from experiment, and there are a few structural clues that can be exploited – some from the water molecules themselves, but more from the water molecules' environments.

4.3.1. Evidence from crystallography: B-factor and occupancy. The B-factor, also known as the temperature or Debye-Waller factor, may be seen as a measure of the atom's mobility or flexibility and is reported in $Å^2$.[227] This displacement can be represented isotropically or anisotropically (i.e., along the Cartesian axes), although the data is too sparse in most biomacromolecular data sets to refine atoms anisotropically. Molecular structure (PDB) files report two kinds of B-factor: an overall average that is the mean of all atom movements in the molecule and individual or "local" B-factors for each atom. With reference to the overall structure's B-factor, the relative flexibility of specific atoms or regions in the protein can be assessed. This also applies to water molecules, where a small B-factor for the water's oxygen atom indicates that it is in a locus of low mobility, thus suggesting that it is cold, while a large B-factor indicates high mobility, and thus a hot water (see Figure 9A). As an illustration, consider the water molecules of HIV-1 protease described above (Figure 1). For the HIV-1/ligand complexes in our data set, the average B-factor of the super cold water **301** is 15 ± 8 $Å^2$, that of the **313/313'** set is 23 ± 8 $Å^2$, and that of **313bis/313bis'** is 17 ± 7 $Å^2$. (High standard deviations are due to the disparate crystallographic conditions for the 23 structures.) For comparison, the five waters (Figure 1A, waters **301**, **313**, **313'**, **313bis**, **313bis'**) present in the unliganded structure and largely retained in the complexes have an average B-factor of 33 ± 10 $Å^2$, the seven water molecules present in the unliganded structure and displaced (waters **a**-**g**) have an average B-factor of 56 ± 16 $Å^2$, supporting the hypothesis that **a**-**g** are hotter waters. The B-factor of the catalytic water (**300**) is an intermediate 42. (See also Table 1.)

A direct indication of water conservation is atom occupancy, which is also reported in PDB files, although there is likely some correlation between occupancies and B-factors.[228] Occupancy represents the completeness of an atom's electron density and ranges between 0 and 100%. Less

than full occupancy can be attributed to thermal disorder or symmetry (in either case two positions for the same atom may be reported) or simply that the atom is not present in all instances of the protein in the crystal. The latter is more common with ligand molecules, especially if the ligand has been "soaked in" rather than co-crystallized. A water molecule (i.e., its oxygen atom) reported to have 100% occupancy is more likely to be cold than one with low occupancy (see Figure 9B). It should be pointed out, however, that assigning occupancy for water molecules in X-ray crystal structures is a lot of extra work, and not usually performed. Indeed, the number of water molecules reported in a crystal structure depends on the resoluteness of the crystallographer as well as the resolution of the data set.[67,229] Some of the early X-ray structures for ligand-bound HIV-1 protease reported water **301** as the only one present. Thus, it is sometimes advisable to characterize a binding site by a software tool that calculates probable locations for water molecules like GRID,[78] HINT[182,183] or Hollow.[230]

4.3.2. Radiolytic protein footprinting. For a few residue types with aromatic and sulfur containing side chains, i.e., phenylalanine, tyrosine, methionine and cysteine, radiolytic labeling under aerobic conditions is mediated by hydroxyl radicals derived from radiolysis of water rather than from molecular oxygen.[231] Thus, exchange with $^{18}OH\cdot$ radicals, generated by exposure of $H_2{}^{18}O$ to X-rays, can reveal whether water molecules associated with those residues are tightly bound (see Figure 9C). Liquid chromatography coupled to mass spectrometry (LC-MS) analysis of peptides obtained by digestion of the irradiated protein/isotopic labeled water sample quantitates the extent of modification with ion current data.[221]

4.3.3. Evidence from NMR: residence time, hydrogen exchange, and generalized order parameter. Regardless of location or how well ordered water molecules appear to be at the protein surface, nuclear magnetic relaxation dispersion (MRD) experiments[232] have greatly illuminated the dynamics of protein–water interactions and demonstrated that *all* water molecules associated with proteins are in constant dynamic exchange (Figure 9D) with the bulk water. Interestingly, many studies have shown that water molecules at protein surfaces exhibit slower correlation times than bulk water, similar to water molecules in confined spaces.[175] Specifically, the more exposed – hot – water molecules on protein surfaces are endowed with a broad distribution of residence times, with medians in the tens of picoseconds, extending to several nanoseconds for a few water molecules in deep surface pockets.[233] By comparison, residence times are in the microsecond regime for deeply buried internal – cold – waters. The hydrogen (proton) exchange between protein protons and water protons is also an indicator of

solvent accessibility, in particular with respect to the highlighting the protons from proteins involved in hydrogen bonds, structurally and dynamically[234] (see Figure 9E). Slow hydrogen exchange, i.e., chemical exchange between $^{1}H/^{2}D$, has been applied in a wide variety of studies including protein folding.[235] Magnetization transfer approaches can also study fast exchange between solvent water and various sites in biomolecules.[236] Another NMR parameter that probes the motional freedom of a water molecule is given by the generalized order parameter, which is usually reported as its mean, squared: S^2.[237] This quantity varies from 0 to 1 as order increases; however, its use in studying water in and around protein structures has been very sparse to date.

4.3.4. Preferred hydration sites. Adapting the strategy used in constructing rotamer libraries[238] that catalogue the frequency of specific sidechain conformations in X-ray crystal structures for use in structure model building, Biedermannova and Schneider used a non-redundant set of 2818 protein crystal structures with resolutions better than 1.8 Å to analyze the extent and structure of the hydration shell of all 20 standard amino-acid residues as function of residue conformation, secondary structure and solvent accessibility.[239] The results show that hydration depends on the amino acid conformation and the environment in which it occurs. In practical terms, distance and angular dependence of the probability of finding waters of hydration were calculated and tabulated for each residue type and each of its previously described[238] sidechain rotameric states. Certainly, water molecules found in the higher "rotamer" probability regions of an X-ray structure would be colder than water molecules located in low-probability regions (see Figure 9F).

4.3.5. Flexibility of the water site. While structural information is the basis of all molecular modeling, crystals are only 3D pictures of a frozen conformation, often at very non-biological cold temperatures. The real situation is that proteins in aqueous media within a cell move, and may be flexible from less than one to hundreds of Å scales. Other than the standard vibrational motions of a H_2O molecule, waters themselves are not flexible. Waters are, however, found in protein regions of low to high flexibility, and this implied flexibility likely impacts their "temperature". Flexibility and the thermal motions represented by B-factors are also related, but flexibility from longer timescale or higher temperature thermal motions may still be possible in crystallographically well-resolved atoms, residues or even domains. There are a number of methods to estimate flexibility, but one very simple metric is the count of rotatable bonds in a residue sidechain, or for all residues within the site[240] (see Figure 9G). However, some residue sidechains may be less flexible than suggested by their number of rotatable bonds because they are "locked" by adjacent

residue sidechains.

4.3.6. Solvent accessibility of water molecules. The solvent-accessible surface area (SASA) concept that has generally been applied to ligand molecules or proteins and their residues is a powerful tool for evaluating water molecules in biomacromolecular structures. Water molecules with low solvent accessibility, i.e., not exposed, are likely to be cold, while those with high solvent accessibility are likely to be in exchange equilibria with other water molecules in the system and more hot (see Figure 9H). Solvent accessibility is a concept developed by Connolly[241] and often reported as the Connolly surface. It can be calculated with a number of easily available tools.[242-244] Table 1 lists the SASA values for the water molecules in unliganded HIV-1, and mostly follow the trends shown by the B-factors.

4.3.7. Rank, HINT score and Relevance. Simple analyses of the geometry of interactions for a water molecule can reveal how stable it is in its site. Water can form up to four hydrogen bonds, with two as a donor and two as an acceptor; these hydrogen bonds are defined in terms of the distances and angles to their partners (Figures 9I). We developed an algorithm called Rank[181] to calculate the number and geometric quality of every potential hydrogen bond for each water molecule to protein and/or ligand atoms within a specific range:

$$\text{Rank} = \sum_{n}^{\leq 4} \{(2.80\ \text{Å}/r_n) + [\sum_{m}^{\leq 6} \cos(\theta_{\text{Td}} - \theta_{nm})]/6\},$$

where r_n is the distance between the water's oxygen and the target heavy atom n (n_{max} = 4). This is scaled relative to 2.8 Å, i.e., an ideal hydrogen bonding distance. Ideality of angles is evaluated by $\cos(\theta_{\text{Td}} - \theta_{nm})/6$, where θ_{Td} is the ideal tetrahedral angle (109.5°) and θ_{nm} is the angle between target atoms n and m, with a maximum number of target pairs (m_{max}) of 6. Any angle less then 60° is rejected along with its associated targets. Rank largely correlates with the number of high-quality hydrogen bonds made by the water and can be calculated without structure optimization, and even before adding protons to the molecular model. The Rank for the water molecules in unliganded HIV-1 protease are shown in Table 1; the water molecules that are displaced by the incoming ligands have for the most part smaller Rank values.

However, the character and strength of hydrogen bonds are not recorded by Rank. We also have applied HINT scoring, wherein the water is treated as a ligand and all surrounding atoms as a pseudo-receptor to evaluate water placement and orientation (Figure 9J). We have previously

reviewed HINT, and its numerous applications in biomolecular structure and drug discovery/design scenarios,[34,184,214,245,246] so only the important operation details will be described here. The HINT score (H_{TOTAL}) provides an empirical, but quantitative, evaluation of a molecular interaction as a sum of all individual atom-atom interactions using:

$$H_{TOTAL} = \sum_i \sum_j b_{ij}; \; b_{ij} = a_i \, a_j \, S_i \, S_j \, T_{ij} \, R_{ij} \; + \; 50 \, r_{ij},$$

where b_{ij} is the interaction score between atoms i and j, a is the hydrophobic atomic constant, i.e., its partial logP for 1-octanol/water – representing chemical properties of hydrophobes, donors and acceptors as well as their charge, S is the solvent accessible surface area, T_{ij} is a logic function assuming +1 or -1 values, depending on the nature of the interacting atoms (when polar), and R_{ij} and r_{ij} are functions of the distance between atoms i and j, an exponential and a Lennard-Jones implementation, respectively. The HINT paradigm is based on the assumption that each b_{ij} is related to a partial δg value and that H_{TOTAL} is directly comparable to the global interaction's ΔG°. In particular for protein-ligand associations, several studies showed that the HINT score could be related to ΔG°: 1 kcal $mol^{-1} \approx -515$ score units.[178,247] Of note is that the HINT scoring function, because of its basis in an experimental measurement, considers both enthalpy and entropy, which is neglected in many scoring functions that attribute binding purely to electrostatic interactions. As with Rank, Table 1 lists the HINT scores for water in the active site of unliganded HIV-1.

We first used Rank and HINT score to characterize water molecules when we tracked them between paired unliganded and ligand-bound crystal structures (Section 4.2.1).[184] Later, we used a Bayesian-like approach to integrate these two metrics into the Relevance metric,[168] which classified waters as Relevant (or "cold") if the metric ≥ 0.5 or Non-Relevant (or "hot") if the metric < 0.5. Our article hypothesized that displacement of what we are referring to in this Perspective as cold waters would likely be energetically neutral or possibly favorable, if performed with polar groups designed for this aim, and able to establish the same hydrogen bonds. On the other hand, waters classified as "Non-Relevant" (i.e., hot) are those near the surface, weakly interacting with the protein and/or mostly involved with other water molecules; these will likely be sterically displaced or otherwise vanish after ligand binding. The Relevance metrics for the water molecules in unliganded HIV-1 are listed in Table 1.

4.3.8. Volume of water site. Calculation of pocket (or cavity) volume can be accomplished by a large number of software tools[248] and is a useful analysis at initiation of drug design

experiments. Obviously, the available volume should match the size of the new ligand. The most common algorithm applied to calculate a cavity volume is the "rolling sphere" method where a sphere of a defined radius rolls on the protein surface – and into cavities. A typical algorithm of this type is *Roll* by Yu *et al*.[249] We described the VICE program,[250] which utilizes integer arithmetic for virtually all calculations. Another set of algorithms calculates volumes by using Voronoi polyhedra, that divide the total volume into subsets with size that depends on how tightly they are packed.[251] Perhaps the most difficult aspect of all such calculations is unambiguously defining the cavity boundary at its opening.

The volume in which a water molecule resides is another indication of its "temperature". Yu and Rick calculated the thermodynamic parameters for single water molecules in both hydrophobic and polar pockets of different volumes[84] (see Figure 9K). Interestingly, while a water molecule in a hydrophobic cavity slightly "cools off" as the size of the cavity increases, there is an optimal "coldest" cavity volume for a water molecule in a polar environment. This latter free energy response can be understood by noting: i) that the removal of a water molecule from small polar cavities is partially compensated by the restoration of some hydrogen bonding between cavity residues, and ii) a water molecule is positioned to make fewer hydrogen bonds as the cavity gets larger. It should also be mentioned that the presence of just one additional water molecule in either type of cavity creates a fundamentally different set of scenarios because of the interactions between the waters.

4.3.9. Molecular dynamics. Simply, molecular dynamics (MD) involves simulation of molecular motion as a function of time and temperature. This can easily be applied to water molecules within a potential binding site and even within the solvent milieu itself. Some of the earliest MD experiments were, in fact, attempts to simulate the behavior of liquid water.[252] The popular WaterMap method[253,254] and others (Section 5.3) use MD simulations as the basis of their water analyses. Here, consider that MD simulations should be very revealing of the "temperature" of a water molecule – as such simulations take into account many of the features of a water molecule and its environment described above. While a large variety of parameters can be extracted from MD, Figure 9L suggests that the characteristics of cold waters include lower potential energy and/or smaller trajectory amplitude (root mean square fluctuation, RMSF), while the characteristics of hot waters include higher potential energy and/or larger trajectory amplitude.

5. COMPUTATIONAL APPROACHES AND TOOLS FOR DEALING WITH WATER

As seen above, considered together, water molecules are a force to be reckoned with. Thermodynamically, the water molecules in a biomacromolecular system usually have a significant, if not dominant, contribution to the free energy of ligand-macromolecule binding. This can be generally presumed to be mostly due to their regaining entropy lost by the ligand upon binding and the release of protein-bound water molecules, both comprising what is called the hydrophobic effect. As we showed above, there are numerous interesting ways that water molecules play roles in structure and function and there are often clues presented by structure that reveal how those water molecules might best be exploited, and both can be differentiated by our "hot" and "cold" water construct.

Experimentally, the standard approach[255] for studying the effect of including or displacing water molecules from a protein's active site is to create modifications to the ligand or protein that will introduce or displace one or more waters, and measure the resulting effects on binding affinity, or occasionally its thermodynamic profile with ITC. Displacing at least a few water molecules is nearly ubiquitous when binding small molecules to proteins, as illustrated above with the hot waters of HIV-1 protease. Conserved cold water molecules are more challenging to displace, like in HIV-1 protease and scytalone dehydratase,[256] and the energy gain may not be as dramatic as envisioned. Deconvolution of the enthalpic and entropic contributions to binding of trimannosides with Concanavalin A with ITC revealed that a compound designed to displace a single water molecule by simply substituting $-CH_2CH_2OH$ for $-OH$ on one of the sugars demonstrated weaker binding affinity than its parent,[257-259] and that enthalpy-entropy compensation (H/SC)[17,61] is the likely cause. Studies performed on a number of benzoarylsulfonamide derivatives at human carbonic anhydrase demonstrated that the water network surrounding the bound ligands contributes significantly to the observed H/SC.[90,260] This case will be highlighted and further discussed below as an example of using cold waters in drug discovery (section 6.2).

Clearly, detailed experimental thermodynamic analyses are very useful for developing an understanding of what happened, i.e., as a post-mortem, but drug discovery really requires a prospective and predictive view. For this, computational experiments using theoretical and/or

empirical models and their associated software tools must be used. As we noted in the introduction, and throughout this Perspective, the rationale (or chemical art) of displacing a particular water molecule may be the core issue in water-aware drug discovery and design. Displacing hot waters from ligand free sites is a simple call and will generally provide some entropy. The more difficult act of displacing (possibly) cold waters from hydrophilic loci by ligand design, such as in lead optimization, is definitely not as straightforward because each water molecule possesses its own complex balance of context-dependent thermodynamic terms. The enthalpies of water molecules binding to a protein pocket do not provide a good measure of displaceability because all waters are not replaced by the same ligand "group"; if they were, tightly bound waters would indeed be hard to displace while loosely bound waters would be easily displaced. Water molecules are typically displaced with ligand groups that make similar interactions with the protein pocket, but that is not always possible. For example, while a water in a hydrophilic pocket is likely to be displaced by a hydrophilic group like a hydroxyl, forming both van der Waals (vdW) contacts and hydrogen bonds with the pocket, a water in a hydrophobic pocket is likely to be displaced by a nonpolar group like a chloro that will presumably make mostly vdW and hydrophobic contacts with the protein. Both of these cases would seem to be energetically favorable, but cases where a water in a hydrophobic pocket is displaced by a hydrophilic group or vice versa also commonly occur.

Altogether, these processes entail very complex and difficult to understand molecular energetics, with many moving parts. To help, a variety of computational tools have been developed. It should come as no surprise that constructing such software tools is similarly challenging. Water molecules in real systems are in constant flux with translations and reorientations along with, as well, the formation and breaking of their hydrogen bonds. The strategies and levels of sophistication employed by the available computational tools for simulation or analysis of these systems are diverse.

5.1. Empirical, phenomenological and knowledge-based methods. As described above, the crystallographic B-factors for water molecules can be revealing metrics concerning its conservation. Because a more labile water is more likely to be displaced than a constrained one, it has long been perceived that there was a correlation between the two, although a clear definition of conservation in this context probably does not exist. Nevertheless, some water molecules occupy the same positions in multiple crystal structures of the same protein, with a variety of

ligands, and even in structurally or functionally related proteins. To evaluate water conversation in proteins, Raymer *et al*. employed a k-nearest-neighbors classifier combined with a genetic algorithm to predict which bound water molecules in unliganded protein structures are conserved in the corresponding ligand-bound structures.[261] The resulting algorithm, *Consolv*, analyzes the environment of each water molecule in the free structure for four environmental features: the water molecule's B-factor, the number of hydrogen bonds between it and the protein, the density of neighboring protein atoms, and their hydrophilicity. One key observation was that the nanoenvironment of each water molecule is a dominant influence. Similarly, García-Sosa *et al*. reported *WaterScor*e,[262] which combined the waters' B-factors, their solvent-contact surface areas, their total hydrogen bond energies and their number of contacts (≤ 3.5 Å) with protein atoms, to discriminate between those conserved and those displaceable. More recently, the PyWATER plug-in for PyMOL [Patel et al. PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering, Bioinformatics, 2014, 30, 2978], which relies on extant structural information to identify conserved water molecules in a structure of interest, was described. PyWATER is a relatively simple tool for water identification in protein structure; in short, structures similar to the query, and having resolutions ≤ 2 Å, are automatically retrieved from the PDB and superimposed to identify consensus waters. The quality of a water molecule thus placed is assessed by calculating the Mobility (a normalized B-factor), as derived by Sanschagrin and Kuhn [Sanschagrin and Kuhn, Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteases and contributions to ligand specificity, Protein Sci, 1998, 7, 2054] and Carugo [Carugo, Correlation between occupancy and B factor of water molecules in protein crystal structures. Protein Eng., 1999, 12, 1021–1024]. In general, water molecules (typically their oxygen atoms) having mobility higher than 2 are discarded. Those remaining are hierarchically clustered and the degree of conservation is determined by the ratio of the number of waters in each cluster to the number of superimposed structures. PyWATER has been validated with a number of proteins previously examined with other analysis tools, and showed consistent results.

We briefly described our work in defining Rank and HINT score in Sections 4.2 and 4.3, above, and alluded to these metrics in other discussion in the Perspective. As with *Consolv* and *WaterScore*, the concept behind Relevance was to define a metric to predict – from the unliganded structure – the conservation/displacement of water molecules. The results we obtained in this regard,[18] were at least comparable to the earlier programs: ~90% accurate for water molecules in

structures with ≤2.0 Å resolution. Relevance relies only on the structural data contained in atom coordinates and atom types, i.e., crystallographic B-factors are not necessary. This is an important distinction because waters placed by modeling, e.g., after dynamics, can also be assessed. Furthermore, it allows Relevance to be used as a metric for map-based searches that computationally hydrate pockets, surfaces, etc.[168,263]

AcquaAlta[264] is a knowledge-based system constructed from detailed analyses of small molecule crystal structures in the Cambridge Structural Database.[265] The AcquaAlta algorithm catalogued the preferred positions and orientations of water molecules as they interact with simplified functional groups such as carbonyl, which includes all aldehydes, amides, carboxylic acids, esters and ketones. A hydration-propensity ranking was established by water interaction energies from *ab initio* calculations on those functional groups. The algorithm then places bridging waters between proteins and their bound ligands based on these propensity rankings. wPMF (water Potential of Mean Forces) is a method based on 3946 non-redundant high resolution X-ray structures from which the structure pattern of water and residue hydrophilicities were extracted.[266] This specialized PMF can be applied to predict the potential hydration sites in protein structures with an accuracy of about 80%. Furthermore, wPMF can also assess whether or not a given water molecule should be targeted for displacement in ligand design, with comparable performances to 3D-RISM (*vide infra*).

5.2. Static molecular mechanics methods. GRID (1985)[267] is one of the oldest programs used to model and understand water molecules in protein environments. Peter Goodford developed the program as a means to identify energetically likely locations for water in protein structures to supplement/validate those identified by X-ray crystallography, which at the time, due to instrumental and computing restraints, often provided incomplete pictures of solvation. Simply, GRID places an exquisitely tuned water probe at the intersection points of grid lattices inside and on the surface of the protein. The water probe is spherical, with an H_2O van der Waals radius and can donate and accept hydrogen bonds. Later, probes for other small molecules and functional groups and a non-molecular "hydrophobic probe" were added. With continual updating, GRID is both still in use, e.g., for stabilizing molecular dynamics simulations,[268] and the core technology of other programs offered by Molecular Discovery, Ltd.,[269] such as WaterFLAP, implemented within the FLAP software.[270] This program predicts water location by taking the energetic minima from the GRID OH2 molecular interaction field followed by subsequent local

optimization. Waters can be iteratively added to a protein target (either unliganded or in ligand optimization scenarios) until the defined site is completely solvated. The free energy of binding for the final water network can be evaluated with the OH2, CRY and ENTR probes that represent hydrophilic, lipophilic, and entropic terms, respectively. Waters can then be classified as structural, displaceable, solvent-like, or as being in low-occupancy locations.

The Multiple Copy Simultaneous Search (MCSS) method from the Karplus group[271,272] differs from GRID in that it randomly places probe molecules at predefined density within a predefined radius around, for example, the pocket centroid. The protein+probe ensemble is subjected to energy minimization using a standard force field commonly used in molecular mechanics. After various pruning operations, the remaining probes represent an energetically reasonable, although purely enthalpic, set of potential locations for that probe molecule in or around the protein. The water probe has been shown to predict positions of bound water molecules quite well.[273] A recent article suggested that surveying a volume in an unliganded pocket containing a collection of water molecules with MCSS could, with multiple probe types, reveal design clues for functional groups to displace those waters.[274]

Docking technologies have also been utilized to study water, especially with respect to the water influence on protein-ligand interactions.[275-280] The ability to accurately and efficiently treat ordered waters has been a major limitation of molecular docking because of the number of states to be investigated. If just four key waters are handled in the most rudimentary way, by toggling them on and off, the docking problem escalates 16-fold. Also, as seen repeatedly above, there are many factors informing the decision of displace vs. retain (or sometimes add) for each water molecule.

The Glide docking program[281-283] from Schrödinger[284] includes implicit solvation in its scoring functions, but deals with explicit water molecules only as part of the site, where users can manually toggle their inclusion in or exclusion from the model. In an extension of FlexX, Rarey *et al.* described a "particle concept" to place explicit waters at precomputed positions if they can form additional hydrogen bonds to the ligand and optimize the ligand orientation in the active site.[285] As an adjunct to the AutoDock Vina docking program,[286] the WaterDock protocol,[287] including a probabilistic water classifier, was developed to place molecules into protein binding sites and evaluate their ease of displacement.

In contrast, Forli and Olson implemented a *ligand-based* hydration model to explicitly account for water during docking [Forli and Olson. A force field with displaceable waters and

desolvation entropy for hydrated ligand docking. J Med Chem 2012, 55(2), 623-638; Forli et al. Computational protein-ligand docking and virtual drug screening with the AutoDock suite, Nature protocols, 2016, 11, 905] in the AutoDock force field [Huey et al. A semiempirical free energy force field with charge- based desolvation. J Comput Chem. 2007; 28:1145–1152]. No prior knowledge of waters at the protein-ligand interface or in the unbound structure is required, since waters are attached to the ligands' polar atoms before docking and their presence and contribution are iteratively evaluated during the search. Both enthalpic and entropic contributions to each water molecule's energetics are considered with a smooth potential sampled during the conformational search. Each water's fate is determined by the balance between its contribution to ligand binding and the hypothetical ligand stabilization gained by its displacement. This approach is based on the facts that hydration patterns change in response to the ligand binding and that not every water in a binding site can interact with every ligand. Only those waters interacting with a specific ligand are modeled and thus no bias from previously determined structures is retained. It would seem to be particularly suitable when no information about the hydration in a binding site is known such as when experimental conditions do not allow the resolution of water positions. Altogether, water-aware docking approaches have improved the fidelity of some ligand geometries, while in other cases, the geometries deteriorated.

5.3. Monte Carlo and molecular dynamics (MD) methods. Monte Carlo (MC) and molecular dynamics (MD) methods are standard tools for simulation studies for modeling protein folding[288,289] and other biomacromolecular processes. MC or MD simulation can be conducted to equilibrate the distribution of water molecules in a binding site. For example, Monte Carlo techniques can be used to simulate the expected probability density of atom-atom contacts, e.g., solvation. Through exhaustive sampling of structure space with random probes, Rakhmanov and Makeev produced knowledge-based potentials for the hydration of protein atoms.[290] This model was able to estimate the atomic hydropathy, quantitatively and in a distance dependent manner, for all types of atoms found in proteins.

The replica exchange thermodynamic integration method (RETI)[291,292] is a Monte Carlo-based method for calculating binding energies. Essex's group incorporated double decoupling[79,293] to yield a more practical methodology for determining the free energy of water binding to proteins and protein-ligand systems in a set containing both displaceable and non-displaceable waters.[80] RETI treats all waters explicitly in all-atom models, which can both

translate and rotate. The protein also is allowed limited flexibility. Another water molecule is used as a probe contained inside an impenetrable sphere. To determine the absolute free energy of water binding, two Monte Carlo simulations are performed simultaneously – the double decoupling: the first simulation is for a water molecule decoupled from the (bulk) solvent (yielding ΔG_{hyd}) and the second is for a water molecule decoupled from the (receptor) protein but held in-place (yielding ΔG_{dec}). To separate van der Waals from electrostatic contributions, calculations that alternately annihilate the van der Waals and electrostatic terms are performed. A key point about RETI is that the method makes fewer assumptions than most others, and energies found using it have been used as a standard.[294] However, the RETI method is very slow. Just Add Water Molecules (JAWS) from the Jorgensen group,[24] also applies the double decoupling approach to compare the removal of a water molecule from the bulk and from the binding site, resulting in an estimate for ΔG_{bind} for that water molecule. JAWS, however, applies a number of well-validated approximations in its more accessible methodology.

MD simulations can be combined with inhomogeneous fluid solvation theory (IFST) to evaluate the binding enthalpies and entropies for interfacial water molecules in protein–ligand complexes.[101,102,255] IFST treats the solute (protein and ligand) as spatially fixed, i.e., an inhomogeneous fluid, and calculates both the solvation energy and entropy by integrating over the solvent-occupied volume. A freely available software package based on MD/IFST is Li and Lazaridis's Solvation Thermodynamics of Ordered Water (STOW).[295] Their implementation used molecular dynamics simulations, e.g., with CHARMM,[296] to sample water configurations. While STOW computes the thermodynamic contributions of individual water molecules to that of the overall complex, it does not estimate the free energy for the liberation of water molecules displaced by ligands. BiKi Hydra[297] is another MD simulations-based tool very recently developed to analyze hydration patterns and assess the persistence of water molecules inside specific binding sites. The methodology quickly estimates water persistence in a given region by using dehydrating bias, acting on a collective variable that mimics the interaction energy of charged particles in an electrolytic solution modeled according to Debye-Hückel theory. Hydra has been validated on the adenosine A_{2A} receptor and tested on other members of the GPCR A-family, providing results in good agreement with experimental data.[298]

The proprietary WaterMap suite[253,299] from Schrödinger[284] builds on the IFST implementation of STOW, with extensions that both locate all water molecules in a protein binding site and evaluate the favorability of their displacement.[253,299] To predict the energetic

contributions of water molecules in protein active sites, WaterMap uses trajectories generated by MD simulations in explicit water; the water molecules surrounding the rigid protein are clustered to obtain hydration sites. Thermodynamic profiles of these hydration sites are then calculated by averaging the solvent-solvent and protein-solvent interaction energies, based on inhomogeneous solvation theory. Water molecules that reveal a significant positive free energy relative to their being in bulk are termed 'unhappy', which suggests a significant free energy gain upon their displacement by a ligand. The positions of 'unhappy' water clusters in a site may reveal hot spots for small molecule ligand binding. With its custom "displaced solvent functional", WaterMap can estimate the liberation free energy for water molecules displaced by a suitable ligand. This functional assumes that water molecules are displaced into a cavity in bulk solvent (previously occupied by the ligand), resulting in the formation of a cavity of identical size and shape in the protein. This is, in other words, an attempt to model the hydrophobic effect, and estimate the associated energetics.

Recently, WaterMap has been used successfully for ligand optimization and for the explanation of the structure-activity relationships in several series of compounds.[253,300-304] WaterMap was also found to ameliorate some of the failings with respect to estimating protein desolvation in the otherwise robust MM-GB/SA scoring functions (molecular mechanics (MM) with Generalized Born (GB) and a hydrophobic solvent accessible surface area (SA) term).[305] In particular, replacing the protein desolvation terms in MM-GB/SA by WaterMap estimates for free energies arising from water displacement after ligand binding provides good results for ranking binding in congeneric series. The most recent Schrödinger development for WaterMap is a docking and scoring method, WScore, which includes an MM-GBSA scoring component, and provides flexible treatment of explicit water molecules.[306] The locations and thermodynamic features of these water molecules are derived from a WaterMap molecular dynamics simulation of the complex that, in turn, supplies atom-level ligand and protein desolvation energies.

5.4. Statistical mechanics methods. Proteins are certainly flexible and the fluctuations of water molecules under real conditions – including changing occupancy, changing hydrogen bonding patterns and other features, especially on protein surfaces – are well documented. Thus, statistical mechanics approaches are a natural fit for evaluating protein hydration. In this context, Beglov and Roux developed a statistical mechanical integral equation theory, now known as 3D-RISM (an extension of the reference interaction site model (RISM) to three dimensions), to model

hydration of protein-ligand-solvent complexes.[307] Their equation is constructed on the basis of the density functional theory of non-uniform polyatomic liquids[308] and is applicable to solvent interactions sites at all points around molecular solutes of arbitrary shape. An important feature is that 3D-RISM can represent water molecules as being non-spherical so that hydrogen bonding is more effectively modeled. 3D-RISM methods include both solute-solvent electrostatic and van der Waals interactions. An implementation of 3D-RISM[309] method is available in the MOE (Molecular Operating Environment) software suite from Chemical Computing Group Inc.[310] The idea behind this method is that waters prefer to have a certain distribution of distances and angles amongst themselves and that polar groups on the protein essentially are trying to mimic waters. The 3D-RISM methods offer two advantages: 1) they are fast compared to many other methods that calculate maps of water locations, and 2) information about hydrogen bond networks formed by water molecules can be captured and exploited.

An alternate statistical mechanics-based approach, SPAM ("maps" spelled backwards), was developed by Cui *et al.* of GlaxoSmithKline.[311] SPAM computes the locations of hydration sites around a protein following a nanosecond+ MD simulation with explicit solvent. The interaction energies of water molecules that travel through potential hydration sites are extracted, and the free energies of these hydration sites (or putative water molecules) are computed, neglecting water-water correlation, by a site partition function. A number of constraints are imposed: the protein conformation is locked by a harmonic potential on its heavy atoms. Finally, hydration sites are identified as peaks in the computed water density map relative to the density of bulk water. This method yields the distribution of interaction energies between water and the surrounding environment at a binding site, and in particular, perturbations to this distribution after events such as ligand binding. SPAM was developed, using simple and accessible statistical mechanics, to be a design tool that can identify hot spots in a target by extracting both average interaction energy and entropy information for discrete water molecules.

5.5. Methods using continuum solvent models. Another class of water methods relies on continuum solvent models such as the Poisson-Boltzmann (PB) or GB equations as the primary compute engine, thus taking advantage of implicit solvation. Some of the methods described above also incorporated elements of PB or GB in their protocols. There is a long-standing debate concerning which of the two continuum models, PB or GB, is preferable: the PB equation has solid theoretical justification, but is computationally expensive. It describes the electrostatic

environment of a solute in a solvent containing ions. PB equation solvers such as DelPhi from the Honig group,[312] Zap from OpenEye[313,314] and others[315,316] have been developed with varying generality and efficiency. The GB model results from an approximation to the exact PB equation (in its linear form) where the solute is modeled as a set of spheres that have dielectric constant different from the external solvent. A key GB parameter is the effective Born radius atoms,[317] which represent their degree of burial inside the solute, i.e., the distance from the atom to the molecular surface. Despite many innovations, speed still remains an issue with PB solvers: performances have not equaled those of the simpler and more commonly used GB approximation, and surprisingly, GB-based results are often quite similar[318,319] and can be superior.[320,321]

Applying the continuum models to characterizing single water molecules seems at first to be almost counterproductive because the point of continuum models is to ameliorate the need for explicit solvent. However, the continuum physics can supplant the less interesting water molecules, and allow a focus on the water molecules of relevance in the system. Both Poisson-Boltzmann[322-325] and Generalized Born[326,327] models have been implemented in this way. However, in high-dielectric environments (where hydrogen bonds are of interest), and in low-dielectric environments like protein binding pockets and membrane interiors, GB-based calculations may be less accurate than other methods.[328]

SZMAP (Solvent Zap MAP) by OpenEye[314] is a fairly new program that combines features from multiple methods. Like RETI,[80] it uses a water probe that consists of three atoms, but it can only be rotated and not translated. Protein atoms are also fixed. The other water molecules in the system are treated implicitly as a Poisson-Boltzmann solvent with Zap[313] surface energy terms. By rotating the probe in many different directions, and calculating energy at each stop, evaluation of the ensemble of these energies yields thermodynamic quantities for the probe's site. SZMAP also simultaneously calculates the thermodynamics for both neutral (water without partial charges) and vacuum (empty water-sized bubble) reference probes. These reference states allow the calculation of thermodynamic properties of water relative to easily imaginable alternatives: e.g., the vacuum probe reveals the differences between a hydrated and an empty site, and the neutral probe reveals the differences between a hydrated site and one occupied by a hydrophobic group. While SZMAP can be used to generate maps of hydrophobic and hydrophilic regions of protein surfaces and reproduce positions of crystallographic waters,[294] it is not as fast as GRID or 3D-RISM. Also, because the method is designed to treat only one water explicitly, SZMAP has difficulty with water networks and can miss the intricate hydrogen bonding patterns in water

clusters.

However, SZMAP quickly and accurately predicts whether a specific crystallographic water molecule is easily displaceable from within a ligand-bound protein pocket, e.g., by "growing" a group onto a nearby ligand – such as in lead optimization. The thermodynamic quantity that is most predictive is the loss of rotational entropy by the water probe due to charge interactions. Waters that lose a lot are hard to displace in this way, while crystallographic waters that *retain* much of their rotational entropy due to charge interactions fall in two groups. More common are waters in hydrophobic pockets, where electrostatic forces are weak. These hot waters have long been known to be easy to displace (*vide supra*), as is predicted by most energy-based methods,[80,329] including the SZMAP energy differences between water and neutral probes.[294] Another, previously underappreciated, class of crystallographic waters that retain rotational entropy are those highly constrained by van der Waals contacts. Even in strong electrostatic fields, switching off the partial charges on these waters (i.e., via the neutral SZMAP probe) does not affect their rotation because they would bump into neighboring groups. These waters are typically displaced by polar ligand groups that mimic the hydrogen bonds they previously had with the protein, so their displacement has no enthalpic penalty. However, once in bulk water, they may get an opportunity to rotate freely, thus gaining entropy.

5.6. Which is the best computational method? The discussion above should make it clear that there are many computational approaches to studying the water problem and most have strengths and more than a few have weaknesses. A clear trend is that, concomitant with the vastly increased access to computing power, the computational tools have become more sophisticated, account for more effects, and purport to be more accurate and precise. Although we categorized the methods into discrete groups, few of them are not influenced or do not incorporate concepts from earlier programs. A better question to ask than "which is best" is, what tools are likely to be used in the future? In that view, Morris's WaterDock,[287] an extension to the widely used and freely available AutoDock, seems like a good bet. The WaterMap method of Schrödinger has appeared in numerous publications since its introduction,[300,330-332] and regular updates indicate that it is an active product. OpenEye's SZMAP product is also of note – it uses a state-of-the-art Poisson-Boltzmann solver in support of a new thermodynamic view of solvation and desolvation. It also is an active product.

However, sometimes simpler can be good too: the GRID program[267] and its successor

WaterFLAP as part of FLAP,[270] are also capable of high accuracy in predicting loci for water molecules in or around a protein and in classifying waters as happy (cold) and unhappy (hot), according to their structural, displaceable and solvent-like character. The Rank, HINT score and Relevance tools we developed[168,181-183] were 90% accurate in predicting conservation/displacement when the target structure was of high resolution. It is also important for medicinal chemists that the results be interpretable and can be fed into their drug discovery efforts. Such a strategy seems to be the purpose behind the OpenEye *gameplan* application.

6. EXPLOITATION OF WATER IN DRUG DISCOVERY: CASE STUDIES

6.1. Src Kinase : bosunitib. The human kinome contains over 500 enzymes that are mostly involved in signal transduction and cell proliferation. With these roles, it is not surprising that multiple kinases have been validated as targets of drugs to treat tumors.[333,334] However, the sheer number of kinases renders targeting them challenging. Thus, to produce high affinity compounds that also possess specificity, a variety of different strategies have been proposed and/or exploited.[335-339] To date, however, most kinase inhibitor drugs bind to the ATP site, which is strongly conserved across the kinome. Inhibitors are known that bind either to the DFG-in conformation (i.e., type I inhibitors) or to the DFG-out conformation (type II), which correspond to active and inactive kinase forms, respectively.

One particularly successful kinase drug is imatinib (Gleevec, **1**, Figure 10A)[340] that targets the aberrantly activated tyrosine kinase BCR-Abl, which has been associated with chronic myelogenous leukemia. Some patients with more advanced progression or acute lymphoblastic leukemia respond poorly to imatinib. Thus, alternative chemical agents were investigated, such as dasatinib, nilotinib and, in particular, a dual Src-Abl inhibitor, SKI-606 (**2**, Figure 10A), which was first reported in 2001,[341] and subsequently developed into the clinical candidate bosutinib.[342-345] Molecular modeling studies suggested that bosutinib's superior activity was due to its ability to bind a BCR-Abl conformation different from that of imatinib,[343] and later crystallographic and spectroscopic analyses[346] confirmed this hypothesis. The 2.4 Å structure of bosutinib bound to Abl shows an inactive DFG-out conformation, whereas the drug bound to Src shows an *active DFG-in* conformation (see Figure 10B). Interestingly, except for the DFG motif portion, the conformation of the activation loop in the bosutinib structure is similar to that of active kinases.

Because it makes limited contact with this activation loop, bosutinib can be accommodated by *both* DFG-in and DFG-out protein conformations, even when tyrosine 393 in the activation loop is phosphorylated. Since this phosphorylation stabilizes the DFG-in conformation, imatinib is unable to bind, which explains the activity of bosutinib against imatinib-resistant mutants of Abl.[346]

Other important questions concerning kinase inhibitors and their selectivity at ATP sites required an even more detailed investigation of structure – in particular, the roles of water. Levinson and Boxer elegantly coupled crystallographic and spectroscopic analyses to study the interaction of bosunitib with Src kinases, as well as other kinases.[193] First, the function of a residue on the kinase hinge region, commonly annotated as the gatekeeper, protects a deep pocket inside the ATP site. Molecules that pass the gatekeeper and bind in this region are generally selective. Bosutinib is selective for kinases possessing a threonine gatekeeper and ineffective against, e.g., the T315I mutation of the BCR-Abl gatekeeper.[347] More interestingly, the structure of the Abl:bosutinib complex has a cavity adjacent to this gatekeeper residue that is large enough for other, more bulky gatekeepers in the inactive DFG-out conformation.[346] This feature was conserved in their 2.1 Å structure of the (active) DFG-in Src:bosutinib complex.[193] Two conserved water molecules, **w1** and **w2**, are located in this unique cavity (see Figure 10C).

The bosutinib inhibitor forms a direct hydrogen bond (~0.4 kcal mol^{-1} by HINT) between its nitrile moiety and one water (**w1**). This is actually one of only two hydrogen bonds formed by the ligand in its interaction with the protein; the other is weaker (~0.2 kcal mol^{-1}) and between the nitrile and the Thr338 gatekeeper. In fact, bosutinib makes few favorable interactions with the site: HINT reports that, absent the support from bridging water, the favorable polar interactions (~0.6 kcal mol^{-1}) and favorable hydrophobic (~3.1 kcal mol^{-1}) are overwhelmed by unfavorable polar (~1.7 kcal mol^{-1}) and hydrophobic-polar (~3.3 kcal mol^{-1}) interactions. In particular, the chloro substituents of bosutinib are both in proximity to carboxylates of Glu310 or Asp404, which suggests that a sizeable dose of entropy from displacement of hot waters may also be at play in bosutinib's binding. The second conserved water molecule, **w2**, is seen to be bridging **w1** to Glu310 (~1.0 kcal mol^{-1}) as well as interacting very weakly with the DFG loop (Asp404-Phe405-Gly406). Mutations of the gatekeeper, which are commonly observed in the aforementioned BCR-Abl, prevent access to the cavity and the **w1**-specific hydrogen bonds. These mutations thus abolish bosunitib's ability to inhibit the enzyme. Altogether, a particularly interesting mechanism for kinase-inhibitor recognition was revealed: these two waters play a key role in dictating inhibitor-enzyme recognition and specificity. Following this observation, it was

recognized that the cavity and its two resident water molecules is a common motif amongst type I inhibitors and that ligand-to-water hydrogen bonding is also conserved, as it was observed in forty different kinases inhibited by 164 different ligands.[193]

6.2. α Human carbonic anhydrase : arylsulfonamides. The carbonic anhydrases comprise a family of metalloenzymes, usually containing a zinc ion coordinated to three histidines, which catalyzes the interconversion of carbon dioxide and water to bicarbonate ion and protons. In humans, this interconversion maintains the acid-base (pH) balance in blood and elsewhere in tissue. Carbonic anhydrases possess some of the fastest reaction rates among all enzymes, with its rate being diffusion limited; typical catalytic rates range between 10^4 and 10^6 s^{-1}.[348] Carbonic anhydrase inhibitors suppress this activity and are clinically used as antiglaucoma agents, diuretics, antiepileptics, or in treatment of altitude sickness, ulcers, gout, neurological disorders, osteoporosis and other indications.[349] George Whitesides and multiple members of his group have studied many aspects of the biochemistry, structure and inhibition of HCA over the past quarter century. A 2008 review[350] is particularly enlightening, and shows that HCA is one of the most chemically and structurally well-defined enzymatic systems known. It is conformationally rigid: extensive literature establishes that it does not undergo conformational changes (>1 Å) upon binding of typical (sulfonamide) ligands.[350] The sulfonamide moiety, through its (deprotonated) $-NH^-$, coordinates to the Zn^{+2} ion, and is further held in-place by several other hydrogen bonds. The sulfonamide ligands are, for all intents and purposes, conformationally immobile and the protein residues embracing these ligands are similarly stationary.

With this system, Whitesides and colleagues recently attempted to experimentally define the thermodynamics of the hydrophobic effect with systematic small chemical perturbations of the bound heterocyclic sulfonamides (**3**, Figure 11A).[110] In particular, benzo derivatives (**3e**-**3h**) were designed to "extend" the hydrophobic portions of the heteroaromatic ligands (**3a**-**3d**, respectively) further into the active site of HCA and near the hydrophobic wall (Phe131, Leu198, Pro201 and Pro202). X-ray crystallographic data were collected and solved for eight of the nine analogues (**3a**-**3h**) and ITC thermodynamic were obtained for all.[110] In addition, the structural models for four of the complexes (**3a**, **3b**, **3e** and **3f**) were subjected to molecular dynamics and WaterMap[253,299] analyses. This setup enabled isolation of the hydrophobic effect due to the benzo extensions. The ITC results showed that the contributions of enthalpy and entropy to the overall $\Delta\Delta G°$ from the benzo substitution were unexpected: the $\Delta\Delta H°$ is favorable (-3 ± 1 kcal mol^{-1}) and

the -TΔΔG° is slightly unfavorable (+1 ± 1 kcal mol^{-1}).[110] In this case, enthalpy is the dominant term in the hydrophobic effects, which is not unheard of, but since the fused cyclohexyl ring in **3i** showed essentially the same thermodynamic parameters as the fused benzo ring in **3f**, these observations are not due to "nonclassical" interactions between the aromatic rings and protein. Structural data further indicate that the atoms of the fused benzo ring are only weakly in contact with residues in the pocket (Figure 11B), which is confirmed by a HINT analysis that the monocyclic ring of **3d** has an interaction with protein of ~0.15 kcal mol^{-1}, while the fused benzo analogue **3h** has an interaction of ~0.50 kcal mol^{-1} – mostly due to an improvement of ~0.30 kcal mol^{-1} in hydrophobic interactions. This small difference is clearly insufficient to explain the observed thermodynamics. However, the thermochemical and structural data are consistent with the presence of additional ordered water molecules in the HCA–bicyclic ligand (**3e**-**3h**) complexes than in the non-extended ligand (**3a**-**3d**) complexes.[110] This indicates the hydrophobic effect in this system is driven by localization of water molecules, but not just those that are in contact with the hydrophobic parts of the ligand. The main conclusion is that the shape of the water in the binding cavity may be as important for the hydrophobic effect as the shape of the cavity.[20,110]

More recently,[260] the Whitesides group applied another subtle structural change – benzo-extended ligands **3h**, **3j** (and others) were fluorinated to yield, e.g., compound **3k** and **3l** (Figure 11A). It was presumed that the two series of compounds would have more or less the same sizes but significantly different electronic characters, which would allow study of the resulting thermodynamic parameters of binding along with changes in the active site's water molecule network. Remarkably, the free energies of binding (ΔG°) for all compounds to HCA are nearly indistinguishable (see Table 2). However, for **3h** and **3k**, the enthalpies of binding (ΔH°) and entropies of binding (–TΔS°) show opposite trends, i.e., H/SC, with **3h** being more *enthalpically* stable by 2.6 ± 1.1 kcal mol^{-1}, while **3k** is the more *entropically* stable by 2.1 ± 1.1 kcal mol^{-1}.[260] Drilling down into the specific effects of the fluorination showed only a minor ligand translation of 0.7 Å, but a substantial difference in the number of water molecules in the active site (Table 2), which can be explained by measured heat capacities. The difference between the benzo and fluorobenzo extensions (**3h** vs. **3k**) appears to be a consequence of differences in Coulombic interactions of each ligand with HCA (enthalpic) and differences in their solvation energies (entropic).

Also interesting is examination of the N-methylpyrrole pair of compounds (**3j** and **3l**) with the thiazole pair (**3h** and **3k**); the methyl ring substitution causes the heterocylic rings of **3j** and **3l**

to rotate about 30° out-of-plane. The virtually identical ΔG° of binding and the insignificant sidechain root mean square deviations (RMSDs) of ~0.2 Å confirms the part played by the aggregated and structured (i.e., cold) water molecules in binding thermodynamics.[260] Lastly, the thermodynamic parameters of binding for seven additional mono-, di-, and tri-fluoro derivatives of **3h** in a variety of configurations were evaluated.[90] Even with such small molecular perturbations, which yielded more or less the same ΔG° for binding as **3h** and **3j**, there were significant differences in their ΔH° and -TΔS° values (see Figure 11C), which can only be due to changes in the binding site's water network. This conclusion was supported by simulation, but as of yet there are no experimental structural data for the intermediately fluorinated ligand-bound complexes. To reemphasize a major point, the hydrophobic effect leading to binding of these ligands to HCA is not attributable to direct hydrophobic interactions between protein and the ligand, but is due to waters displaced into bulk from the binding pocket.

6.3. Hot waters identify druggable cavities in GPCRs : Adenosine A_{2A}. Cold water molecules have perhaps received more attention due to their fundamental role in protein-ligand binding, and thus in drug discovery/design. Their displacement and/or participation in bridging interactions make thermodynamic sense, notwithstanding that the underlying entropy and enthalpy may not be as first presumed. Hot waters seem to be more involved in supporting roles of protein motion and function; they are, by our definition, less stable water molecules that are "unhappy" in their position in or around the binding site, and should be easily displaced. Hot waters represent an easy way to improve the free energy of binding, as their removal can give both entropic and enthalpic gains, since the same waters in the bulk would be more flexible and able to form more hydrogen bonds. Such water molecules occupy what are termed hydrophobic hot spots.[254]

The adenosine receptor A_{2A} belongs to class A of the GPCR family and is involved in the regulation of a number of different physiological functions such as myocardial oxygen consumption and blood flow and glutamate and dopamine release in the brain. A_{2A} is ubiquitously expressed in humans and has been targeted for the development of agonists acting as anti-inflammatory agents and binding the active state of the receptor, and for antagonists that are able to stabilize the inactive conformation and thus used in the treatment of neurodegenerative disorders such as Parkinson's disease.[351] In the basal ganglia, the receptor counteracts the action of the dopamine D_2 receptor, thus, its inhibition emerged as a potential non-dopaminergic therapy

for this disorder.[352] A_{2A} is thus widely recognized as an excellent drug target.[353] In recent years, intensive crystallization efforts involving fusion proteins and thermostabilization have resulted in a number of GPCR structures in both active and inactive states, with close to 200 currently available in the PDB. This improved structural understanding now strongly supports structure-based drug discovery and design.[354] GPCR binding sites contain a highly variable number of water molecules that are often exploited by ligands to better target the receptor, and thus participate in the overall binding process. Yet, despite the progress in GPCR structural biology, the proper solvation of their models remains difficult and computational methods able to create and score water networks dynamically are very useful tools. Mason and co-workers have extensively investigated the water networks of the adenosine A_{2A} receptor and their perturbations resulting from ligand binding. A key observation is that lipophilic hot spots, often occupied by hot waters, can be key drivers for ligand design.[254] These waters are easily and favorably displaced by incoming ligands, thus improving the free energy of binding. This dewetting scenario (Section 4.1.3) may in many cases be a more successful strategy than attempting to replace cold waters, which, as seen above, is difficult to predict. Lipophilic hot spots containing hot waters were also found in the unliganded structure of both the C-C chemokine receptor type 5 (CCR5)[355] and the chemokine CXCR4 receptor.[356] Binding of maraviroc to CCR5 and IT1t to CXRC4 displaced the hot waters, and favorably contributed to binding.

In the adenosine A_{2A} case, molecular dynamic-based predictions of the A_{2A} binding site solvation did not succeed, even with a default and equilibrated box of TIP3P waters, and resulted in a dewetted binding pocket. In contrast, starting with a GRID (WaterFLAP)-based[269,270] or WaterMap-based[253,284,299] water placement, followed by MD, provided a more reliable representation of a solvated binding site. As expected, water rearrangements are dictated by the ligand nature and even small ligand modifications often modify the number and orientation of the surrounding waters.

Complexes formed by the A_{2A} receptor with potent leads from chromone and triazine series[357] were studied in more detail. Co-crystals of the A_{2A} selective receptor antagonist, ZM241385 (**4a**, Figure 12A)[358] have been obtained and a number of resulting structures have been reported.[137,359,360] From this, virtual screening discovered the first chromone derivative (**5a**, Figure 12A), which was later optimized with structure-based tools, MD simulations and biophysical techniques. In particular, the carboxylic moiety, lying in a hydrophobic hotspot surrounded by Ile66, Leu167and Met270, was removed for substitution with an ester, an acetate,

different acyl groups and, finally, by a propyl, giving the compound (chromone12, **4b**, Figure 12A) with the highest ligand efficiency. Simulations of the unliganded pseudo-apo form (modeled from the high resolution (inactive) A_{2A} crystal structure (pdb: 4eiy)[137]) receptor with GRID (WaterFLAP) and WaterMap calculations showed that hot waters occupying this region were easily displaced by the **4b** ligand (Figure 12B).[354] Similarly, a methyl group located on the thiazole ring on the opposite side of the molecule was located in a small lipophilic pocket, lined by Met177 and Leu249 (Figure 12B). The ligand **4b** showed a pK_i of 8.5, but its des-methyl derivative **4c** had a 33-fold lower affinity.[357] The decreased affinity is suggestive of this being a "magic methyl", as defined by Lunn.[361] Mason's analysis of water network perturbations[254] showed that both WaterFLAP and WaterMap placed a very hot water in the small lipophilic pocket in the pseudo-apo receptor. Its displacement upon binding of **4b** gave a positive entropic gain and a negative enthalpic gain, resulting in an overall better binding free energy. However, the removal of the magic methyl, i.e., with **5c**, induced slight movement of the ligand within the site and allowed that water's return to the site, as indicated by the X-ray structure (Figure 12B),[254] after only 100 ps of MD simulation.

In a second ligand class, structure-based studies identified the 1,3,5-triazine series as potent and selective adenosine A_{2A} receptor antagonists.[362] The binding mode of the compounds was first predicted by molecular modeling and BioPhysical Mapping (BPM) analyses,[363] and then confirmed by X-ray crystallography.[364] Starting from the commercial compound **5a** (Figure 12C), more potent derivatives were synthesized, with compound **5b** showing the highest activity, an observation supported by its, and the close analogue **5c**'s, binding modes (Figure 12D). Interestingly, **5b** is able to more directly interact with the sidechain of His278 via its chlorophenolic hydroxyl on the chloro-phenol substituent compared to **5c**, which pulls it about 1.2 Å deeper into the receptor. Here, again, WaterMap predictions suggest that these compounds displace a cluster of hot water molecules present in the unliganded form of the receptor. Closer evaluation of the water networks with WaterFLAP showed that significant changes in orientation and energetics might be attributed and correlated with the ligands' structure, potency and residence times. While **5b** directly interacts with His278, **5c** contacts the histidine thanks to a (likely cold) water molecule, not present in the **5b** structure (Figure 12D). Completely replacing the pyridine with a phenyl (i.e., **5d**) renders impossible the formation of any hydrogen bond and allows hot waters into the site where they are trapped at the protein-ligand interface. Residence time measurements for the ligands **5b**, **5c**, and **5d**, estimated to be 990 s, 87 s and 0 s,

respectively,[365] are in agreement with the increased number of hot waters at the protein-ligand interface. In summary, careful analyses (and predictions) of water network perturbations, with tools such as WaterMap, WaterFLAP and others, can yield insight into hot water structure, and the consequent effect on the binding free energy of ligands. An increased affinity can be explained by the displacement of hot waters, while a decreased affinity may be associated with hot waters trapped at the protein-ligand interface.

6.4. Thermolysin : hot water compensates for changes in ligand hydrophobicity. At its most visceral level, the terminology we use of "hot" water molecules refers to those that are labile and thus easily displaced. However, these waters also play many other roles as described in Section 4.1, and can be characterized as suggested in Section 4.3. Water molecules that are solvent exposed, interacting wholly or mostly with other water molecules (i.e., not with protein or ligand), or can facilely rearrange their positions and/or bonding patterns with changes in ligand identity could also be considered hot. Klebe and colleagues have been investigating a tight congeneric series of thermolysin inhibitors in which only one substituent, located in a solvent accessible pocket is varied.[366] Thermolysin is a thermostable neutral metalloproteinase enzyme that is produced by *Bacillus thermoproteolyticus*. It contains one Zn^{+2} ion in the active site and four Ca^{+2} ions with structural roles. Thermolysin is remarkable for its thermal stability,[367] which appears to be due to the Ca^{+2} ions and a small cluster of N-terminal amino acids on its surface, and has thus been often used in experimental and computational structural studies. It is for the most part rigid, easily crystallized and generally yields X-ray crystal structures with resolutions of 1.1–1.6 Å.[366] Klebe and colleagues determined the complex crystal structures (resolution 1.14 to 1.45 Å) and measured the thermodynamic profiles for nine analogues based on a carboxybenzyl-Gly-(PO_2^-)-L-Leu-NH_2- scaffold (TLN1-TLN9, **6a**-**6i**, Figure 13A). All ligands superimpose nearly perfectly, except for the hydrocarbon substituent being varied. More interesting is that, while the solvent structure is also nearly identical at the conserved portions of the structures, these ligands induce dramatic changes in the waters surrounding the different substituent groups. Of further note is that the subpocket binding the unique **6a**-**6i** substituents is a shallow, bowl-shaped hydrophobic depression, and is thus very accessible to water molecules.

Because of this accessibility, the water molecules are free to individually adapt to each of the hydrophobic ligands' sidechains. This is manifested in water chains that encircle each ligand, and make few productive interactions with the ligands, only modest interactions with the protein,

but engage with each other rather robustly. Klebe's careful progression from methyl (**6a**) to phenylethyl (**6i)**[366] tells a more complete story than we can here, but looking at a few of the nine is very instructive. The terminal methyl group of **6a** is framed by six water molecules within 4.0 Å of it (Figure 13B, panel 1). They are arranged in a vaguely hexagonal open chain possessing inter-water distances of 2.6–3.1 Å, with the two terminal waters of the chain forming hydrogen bonds with the Asn111 backbone carbonyl oxygen and the Asn112 sidechain ND2. The most conserved waters across the series are numbered. The **6b** ethyl derivative (Figure 13B, panel 2) engineers the removal of two water molecules from the methyl case (**w"** and **w'"**), but two new water molecules, **w4** and a water that caps the network that we are labeling **w0**, are within 3.5-4.3 Å. Ligands **6c** and **6d** are continuing variations on this theme,[366] but the 2-methylbutyl substituent of **6e** buries its chiral carbon deeper into the pocket, leaving the ethyl (and especially *its* methyl) more exposed. An evolved, largely pentagonal, water network is thus created (Figure 13B, panel 3) with **w2** and **w3** shifting to new positions (**w2'** and **w3'**), **w0**, **w1**, **w5** and **w6** remaining unchanged in position, and **w4** expelled. Remnants of the pentagonal water chain are retained for **6f** and **6g**, but for **6h** (Figure 13B, panel 4) and **6i** only a portion of the original water chain structure survives. Many other water molecules can be crystallographically detected, but these are outside the normal van der Waals interaction range of the ligands. The **6i** case is complicated by the presence of at least three distinct phenylethyl conformations in the electron density map.[366]

Thermodynamic profiles[366,368] measured by ITC are shown for ligands **6a**-**6i** in Figure 13C. H/SC is clearly at work in this series. The ranges of ΔΔH and -TΔΔS are much larger than that of ΔΔG, but there are notable trends in all three: ΔG reaches a minimum for **6e** (the 2-methylbutyl analogue) and then rebounds to near its **6a** value for **6i**; ΔH is more negative than **6a** for **6b**-**6f** but more positive for **6g**-**6i**, while –TΔS shows essentially the opposite trends. Buried hydrophobic surface areas increase from **6a** to **6e,**[366] which conflicts with the classical view that increased burial should correlate with increased entropy due to water displacement. Ligands **6f** to **6i** have buried hydrophobic surface areas nearly identical to **6e** probably because **6e** has already buried as much hydrophobic structure as room is available in the pocket. HINT calculations support this view: the hydrophobic-hydrophobic component of the HINT scores (H_{HH}) for molecular fragments containing the substituents increase gradually from -0.38 kcal mol^{-1} for **6a** to -1.16 kcal mol^{-1} for **6f,** and then decrease to -0.56 kcal mol^{-1} for **7i** (with **6h** being intermediate between **6e** and **6f**). As a consequence, an increasing amount of hydrophobic surface is exposed by each ligand in the

series (except for **6h**, apparently more similar to **6e**), which in turn highlights the roles of water molecules and their rearrangements surrounding these groups.

Binding of ligands locally alters the solvent-accessible surface of the complex[368,369] and changes the relationship amongst the protein, ligand and surrounding solvent, with either stabilization or destabilization of water molecules being possible. A stabilization is enthalpically beneficial and entropically detrimental as more waters become ordered; destabilization of these solvent molecules should produce an enthalpic penalty through loss of hydrogen bonding or Coulombic interactions coupled with an entropic gain in degrees of freedom. As noted above, the smaller ligands, i.e., **6a**-**6e**, engineered an impressive water cage, which had the effect of "cooling" a number of water molecules that would otherwise be in bulk and possess higher mobility. Hydrophobic (or alternatively van der Waals) interactions between the substituents and the pocket explain the enthalpic behavior. The clear message from Klebe's insightful study is that one-dimensional metrics that represent changes due to hydrophobic substitutions are often not sufficient in the three-dimensional solvated world of biological structure and function. The formation or destruction of "perfect" water networks, regardless of whether they are cold or hot, delivers a formidable thermodynamic force. It is also significant to note that hot waters in this case, as they cool to form the water network, are providing enthalpic as well as the expected entropic contributions to binding free energy.

7. WHERE THERE IS WATER, THERE IS OPPORTUNITY

On the macro scale there is likely not a more true statement than that above – since life cannot exist without water, and the opportunities proffered by the presence of water have, for humans, led to our form of civilization. For medicinal chemists, opportunities for drug discovery and design are also enhanced by the presence of water in and around an active site. Despite its simple molecular structure, or perhaps because of it, water can perform numerous, and sometimes unexpected, structural and chemical roles in biology. Even more interesting is the energetic versatility of water: water molecules impact binding free energy both enthalpically and entropically – in fascinating and situation-specific ways. In the paragraphs above, we outlined and described a wide number of such scenarios and showed how exploiting water has engineered new, highly active and selective small molecules, of which some are in clinical use. It should be clear, however, that the understanding of water roles in the underlying structural ligand-protein

complexes came at the expense of very detailed and careful dissections of the relevant scenarios with crystallographic, isothermal calorimetric, spectroscopic and computational experiments. Which means, of course, that key aspects of these lovely cases were seldom, if at all, predicted *de novo*. In a recent Perspective, the Roche Pharmaceutical Research and Early Development group[370] noted that calculation of water positions and associated energies, especially those involving entropy, often cannot be validated with experiment, and are often most valuable as *post hoc* rationalizations. Thus, drug discovery hypotheses based on such results and analyses are quite difficult to formalize and apply.

Nevertheless, there is much that can be learned by paying attention to the water molecules in a structure. We proposed here the somewhat simplistic terminology of "hot" and "cold" waters as a starting place to define the roles and potential exploitability of individual water molecules. It is, for the most part, similar to another phraseology – "unhappy" and "happy" waters[299] – but has the advantages of providing a language with which to deal with intermediate cases, which is probably lacking on an emotion-based scale, and having a scope for quantitation. While the language of describing water as hot and cold is fairly clear, the actual definitions *do* remain somewhat ambiguous. The authors of this Perspective had several discussions about characterizing the waters in some of the cases above before reaching consensus on terminology. Section 4.3 offered suggestions of about a dozen methods to take the temperature of a water molecule. There are certainly more. It was by design that the progression of methods outlined in section 4.3 started from those most experimentally based and ended with those relying most on computational tools. Correspondingly, the reliability of water "temperature" estimates should be considered in the context of the "quality" of the structural data. Clearly, with a high-resolution X-ray crystallographic structure, where care was taken to properly assign relevant electron density to water molecules, the waters' B-factors and occupancies are very good indicators of temperature. These B-factors should be examined with respect to the overall structure's B-factor to determine relative movement/flexibility or, in our vernacular, temperature. Similar trust can be placed in NMR measurements of residence times or exchange rates, but these data are more difficult to obtain for water in protein environments and subject to the general limitations of NMR experiments.

One overarching goal of water-aware drug discovery/design is to determine which water molecules are the actors in the ligand binding drama and which are bystanders. This is not necessarily the same as waters being hot or cold: indeed cold waters could just as easily be

bystanders or key participants. The true actors must be considered as explicitly as the protein and ligand. Much can be learned from comparing the X-ray crystal structures of protein-ligand complexes and their unliganded counterparts, when available. Tracking the "before" and "after" roles of water molecules certainly aids in determining those important from those unimportant to the ligand binding scheme employed by the protein, which are data also illuminating for drug optimization experiments. Similarly, when possible, comparisons of water structure in models of complexes where different ligands are bound to the same enzyme or receptor can differentiate between conserved and casual water molecules. This latter "trick" was applied by Levinson and Boxer,[193] Whitesides,[90,110] Mason[254] and Klebe,[366] in all four of the case studies above. However, two of those stories, for human carbonic anhydrase and thermolysin, were largely *post facto* and academically driven and executed. This kind of exhaustive study is an expense of time and resources that is not likely to be borne by the pharma industry, so the lessons learned from such studies must be translated to new situations.

This is the place for computational tools. Ideally, such tools would already detect the energetic and thermodynamic profiles displayed by all of the scenarios above and many others, or such enhancements would be forthcoming as the software evolves. However, this is all so **very** complicated! We mentioned in the Introduction that there are many moving parts in a protein-small molecule binding event, and some phenomena like the hydrophobic effect are emergent properties of the entire ensemble. How much can reasonably and accessibly simulated? There is significant and unquestionable value in computational analyses that explicitly consider water and its roles. We reviewed the current status of such software in section 5 of the Perspective. However, the Roche group makes a strong argument that the "best practice" in molecular design is to focus on clarity, simplicity, and good experimental design.[370] In this view, it is particularly important that medicinal chemists use computational tools to *benefit* their creativity,[371] and not as a replacement for it. Thus, simple tools and computational models can be good – if their results inspire a great experiment. It is also encouraging that the emerging drug discovery mindset includes a reemphasis on structure-based tools and approaches,[372] and the explicit consideration of water is now a key feature of SBDD.

Lastly, returning to the water molecule's view of protein/small molecule associations, a number of generalizations and design suggestions can be made: 1) displacement of hot waters (often in hydrophobic environments) is usually a good call, and probably unavoidable in many situations – their interactions with protein (and/or ligand) are weak, and easily reprised or

improved in bulk or manifesting a gain in entropy; 2) displacement of cold waters is more difficult and often costly – the gain in entropy may not offset the loss in enthalpy – and should be executed with some caution; which leads to 3) the amphiprotic and other unique properties of water molecules yield enthalpy/entropy compensation – that often makes detailed drug design efforts seemingly futile because ΔG changes little even though the underlying ΔH and TΔS values are changing dramatically; 4) the importance of maintaining a quality water network around the surface of a protein–ligand complex is profound on the overall stability of a complex.[344] In fact, hot waters can contribute in this way by encaging hydrophobic substituents, even in exposed crevices; 5) using cold water molecules *within* the active site to bridge between proteins and appropriate ligand functional groups can also be a very effective design strategy, and in some cases may be more thermodynamically favorable than displacing them; 6) the presence of water at protein-protein and other interfaces, whether hot or cold, suggests novel opportunities for drugs designed to disrupt these associations; and 7) some of the less obvious roles of water, such as its influence on the complex's overall dielectric (sum of individual water molecule's dipoles) may also be exploitable in novel, creative and productive ways.

Figure Captions

Figure 1. Water molecules in HIV-1 protease. A) In the unliganded protease (pdb: 1g6l), the catalytic water (**300**), five structural waters (**301**, **313**, **313'**, **313bis** and **313bis'**) and a number of (hot) waters appearing to simply occupy the active site (**a**-**g**) were located by X-ray crystallography; B) The pseudosymmetric peptidomimetic inhibitor CGP 53820 displaces waters **300** and **a**-**g** from the protease active site (pdb: 1hih). Its hydroxyl group fills the role of water **300**, while the others are generally displaced sterically; C) The cyclic sulfamide ligand AHA001 (pdb: 1ajx) further displaces water **301** by using its sulfoxide oxygens to act as hydrogen bond acceptors from the backbone NH groups of Ile50 and Ile50', a role previously played by water **301**.

Figure 2. Various roles of water in the biological environment. (General color-coding: green – hydrophobic; red – hydrogen bond acceptor; blue – hydrogen bond donor.) A) Water molecules act as functional group mimics; B) Water molecules act as extenders – either Lewis acid or Lewis base properties can be extended by about 3.0 Å to form an interaction; C) Water molecules can reverse Lewis acid or base properties, which is a way for water molecules to buffer changes in pH; D) Water molecules can occupy voids, e.g., in volumes created during protein folding; E) Water molecules can occupy voids between interacting protein surfaces; F) Water molecules solubilize small polar molecules and can charge separate organic and inorganic salts; G) Water molecules are reactants and/or catalysts in important biochemical reactions, such as the ATP hydrolysis reaction; H) Water molecules can act as a lubricant, e.g., allowing two hydrophobic surfaces to slide rather than associate; I) Water molecules can stabilize intermediate states and thus facilitate structural transformations, such as between the open turn β strand and the α helix. The reverse turn conformations are found in the region between those of β strand and α helix in Ramachandran ϕ-ψ space; J) Water acts as a medium for transport of protons (left-to-right) by alternately accepting a proton to form hydronium (H_3O^+) and donating it (or another proton) to a neighboring water molecule. Thus, a proton can move great distances through a channel containing such a "water wire".

Figure 3. The hydrophobic effect. A) Water molecules in bulk form extended hydrogen bonding networks (H-bonds shown as blue dotted lines); B) Polar molecules, i.e., those possessing hydrogen bond donors and/or acceptors, are facilely inserted into these networks; C)

Hydrophobic molecules may be surrounded or encaged by water molecules that form a network around them. The driving force (magenta arrows) for water molecules to maximize their number of polar interactions also, as a consequence, tends to push hydrophobic species together (green arrows); D) Ultimately, hydrophobic molecules congregate with an apparent hydrophobic "force" (diffuse green line); E) On a pairwise basis, van der Waals' (London) forces are energetically favorable for hydrophobic-hydrophobic interactions, as indicated by the methyl-methyl interaction shown here. The electrons in the orbitals of each methyl are polarized towards the nuclei of the other methyl, causing in effect an attractive force; F) In contrast, on a pairwise basis, Coulombic forces generally are energetically unfavorable for hydrophobic-hydrophobic interactions such as between two methyls, which are electropositive relative to heteroatoms. The total partial charge of atoms comprising most hydrophobic entities is small and positive, thus there likely is a small repulsive force between them.

Figure 4. Desolvation and molecule-molecule interactions; enthalpy-entropy compensation. (General color-coding: green – hydrophobic; red – hydrogen bond acceptor; blue – hydrogen bond donor.) A) A small molecule in a water matrix. Polar functional groups are hydrogen-bonded (blue dotted lines) to nearby water molecules; B) A protein "receptor" site or cavity filled by a number of water molecules that interact with acidic and basic residues on the site's interior to form hydrogen bonds; C) Water molecules under green hatch are easily stripped as they are only associated with other water molecules, and water molecules in red circles are potentially retained and/or may be removed with a desolvation energy cost; D) Removal of water from cavity/receptor site – waters marked with green X are easily stripped as they have few hydrogen bonds with other water molecules and none with the site, water molecules in red circles interact with the site and are potentially retained and/or may be removed with a desolvation energy cost, and waters in blue boxes are sterically redundant and only interact with other water molecules; E) The final complex, where two waters are retained in the site to bridge between protein and ligand, and hydrophobic interactions are indicated by diffuse green lines; F) Enthalpy-entropy compensation: 1 – compound (a) bound in site with moderately strong interactions (enthalpy) possesses a fair amount of internal entropy manifested by flexibility in the site, but a ligand modification that improves enthalpy (b) through an additional hydrogen bond (replacing methyl with hydroxyl) also can reduce that entropy, 2 – tightly bound compound (a) in site has little internal entropy, but a ligand modification (methyl substitution) that reduces enthalpy (b) by

degrading its fit in the site also can increase the ligand's internal entropy. Note that, in this case, the modified ligand also expels a water molecule from the site, which is a further confounding factor for thermodynamic analysis.

Figure 5. Hot water and cavity effects. (General color-coding: green – hydrophobic; red – hydrogen bond acceptor; blue – hydrogen bond donor.) A) Largely hydrophobic cavities can be sparsely populated by water molecules that are mobile (small arrows), and are stabilized by only a few inter-water hydrogen bonds. Ligand binding may release these hot waters to bulk (bold red arrows), where they are more stabilized by joining the polar network and engaging in hydrogen bonding; B) Water molecules that are mobile within a cavity with few inter-water hydrogen bonds before ligand binding may be stabilized by acting as bridging water molecules in the complex, and thereby engage in more productive hydrogen bonding; C) Water molecules that were very mobile prior to ligand binding (1) may reorganize (red arrows) in the resulting complex (2) such that they participate in a robust protein-ligand-water network even if they (red-shaded waters) are not actually bridging between or interacting with either the protein or ligand; D) Fairly large hydrophobic pockets, often cylindrical in shape, can possess small openings that block entry of water molecules (red arrows) and other polar species, but allow entry for hydrophobic ligands (large green arrow). Such pockets can be devoid of water molecules – even when unoccupied by the ligand – so that the ligand would not have wait outside for the pocket to empty through the small opening.

Figure 6. Dewetting of cavities; Protein-protein interactions. (General color-coding: green – hydrophobic; red – hydrogen bond acceptor; blue – hydrogen bond donor.) A) Large cavities (or protein surfaces/shallow pockets) have a more flat topography and water molecules at or near these surfaces, even if hydrophobic, are better able to network into large clusters with larger numbers of hydrogen bonds per water molecule. Thus, breaking up of these clusters suggests that dewetting is energetically disfavored; B) Narrow cavities or tunnels allow fewer hydrogen bonds per water molecule and dewetting is more energetically favored; C) Dewetting mechanisms can also use water channels and tunnels to the active site to both exchange (red arrows) water molecules between the site and bulk, and store (blue arrows) them near the site; D) Interactions and various roles of water at a protein-protein interface (two proteins highlighted in different shades of green): 1 – waters can bridge between proteins and possess high solvent

accessible surface area around periphery of interface, 2 – particularly strong hydrogen bonding interactions between proteins, often including shape complementarity, are sometimes termed "hot spots", 3 – waters buried within the interface can bridge between proteins with effective, favorable interactions with both, 4 – more than half of waters at protein-protein interfaces have favorable interactions with only one of the two proteins, 5 – hydrophobic interaction in interface between proteins, 6 – "hydrophobic bubbles" can be formed when one or more water molecules are trapped at the interface but possess no favorable interactions with either protein, and 7 – one of many other hydrogen bonding interactions between the proteins.

Figure 7. Hot water molecules in G protein-coupled receptors and ion channels. (General color-coding: green – hydrophobic; red – hydrogen bond acceptor; blue – hydrogen bond donor.) A) GPCRs in a deactivated state (A_1) do not allow the through-transport of water due to one or more hydrophobic layers (green box) formed by the hydrophobic interactions of residues from two or more helices. In particular, the two (magenta and orange) helices in the foreground have an interlocking pattern of hydrophobic residues. The water is a contributing driving force to the movement and reorientation of the helices (indicated by like-colored arrows) leading to the activated state (A_2). Some simulations and crystal structures suggest the formation of a channel that permits through-transport of water through the activated state GPCR; B) Two models proposed for ion conduction through a channel: 1 – "knock-on" model where K^+ ions (purple spheres) pass through the channel from intracellular side to extracellular side in a single file queue while alternately occupying the channel with K^+ ions and H_2O molecules the four sites (S1-S4) in the selectivity filter, 2 – "hard-knock" model does not require water for ion conduction, but relies completely on the Coulombic repulsion between K^+ ions for translocation.

Figure 8. Cold water molecules in protein structure. (General color-coding: green – hydrophobic; red – hydrogen bond acceptor; blue – hydrogen bond donor.) A) Protein structure (HIV-1, color coded by cavity depth: red (deepest) to green (most shallow); shades of blue indicate peaks) illustrating water molecules in key motifs: 1 – first-shell (possibly hot) on exterior of protein, 2 – in active site (possibly cold) that could potentially bridge ligand or be displaced, 3 – buried within active site and likely non-displaceable (likely cold), 4 – inaccessible (cold) and trapped in a small cavity; B) Using water molecules to tune specificity of ligands for active sites, i.e., in different, but related, proteins, with a common substrate (block shape) such as ATP. The

four active sites (B_1-B_4) are subtly different in terms of the character and shape of the site residues, but also have different water molecules in their active sites that can be used as design cues for substrate analogue ligands with high specificity; C) Conserved water molecules in binding site of fatty acid binding protein (FABP) can impact its substrate specificity. Also, these waters align their dipoles with the protein's electric field: large orange arrow – average protein electric field, small orange arrows – electric field calculated at each water's location, magenta arrows – dipole moments of each water.

Figure 9. Characterization of water "temperature". A) The crystallographic B-factor includes, among other more esoteric contributions, the mobility of an atom as defined by how much the electron density is smeared: 1 – small isotropic B-factor indicates low mobility and sharp positional definition corresponding to a water likely to be cold, 2 – anisotropic B-factors are seldom reported in protein structures but define more precisely the shape of the electron density envelope and preferred directions for mobility, 3 – an intermediate B-factor, 4 – high B-factor indicates large mobility and more uncertain atomic coordinates, corresponding to a water likely to be hot; B) occupancy of atom, when defined, reports cases where some fraction of the many protein structures in the crystal are not occupied by ligand or potentially water: 1 – full 100% occupancy, 2 – occupancy where two or more conformations due to symmetry or disorder suggests water may not be cold, 3 and 4 – low occupancy and very low occupancy, respectively, corresponds to low crystallographic certainty and perhaps a hot water; C) radiolytic footprinting can identify tightly bound waters associated with Phe, Tyr, Cys and Phe: $^{18}OH\cdot$ radicals are generated by X-rays, these radicals interact with water molecules at these residues, and the mass spectrometry signals for the ^{18}O modified peptides indicate whether the water was cold or hot; D) NMR-measured water molecule residence time: 1 – deeply buried (cold) water molecules have residence times on the millisecond to microsecond time scales, 2 – partially buried waters or those in surface pockets have microsecond to nanosecond residence times, 3 – surface water molecules have variable residence times in the tens of picoseconds, 4 – (hot) water in bulk solvent has even shorter residence times; E) NMR-measured proton exchange for D_2O/H_2O with amide ^{15}N. Slow exchange rates, which reveal two sharp singlet peaks, suggest association with a cold water; intermediate exchange rates display a single broad singlet; fast exchange rates with a single sharp singlet peak suggest hot water. Figure adapted from ref. 373; F) detailed crystallographic analysis of high-resolution structures has revealed preferred distances and torsion angles for hydration

sites associated with amino acid residues, similar to rotamer libraries for sidechain conformations. Water molecules in these preferred locations are likely to be cold, whereas those in non-preferred locations could possibly be hot. Distance dependence for Glu (red) and Trp (green) is plotted on the right and bottom axes; torsion angular dependence for Thr (blue) is plotted on the left and top axes. Figure adapted from ref. 239; G) flexibility in protein structure: 1 – more rigid residue sidechains (or those locked in-place by neighbors) are more likely to interact with cold water molecules, 2 – highly flexible sidechains like lysine are probably bound with more mobile, transient and hot waters; H) Solvent-accessible surface area (SASA) is the area of an atom, molecule, sidechain, protein or water exposed (or not occluded) to the (water solvent). It is defined as the van der Waals envelope (green surface) of the molecule expanded by the radius of the solvent sphere (red, usually the radius of water, 1.4 Å) about each atom center. The SASA (blue) is the surface generated by the center of the solvent sphere rolling on the vdW surface: 1 – low solvent accessibility suggests a cold water, 2 – intermediate solvent accessibility, 3 – high accessibility indicates an exposed, and likely to be hot, water; I) Rank calculates the count and quality of hydrogen bonds potentially made by a water molecule: 1 – high Rank water has optimal numbers of donor (2, blue) and acceptor (2, red) atoms, optimal heavy atom-heavy atom distances (2.8 Å) for hydrogen bonding, optimal (tetrahedral) geometry, and is likely to be cold, 2 – intermediate Rank water has less optimal set of doneptors (i.e., donors *or* acceptors) and geometry, 3 – water molecules not hydrogen bonded to the protein may be hot; J) The HINT score[182,183] for a water molecule calculates the hydropathic compatibility of each atom-atom pair in an interaction between the water and a pseudo-receptor of atoms surrounding the atom (green, hydrophobic, red H-bond acceptor, blue H-bond donor). The hydropathic atom constant (a) is the partial log $P_{octanol/water}$ for the atom, r is the distance between interacting atoms and R is an adaptation of the Lennard-Jones function: 1 – a strongly favorable interaction with four hydrogen bonds (black arrows), suggesting that the water is cold, 2 – an intermediate interaction with a mixture of favorable (black arrows) and unfavorable (red arrows) terms, 3 – a water molecule incompatible with its hydrophobic environment that is likely to be hot; K) The cavity volume encaging a single water molecule is related to its free energy of displacement: water molecules in hydrophobic (green) and polar (purple) cavities have different behaviors (see text). Data as reported by Yu and Rick;[84] L) Molecular dynamics simulations are in effect modeling the temperature of molecules such as water. Small amplitudes in MD trajectories (i.e., between black arrows) and lower potential energies (E_{pot}) should be indicative of cold water with little thermal

motion, while larger amplitudes and higher potential energies should be indicative of hot water with more extreme thermal motion.

Figure 10. Src kinase : bosutinib. A) Two Src kinase inhibitors in clinical usage: **1** – imatinib, **2** – bosutinib; B) The Abl and Src kinase DFG motif: purple – active DFG-in conformation for bosutinib-bound Src (pdb: 4mxo), green – inactive DFG-out conformation for bosutinib-bound Abl (pdb: 3ue4). Figure adapted from ref. 193; C) View of the two water molecules (**w1** and **w2**) aiding in binding of bosutinib and in binding of many other kinase inhibitors.[193] Only **w1** makes direct interaction with the ligand, while **w2** stabilizes **w1**. Figure adapted from ref. 193.

Figure 11. Human carbonic anhydrase : arylsulfonamides. A) arylsulfonamide-based inhibitors of human carbonic anhydrase; B) X-ray structures showing effect of benzo- ligand extensions on ligand-protein interactions and structured water: 1 – ligand **3d** with monocyclic ring makes few contacts with enzyme residues and surrounding waters (**w1**-**w5**) that are only modestly ordered, 2 – ligand **3h** with fused bicyclic ring displaces **w5**, makes only slightly better (hydrophobic) interactions with residues in binding pocket, but forces more structure on waters **w1**-**w4** and others in the complex; C) The thermodynamic profiles for HCA inhibitors des-fluoro **3h**, tetra-fluoro **3k** and seven partially fluorinated derivatives where ΔΔG (purple), ΔΔH (blue) and -TΔΔS (red) values are calculated with reference to those of the des-fluoro derivative **3h**. Figure based on ref. 260.

Figure 12. Druggable cavities in GPCRs : Adenosine A_{2A}. A) Inhibitors based on chromones; B) The compound chromone12 (**4b**) has 33-fold better binding affinity (pK_i) than its des-methyl derivative (**4c**), suggesting a "magic methyl": 1 – in **4b**, the ligand's methyl remains in the hydrophobic pocket, even after 100 ps MD, 2 – in **4c**, an additional "hot" water (circled in red) slips into the pocket causing the ligand to move away from the hydrophobic pocket, and the interaction with Asn253 to be weakened. This particular hot water was also seen by MD in the unliganded pocket, and is made even hotter by its association with the lipophilic thiazole carbon rather than other water molecules in the unliganded pocket. Figure adapted from ref. 254 with the assistance of J. S. Mason; C) Inhibitors based on triazines; D) Structure models for compounds **5b**, **5c** and **5d** bound at adenosine A_{2A}, with waters modeled using WaterFLAP, followed by all-atom MD optimization: 1 – in most potent inhibitor complex (based on pdb 3uzc), **5b**, with longest

residence time, the *m*-chloro draws ligand deeper into hydrophobic region of pocket and *p*-phenol makes a direct hydrogen bond with His278, 2 – cold water (blue asterisk) forms bridging hydrogen bond between ligand **5c** and His278 (based on pdb 3uza) , 3 – in **5d** complex, phenyl replacement of ring eliminates hydrogen bond, and at least two hot waters (e.g., red asterisks) now occupy ligand-receptor interface, leading to weak binding and negligible residence time (initial model created by docking **5d** into 3uza). Figure adapted from ref. 364 with the assistance of J. S. Mason.

Figure 13. Thermolysin : hot water compensates for changes in ligand hydrophobicity. A) Carboxybenzyl-Gly-($PO2^-$)-L-Leu-NH2-based thermolysin inhibitors; B) X-ray crystal structure models for four members of the inhibitor series: 1 – in the methyl derivative (**6a**), the terminal methyl is surrounded by a chain of six water molecules, **w1**, **w2**, **w3**, **w'**, **w"**, and **w'"**, all ≤ 4.0 Å from the methyl, but within ideal hydrogen bonding distance (2.6-3.1 Å) to the other waters and/or the Asn111 or Asn112 termini of the chain, 2 – the ethyl of the **6b** derivative is surrounded by seven water molecules in a chain (**w1**, **w2**, **w3**, **w4**, **w5**, **w6** and **w'**) and a capping water, **w0**, that bridges between **w1** and **w6** , while hot waters **w"** and **w'"** have been sterically displaced by the larger ethyl, 3 – the even larger 2-methylbutyl of the **6e** derivative displaces **w4** and moves **w2** and **w3** to new positions **w2'** and **w3'**, respectively (**w***, somewhat cooler than other waters in the complex, appears to anchor the water network), 4 – the still larger *neo*-pentyl of **6h** further disrupts the water network by displacing **w2'** and **w3'**. Figure adapted from ref. 366; C) The thermodynamic profiles for thermolysin inhibitors **6a** -**6i**. ΔΔG (purple), ΔΔH (blue) and -TΔΔS (red) values are calculated with reference to those of the methyl derivative **6a** (data from ref. 366).

AUTHOR INFORMATION

Corresponding Authors

*E-mail: andrea.mozzarelli@unipr.it. Phone: +39 0521 905138.

*E-mail: glen.kellogg@vcu.edu. Phone: +01 804 828-6452.

◊ A.S.B. present address: IDEAYA Biosciences, 280 Utah Ave, Suite 250, South San Francisco, California 94080, United States

Author Contributions

The manuscript was written through the contributions of all authors. A.M. and G.E.K. developed the manuscript concept and constructed the working outline. G.E.K. performed overall manuscript organization and final editing. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

Biographies

Francesca Spyrakis received her first degree (2003) in medicinal chemistry and her Ph.D. (2007) in biochemistry and molecular biology from the University of Parma, Italy, with additional training at the University of Perugia and the University of Barcelona. Starting October 2016 she is Associate Professor of Medicinal Chemistry at the University of Turin. Her studies are focused on development and application of computational methodologies for identifying ligands that are active at relevant pharmaceutical targets. Particular attention is given to the roles played by flexibility and by water molecules that mediate biological interactions, with the overall goal of integrating molecular dynamics and virtual screening. Her scientific activity includes more than 50 publications in international peer-reviewed journals, 10 book chapters and more than 70 communications at congresses and meetings.

Mostafa H. Ahmed received his Bachelors of Science in Pharmaceutical Sciences from the School of Pharmacy, Misr International University in Cairo, Egypt in 2006. After a few years of working as teaching assistant at the School of Pharmacy, Misr International University in Cairo, Egypt, he was awarded a Fulbright scholarship for a Master of Science degree in Pharmaceutical Sciences at

Virginia Commonwealth University, which he completed in 2011. In 2014, he completed his Ph.D. from the department of Medicinal Chemistry at Virginia Commonwealth University, where he is currently a postdoctoral fellow. His research interests center around developing computational tools for protein structure prediction and drug discovery as well as X-ray crystallography. He is also involved in multiple drug discovery projects for diseases like sickle cell disease and cancer.

Alexander S. Bayden received his B.S. in Chemistry from Virginia Polytechnic Institute and State University, Blacksburg, in 2000, following which he earned a Ph.D. in Chemistry from the University of Pittsburgh in 2005. He was a postdoctoral at Virginia Commonwealth University, Richmond, where he developed models for predicting selectivity of tyrosine nitration and ionization state optimization for protein-ligand complexes, and at AstraZeneca (Boston), where he investigated the role of water in protein-ligand interactions. Prior to his current position as Senior Research Scientist at IDEAYA Biosciences, South San Francisco, California, he was Computational Chemist at CMDBioscience, Inc., New Haven, Connecticut, where he developed new tools for molecular modeling of peptides and protein-peptide interactions.

Pietro Cozzini received his Laurea degree in Chemistry at the University of Parma, Italy in 1978. After a few years in the international software industry, he returned to the University of Parma as a developer of chemistry databases. Currently, he is Associate Professor in General Chemistry and head of the Molecular Modeling Laboratory at the Department of Food and Drug Sciences, University of Parma. His research interests focus on the development of computational methods for evaluating the energetics in molecular associations, with particular consideration of the role of water molecules; on the identification of new food pollutants that act as endocrine disruptors; the design of new chemosensors using *in silico* screening; and with colleagues at the University of Barcelona, exploring the quantum chemistry basis for hydrophobic descriptors.

Andrea Mozzarelli received his degree in Chemistry (1974) at the University of Parma, Italy and was a Fogarty Fellow in the NIDDK Laboratory of Chemical Physics of the US NIH (1984-1985). Currently, he is Full Professor in Biochemistry in the Department of Food and Drug Sciences at the University of Parma. He is serving as the Protein Group Coordinator of the Italian Biochemical Society. His research is aimed at understanding protein structure-dynamics-function-regulation relationships by exploiting biochemical and biophysical approaches and techniques, including protein entrapment in silica gels, stopped-flow and single crystal microspectrophotometry. His application of complementary experimental and computational methods, specifically HINT, has

identified enzyme inhibitors with potential antibiotic activity. The proteins of particular interest to him have been pyridoxal 5'-phosphate- and NAD-dependent enzymes and hemoglobin.

Glen E. Kellogg received his B.S. (1979) in chemistry at the University of New Mexico and his Ph.D. (1985) from the University of Arizona. After a postdoctoral at Northwestern University, he joined the Medicinal Chemistry Department of Virginia Commonwealth University in 1988, where he is now Professor and Assistant Chair. He is a coauthor of the HINT (Hydropathic INTeractions) interaction model and related computational tools. His research has focused on – from a unique empirical and phenomenological approach – understanding and modeling the hydrophobic effect and related phenomena that involve water in the biological environment. Recently, he and his colleagues have reported a new paradigm for 3D protein structure prediction based on interaction homology and hydropathic valence, and have become interested in modeling and inhibiting protein-protein interactions.

ACKNOWLEDGMENTS

Protein structural data that have not been deposited in the PDB for adenosine A_{2A} were generously provided by Dr. Jonathan S. Mason (Heptares Therapeutics Ltd). The authors wish to express our gratitude to our many colleagues and friends who have offered advice, encouragement, commentary and technical assistance during the process of preparing this manuscript. In particular, we thank Dr. Donald J. Abraham, Dr. Paolo Benedetti, Dr. Gabriele Cruciani, Dr. Simon Cross, Dr. Sergio Decherchi, Dr. Keith C. Ellis, Dr. Richard A. Glennon, Dr. Gianluigi Ingletto, Dr. Philip D. Mosier, Mr. Ahmad J. Obaidullah, Dr. Martin K. Safo, Dr. J. Neel Scarsdale, Dr. Richard B. Westkaemper, and Dr. Yan Zhang. In addition, P.C., A.M., and G.E.K. thank our many former students and postdoctoral associates who have extensively contributed to our studies of the hydrophobic effect and water over the last two decades.

ABBREVIATIONS USED

AZT, azidothymidine; BLG, β-lactoglobulin; CCR5, C-C chemokine receptor type 5; CXCR4, C-X-C chemokine receptor type 4; DFG, Asp-Phe-Gly; FEP, free energy perturbation; GB, Generalized Born; H-FABP, heart fatty acid binding protein; H/SC, enthalpy–entropy compensation; HCA, human carbonic anhydrase; HINT, Hydropathic INTeractions; HT, high-throughput; ICF,

immunodeficiency, centromeric instability and facial abnormalities; IFST, inhomogeneous fluid solvation theory; ITC, isothermal titration calorimetry; LIE, linear interaction energy; MAD, multiple-wavelength anomalous dispersion; MC, Monte Carlo; MD, molecular dynamics; MOR, μ-opioid receptor; MRD, magnetic relaxation dispersion; MUP, major urinary protein; OppA, oligopeptide binding protein; PB, Poisson-Boltzmann; PMF, potential of mean forces; PPI, protein-protein interface; RETI, replica exchange thermodynamic integration; RISM, reference interaction site model; SA, surface area; SAD, single wavelength anomalous dispersion; SASA, solvent-accessible surface area; SBDD, structure-based drug discovery/design; Tgt, tRNA-guanine transglycosylase; TI, thermodynamic integration; vdW, van der Waals; ZDV, zidovudine

REFERENCES

1. Abraham, D. J. The potential role of single crystal X-ray diffraction in medicinal chemistry. *Intra-Sci. Chem. Rep.* **1974,** 8, 1-9.
2. Kuhn, P.; Wilson, K.; Patch, M. G.; Stevens, R. C. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr. Opin. Chem. Biol.* **2002,** 6, 704-710.
3. Babu, Y. S.; Chand, P.; Bantia, S.; Kotian, P.; Dehghani, A.; El-Kattan, Y.; Lin, T. H.; Hutchison, T. L.; Elliott, A. J.; Parker, C. D.; Ananth, S. L.; Horn, L. L.; Laver, G. W.; Montgomery, J. A. BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J. Med. Chem.* **2000,** 43, 3482-3486.
4. Regan, J.; Breitfelder, S.; Cirillo, P.; Gilmore, T.; Graham, A. G.; Hickey, E.; Klaus, B.; Madwed, J.; Moriak, M.; Moss, N.; Pargellis, C.; Pav, S.; Proto, A.; Swinamer, A.; Tong, L.; Torcellini, C. Pyrazole urea-based inhibitors of p38 MAP kinase: from lead compound to clinical candidate. *J. Med. Chem.* **2002,** 45, 2
994-3008.
5. Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005,** 48, 1088-1097.
6. Verlinde, C. L.; Hol, W. G. Structure-based drug design: progress, results and challenges. *Structure* **1994,** 2, 577-587.
7. Hardy, L. W.; Abraham, D. J.; Safo, M. K. Structure-Based Drug Design. In *Burger's Medicinal Chemistry and Drug Discovery*, Abraham, D. J., Ed. Wiley-Interscience: New York, 2003; pp 417-469.
8. Acharya, K. R.; Sturrock, E. D.; Riordan, J. F.; Ehlers, M. R. Ace revisited: a new target for structure-based drug design. *Nat. Rev. Drug Discov.* **2003,** 2, 891-902.
9. Cohen, P. Protein kinases--the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **2002,** 1, 309-315.
10. Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med. (Berl.)* **2000,** 78, 269-281.
11. Keseru, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discov. Today* **2006,** 11, 741-8.

12. Kennedy, J. P.; Williams, L.; Bridges, T. M.; Daniels, R. N.; Weaver, D.; Lindsley, C. W. Application of combinatorial chemistry science on modern drug discovery. *J. Comb. Chem.* **2008,** 10, 345-54.

13. Messeguer, A.; Cortes, N. Combinatorial chemistry in cancer research. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **2007,** 9, 83-92.

14. Anon. A decade of drug-likeness. *Nat. Rev. Drug Discov.* **2007,** 6, 853-853.

15. Dill, K. A. Additivity Principles in Biochemistry. *J. Biol. Chem.* **1997,** 272, 701-704.

16. Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discov.* **2015,** 14, 95-110.

17. Geschwindner, S.; Ulander, J.; Johansson, P. Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip? *J. Med. Chem.* **2015,** 58, 6321-6335.

18. Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010,** 53, 5061-84.

19. Homans, S. W. Water, water everywhere--except where it matters? *Drug Discov. Today* **2007,** 12, 534-539.

20. Snyder, P.; Lockett, M.; Moustakas, D.; Whitesides, G. Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Special Topics* **2013**, 1-39.

21. Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today* **2006,** 11, 580-94.

22. Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007,** 47, 668-75.

23. Ladbury, J. E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996,** 3, 973-980.

24. Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.* **2009,** 131, 15403-15411.

25. Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc. Natl. Acad. Sci. USA* **1988,** 85, 6612-6616.

26. Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B.; Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science* **1989,** 246, 1149-1152.

27. Brik, A.; Wong, C. H. HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.* **2003,** 1, 5-14.

28. Lam, P. Y.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994,** 263, 380-384.

29. Grzesiek, S.; Bax, A.; Nicholson, L. K.; Yamazaki, T.; Wingfield, P.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Hodge, C. N. NMR Evidence for the Displacement of a Conserved Interior Water Molecule in HIV Protease by a Non-Peptide Cyclic Urea-Based Inhibitor. *J. Am. Chem. Soc.* **1994,** 116, 1581-1582.

30. Hodge, C. N.; Aldrich, P. E.; Bacheler, L. T.; Chang, C. H.; Eyermann, C. J.; Garber, S.; Grubb, M.; Jackson, D. A.; Jadhav, P. K.; Korant, B.; Lam, P. Y.; Maurin, M. B.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Reid, C.; Sharpe, T. R.; Shum, L.; Winslow, D. L.; Erickson-Viitanen, S. Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chem. Biol.* **1996,** 3, 301-314.

31. Pillai, B.; Kannan, K. K.; Hosur, M. V. 1.9 A x-ray study shows closed flap conformation in crystals of tethered HIV-1 PR. *Proteins* **2001,** 43, 57-64.

32. Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989,** 245, 616-621.

33. Jhoti, H.; Singh, O. M.; Weir, M. P.; Cooke, R.; Murray-Rust, P.; Wonacott, A. X-ray crystallographic studies of a series of penicillin-derived asymmetric inhibitors of HIV-1 protease. *Biochemistry* **1994,** 33, 8417-8427.

34. Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J. Med. Chem.* **2004,** 47, 4507-4516.

35. Priestle, J. P.; Fassler, A.; Rosel, J.; Tintelnot-Blomley, M.; Strop, P.; Grutter, M. G. Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor. *Structure* **1995,** 3, 381-389.

36. Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; Hulten, J.; Bonham, N. M.; Schaal, W.; Karlen, A.; Hallberg, A. Unexpected binding mode of a cyclic sulfamide HIV-1 protease inhibitor. *J. Med. Chem.* **1997,** 40, 898-902.

37. Barron, L. D.; Hecht, L.; Wilson, G. The lubricant of life: a proposal that solvent water promotes extremely fast conformational fluctuations in mobile heteropolypeptide structure. *Biochemistry* **1997,** 36, 13143-13147.

38. Panman, M. R.; Bakker, B. H.; den Uyl, D.; Kay, E. R.; Leigh, D. A.; Buma, W. J.; Brouwer, A. M.; Geenevasen, J. A.; Woutersen, S. Water lubricates hydrogen-bonded molecular machines. *Nature chemistry* **2013,** 5, 929-934.

39. Kier, L. B.; Cheng, C.-K.; Testa, B. Studies of Ligand Diffusion Pathways over a Protein Surface. *J. Chem. Inform. Comput. Sci.* **2003,** 43, 255-258.

40. Kier, L. B.; Hall, L. H. The creation of proton hopping from a drug-receptor encounter. *Chem. Biodivers.* **2013,** 10, 2221-2225.

41. Einstein, A. *Autobiographical notes*. Open Court: The University of California, 1949.

42. Gilli, P.; Ferretti, V.; Gilli, G.; Borea, P. A. Enthalpy-entropy compensation in drug-receptor binding. *J. Phys. Chem.* **1994,** 98, 1515-1518.

43. Freed, K. F. Entropy-enthalpy compensation in chemical reactions and adsorption: an exactly solvable model. *J. Phys. Chem. B* **2011,** 115, 1689-1692.

44. Cornish-Bowden, A. Enthalpy-entropy compensation: a phantom phenomenon. *J. Biosci.* **2002,** 27, 121-126.

45. Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nollmann, M. Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* **2001,** 93, 215-230.

46. Linert, W.; Kudrjawtsev, A. B.; Schmid, R. Concerning the problem of the isokinetic relationship. I. A statistical mechanical model. *Aust. J. Chem.* **1983,** 36, 1903-1912.

47. Hansch, C. Hydrophobic effect in host-guest interactions. Hydrolysis of nitrophenyl carboxylates. *J. Org. Chem.* **1978,** 43, 4889-4890.

48. Murakami, Y.; Aoyama, Y.; Kida, M.; Nakano, A. A Macrocyclic Enzyme Model System. Deacylation of p-Nitrophenyl Carboxylates as Effected by a 20]Paracyclophane Bearing an Imidazole Group. *BCSJ* **1977,** 50, 3365-3371.

49. Schneider, H.-J. Mechanisms of Molecular Recognition : Investigations of Organic Host–Guest Complexes. *Angew. Chem. Int. Ed. Engl.* **1991,** 30, 1417-1436.

50. Ferguson, S. B.; Seward, E. M.; Diederich, F.; Sanford, E. M.; Chou, A.; Inocencio-Szweda, P.; Knobler, C. B. Strong enthalpically driven complexation of neutral benzene guests in aqueous solution. *J. Org. Chem.* **1988,** 53, 5593-5595.

51. Harata, K.; Tsuda, K.; Uekama, K.; Otagiri, M.; Hirayama, F. Complex formation of hexakis(2,3,6-tri-O-methyl)-α-cyclodextrin with substituted benzenes in aqueous solution. *J. Incl. Phenom. Macrocycl. Chem. (DNLM)* **1988,** 6, 135-142.

52. Sciortino, F.; Geiger, A.; Stanley, H. E. Isochoric differential scattering functions in liquid water: The fifth neighbor as a network defect. *Phys. Rev. Lett.* **1990,** 65, 3452-3455.

53. Giguère, P. A. Bifurcated hydrogen bonds in water. *J. Raman Spectrosc.* **1984,** 15, 354-359.

54. Nucci, N. V.; Pometun, M. S.; Wand, A. J. Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* **2011,** 18, 245-249.

55. Ghai, R.; Falconer, R. J.; Collins, B. M. Applications of isothermal titration calorimetry in pure and applied research--survey of the literature from 2010. *J. Mol. Recogn.* **2012,** 25, 32-52.

56. Ringe, D. What makes a binding site a binding site? *Curr. Opin. Struct. Biol.* **1995,** 5, 825-829.

57. Tame, J. R.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* **1996,** 3, 998-1001.

58. Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.* **1999,** 291, 393-415.

59. Berntsson, R. P.; Doeven, M. K.; Fusetti, F.; Duurkens, R. H.; Sengupta, D.; Marrink, S. J.; Thunnissen, A. M.; Poolman, B.; Slotboom, D. J. The structural basis for peptide selection by the transport receptor OppA. *EMBO J.* **2009,** 28, 1332-1340.

60. Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **1994,** 264, 670.

61. Dunitz, J. D. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995,** 2, 709-712.

62. Cooper, A. Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.* **2005,** 115, 89-97.

63. Petrova, T.; Steuber, H.; Hazemann, I.; Cousido-Siah, A.; Mitschler, A.; Chung, R.; Oka, M.; Klebe, G.; El-Kabbani, O.; Joachimiak, A.; Podjarny, A. Factorizing selectivity determinants of inhibitor binding toward aldose and aldehyde reductases: structural and thermodynamic properties of the aldose reductase mutant Leu300Pro-fidarestat complex. *J. Med. Chem.* **2005,** 48, 5659-5665.

64. Kuntz, I. D. Protein folding. *J. Am. Chem. Soc.* **1972,** 94, 4009-4012.

65. Kuntz, I. D., Jr.; Kauzmann, W. Hydration of proteins and polypeptides. *Adv. Protein Chem.* **1974,** 28, 239-345.

66. Hagler, A. T.; Moult, J. Computer simulation of the solvent structure around biological macromolecules. *Nature* **1978,** 272, 222-226.

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

68. Daggett, V.; Levitt, M. Realistic Simulations of Native-Protein Dynamics in Solution and Beyond. *Annu. Rev. Biophys. Biomol. Struct.* **1993,** 22, 353-380.

69. Cho, C. H.; Singh, S.; Robinson, G. W. Liquid water and biological systems: the most important problem in science that hardly anyone wants to see solved. *Faraday Discuss.* **1996,** 103, 19-27.

70. Garcia, A. E.; Onuchic, J. N. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA* **2003,** 100, 13898-13903.

71. Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes. *J. Med. Chem.* **1995,** 38, 4953-4967.

72. Tame, J. R. Scoring functions: A view from the bench. *J. Comput.-Aided Mol. Des.* **1999,** 13, 99-108.

73. Wong, C. F.; McCammon, J. A. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* **1986,** 108, 3830-3832.

74. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. Free energy calculations by computer simulation. *Science* **1987,** 236, 564-568.

75. Merz, K. M.; Kollman, P. A. Free energy perturbation simulations of the inhibition of thermolysin: prediction of the free energy of binding of a new inhibitor. *J. Am. Chem. Soc.* **1989,** 111, 5649-5658.

76. Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002,** 35, 358-365.

77. Helms, V.; Wade, R. C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study. *Biophys. J.* **1995,** 69, 810-824.

78. Wade, R. C. Solvation of the active site of cytochrome P450-cam. *J. Comput.-Aided Mol. Des.* **1990,** 4, 199-204.

79. Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004,** 126, 7683-7689.

80. Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of Water Molecules in Protein Binding Sites. *J. Am. Chem. Soc.* **2007,** 129, 2577-2587.

81. Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins* **1996,** 24, 433-438.

82. Olano, L. R.; Rick, S. W. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.* **2004,** 126, 7991-8000.

83. Huggins, D. J. Quantifying the entropy of binding for water molecules in protein cavities by computing correlations. *Biophys. J.* **2015,** 108, 928-936.

84. Yu, H.; Rick, S. W. Free energy, entropy, and enthalpy of a water molecule in various protein environments. *J. Phys. Chem. B* **2010,** 114, 11552-11560.

85. Pan, A.; Biswas, T.; Rakshit, A. K.; Moulik, S. P. Enthalpy-Entropy Compensation (EEC) Effect: A Revisit. *J. Phys. Chem. B* **2015,** 119, 15876-15884.

86. Chodera, J. D.; Mobley, D. L. Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys.* **2013,** 42, 121-142.

87. Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Mario Amzel, L.; Freire, E. Compensating enthalpic and entropic changes hinder binding affinity optimization. *Chem. Biol. Drug. Des.* **2007,** 69, 413-422.

88. Freire, E. Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today* **2008,** 13, 869-874.

89. Ahmad, M.; Helms, V.; Lengauer, T.; Kalinina, O. V. Enthalpy-Entropy Compensation upon Molecular Conformational Changes. *J. Chem. Theory Comput.* **2015,** 11, 1410-1418.

90. Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M. Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J. Am. Chem. Soc.* **2013,** 135, 15579-15584.

91. Searle, M. S.; Westwell, M. S.; Williams, D. H. Application of a generalised enthalpy-entropy relationship to binding co-operativity and weak associations in solution. *J. Chem. Soc., Perkin Trans. 2* **1995**, 141-151.

92. Sondergaard, C. R.; Garrett, A. E.; Carstensen, T.; Pollastri, G.; Nielsen, J. E. Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J. Med. Chem.* **2009,** 52, 5673-84.

93. Zheng, H.; Hou, J.; Zimmerman, M. D.; Wlodawer, A.; Minor, W. The future of crystallography in drug discovery. *Exp. Op. Drug Disc.* **2013,** 9, 125-137.

94. Ishida, N.; Inoue, T.; Miyahara, M.; Higashitani, K. Nano Bubbles on a Hydrophobic Surface in Water Observed by Tapping-Mode Atomic Force Microscopy. *Langmuir* **2000,** 16, 6377-6380.

95. Poynor, A.; Hong, L.; Robinson, I. K.; Granick, S.; Zhang, Z.; Fenter, P. A. How water meets a hydrophobic surface. *Phys. Rev. Lett.* **2006,** 97, 266101.

96. Steitz, R.; Gutberlet, T.; Hauss, T.; Klösgen, B.; Krastev, R.; Schemmel, S.; Simonsen, A. C.; Findenegg, G. H. Nanobubbles and Their Precursor Layer at the Interface of Water Against a Hydrophobic Substrate. *Langmuir* **2003,** 19, 2409-2418.

97. Tyrrell, J. W.; Attard, P. Images of nanobubbles on hydrophobic surfaces and their interactions. *Phys. Rev. Lett.* **2001,** 87, 176104.

98. Berne, B. J.; Weeks, J. D.; Zhou, R. Dewetting and hydrophobic interaction in physical and biological systems. *Annu. Rev. Phys. Chem.* **2009,** 60, 85-103.

99. Huang, X.; Zhou, R.; Berne, B. J. Drying and Hydrophobic Collapse of Paraffin Plates. *J. Phys. Chem. B* **2005,** 109, 3546-3552.

100. Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999,** 103, 4570-4577.

101. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998,** 102, 3531-3541.

102. Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998,** 102, 3542-3550.

103. Rasaiah, J. C.; Garde, S.; Hummer, G. Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond. *Annu. Rev. Phys. Chem.* **2008,** 59, 713-740.

104. Wallqvist, A.; Berne, B. J. Computer Simulation of Hydrophobic Hydration Forces on Stacked Plates at Short Range. *J. Phys. Chem.* **1995,** 99, 2893-2899.

105. Olsson, T. S. G.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. The Thermodynamics of Protein–Ligand Interaction and Solvation: Insights for Ligand Design. *J. Mol. Biol.* **2008,** 384, 1002-1017.

106. Barratt, E.; Bingham, R. J.; Warner, D. J.; Laughton, C. A.; Phillips, S. E. V.; Homans, S. W. Van der Waals Interactions Dominate Ligand−Protein Association in a Protein Binding Site Occluded from Solvent Water. *J. Am. Chem. Soc.* **2005,** 127, 11827-11834.

107. Malham, R.; Johnstone, S.; Bingham, R. J.; Barratt, E.; Phillips, S. E.; Laughton, C. A.; Homans, S. W. Strong solute-solute dispersive interactions in a protein-ligand complex. *J. Am. Chem. Soc.* **2005,** 127, 17061-7.

108. Oas, T. G.; Toone, E. J. Thermodynamic Solvent Isotope Effects and Molecular Hydrophobicity. *Adv. Biophys. Chem.* **1997,** 6, 1-52.

109. Williams, B. A.; Chervenak, M. C.; Toone, E. J. Energetics of lectin-carbohydrate binding. A microcalorimetric investigation of concanavalin A-oligomannoside complexation. *J. Biol. Chem.* **1992,** 267, 22907-11.

110. Snyder, P. W.; Mecinović, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Héroux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. USA* **2011,** 108, 17889-17894.

111. Akerstrom, B.; Flower, D. R.; Salier, J. P. Lipocalins: unity in diversity. *Biochim. Biophys. Acta* **2000,** 1482, 1-8.

112. Modig, K.; Rademacher, M.; Lucke, C.; Halle, B. Water dynamics in the large cavity of three lipid-binding proteins monitored by (17)O magnetic relaxation dispersion. *J. Mol. Biol.* **2003,** 332, 965-77.

113. Qvist, J.; Davidovic, M.; Hamelberg, D.; Halle, B. A dry ligand-binding cavity in a solvated protein. *Proc. Natl. Acad. Sci. USA* **2008,** 105, 6296-301.

114. Desvaux, H.; Dubois, L.; Huber, G.; Quillin, M. L.; Berthault, P.; Matthews, B. W. Dynamics of Xenon Binding Inside the Hydrophobic Cavity of Pseudo-Wild-type Bacteriophage T4 Lysozyme Explored through Xenon-Based NMR Spectroscopy. *J. Am. Chem. Soc.* **2005,** 127, 11676-11683.

115. Clore, G. M.; Bax, A.; Wingfield, P. T.; Gronenborn, A. M. Identification and localization of bound internal water in the solution structure of interleukin 1.beta. by heteronuclear three-dimensional proton rotating-fram Overhauser nitrogen-15-proton multiple quantum coherence NMR spectroscopy. *Biochemistry* **1990,** 29, 5671-5676.

116. Otting, G.; Liepinsh, E.; Halle, B.; Frey, U. NMR identification of hydrophobic cavities with ow water occupancies in protein structures using small gas molecules. *Nat. Struct. Biol.* **1997,** 4, 396-404.

117. Matthews, B. W.; Liu, L. A review about nothing: Are apolar cavities in proteins really empty? *Protein Sci.* **2009**, NA-NA.

118. Barrow, J. D. *The Book of Nothing: Vacuums, Voids, and the Latest Ideas about the Origins of the Universe*. Knopf Doubleday Publishing Group: 2009.

119. Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers* **1991,** 31, 919-931.

120. Hubbard, S. J.; Gross, K.-H.; Argos, P. Intramolecular cavities in globular proteins. *Protein Eng.* **1994,** 7, 613-626.

121. Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. Buried waters and internal cavities in monomeric proteins. *Protein Sci.* **1994,** 3, 1224-1235.

122. Ball, P. Chemical physics: How to keep dry in water. *Nature* **2003,** 423, 25-26.

123. Huang, X.; Margulis, C. J.; Berne, B. J. Dewetting-induced collapse of hydrophobic particles. *Proc. Natl. Acad. Sci. USA* **2003,** 100, 11953-11958.

124. Hummer, G.; Garde, S. Cavity Expulsion and Weak Dewetting of Hydrophobic Solutes in Water. *Phys. Rev. Lett.* **1998,** 80, 4193-4196.

125. Leung, K.; Luzar, A.; Bratko, D. Dynamics of Capillary Drying in Water. *Phys. Rev. Lett.* **2003,** 90.

126. Liu, P.; Huang, X.; Zhou, R.; Berne, B. J. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* **2005,** 437, 159-162.

127. Giovambattista, N.; Lopez, C. F.; Rossky, P. J.; Debenedetti, P. G. Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc. Natl. Acad. Sci. USA* **2008,** 105, 2274-2279.

128. Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. USA* **2011,** 108, 13118-13123.

129. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011,** 133, 9181-9183.

130. Benson, S. P.; Pleiss, J. Solvent Flux Method (SFM): A Case Study of Water Access toCandida antarcticaLipase B. *J. Chem. Theory Comput.* **2014,** 10, 5206-5214.

131. Kalyaanamoorthy, S.; Chen, Y.-P. P. Ligand release mechanisms and channels in histone deacetylases. *J. Comput. Chem.* **2013,** 34, 2270-2283.

132. Krem, M. M.; Di Cera, E. Conserved water molecules in the specificity pocket of serine proteases and the molecular mechanism of Na+ binding. *Proteins* **1998,** 30, 34-42.

133. Kurisaki, I.; Barberot, C.; Takayanagi, M.; Nagaoka, M. Dewetting of S1-Pocket via Water Channel upon Thrombin–Substrate Association Reaction. *J. Phys. Chem. B* **2015,** 119, 15807-15812.

134. Angel, T. E.; Chance, M. R.; Palczewski, K. Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc. Natl. Acad. Sci. USA* **2009,** 106, 8555-8560.

135. Choe, H.-W.; Kim, Y. J.; Park, J. H.; Morizumi, T.; Pai, E. F.; Krauß, N.; Hofmann, K. P.; Scheerer, P.; Ernst, O. P. Crystal structure of metarhodopsin II. *Nature* **2011,** 471, 651-655.

136. Jastrzebska, B.; Palczewski, K.; Golczak, M. Role of Bulk Water in Hydrolysis of the Rhodopsin Chromophore. *J. Biol. Chem.* **2011,** 286, 18930-18937.

137. Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; Ijzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science* **2012,** 337, 232-236.

138. Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **2011**, *477*, 549-555.

139. Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular signatures of G-protein-coupled receptors. *Nature* **2013,** 494, 185-194.

140. Yuan, S.; Filipek, S.; Palczewski, K.; Vogel, H. Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.* **2014,** 5, 4733.

141. Yuan, S.; Vogel, H.; Filipek, S. The Role of Water and Sodium Ions in the Activation of the μ-Opioid Receptor. *Angew. Chem. Int. Ed.* **2013,** 52, 10112-10115.

142. Garczarek, F.; Gerwert, K. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature* **2006,** 439, 109-112.

143. Blankenship, E.; Vahedi-Faridi, A.; Lodowski, David T. The High-Resolution Structure of Activated Opsin Reveals a Conserved Solvent Network in the Transmembrane Region Essential for Activation. *Structure* **2015,** 23, 2358-2364.

144. Sun, X.; Agren, H.; Tu, Y. Functional water molecules in rhodopsin activation. *J. Phys. Chem. B* **2014,** 118, 10863-10873.

145. Gerwert, K.; Freier, E.; Wolf, S. The role of protein-bound water molecules in microbial rhodopsins. *Biochim. Biophys. Acta* **2014,** 1837, 606-613.

146. Huang, W.; Manglik, A.; Venkatakrishnan, A. J.; Laeremans, T.; Feinberg, E. N.; Sanborn, A. L.; Kato, H. E.; Livingston, K. E.; Thorsen, T. S.; Kling, R. C.; Granier, S.; Gmeiner, P.; Husbands, S. M.; Traynor, J. R.; Weis, W. I.; Steyaert, J.; Dror, R. O.; Kobilka, B. K. Structural insights into μ-opioid receptor activation. *Nature* **2015,** 524, 315-321.

147. Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Mathiesen, J. M.; Sunahara, R. K.; Pardo, L.; Weis, W. I.; Kobilka, B. K.; Granier, S. Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature* **2012**, *485*, 321-326.

148. Manglik, A.; Lin, H.; Aryal, D. K.; McCorvy, J. D.; Dengler, D.; Corder, G.; Levit, A.; Kling, R. C.; Bernat, V.; Hübner, H.; Huang, X.P.; Sassano, M. F.; Giguère, P. M.; Löber, S.; Duan, D.; Scherrer, G.; Kobilka, B. K.; Gmeiner, P.; Roth, B. L.; Shoichet, B. K. Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **2016**, *537*, 185–190.

149. Sansom, M. S.; Shrivastava, I. H.; Ranatunga, K. M.; Smith, G. R. Simulations of ion channels--watching ions and water move. *Trends Biochem. Sci.* **2000,** 25, 368-74.

150. Breed, J.; Sankararamakrishnan, R.; Kerr, I. D.; Sansom, M. S. Molecular dynamics simulations of water within models of ion channels. *Biophys. J.* **1996,** 70, 1643-1661.

151. Hodgkin, A. L.; Keynes, R. D. The potassium permeability of a giant nerve fibre. *J. Physiol.* **1955,** 128, 61-88.

152. Doyle, D. A. The Structure of the Potassium Channel: Molecular Basis of K+ Conduction and Selectivity. *Science* **1998,** 280, 69-77.

153. Zhou, Y.; MacKinnon, R. The Occupancy of Ions in the K+ Selectivity Filter: Charge Balance and Coupling of Ion Binding to a Protein Conformational Change Underlie High Conduction Rates. *J. Mol. Biol.* **2003,** 333, 965-975.

154. Berneche, S.; Roux, B. A microscopic view of ion conduction through the K+ channel. *Proc. Natl. Acad. Sci. USA* **2003,** 100, 8644-8648.

155. Iwamoto, M.; Oiki, S. Counting Ion and Water Molecules in a Streaming File through the Open-Filter Structure of the K Channel. *J. Neurosci.* **2011,** 31, 12180-12188.

156. Fowler, P. W.; Abad, E.; Beckstein, O.; Sansom, M. S. P. Energetics of Multi-Ion Conduction Pathways in Potassium Ion Channels. *J. Chem. Theory Comput.* **2013,** 9, 5176-5189.

157. Kopfer, D. A.; Song, C.; Gruene, T.; Sheldrick, G. M.; Zachariae, U.; de Groot, B. L. Ion permeation in K+ channels occurs by direct Coulomb knock-on. *Science* **2014,** 346, 352-355.

158. Kratochvil, H. T.; Carr, J. K.; Matulef, K.; Annen, A. W.; Li, H.; Maj, M.; Ostmeyer, J.; Serrano, A. L.; Raghuraman, H.; Moran, S. D.; Skinner, J. L.; Perozo, E.; Roux, B.; Valiyaveetil, F. I.; Zanni, M. T. Instantaneous ion configurations in the K+ ion channel selectivity filter revealed by 2D IR spectroscopy. *Science* **2016,** 353, 1040-1044.

159. Ostmeyer, J.; Chakrapani, S.; Pan, A. C.; Perozo, E.; Roux, B. Recovery from slow inactivation in K+ channels is controlled by water molecules. *Nature* **2013,** 501, 121-124.

160. Chakrapani, S.; Cordero-Morales, J. F.; Perozo, E. A Quantitative Description of KcsA Gating I: Macroscopic Currents. *J. Gen. Physiol.* **2007,** 130, 465-478.

161. Chakrapani, S.; Cordero-Morales, J. F.; Perozo, E. A Quantitative Description of KcsA Gating II: Single-Channel Currents. *J. Gen. Physiol.* **2007,** 130, 479-496.

162. Podobnik, M.; Krasevec, N.; Bedina Zavec, A.; Naneh, O.; Flasker, A.; Caserman, S.; Hodnik, V.; Anderluh, G. How to Study Protein-protein Interactions. *Acta Chim Slov* **2016,** 63, 424-39.

163. Hayes, S.; Malacrida, B.; Kiely, M.; Kiely, P. A. Studying protein-protein interactions: progress, pitfalls and solutions. *Biochem. Soc. Trans.* **2016,** 44, 994-1004.

164. Modell, A. E.; Blosser, S. L.; Arora, P. S. Systematic Targeting of Protein-Protein Interactions. *Trends Pharmacol. Sci.* **2016,** 37, 702-13.

165. Huggins, D. J.; Marsh, M.; Payne, M. C. Thermodynamic Properties of Water Molecules at a Protein-Protein Interaction Surface. *J. Chem. Theory Comput.* **2011,** 7, 3514-3522.

166. Zafra-Ruano, A.; Luque, I. Interfacial water molecules in SH3 interactions: Getting the full picture on polyproline recognition by protein-protein interaction domains. *FEBS Lett.* **2012,** 586, 2619-30.

167. Ahmed, M. H.; Spyrakis, F.; Cozzini, P.; Tripathi, P. K.; Mozzarelli, A.; Scarsdale, J. N.; Safo, M. A.; Kellogg, G. E. Bound Water at Protein-Protein Interfaces: Partners, Roles and Hydrophobic Bubbles as a Conserved Motif. *PLoS ONE* **2011,** 6, e24712.

168. Amadasi, A.; Surface, J. A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust classification of "relevant" water molecules in putative protein binding sites. *J. Med. Chem.* **2008,** 51, 1063-7.

169. Meenan, N. A. G.; Sharma, A.; Fleishman, S. J.; MacDonald, C. J.; Morel, B.; Boetzel, R.; Moore, G. R.; Baker, D.; Kleanthous, C. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc. Natl. Acad. Sci. USA* **2010,** 107, 10080-10085.

170. Sundaralingam, M.; Sekharudu, Y. Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science* **1989,** 244, 1333-1337.

171. Vaynberg, J.; Qin, J. Weak protein–protein interactions as probed by NMR spectroscopy. *Trends Biotechnol.* **2006,** 24, 22-27.

172. Jamadagni, S. N.; Godawat, R.; Garde, S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annu. Rev. Chem. Biomol. Eng.* **2011,** 2, 147-171.

173. Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *J. Phys. Chem. B* **2010,** 114, 1632-1637.

174. Franzosa, E. A.; Xia, Y. Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. *Mol. Biol. Evol.* **2009,** 26, 2387-2395.

175. Mattos, C. Protein-water interactions in a dynamic world. *Trends Biochem. Sci.* **2002,** 27, 203-208.

176. Raschke, T. M. Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol.* **2006,** 16, 152-159.

177. Moreira, I. S. The Role of Water Occlusion for the Definition of a Protein Binding Hot-Spot. *Curr. Top. Med. Chem.* **2015,** 15, 2068-2079.

178. Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. *Curr. Med. Chem.* **2004,** 11, 3093-3118.

179. Carugo, O. Statistical survey of the buried waters in the Protein Data Bank. *Amino Acids* **2016,** 48, 193-202.

180. Liu, Q.; Li, Z.; Li, J. Use B-factor related features for accurate classification between protein binding interfaces and crystal packing contacts. *BMC Bioinformatics* **2014,** 15, S3-S3.

181. Kellogg, G. E.; Chen, D. L. The importance of being exhaustive. Optimization of bridging structural water molecules and water networks in models of biological systems. *Chem. Biodivers.* **2004,** 1, 98-105.

182. Sarkar, A.; Kellogg, G. E. Hydrophobicity--shake flasks, protein folding and drug discovery. *Curr. Top. Med. Chem.* **2010,** 10, 67-83.

183. Kellogg, G. E.; Abraham, D. J. Hydrophobicity: is LogP(o/w) more than the sum of its parts? *Eur. J. Med. Chem.* **2000,** 35, 651-61.

184. Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Mapping the energetics of water-protein and water-ligand interactions with the "natural" HINT forcefield: predictive tools for characterizing the roles of water in biomolecules. *J. Mol. Biol.* **2006,** 358, 289-309.

185. Wang, Y.-X.; Freedberg, D. n. I.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Bhat, T. N.; Erickson, J. W.; Torchia, D. A. Bound Water Molecules at the Interface between the HIV-1 Protease and a Potent Inhibitor, KNI-272, Determined by NMR. *J. Am. Chem. Soc.* **1996,** 118, 12287-12290.

186. Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H. O.; Danielson, U. H.; Classon, B.; Unge, T.; Samuelsson, B.; Hulten, J.; Hallberg, A.; Karlen, A. Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors. *J. Med. Chem.* **2001,** 44, 155-169.

187. Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *J. Med. Chem.* **2002,** 45, 2469-83.

188. García-Sosa, A. T. Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies. *J. Chem. Inf. Model.* **2013,** 53, 1388-1405.

189. Meyer, E. Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. *Protein Sci.* **1992,** 1, 1543-1562.

190. Schiebel, J.; Chang, A.; Merget, B.; Bommineni, G. R.; Yu, W.; Spagnuolo, L. A.; Baxter, M. V.; Tareilus, M.; Tonge, P. J.; Kisker, C.; Sotriffer, C. A. An ordered water channel in Staphylococcus aureus FabI: unraveling the mechanism of substrate recognition and reduction. *Biochemistry* **2015,** 54, 1943-1955.

191. Barillari, C.; Duncan, A. L.; Westwood, I. M.; Blagg, J.; van Montfort, R. L. Analysis of water patterns in protein kinase binding sites. *Proteins* **2011,** 79, 2109-2121.

192. Knight, J. D.; Hamelberg, D.; McCammon, J. A.; Kothary, R. The role of conserved water molecules in the catalytic domain of protein kinases. *Proteins* **2009,** 76, 527-535.

193. Levinson, N. M.; Boxer, S. G. A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nat. Chem. Biol.* **2014,** 10, 127-132.

194. Barman, A.; Smitherman, C.; Souffrant, M.; Gadda, G.; Hamelberg, D. Conserved Hydration Sites in Pin1 Reveal a Distinctive Water Recognition Motif in Proteins. *J. Chem. Inf. Model.* **2016,** 56, 139-147.

195. Rondelet, G.; Dal Maso, T.; Willems, L.; Wouters, J. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *J. Struct. Biol.* **2016,** 194, 357-367.

196. Salie, Z. L.; Kirby, K. A.; Michailidis, E.; Marchand, B.; Singh, K.; Rohan, L. C.; Kodama, E. N.; Mitsuya, H.; Parniak, M. A.; Sarafianos, S. G. Structural basis of HIV inhibition by translocation-defective RT inhibitor 4'-ethynyl-2-fluoro-2'-deoxyadenosine (EFdA). *Proc. Natl. Acad. Sci. USA* **2016,** 113, 9274-9279.

197. Silva, J. L.; Foguel, D. Hydration, cavities and volume in protein folding, aggregation and amyloid assembly. *Phys. Biol.* **2009,** 6, 015002-3975/6/1/015002.

198. Szep, S.; Park, S.; Boder, E. T.; Van Duyne, G. D.; Saven, J. G. Structural coupling between FKBP12 and buried water. *Proteins* **2009,** 74, 603-611.

199. Brotzakis, Z. F.; Groot, C. C.; Brandeburgo, W. H.; Bakker, H. J.; Bolhuis, P. G. Dynamics of Hydration Water around Native and Misfolded alpha-Lactalbumin. *J. Phys. Chem. B* **2016,** 120, 4756-4766.

200. Silva, J. L.; Vieira, T. C. R. G.; Gomes, M. P. B.; Bom, A. P. A.; Lima, L. M. T. R.; Freitas, M. S.; Ishimaru, D.; Cordeiro, Y.; Foguel, D. Ligand Binding and Hydration in Protein Misfolding: Insights from Studies of Prion and p53 Tumor Suppressor Proteins. *Acc. Chem. Res.* **2010,** 43, 271-279.

201. Dalal, V.; Arya, S.; Mukhopadhyay, S. Confined Water in Amyloid-Competent Oligomers of the Prion Protein. *Chemphyschem* **2016,** 17, 2804-2807.

202. Loo, T. W.; Clarke, D. M. Chemical and pharmacological chaperones as new therapeutic agents. *Expert Rev. Mol. Med.* **2007,** 9, 1-18.

203. Gupta, A. N.; Neupane, K.; Rezajooei, N.; Cortez, L. M.; Sim, V. L.; Woodside, M. T. Pharmacological chaperone reshapes the energy landscape for folding and aggregation of the prion protein. *Nat. Commun.* **2016,** 7, 12058.

204. Janin, J. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure* **1999,** 7, R277-9.

205. Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of protein-protein interfaces. *Proteins* **2005,** 60, 36-45.

206. Hong, S.; Kim, D. Interaction between bound water molecules and local protein structures: A statistical analysis of the hydrogen bond structures around bound water molecules. *Proteins* **2016,** 84, 43-51.

207. Visscher, K. M.; Kastritis, P. L.; Bonvin, A. M. Non-interacting surface solvation and dynamics in protein-protein interactions. *Proteins* **2015,** 83, 445-458.

208. Jakobi, S.; Nguyen, P. T.; Debaene, F.; Cianferani, S.; Reuter, K.; Klebe, G. What Glues a Homodimer Together: Systematic Analysis of the Stabilizing Effect of an Aromatic Hot Spot in the Protein-Protein Interface of the tRNA-Modifying Enzyme Tgt. *ACS Chem. Biol.* **2015,** 10, 1897-1907.

209. Selkoe, D. J. Translating cell biology into therapeutic advances in Alzheimer's disease. *Nature* **1999,** 399, A23-31.

210. Chong, S. H.; Ham, S. Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis. *Acc. Chem. Res.* **2015,** 48, 956-965.

211. Pradhan, M. R.; Pal, A.; Hu, Z.; Kannan, S.; Chee Keong, K.; Lane, D. P.; Verma, C. S. Wetting of nonconserved residue-backbones: A feature indicative of aggregation associated regions of proteins. *Proteins* **2016,** 84, 254-266.

212. Schwabe, J. W. The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **1997,** 7, 126-134.

213. Jayaram, B.; Jain, T. The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** 33, 343-361.

214. Spyrakis, F.; Cozzini, P.; Bertoli, C.; Marabotti, A.; Kellogg, G. E.; Mozzarelli, A. Energetics of the protein-DNA-water interaction. *BMC Struct. Biol.* **2007,** 7, 4.

215. Naughton, B.; Pollak, A.; Han, S.; Reich, N. The role of water dynamics in mediating protein-DNA interactions (549.4). *FASEB J.* **2014,** 28.

216. Barik, A.; Bahadur, R. P. Hydration of protein-RNA recognition sites. *Nucleic Acids Res.* **2014,** 42, 10148-10160.

217. Barik, A.; Nithin, C.; Karampudi, N. B.; Mukherjee, S.; Bahadur, R. P. Probing binding hot spots at protein-RNA recognition sites. *Nucleic Acids Res.* **2016,** 44, e9.

218. Hofmann, K. P.; Scheerer, P.; Hildebrand, P. W.; Choe, H.-W.; Park, J. H.; Heck, M.; Ernst, O. P. A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **2009,** 34, 540-552.

219. Pardo, L.; Deupi, X.; Dölker, N.; López-Rodríguez, M. L.; Campillo, M. The Role of Internal Water Molecules in the Structure and Function of the Rhodopsin Family of G Protein-Coupled Receptors. *ChemBioChem* **2007,** 8, 19-24.

220. Katrich, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Allosteric Sodium in Class A GPCR Signaling. Trends Biochem. Sci. **2014**, *39*, 233-244.

221. Angel, T. E.; Gupta, S.; Jastrzebska, B.; Palczewski, K.; Chance, M. R. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl. Acad. Sci. USA* **2009,** 106, 14367-14372.

222. Orban, T.; Gupta, S.; Palczewski, K.; Chance, M. R. Visualizing water molecules in transmembrane proteins using radiolytic labeling methods. *Biochemistry* **2010,** 49, 827-834.

223. Howard, E. I.; Guillot, B.; Blakeley, M. P.; Haertlein, M.; Moulin, M.; Mitschler, A.; Cousido-Siah, A.; Fadel, F.; Valsecchi, W. M.; Tomizaki, T.; Petrova, T.; Claudot, J.; Podjarny, A. High-resolution neutron and X-ray diffraction room-temperature studies of an H-FABP-oleic acid complex: study of the internal water cluster and ligand binding by a transferred multipolar electron-density distribution. *IUCrJ* **2016,** 3, 115-126.

224. Matsuoka, S.; Sugiyama, S.; Matsuoka, D.; Hirose, M.; Lethu, S.; Ano, H.; Hara, T.; Ichihara, O.; Kimura, S. R.; Murakami, S.; Ishida, H.; Mizohata, E.; Inoue, T.; Murata, M. Water-mediated recognition of simple alkyl chains by heart-type fatty-acid-binding protein. *Angew. Chem. Int. Ed. Engl.* **2015,** 54, 1508-1511.

225. Milenkovic, S.; Bondar, A. N. Mechanism of conformational coupling in SecA: Key role of hydrogen-bonding networks and water interactions. *Biochim. Biophys. Acta* **2016,** 1858, 374-385.

226. Sasikala, W. D.; Mukherjee, A. Single water entropy: hydrophobic crossover and application to drug binding. *J. Phys. Chem. B* **2014,** 118, 10553-10564.

227. Giacovazzo, C. *Fundamentals of Crystallography*. Oxford University Press: 2002.

228. Carugo, O. Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Eng.* **1999,** 12, 1021-1024.

229. Carugo, O.; Bordo, D. How many water molecules can be detected by protein crystallography? *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **1999,** 55, 479-83.

230. Ho, B. K.; Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **2008,** 8, 49.

231. Xu, G.; Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **2007,** 107, 3514-3543.

232. Halle, B.; Denisov, V. P. Magnetic relaxation dispersion studies of biomolecular solutions. *Methods Enzymol.* **2001,** 338, 178-201.

233. Halle, B. Water in biological systems: the NMR picture. In *Hydration Processes in Biology: Theoretical and Experimental Approaches*, Bellissent-Funel, M., Ed. IOS Press: 1999; pp 221-232.

234. Uchida, K.; Markley, J. L.; Kainosho, M. Carbon-13 NMR method for the detection of correlated hydrogen exchange at adjacent backbone peptide amides and its application to hydrogen exchange in five antiparallel beta strands within the hydrophobic core of Streptomyces subtilisin inhibitor (SSI). *Biochemistry* **2005,** 44, 11811-11820.

235. Krishna, M. M.; Hoang, L.; Lin, Y.; Englander, S. W. Hydrogen exchange methods to study protein folding. *Methods (San Diego, Calif.)* **2004,** 34, 51-64.

236. Balaban, R. S. Magnetization Transfer between Water and Macromolecules in Proton MRI. In *eMagRes*, John Wiley & Sons, Ltd: 2007.

237. Wand, A. J. Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struct. Biol.* **2001,** 8, 926-931.

238. Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins: Struct., Funct., Genet.* **2000,** 40, 389-408.

239. Biedermannova, L.; Schneider, B. Structure of the ordered hydration of amino acids in proteins: analysis of crystal structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2015,** 71, 2192-2202.

240. Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004,** 47, 45-55.

241. Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983,** 221, 709-13.

242. Voss, N. R.; Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010,** 38, W555-W562.

243. Ali, S. A.; Hassan, M. I.; Islam, A.; Ahmad, F. A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. *Curr. Protein Peptide Sci.* **2014,** 15, 456-476.

244. Hubbard, S.; Thornton, J. *NACCESS, Computer Program*, Department of Biochemistry Molecular Biology, University College London, 1993.

245. Marabotti, A.; Spyrakis, F.; Facchiano, A.; Cozzini, P.; Alberti, S.; Kellogg, G. E.; Mozzarelli, A. Energy-based prediction of amino acid-nucleotide base recognition. *J. Comput. Chem.* **2008,** 29, 1955-69.

246. Burnett, J. C.; Kellogg, G. E.; Abraham, D. J. Computational methodology for estimating changes in free energies of biomolecular association upon mutation. The importance of bound water in dimer-tetramer assembly for beta 37 mutant hemoglobins. *Biochemistry* **2000,** 39, 1622-33.

247. Burnett, J. C.; Botti, P.; Abraham, D. J.; Kellogg, G. E. Computationally accessible method for estimating free energy changes resulting from site-specific mutations of biomolecules: systematic model building and structural/hydropathic analysis of deoxy and oxy hemoglobins. *Proteins* **2001,** 42, 355-77.

248. Henry, V. J.; Bandrowski, A. E.; Pepin, A. S.; Gonzalez, B. J.; Desfeux, A. OMICtools: an informative directory for multi-omic data analysis. *Database (Oxford)* **2014,** 2014, bau069.

249. Yu, J.; Zhou, Y.; Tanaka, I.; Yao, M. Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics (Oxford, England)* **2010,** 26, 46-52.

250. Tripathi, A.; Kellogg, G. E. A novel and efficient tool for locating and characterizing protein cavities and binding sites. *Proteins* **2010,** 78, 825-842.

251. Gerstein, M.; Tsai, J.; Levitt, M. The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **1995,** 249, 955-966.

252. Stillinger, F. H.; Rahman, A. Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* **1974,** 60, 1545-1557.

253. Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008,** 130, 2817-2831.

254. Mason, J. S.; Bortolato, A.; Weiss, D. R.; Deflorian, F.; Tehan, B.; Marshall, F. H. High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks. *In Silico Pharmacol.* **2013,** 1, 23.

255. Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007,** 9, 573-81.

256. Chen, J. M.; Xu, S. L.; Wawrzak, Z.; Basarab, G. S.; Jordan, D. B. Structure-Based Design of Potent Inhibitors of Scytalone Dehydratase: Displacement of a Water Molecule from the Active Site. *Biochemistry* **1998,** 37, 17735-17744.

257. Loris, R.; Maes, D.; Poortmans, F.; Wyns, L.; Bouckaert, J. A Structure of the Complex between Concanavalin A and Methyl-3,6-di-O-(-D-mannopyranosyl)- -D-mannopyranoside Reveals Two Binding Modes. *J. Biol. Chem.* **1996,** 271, 30614-30618.

258. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. Involvement of Water in Carbohydrate−Protein Binding. *J. Am. Chem. Soc.* **2001,** 123, 12238-12247.

259. Mikol, V.; Papageorgiou, C.; Borer, X. The Role of Water Molecules in the Structure-Based Design of (5-Hydroxynorvaline)-2-cyclosporin: Synthesis, Biological Activity, and Crystallographic Analysis with Cyclophilin A. *J. Med. Chem.* **1995,** 38, 3361-3367.

260. Lockett, M. R.; Lange, H.; Breiten, B.; Heroux, A.; Sherman, W.; Rappoport, D.; Yau, P. O.; Snyder, P. W.; Whitesides, G. M. The binding of benzoarylsulfonamide ligands to human carbonic anhydrase is insensitive to formal fluorination of the ligand. *Angew. Chem. Int. Ed. Engl.* **2013,** 52, 7714-7717.

261. Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997,** 265, 445-464.

262. García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *J. Mol. Model.* **2003,** 9, 172-182.

263. Kellogg, G. E.; Fornabaio, M.; Chen, D. L.; Abraham, D. J. New Application Design for a 3D Hydropathic Map–Based Search for Potential Water Molecules Bridging between Protein and Ligand. *Internet Electron. J. Mol. Des.* **2005,** 4, 194-209.

264. Rossato, G.; Ernst, B.; Vedani, A.; Smieško, M. AcquaAlta: A Directional Approach to the Solvation of Ligand–Protein Complexes. *J. Chem. Inf. Model.* **2011,** 51, 1867-1881.

265. Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. Sect. B: Struct. Sci.* **2002,** 58, 380-388.

266. Zheng, M.; Li, Y.; Xiong, B.; Jiang, H.; Shen, J. Water PMF for predicting the properties of water molecules in protein binding site. *J. Comput. Chem.* **2013,** 34, 583-592.

267. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985,** 28, 849-857.

268. Wallnoefer, H. G.; Liedl, K. R.; Fox, T. A GRID-Derived Water Network Stabilizes Molecular Dynamics Computer Simulations of a Protease. *J. Chem. Inf. Model.* **2011,** 51, 2860-2867.

269. *Molecular Discovery, Ltd.*, 215 Marsh Road, 1st Floor, HA5 5NE, Pinner, Middlesex, UK.

270. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007,** 47, 279-294.

271. Miranker, A.; Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins: Struct., Funct., Genet.* **1991,** 11, 29-34.

272. Evensen, E.; Joseph-McCarthy, D.; Weiss, G. A.; Schreiber, S. L.; Karplus, M. Ligand design by a combinatorial approach based on modeling and experiment: application to HLA-DR4. *J. Comput.-Aided Mol. Des.* **2007,** 21, 395-418.

273. Bitetti-Putzer, R.; Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Functional group placement in protein binding sites: a comparison of GRID and MCSS. *J. Comput. Aided Mol. Des.* **2001,** 15, 935-60.

274. Haider, K.; Huggins, D. J. Combining solvent thermodynamic profiles with functionality maps of the Hsp90 binding site to predict the displacement of water molecules. *J. Chem. Inf. Model.* **2013,** 53, 2571-2586.

275. Roberts, B. C.; Mancera, R. L. Ligand-protein docking with water molecules. *J. Chem. Inf. Model.* **2008,** 48, 397-408.

276. Thilagavathi, R.; Mancera, R. L. Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Model.* **2010,** 50, 415-421.

277. Lie, M. A.; Thomsen, R.; Pedersen, C. N.; Schiott, B.; Christensen, M. H. Molecular docking with ligand attached water molecules. *J. Chem. Inf. Model.* **2011,** 51, 909-917.

278. Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005,** 48, 6504-6515.

279. Lemmon, G.; Meiler, J. Towards ligand docking including explicit interface water molecules. *PloS One* **2013,** 8, e67536.

280. Huang, N.; Shoichet, B. K. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008,** 51, 4862-4865.

281. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** 47, 1739-1749.

282. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004,** 47, 1750-1759.

283. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006,** 49, 6177-6196.

284. *Schrödinger, Inc.*, 120 West 45th Street, New York, 10036 USA.

285. Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **1999,** 34, 17-28.

286. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** 31, 455-61.

287. Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* **2012,** 7, e32036.

288. Tanaka, S.; Scheraga, H. A. Model of protein folding: inclusion of short-, medium-, and long-range interactions. *Proc. Natl. Acad. Sci. USA* **1975,** 72, 3802-3806.

289. Shimada, J.; Shakhnovich, E. I. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. USA* **2002,** 99, 11175-11180.

290. Rakhmanov, S. V.; Makeev, V. J. Atomic hydration potentials using a Monte Carlo Reference State (MCRS) for protein solvation modeling. *BMC Struct. Biol.* **2007,** 7, 19.

291. Woods, C. J.; Essex, J. W.; King, M. A. The Development of Replica-Exchange-Based Free-Energy Methods. *J. Phys. Chem. B* **2003,** 107, 13703-13710.

292. Woods, C. J.; Essex, J. W.; King, M. A. Enhanced Configurational Sampling in Binding Free-Energy Calculations. *J. Phys. Chem. B* **2003,** 107, 13711-13718.

293. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997,** 72, 1047-1069.

294. Bayden, A. S.; Moustakas, D. T.; Joseph-McCarthy, D.; Lamb, M. L. Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *J. Chem. Inf. Model.* **2015,** 55, 1552-1565.

295. Li, Z.; Lazaridis, T. Computing the thermodynamic contributions of interfacial water. *Methods Mol. Biol.* **2012,** 819, 393-404.

296. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** 30, 1545-1614.

297. *BiKi Hydra*, BiKi Technologies s.r.l.., Via XX Settembre: n. 33/10, 16121 Genova, Italy.

298. Zia, S. R.; Gaspari, R.; Decherchi, S.; Rocchia, W. Probing Hydration Patterns in Class-A GPCRs via Biased MD: The A2A Receptor. *J. Chem. Theory Comput.* **2016,** 12, 6049-6061.

299. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci. USA* **2007,** 104, 808-813.

300. Horbert, R.; Pinchuk, B.; Johannes, E.; Schlosser, J.; Schmidt, D.; Cappel, D.; Totzke, F.; Schächtele, C.; Peifer, C. Optimization of Potent DFG-in Inhibitors of Platelet Derived Growth Factor Receptorβ (PDGF-Rβ) Guided by Water Thermodynamics. *J. Med. Chem.* **2015,** 58, 170-182.

301. Myrianthopoulos, V.; Kritsanida, M.; Gaboriaud-Kolar, N.; Magiatis, P.; Ferandin, Y.; Durieu, E.; Lozach, O.; Cappel, D.; Soundararajan, M.; Filippakopoulos, P.; Sherman, W.; Knapp, S.; Meijer, L.; Mikros, E.; Skaltsounis, A.-L. Novel Inverse Binding Mode of Indirubin Derivatives Yields Improved Selectivity for DYRK Kinases. *ACS Med. Chem. Lett.* **2013,** 4, 22-26.

302. Brodney, M. A.; Barreiro, G.; Ogilvie, K.; Hajos-Korcsok, E.; Murray, J.; Vajdos, F.; Ambroise, C.; Christoffersen, C.; Fisher, K.; Lanyon, L.; Liu, J.; Nolan, C. E.; Withka, J. M.; Borzilleri, K. A.; Efremov, I.; Oborski, C. E.; Varghese, A.; O'Neill, B. T. Spirocyclic Sulfamides as β-Secretase 1 (BACE-1) Inhibitors for the Treatment of Alzheimer's Disease: Utilization of Structure Based Drug Design, WaterMap, and CNS Penetration Studies To Identify Centrally Efficacious Inhibitors. *J. Med. Chem.* **2012,** 55, 9224-9239.

303. Shah, F.; Gut, J.; Legac, J.; Shivakumar, D.; Sherman, W.; Rosenthal, P. J.; Avery, M. A. Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure–Activity

Relationships, Hydration Site Thermodynamics, and Reactivity Analysis. *J. Chem. Inf. Model.* **2012,** 52, 696-710.

304. Higgs, C.; Beuming, T.; Sherman, W. Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Med. Chem. Lett.* **2010,** 1, 160-164.

305. Guimarães, C. R. W.; Mathiowetz, A. M. Addressing Limitations with the MM-GB/SA Scoring Procedure using the WaterMap Method and Free Energy Perturbation Calculations. *J. Chem. Inf. Model.* **2010,** 50, 547-559.

306. Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand–Receptor Docking. *J. Med. Chem.* **2016,** 59, 4364-4384.

307. Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997,** 101, 7821-7826.

308. Chandler, D.; McCoy, J. D.; Singer, S. J. Density functional theory of nonuniform polyatomic systems. I. General formulation. *J. Chem. Phys.* **1986,** 85, 5971.

309. Truchon, J. F.; Pettitt, B. M.; Labute, P. A Cavity Corrected 3D-RISM Functional for Accurate Solvation Free Energies. *J. Chem. Theory Comput.* **2014,** 10, 934-941.

310. *Chemical Computing Group Inc.*, 1010 Sherbrooke St. W, Suite 910, Montreal, Quebec, CANADA H3A 2R7.

311. Cui, G.; Swails, J. M.; Manas, E. S. SPAM: A Simple Approach for Profiling Bound Water Molecules. *J. Chem. Theory Comput.* **2013,** 9, 5539-5549.

312. Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002,** 23, 128-137.

313. Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson–Boltzmann solvation methods. *J. Comput. Chem.* **2001,** 22, 608-640.

314. OpenEye Scientific Software. In 9 Bisbee Court, Suite D, Santa Fe, New Mexico 87508 USA.

315. Shestakov, A. I.; Milovich, J. L.; Noy, A. Solution of the nonlinear Poisson-Boltzmann equation using pseudo-transient continuation and the finite element method. *J. Colloid Interface Sci.* **2002,** 247, 62-79.

316. Lu, B.; Zhang, D.; McCammon, J. A. Computation of electrostatic forces between solvated molecules determined by the Poisson-Boltzmann equation using a boundary element method. *J. Chem. Phys.* **2005,** 122, 214102.

317. Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **2002,** 23, 1297-1304.

318. Liu, H. Y.; Zou, X. Electrostatics of ligand binding: parametrization of the generalized Born model and comparison with the Poisson-Boltzmann approach. *J. Phys. Chem. B* **2006,** 110, 9304-9313.

319. Lee, M. C.; Yang, R.; Duan, Y. Comparison between Generalized-Born and Poisson-Boltzmann methods in physics-based scoring functions for protein structure prediction. *J. Mol. Model.* **2005,** 12, 101-110.

320. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011,** 32, 866-877.

321. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Exp. Op. Drug Disc.* **2015,** 10, 449-461.

322. Buyukdagli, S.; Ala-Nissila, T. Microscopic formulation of nonlocal electrostatics in polar liquids embedding polarizable ions. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* **2013,** 87, 063201.

323. Abrashkin, A.; Andelman, D.; Orland, H. Dipolar Poisson-Boltzmann equation: ions and dipoles close to charge interfaces. *Phys. Rev. Lett.* **2007,** 99, 077801.

324. Huggins, D. J.; Tidor, B. Systematic placement of structural water molecules for improved scoring of protein-ligand interactions. *Protein Eng. Des. Sel.* **2011,** 24, 777-789.

325. Ritchie, A. W.; Webb, L. J. Optimizing electrostatic field calculations with the Adaptive Poisson-Boltzmann Solver to predict electric fields at protein-protein interfaces II: explicit near-probe and hydrogen-bonding water molecules. *J. Phys. Chem. B* **2014,** 118, 7692-7702.

326. Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 Implicit Solvation Model. *J. Chem. Theory Comput.* **2009,** 5, 2544-2564.

327. Lee, M. S.; Salsbury, F. R., Jr.; Olson, M. A. An efficient hybrid explicit/implicit solvent method for biomolecular simulations. *J. Comput. Chem.* **2004,** 25, 1967-1978.

328. Zhang, H.; Tan, T.; van der Spoel, D. Generalized Born and Explicit Solvent Models for Free Energy Calculations in Organic Solvents: Cyclodextrin Dimerization. *J. Chem. Theory Comput.* **2015,** 11, 5103-5113.

329. Yang, Y.; Lightstone, F. C.; Wong, S. E. Approaches to efficiently estimate solvation and explicit water energetics in ligand binding: the use of WaterMap. *Exp. Op. Drug Disc.* **2013,** 8, 277-287.

330. Lenselink, E. B.; Louvel, J.; Forti, A. F.; van Veldhoven, J. P. D.; de Vries, H.; Mulder-Krieger, T.; McRobb, F. M.; Negri, A.; Goose, J.; Abel, R.; van Vlijmen, H. W. T.; Wang, L.; Harder, E.; Sherman, W.; Ijzerman, A. P.; Beuming, T. Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. *ACS Omega* **2016,** 1, 293-304.

331. Kamps, J. J. A. G.; Huang, J.; Poater, J.; Xu, C.; Pieters, B. J. G. E.; Dong, A.; Min, J.; Sherman, W.; Beuming, T.; Matthias Bickelhaupt, F.; Li, H.; Mecinović, J. Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. *Nat. Commun.* **2015,** 6, 8911.

332. Fox, J. M.; Kang, K.; Sherman, W.; Héroux, A.; Sastry, G. M.; Baghbanzadeh, M.; Lockett, M. R.; Whitesides, G. M. Interactions between Hofmeister Anions and the Binding Pocket of a Protein. *J. Am. Chem. Soc.* **2015,** 137, 3859-3866.

333. Duong-Ly, K. C.; Peterson, J. R. The human kinome and kinase inhibition. *Current protocols in pharmacology* **2013,** Chapter 2, Unit2.9.

334. Gross, S.; Rahal, R.; Stransky, N.; Lengauer, C.; Hoeflich, K. P. Targeting cancer with kinase inhibitors. *J. Clin. Invest.* **2015,** 125, 1780-1789.

335. Harrison, S.; Das, K.; Karim, F.; Maclean, D.; Mendel, D. Non-ATP-competitive kinase inhibitors - enhancing selectivity through new inhibition strategies. *Exp. Op. Drug Disc.* **2008,** 3, 761-774.

336. Licht-Murava, A.; Eldar-Finkelman, H. Exploiting substrate recognition for selective inhibition of protein kinases. *Curr. Pharm. Des.* **2012,** 18, 2914-2920.

337. Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011,** 29, 1046-1051.

338. Lamba, V.; Ghosh, I. New directions in targeting protein kinases: focusing upon true allosteric and bivalent inhibitors. *Curr. Pharm. Des.* **2012,** 18, 2936-2945.

339. Garske, A. L.; Peters, U.; Cortesi, A. T.; Perez, J. L.; Shokat, K. M. Chemical genetic strategy for targeting protein kinases based on covalent complementarity. *Proc. Natl. Acad. Sci. USA* **2011,** 108, 15046-15052.

340. Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* **2002,** 1, 493-502.

341. Boschelli, D. H.; Ye, F.; Wang, Y. D.; Dutia, M.; Johnson, S. L.; Wu, B.; Miller, K.; Powell, D. W.; Yaczko, D.; Young, M.; Tischler, M.; Arndt, K.; Discafani, C.; Etienne, C.; Gibbons, J.; Grod, J.; Lucas, J.; Weber, J. M.; Boschelli, F. Optimization of 4-phenylamino-3-quinolinecarbonitriles as potent inhibitors of Src kinase activity. *J. Med. Chem.* **2001,** 44, 3965-3977.

342. Golas, J. M.; Arndt, K.; Etienne, C.; Lucas, J.; Nardin, D.; Gibbons, J.; Frost, P.; Ye, F.; Boschelli, D. H.; Boschelli, F. SKI-606, a 4-anilino-3-quinolinecarbonitrile dual inhibitor of Src and Abl kinases, is a potent antiproliferative agent against chronic myelogenous leukemia cells in culture and causes regression of K562 xenografts in nude mice. *Cancer Res.* **2003,** 63, 375-381.

343. Puttini, M.; Coluccia, A. M.; Boschelli, F.; Cleris, L.; Marchesi, E.; Donella-Deana, A.; Ahmed, S.; Redaelli, S.; Piazza, R.; Magistroni, V.; Andreoni, F.; Scapozza, L.; Formelli, F.; Gambacorti-Passerini, C. In vitro and in vivo activity of SKI-606, a novel Src-Abl inhibitor, against imatinib-resistant Bcr-Abl+ neoplastic cells. *Cancer Res.* **2006,** 66, 11314-11322.

344. Vultur, A.; Buettner, R.; Kowolik, C.; Liang, W.; Smith, D.; Boschelli, F.; Jove, R. SKI-606 (bosutinib), a novel Src kinase inhibitor, suppresses migration and invasion of human breast cancer cells. *Mol. Cancer Ther.* **2008,** 7, 1185-1194.

345. Cortes, J. E.; Kantarjian, H. M.; Brummendorf, T. H.; Kim, D. W.; Turkina, A. G.; Shen, Z. X.; Pasquini, R.; Khoury, H. J.; Arkin, S.; Volkert, A.; Besson, N.; Abbas, R.; Wang, J.; Leip, E.; Gambacorti-Passerini, C. Safety and efficacy of bosutinib (SKI-606) in chronic phase Philadelphia chromosome-positive chronic myeloid leukemia patients with resistance or intolerance to imatinib. *Blood* **2011,** 118, 4567-4576.

346. Levinson, N. M.; Boxer, S. G. Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. *PloS One* **2012,** 7, e29828.

347. Redaelli, S.; Piazza, R.; Rostagno, R.; Magistroni, V.; Perini, P.; Marega, M.; Gambacorti-Passerini, C.; Boschelli, F. Activity of bosutinib, dasatinib, and nilotinib against 18 imatinib-resistant BCR/ABL mutants. *J. Clin. Oncol.* **2009,** 27, 469-471.

348. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **1997,** 74, 1-20.

349. Supuran, C. T.; Scozzafava, A.; Casini, A. Carbonic anhydrase inhibitors. *Med. Res. Rev.* **2003,** 23, 146-189.

350. Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **2008,** 108, 946-1051.

351. Jenner, P. A2A antagonists as novel non-dopaminergic therapy for motor dysfunction in PD. *Neurology* **2003,** 61, S32-8.

352. Cristalli, G.; Lambertucci, C.; Marucci, G.; Volpini, R.; Dal Ben, D. A2A adenosine receptor and its modulators: overview on a druggable GPCR and on structure-activity relationship analysis and binding requirements of agonists and antagonists. *Curr. Pharm. Des.* **2008,** 14, 1525-52.

353. Congreve, M.; Langmead, C. J.; Mason, J. S.; Marshall, F. H. Progress in structure based drug design for G protein-coupled receptors. *J. Med. Chem.* **2011,** 54, 4283-311.

354. Mason, J. S.; Bortolato, A.; Congreve, M.; Marshall, F. H. New insights from structural biology into the druggability of G protein-coupled receptors. *Trends Pharmacol. Sci.* **2012,** 33, 249-60.

355. Tan, Q.; Zhu, Y.; Li, J.; Chen, Z.; Han, G. W.; Kufareva, I.; Li, T.; Ma, L.; Fenalti, G.; Li, J.; Zhang, W.; Xie, X.; Yang, H.; Jiang, H.; Cherezov, V.; Liu, H.; Stevens, R. C.; Zhao, Q.; Wu, B. Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* **2013,** 341, 1387-90.

356. Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010,** 330, 1066-71.

357. Andrews, S. P.; Mason, J. S.; Hurrell, E.; Congreve, M. Structure-based drug design of chromone antagonists of the adenosine A2A receptor. *MedChemComm* **2014,** 5, 571-575.

358. Poucher, S. M.; Keddie, J. R.; Singh, P.; Stoggall, S. M.; Caulkett, P. W.; Jones, G.; Coll, M. G. The in vitro pharmacology of ZM 241385, a potent, non-xanthine A2a selective adenosine receptor antagonist. *Br. J. Pharmacol.* **1995,** 115, 1096-1102.

359. Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 Å Crystal Structure of a Human A(2A) Adenosine Receptor Bound to an Antagonist. *Science* **2008,** 322, 1211-1217.

360. Segala, E.; Guo, D.; Cheng, R. K. Y.; Bortolato, A.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Heitman, L. H.; Ijzerman, A. P.; Marshall, F. H.; Cooke, R. M. Controlling the Dissociation of Ligands

from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J. Med. Chem.* **2016,** 59, 6470-6479.

361. Lunn, G.; Banks, B. J.; Crook, R.; Feeder, N.; Pettman, A.; Sabnis, Y. Discovery and synthesis of a new class of opioid ligand having a 3-azabicyclo[3.1.0]hexane core. An example of a 'magic methyl' giving a 35-fold improvement in binding. *Bioorg. Med. Chem. Lett.* **2011,** 21, 4608-11.

362. Langmead, C. J.; Andrews, S. P.; Congreve, M.; Errey, J. C.; Hurrell, E.; Marshall, F. H.; Mason, J. S.; Richardson, C. M.; Robertson, N.; Zhukov, A.; Weir, M. Identification of novel adenosine A(2A) receptor antagonists by virtual screening. *J. Med. Chem.* **2012,** 55, 1904-9.

363. Zhukov, A.; Andrews, S. P.; Errey, J. C.; Robertson, N.; Tehan, B.; Mason, J. S.; Marshall, F. H.; Weir, M.; Congreve, M. Biophysical mapping of the adenosine A2A receptor. *J. Med. Chem.* **2011,** 54, 4312-23.

364. Congreve, M.; Andrews, S. P.; Dore, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H. Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J. Med. Chem.* **2012,** 55, 1898-903.

365. Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water network perturbation in ligand binding: adenosine A(2A) antagonists as a case study. *J Chem Inf Model* **2013,** 53, 1700-13.

366. Krimmer, S. G.; Betz, M.; Heine, A.; Klebe, G. Methyl, Ethyl, Propyl, Butyl: Futile But Not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors. *ChemMedChem* **2014,** 9, 833-846.

367. Eijsink, V. G.; Veltman, O. R.; Aukema, W.; Vriend, G.; Venema, G. Structural determinants of the stability of thermolysin-like proteinases. *Nat. Struct. Biol.* **1995,** 2, 374-9.

368. Biela, A.; Betz, M.; Heine, A.; Klebe, G. Water makes the difference: rearrangement of water solvation layer triggers non-additivity of functional group contributions in protein-ligand binding. *ChemMedChem* **2012,** 7, 1423-34.

369. Biela, A.; Nasief, N. N.; Betz, M.; Heine, A.; Hangauer, D.; Klebe, G. Dissecting the Hydrophobic Effect on the Molecular Level: The Role of Water, Enthalpy, and Entropy in Ligand Binding to Thermolysin. *Angew. Chem. Int. Ed.* **2013,** 52, 1822-1828.

370. Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Körner, M.; Kuglstatter, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.;

Woltering, T.; Stahl, M. A Real-World Perspective on Molecular Design. *J. Med. Chem.* **2016,** 59, 4087-4102.

371. Stahl, M.; Baier, S. How Many Molecules Does It Take to Tell a Story? Case Studies, Language, and an Epistemic View of Medicinal Chemistry. *ChemMedChem* **2015,** 10, 949-956.

372. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current Progress in Structure-Based Rational Drug Design Marks a New Mindset in Drug Discovery. *Comp. Struct. Biotech. J.* **2013**, *5*, e2013202011.

373. Hanashima, S.; Kato, K.; Yamaguchi, Y. 13C-NMR quantification of proton exchange at LewisX hydroxyl groups in water. *Chem. Commun.* **2011,** 47, 10800-10802.