Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Water Molecules at Protein-Drug Interfaces: Computational Prediction and Analysis Methods

Marley L. Samways,[a] Richard D. Taylor,[b] Hannah E. Bruce Macdonald,[c,d] and Jonathan W. Essex*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The fundamental importance of water molecules at drug-protein interfaces is now widely recognised and a significant feature in structure-based drug design. Experimental methods for analysing the role of water in drug binding have many challenges, including the accurate location of bound water molecules in crystal structures, and problems in resolving specific water contributions to binding thermodynamics. Computational analyses of binding site water molecules provide an alternative, and in principle complete, structural and thermodynamic picture, and their use is now commonplace in the pharmaceutical industry. In this review, we describe the computational methodologies that are available and discuss their strengths and weaknesses. Additionally, we provide a critical analysis of the experimental data used to validate the methods, regarding the type and quality of experimental structural data. We also discuss some of the fundamental difficulties of each method and suggest directions for future study.

1 Introduction

1.1 Background

After many years of rising interest, the importance of water molecules in drug binding is now widely accepted, and, given the prevalence of water-mediated interactions between small molecules and their protein targets,[1–4] is of fundamental importance in modern structure-based drug design programmes. There are many examples in the literature where compounds designed to displace a given water molecule have shown an increase[5,6] or decrease[7,8] in binding affinity. In addition, the hydration pattern of a binding site can strongly influence the selectivity[9,10] or promiscuity[11] of the site for small molecules.

Water molecules which are bound to protein cavities are more restricted than those in bulk solvent, and as such, their binding to the protein typically comes at an entropic cost — the maximum amount of entropy that can be lost by constraining a water molecule has been estimated by Dunitz to be around 2 kcal mol^{-1} at room temperature, based on the entropy of ice formation.[12] When a water molecule is displaced from the binding site by a small molecule, this entropy is relased and contributes favourably to the ligand binding affinity. However, if the interactions that the water makes with the protein are not recovered by the ligand, the resulting unfavourable enthalpic effect may outweigh the favourable entropic effect of water displacement — in such a case, ligand binding might be better served by stabilisation of the water, rather than displacement. The relative magnitudes of these effects are not known *a priori*, and it is therefore difficult to know which approach to take. As a result of differences in water binding, sometimes very similar compounds can bind with similar free energies, but significantly different enthalpies and entropies.[13]

One of the earliest examples of the importance of water molecules in protein complexes is seen in HIV-1 protease, where a key feature in molecular design is the water molecule which forms four hydrogen bonds with the ligand and Ile50 and Ile50' residues (Fig. 1).[14] Further examples include small molecules that target TYK2 and JAK3 kinases,[15] PDE4 inhibitors[16] and Adenosine A2A inhibitors.[17] A notable example currently in clinical trials is the molecule GLP1690, developed by Galapagos, which is a first-in-class molecule for idiopathic pulmonary fibrosis, where the effect of the nitrile group is thought to be governed by water displacement.[18] Additionally, a series of c-KIT inihibitors have been found to exhibit selectivity over an important off-target via the stabilisation of a water network by a triazole group.[19,20]

To fully account for the effects of water molecules in the binding site of a therapeutic compound, a medicinal chemist would generally consider the following questions:

- Where at the target-ligand interface do water molecules bind?

[a] *School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom. E-mail: J.W.Essex@soton.ac.uk*
[b] *Computer-Aided Drug Design, UCB, 216 Bath Road, Slough, SL1 3WE, United Kingdom.*
[c] *Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, 10065, United States of America.*
[d] *MSD, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, United Kingdom*

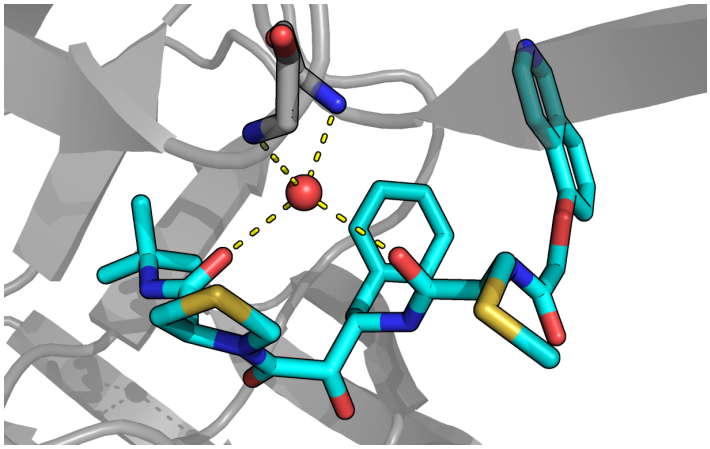

Fig. 1 Image of the crystal structure of HIV-1 protease in complex with the inhibitor KNI-272 (PDB: 1HPX), with the key water molecule shown. This water molecule is known to offer a significant affinity boost upon displacement. The backbone atoms of the Ile50 and Ile50' residues are also shown as sticks.

- Which of these water molecules are important in mediating binding and selectivity?
- For each of these water sites, will optimisation of the bound ligand be better served by displacement or stabilisation?

This review will discuss how these questions may be answered *a priori*, and in particular the role of computational approaches.

1.2 Experiment

Structural water molecules are difficult to study experimentally, with X-ray crystallography the method most commonly used to identify water binding sites. Once a suitable single crystal of the structure of interest has been obtained (a process which is typically far from trivial), an X-ray diffraction pattern is collected. The intensities of the reflections are converted into structure-factors, which are then subjected to a Fourier transform (after their phases have been determined), yielding a three-dimensional map of the electron density.[21] An initial structural model is then proposed, refined to better fit the electron density, and the model is then iteratively rebuilt and refined. It then becomes possible to assign regions of electron density that may have initially been unclear, including water molecules.[21]

One potential issue, common to all protein crystallography, is that the experimental conditions required to precipitate crystal formation (such as pH, salt content, temperature, etc.) may differ significantly from physiological conditions, and could distort the protein structure.[22] Two key issues arise when determining water locations in protein crystallography. First, if the water is disordered, or the site partially occupied, the electron density may be difficult to resolve, making the refinement process problematic. Second, the electron density of water can be similar to that of small, isoelectronic ions (such as sodium or ammonium), which may well be present in the crystal structure, and could result in incorrect assignment.[21] Additionally, water molecules may be misassigned to noise in the electron density, or used to 'absorb' unexplained density.[23] The issues associated with the identification of water molecules in protein crystal structures has been highlighted by studies in which the same, or very similar, structures have been solved by different groups, with different water locations.[24,25]

Some of the aforementioned ambiguity in the analysis of crystallographic water molecules could be resolved by using neutron, rather than X-ray, diffraction to analyse the structures. Unlike X-ray diffraction, neutron diffraction is able to identify the locations of hydrogen (or deuterium) atoms, even at modest resolutions, owing to their increased scattering of neutrons, relative to X-rays.[26,27] However, it should be noted that neutron diffraction structures are much less common, with only 179 of these present in the Protein Data Bank[28,29] which contains over 176,000 structures (as of 1st April 2021) — this is partly caused by the need for much larger crystals than are required for X-ray diffraction.[26]

Nuclear magnetic resonance (NMR) spectroscopy can also be used to indicate the locations of water molecules within protein structures.[30–33] This typically involves the use of the nuclear Overhauser effect (NOE) and/or rotating-frame Overhauser effect (ROE) to infer short-range interactions between water protons and protein nuclei, and therefore the approximate locations of water molecules. An advantage of this approach is that it can be used with solution NMR spectroscopy, which avoids the need to crystallise the protein complex, and that combination of the NOE with the ROE can be used to indicate the relative persistence of different water-protein interactions, as well as distinguish water sites proximal to the protein from those undergoing proton exchange, based on the signs of these effects.[34] However, this analysis is hampered by the typically short-lived nature of protein-water interactions,[35] by hydrogen exchange with labile groups on the protein[30,31,35] and also by long-range dipole coupling.[36] It should be noted that some of these effects can be reduced by encapsulating the protein complex in a reverse micelle, which slows the hydration dynamics and reduces long-range dipole effects.[34] Additionally, NMR can be used via nuclear magnetic relaxation dispersion (NMRD)[32] and Overhauser dynamic nuclear polarisation (ODNP)[37,38] to study residence times and dynamics of buried water molecules.[39] Nevertheless use of these methods remains significantly less common than X-ray diffraction, and as such, there is comparatively little experimental data available for protein-ligand complexes.

The experimental methods discussed thus far have all been structural techniques, which seek to identify the locations of water molecules within a macromolecular structure. As well as the location where a water molecule binds, the enthalpic and entropic components of binding are vital in understanding the role of a water within a system, and its impact on the drug design process. However, it is currently not possible to carry out thermodynamic measurements for a single water molecule within a protein system experimentally, as the effects cannot be isolated from all other water molecules. It is possible to compare rigorous affinity measurements of very similar structures with different degrees of binding site hydration.[40] However, this comparison still does not isolate the effects of the water molecule, as there will likely be many other factors influencing the difference in measurement. For this reason, only structural experimental analyses (primarily X-ray crystallography) will be discussed in this review.

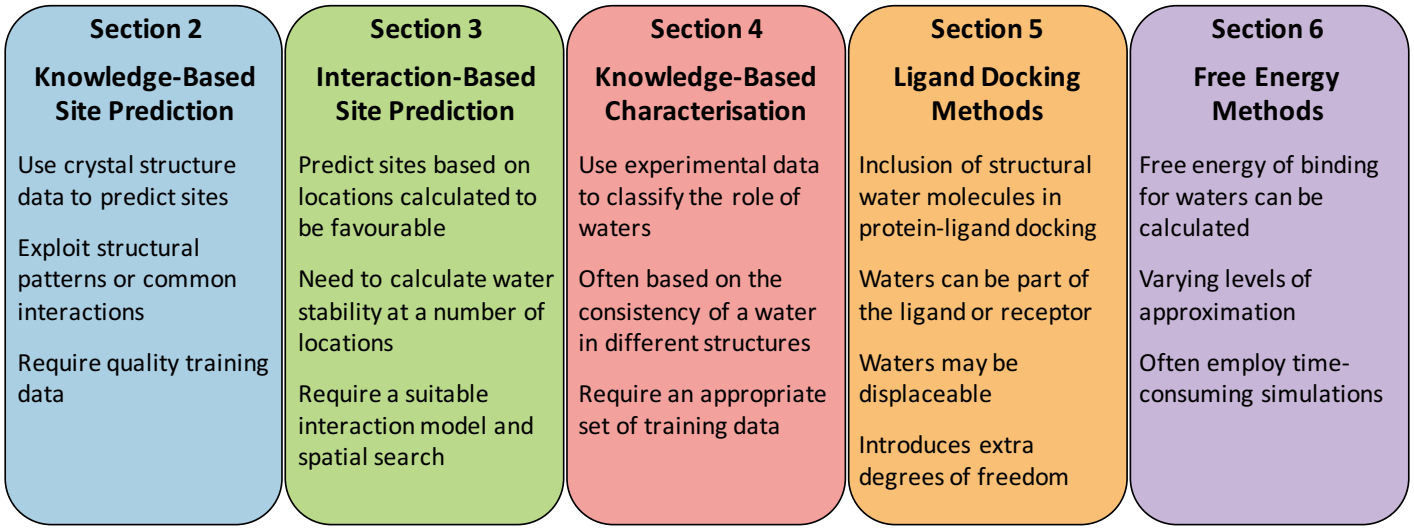

Fig. 2 Brief summaries of the five classes of methods discussed in this review, along with the sections in which they are contained.

1.3 Computation

Owing to the aforementioned difficulties associated with experimental measurements in this area, there is an opportunity for computational techniques to offer unique insight. Experimental limitations have been discussed for systems where the structure of a given protein-ligand complex is available. However, the structure may not be available: either it is difficult to crystallise or solve, or the ligand of interest may not yet have been synthesised. For these cases, computational methods may be able to predict the locations of active site water molecules.

Alongside the capacity to act as a molecular-level microscope, these techniques also permit the application of theory to allow the binding free energy of individual water molecules to be calculated, which cannot be directly measured by experiment.[41] Understanding the thermodynamics of an active-site water molecule can provide guidance as to whether ligand optimisation should aim to displace or retain the water molecule, as the energetic cost of displacement would need to be recovered by the ligand.

There are numerous methods and software packages, both open source and commercial, which offer prediction and/or characterisation of hydration sites. This review will describe the theoretical underpinning of the main methods, along with a discussion of how their performances have been validated. Our intention is that this review will serve as a reference for those seeking to use these methods, such that their strengths and limitations can be understood.

Choosing the appropriate water analysis method for a given project will depend on various factors: the speed of computation; availability of the software; detail of information afforded; and the reliability of the method. While there is an abundance of publications describing and evaluating particular methods, there are few publications comparing the relative successes or merits of different methods, and those that exist often compare only a few.[42] As comparative studies are limited, how the performance of a method is determined is important; differing criteria, or thresholds, may be used to classify predictions of water locations as either correct or incorrect. Here we summarise and discuss both the success rates quoted, and the differences in how methods are scored. Those approaches which characterise water sites via conservation probabilities or binding free energies, are much more difficult to quantitatively assess, owing to the inability to make a direct comparison with experimental measurements. Further, the treatments of water molecules in holistic methods (which focus on multiple aspects of a structure simultaneously), such as protein-ligand docking, are very difficult to compare, as there will be many other differences between two methods, making it challenging to attribute performance to a single aspect. For this reason, this review will focus on the validation of water location predictions.

To determine the success rate of a computational method for water prediction, the computationally predicted water locations are often compared to crystallographic locations. These water locations can be limited in their accuracy, for the reasons discussed previously. For a robust comparison, as many high- quality crystallographic structures as possible should be used for assessment. Where possible, we have collated data on the structures used to validate each computational approach. The quality and diversity of the structures used for validation will also be considered. On the one hand, repetition of the same structure with various methods provides consistency, but over-use of particular complexes or families of structures may result in methods that are over-trained and not broadly applicable.

This review separates the methods discussed into five groups, as indicated by Fig. 2, with the sections organised as follows: section 2, knowledge-based site prediction; section 3, interaction-based site prediction; section 4, knowledge-based characterisation; section 5, ligand docking methods; section 6, free energy methods. We also discuss the domain of applicability for each method class, as well as unsolved challenges for the field in section 7, including an analysis of the structural data used to validate algorithms, and apparent biases in the proteins used for testing.

Table 1 Brief descriptions of the knowledge-based site prediction methods, along with their reported performances and the number of crystal structures used in their testing (the number in parentheses indicates the estimated number of distinct proteins covered).

Method	Description	Validation	Structures
AcquaAlta [43]	Observed hydrogen bond geometries are used to map water positions onto a structure, with ab initio results used to prioritise the sites by interaction strength.	For one test set, 76 % of crystallographic waters are located, and 66 % were identified in a second set, using an accuracy threshold of 1.4 Å.	34 (18)
AQUARIUS [44,45]	Observed distributions of water sites around amino acids are mapped onto a structure.	The first version predicted 41 % of crystallographic sites to within 1.0 Å (73 % within 1.8 Å). The second version predicted 66 % to within 1.0 Å.	7 (7)
MCRS [46]	Knowledge-based potentials are constructed for water interactions with different protein atoms, using a Monte Carlo reference state (as described in the main text).	For each structure tested, the locations of between 75 and 90 % of crystallographic water molecules were reproduced (the accuracy threshold was not reported).	16 (16)
SuperStar [47,48]	Scatter plots of water positions around functional groups from the IsoStar knowledge base are mapped onto a structure.	88 % of crystallographic water molecules have greater interaction propensities than obtained from random sampling.	50
WarPP [49]	Points around the protein and ligand forming hydrogen bonds with near-ideal geometries are selected, then refined, clustered and optimised.	80 % of crystallographic sites with two or more hydrogen bonds are predicted to within 1.0 Å.	1509 (690)
wPMF [50]	Interaction potentials are constructed using radial distribution functions from crystal structures, then used to identify favourable regions for water binding.	The accuracy of predicted sites is greater for sites with higher scores. The scores are also highly correlated with temperature factor.	237 (129)
Xiao *et al.* [51]	Protein atom triplets are identified within a structure and waters are inserted to create tetrahedral units, based on observed tetrahedra in crystal structures.	88 % of crystallographic water sites are correctly predicted to within 2.0 Å (73 % within 1.5 Å).	193 (145)

2 Knowledge-based site prediction

There are a large number of methods in the literature which predict the locations of crystallographic water sites using similar crystallographic data. These methods are useful when the protein conformation is available, but the locations of water molecules are unclear, such as when using homology models, or if the crystallographic resolution of water locations is poor. These methods typically exploit structural patterns around crystallographic water sites, as observed in a 'knowledge base' of structures, to locate likely regions of water binding in a locally similar structure. Another common approach is the generation of a knowledge-based potential (KBP), which quantitatively ranks different interactions based on the frequency with with they are observed in a training database, where more frequently observed interactions are assumed to be more stable. [52] The methods discussed in this section are summarised in Table 1.

A number of techniques are used to apply structural patterns from training data to a query structure. AQUARIUS is a method which superimposes observed distributions of crystallographic water molecules around amino acids onto a structure, giving a number of predicted water sites (as discrete points) distributed around the protein, [44] or alternatively, a probability distribution of water. [45] SuperStar follows a somewhat similar approach, mapping a scatterplot of waters onto a query structure, using fragment-fragment interaction distributions from the IsoStar database, [53] generating a water propensity map around the structure of interest. [47,48] AcquaAlta uses observed hydrogen-bonding geometries from the Cambridge Structural Database (CSD) [54] to locate bridging water molecules, with data from quantum mechanical calculations used to prioritise different hydrogen bond types. [43] The method published by Xiao *et al.* uses a statistical tetrahedral model to locate water molecules, based on triplets of protein atoms and their observed interaction geometries with waters from a training database. [51] The WarPP method searches the free space around hydrogen bonding groups, and ranks locations based on the quality of the hydrogen bonds that would be formed, relative to observed hydrogen bond geometries. [49] These points are then refined and clustered, and the inserted water molecules are optimised together.

The Monte Carlo reference state (MCRS) method constructs KBPs between water molecules and each type of protein atom, with pseudorandom sampling used as a reference null-interaction, and then samples a grid with these potentials to identify the most favourable locations for water molecules. [46] The water PMF (wPMF) method also uses a KBP between water and protein atoms, but based on observed radial distribution functions (RDFs), and clusters favourable locations (after sampling points on a lattice) to identify water sites. [50]

These methods are typically executed rapidly (which is ideal for screening) and are often able to achieve good agreement between predicted and observed water binding sites for the systems which have been tested. This is because the patterns typically observed in crystal structures will be reproduced. However, it is important to note that these methods can be limited by the quality and relevance of the structures comprising the knowledge base. For example, if a particular structural motif is underrepresented in the training data, predictions surrounding these motifs will likely be less accurate than others — this may prove a problem with small molecules containing uncommon functional groups, for which the hydrogen bonding has not been well studied. Additionally, these methods will typically only be able to predict the stable, ordered water sites which are well resolved in X-ray crystallography, whilst

Table 2 Brief descriptions of the interaction-based site prediction methods, along with their reported performances and the number of structures used in their testing (the number in parentheses indicates the estimated number of distinct proteins covered).

Method	Description	Validation	Structures
3D-RISM [55–57]	A static, continuous water distribution is solved for a given region, based on statistical mechanics, using integral equation theory.	Peaks in the resulting solvent distribution are reported to correspond to crystallographic water sites.	9 (1)
Ben-Shalom *et al.* [58]	Monte Carlo translations of water molecules between points on a grid (which covers both the protein and bulk solvent) are attempted during molecular dynamics simulations.	Transitions between occluded protein binding sites are observed for the MC/MD simulations which are not observed over very long MD simulations. The aim is to generate the correct water distribution, rather than specific individual sites.	2 (2)
Dowser [59–61]	Water molecules are inserted into cavities and optimised, before keeping/discarding sites based on interaction energy.	The predictive accuracy increased with each methodological development (Dowser, Dowser+, Dowser++) from 63 to 74, to 85 % (within 2.0 Å) for the same test set.	15 (2)
Fold-X [62]	A binding site is flooded with water particles and the positions are then optimised.	Crystallographic water sites that coordinate to two or more protein atoms were identified in 76 % of cases (no accuracy threshold given). These sites were predicted with a root mean square error of 0.7 Å.	50
GAsol [63]	A genetic algorithm is used to generate a set of water sites which best fit the distribution calculated with 3D-RISM.	95 % of binding site water molecules are correctly predicted within 2.0 Å.	187 (4)
GRID [64]	A water probe is used to sample positions on a grid, identifying regions of favourable interactions.	Points of most favourable water interactions correspond to crystallographic water positions.	1 (1)
MCSS [65]	Many copies of a given probe are inserted into a structure, then gradually optimised whilst removing overlap.	No validation reported for water probes.	N/A
Placevent [66]	A solvent distribution calculated with 3D-RISM is used to place water sites at peaks in the density.	Crystallographic water molecules in the binding site were well reproduced, with an average prediction error of 0.5 Å.	3 (2)
Setny and Zacharias [67]	A semi-explicit water model is used to sample points on a grid, and then a solvent distribution is evolved via a cellular automata approach.	64 % of grid points within 1.4 Å of a crystallographic water site were correctly predicted as being hydrated.	16 (15)
WaterDock [68,69]	Water molecules are docked into the binding site many times, with the results clustered into hydration sites.	97 % of water sites are correctly predicted by the first version (2.0 Å). The second version predicted 91 % for the same test set, with a reduction in false positive predictions.	37 (12)
Water Flooding [70,71]	A large number of water configurations are generated by filling cavities and then post-processing with Monte Carlo sampling.	Variation in the number of water sites with chemical potential showed similar results to a more rigorous Monte Carlo simulation.	2 (2)
WATGEN [72]	A binding site is flooded with water, then the sites forming the best interactions with the protein are kept, removing any clashes.	88 % of crystallographic water sites were correctly predicted (within 2.0 Å), compared to an expected accuracy of 40 % for random placement.	101 (89)

more disordered sites may be missed.

3 Interaction-based site prediction

As well as the knowledge-based methods used to predict the locations of water molecules within a structure, a large number of methods have been developed to predict favourable water binding locations based on models used to calculate interaction strengths. These models vary significantly in complexity, from simple, largely heuristic models to rigorous, physics-based force fields.[73] Additionally, the methods differ in how the space in and around the structure is sampled, with many of the methods divided into two groups — those which involve probing points on a lattice, and those which involve flooding cavities. The methods discussed in this section are summarised in Table 2.

Grid-based probe sampling methods have existed for a long time, and are one of the oldest approaches in the prediction of water locations. This began with the publication of the GRID method by Goodford in 1985, in which a water probe was used to sample points on a grid, with favourable regions contoured by the interaction energy.[64] Setny and Zacharias published an approach which marks grid cells as occupied or unoccupied based on the interaction strength (with a continuum model used for water-water interactions), and iteratively refines the solvent distribution.[67] The 3D reference interaction site model (3D-RISM) uses an integral equation theory to resolve a continuous solvent distribution onto a grid, based on statistical mechanics.[55–57] The Placevent method uses the 3D-RISM distribution to identify likely points for water molecules and iteratively places water sites at the most probable locations.[66] Somewhat similarly, the GAsol program uses a genetic algorithm to generate a set of water molecules

which best represents the 3D-RISM distribution.[63] Ben-Shalom *et al.* have recently developed a Monte Carlo method which attempts translations of water molecules between sites on a grid to allow transitions between occluded sites and bulk water during molecular dynamics simulations.[58,74]

Flooding-based sampling methods are based on filling cavities or regions in the protein with water molecules which are then filtered or refined via some form of interaction or energy-based threshold. The multiple copy simultaneous search (MCSS) method, after inserting a large number of probes, discards those with insufficient interaction energies, then optimises each probe separately, discarding any which overlap.[65] The Fold-X force field[83] has been applied in this context by optimising a cloud of water molecules (generated from crystallographic water binding patterns around protein atoms).[62] The WATGEN method first scores the inserted water molecules based on the hydrogen bonds formed, and then selects water molecules in decreasing order of score, eliminating any which clash with higher scoring sites.[72] The Dowser method first minimises the water molecules and each is kept if the interaction energy is sufficiently negative.[59] Various parameter refinements led to Dowser+,[60] while in Dowser++[61] the flooding method of insertion was replaced with the WaterDock protocol.[68] The Water Flooding (WF) approach generates a large number of water configurations for a particular cavity and then constructs a most probable arrangement of waters, using the Monte Carlo method to sample from the set of configurations generated.[70,71] The WaterDock method uses AutoDock Vina[84] to repeatedly dock water molecules into a structure, and after filtering waters based on docking score, clusters the inserted waters into discrete hydration sites.[68,69]

The methods presented in this section allow a significant degree of control over the rigour of the prediction procedure, and therefore, the time required for execution. This could be achieved by varying the extent of the spatial sampling around the protein, or by using more or less elaborate functions to describe the interaction energies. A common issue shared by many of the methods described in this section is the neglect of explicit water-water interactions. Some of these approaches (Dowser, GRID and WaterDock) consider a single water at a time, and therefore may miss any sites for which hydrogen bonding between waters is important. This problem might be alleviated by running the prediction procedure multiple times, after predicting locations for each hydration shell, but this would still have issues with co-dependent water molecules. It should also be noted that the results from methods which produce solvent distributions (GRID, 3D-RISM and the method by Setny and Zacharias) rather than discrete water sites may be more difficult to interpret in a drug design context, especially when the peaks of these distributions are non-spherical or indicate partial occupancy. Conversely, the grid-based nature of the results provides some additional practicality, such as when incorporating these data into docking programs which utilise interaction grids.[85]

4 Knowledge-based characterisation

Whilst crystallographic data are primarily used to identify the locations of water molecules in a structure, there are a number of approaches which seek to characterise water molecules using these data. These methods generally seek to predict the fate of a crystallographic water molecule upon the binding of a small molecule, i.e. whether the water molecule is likely to be conserved or displaced. There appear to be two distinct approaches to this problem — one is based on the presence or absence of a water molecule in similar structures, and the other is to model each water site with a set of chemical descriptors. The methods discussed in this section are summarised in Table 3.

A common approach in structural comparison for the purpose of characterising water molecules is to cluster water positions from aligned structures with some degree of sequence similarity. The probability of each water being conserved is then inferred from the cluster occupancy (the proportion of the structures which contain a given water molecule). There are a number of similar approaches which employ this method, which differ slightly in implementation and the clustering protocol used. These methods include WatCH,[81] PyWATER,[80] ProBiS H2O[79] and that published by Bottoms *et al.*[77]

Consolv predicts whether a water molecule in an *apo* structure is likely to be conserved or displaced upon ligand binding using a *k*-nearest neighbours algorithm, where each water molecule is described by the atomic density, hydrophilicity, number of hydrogen bonds and temperature factor.[78] The *k* most similar water molecules (based on these parameters) in a knowledge base are used to predict the fate of a specific water site based on whether more of the *k* waters are conserved or displaced. The WaterScore method similarly predicts the probability of waters in an *apo* structure being conserved in the corresponding *holo* structure, using a logistic regression of the temperature factor, number of protein contacts and solvent accessible surface area.[82] Amadasi *et al.* published two closely related methods employing the HINT[86–88] and Rank[89] functions to characterise water sites. These methods used slightly different models to characterise water molecules, based on whether they would be likely conserved, or displaced sterically (by a hydrophobic group) or functionally (by a hydrophilic group).[75,76] Ross *et al.* presented a tree-based machine learning model to predict water conservation, trained using a hydrogen bonding term[84] and heuristic models[90] for hydrophilicity and lipophilicity.[68]

Similar to the knowledge-based location prediction methods, these approaches are often quickly executed, owing to the rapid speed with which many of the necessary calculations can be performed. The accuracy of these methods can be difficult to assess as many of the predictions made are difficult to objectively and independently validate. For example, conservation probabilities can only be compared to the structures currently available, many or all of which will have been used in the training. Again, the quality, and therefore utility, of the results obtained is highly dependent on the data used in the training stage. If the knowledge base is overpopulated with structures from a compound series which displace or conserve a particular water site, the results concerning this site will be biased. Additionally, it should be noted that the methods which employ the temperature factor of a water site can only be applied to crystallographic sites, and not predictions.

Table 3 Brief descriptions of the knowledge-based site characterisation methods, along with their reported performances and the number of structures used in their testing (the number in parentheses indicates the estimated number of distinct proteins covered).

Method	Description	Validation	Structures
Amadasi *et al.* [75,76]	The HINT and Rank scoring functions are combined to characterise water sites. The method was later extended using nonlinear regression and pseudo-Bayesian analysis.	Initially, 76 % of waters were correctly characterised as conserved/functionally displaced or sterically displaced/missing. With a different test set, the second model correctly characterised 87 % of sites.	33 (13)
Bottoms *et al.* [77]	Water sites from equivalent structures are clustered to predict the conservation probability, based on the cluster occupancies.	Water sites with higher conservation probabilities are typically found in regions of high sequence conservation.	244 (235)
Consolv [78]	An optimised k-nearest neighbours algorithm predicts whether waters are likely to be conserved or displaced, based on a set of four descriptors.	75 % of apo water sites are correctly identified as conserved or displaced.	10 (7)
Ross *et al.* [68]	A decision tree machine learning algorithm is used to predict whether water sites will be conserved or displaced upon ligand binding, using three descriptors.	75 % of water sites are correctly identified as conserved or displaced. 80 % of displaced sites were correctly predicted to have been displaced by a polar or nonpolar group.	85 (85)
ProBiS H2O [79]	Water sites from similar structures are clustered to predict the conservation probability.	The conservation scores for the water sites were noted to be consistent with results from PyWATER.	19 (17)
PyWATER [80]	Water sites from similar structures are clustered to predict the conservation probability.	Key water sites from previous studies were identified as conserved.	(>4)
WatCH [81]	Water sites from similar structures are clustered to predict the conservation probability.	Higher conservation probability is observed for sites with lower thermal mobility and in more hydrophilic environments.	19 (3)
WaterScore [82]	The probability of waters being conserved/displaced is modelled using a logistic regression, based on three descriptors.	Between 67 % and 72 % of water sites are correctly predicted as conserved or displaced, depending on the threshold used.	8 (4)

5 Ligand docking methods

The docking of small molecules to macromolecular targets has long been used as a rapid virtual screening technique in the pharmaceutical industry.[85] As the influence of water molecules on ligand binding became recognised, docking programs began to model these effects. Many different approaches are taken, regarding the flexibility of the water molecules (in terms of translational and rotational freedom) and their environment, as well as the displaceability of the water sites. There are also distinct approaches as to whether the water molecules are considered as part of the receptor or the ligand. The methods discussed in this section are summarised in Table 4.

The majority of docking programs which have accounted for water molecules typically do this in a protein-centric fashion, such that the water molecules are considered a part of the receptor, often requiring predetermined water locations. The FlexX method [108] allows fixed water molecules to influence the docking of small molecule fragments, as they are grown into a ligand, with the water molecules included/excluded at each stage, whichever produces a better score.[99] The SLIDE method uses Consolv [78] to predict the conservation probability of each water site, and if a ligand-water clash cannot be resolved by moving the water molecule, it is considered displaced and a penalty is applied, depending on the conservation probability.[106] Whilst Glide originally only allowed rigid, non-displaceable waters in docking,[100,101] Glide XP used grid sampling to add water molecules to a complex after docking the ligand, with penalties applied depending on the polarity of the groups hydrated.[102] The GOLD program [109] was extended to allow rotationally flexible water molecules to be conserved or displaced during docking, with an entropic penalty applied to conserved water sites.[103] Similarly, FITTED allows discrete, flexible water sites to be conserved or displaced, applying an entropic penalty to conserved waters.[98] The DOCK program [110] uses separate interaction grids for each water site, against which the ligand is scored to determine if the site should be conserved or displaced.[95] Additionally, a function to model the unbinding of water molecules, based on grid-based inhomogeneous solvation theory (GIST, discussed in the following section),[111] has been implemented alongside DOCK by Sun *et al.* [96] and also separately by Balius *et al.* [97] to account for water displacement. GIST has also been used alongside AutoDock,[112] where the GIST results are added to the interaction grids to influence the docking process, and a desolvation penalty is applied for grid points covered by ligand atoms.[92] The WScore method, based on Glide XP,[102] allows flexible waters to be displaced, using results from a WaterMap analysis [113–115] (discussed in the following section) to determine the impact on complex stability.[107]

A ligand-centric approach to including water typically involves attaching water molecules to the ligand, which does not require the water positions in the binding site to be known (or predicted) beforehand. In AutoDock,[112] water molecules are constrained with respect to the ligand, and at each stage in docking are assessed as to whether they should be conserved or displaced, including an entropic contribution.[91] Lie *et al.* published a very similar approach, in which the water molecules are made flexible by allowing their rotation about hydrogen bonding groups on the ligand, also applying an entropic penalty to conserved waters.[104]

The RosettaLigand docking tool [116] can be used to include wa-

Table 4 Brief descriptions of the ligand docking methods, along with their reported performances and the number of structures used in their testing (the number in parentheses indicates the estimated number of distinct proteins covered).

Method	Description	Validation	Structures
AutoDock [91]	Displaceable, rigid water molecules are attached to, and docked alongside the ligand.	Improvements in docking range from none to very significant. Accuracy was not noted to deteriorate in any case.	1447 (208)
AutoDock-GIST [92]	A grid-based inhomogeneous solvation theory (GIST) analysis is performed to generate data regarding the desolvation penalty for sites around the protein, and included alongside other interaction grids in AutoDock.	Including the GIST results improves the percentage of ligands docked to within 2.0 Å of the crystallographic pose from 83 % to 96 % or 90 % (depending on the parameters used). Enrichment factors of docking are also improved.	23 (1)
DeepWATsite [93]	Docked poses are rescored using a convolutional neural network model. Thermodynamic details of water sites can be obtained using WATsite [94] and incorporated into the rescoring.	The crystallographic binding pose is assigned the highest rank in 89 % of cases. The inclusion of the water data offers a significant improvement over neural networks based only on the protein-ligand interactions.	2406 (703)
DOCK [95]	Displaceable water molecules are included in docking as part of a flexible receptor.	Docking improves for all but one case with displaceable water. When water is non-displaceable, the accuracy deteriorates significantly.	24 (24)
DOCK-GIST [96,97]	A function is used to model the displacement of water molecules from their environment to bulk water, based on grid-based inhomogeneous solvation theory (GIST).	62 % of water sites are predicted to within 1.0 Å (75 % within 1.5 Å) for 124 structures. Docking success rate increases from 45 % to 56 % when including the GIST-based function. [96] In prospective screening, the inclusion of the GIST term was found to significantly improve the hit rates and ligand geometries. [97]	314 (100)
FITTED [98]	Displaceable water molecules are included in docking as part of a flexible receptor.	79 % of ligands are self-docked to a rigid protein within an RMSD of 2.0 Å of the experimental pose when including displaceable water (an improvement from 67 % with no water). 82 % of crystallographic water sites are correctly conserved or displaced.	33 (5)
FlexX [99]	Spherical, displaceable water molecules are included in docking as part of a flexible receptor.	28 % of docking test cases improve when displaceable water is included and 23 % deteriorate. 35 % of bridging waters are predicted (those with fewer than two protein contacts cannot be predicted).	200 (120)
Glide [100–102]	Crystallographic water molecules were initially included in docking, and later made flexible.	When waters are rigid, success can depend on the waters selected. When the sites were made flexible, an increase in docking performance was observed.	288 (147)
GOLD [103]	Displaceable, semi-flexible water molecules are included in docking as part of the receptor environment.	Use of displaceable water sites causes docking to improve in one test set and deteriorate in two others.	186 (127)
Lie *et al.* [104]	Displaceable, rigid water molecules are attached to, and docked alongside the ligand.	Docked water molecules correspond to crystallographic sites.	24 (20)
RosettaLigand [105]	Water molecules can be included in docking either as part of the ligand or the receptor.	Inclusion of water molecules gives a net increase in docking accuracy. The ratio of improved to worsened cases ranged from 1.2:1 to 9.0:1.	351 (207)
SLIDE [106]	Displaceable water molecules are included in docking, using Consolv to determine whether or not to penalise displacement.	The effects of water inclusion were not reported.	2 (2)
WScore [107]	Displaceable, flexible water molecules are included in docking, using WaterMap results for guidance.	WScore was observed to outperform both Glide SP and Gide XP in pose prediction, affinity prediction and screening enrichment.	991 (22)

ters in docking in either a protein- or ligand-centric manner. [105] In the protein-centric option, the waters are allowed to move around their initial position as part of the flexible receptor, whereas in the ligand-centric mode, they are allowed to independently reorganise after the ligand is docked.

Whilst not strictly a docking method, the DeepWATsite method can be used to rank poses obtained from docking using convolutional neural networks, based on 3D atomic density grids. [93] This method places water sites using a combination of 3D-RISM [55–57] and GAsol, [63] and then calculates their thermodynamics using the WATsite method [94] (discussed in the following section), which are then fed into the neural network, alongside the structural data of the protein-ligand complex. The authors report that the inclusion of the water features drastically improves the pose ranking of this deep learning method. [93]

Docking calculations are very fast, and, as such, are widely used in initial screening of small molecules. A range of different approaches are taken to include the influence of water molecules in

this process, allowing varying degrees of flexibility in the water molecules, to limit the negative impact of too many additional degrees of freedom on efficiency. It is difficult to compare the accuracies of these methods, as any differences in performance are affected by not only the treatment of water molecules, but also the search algorithm and scoring function (including associated parameters) used by the docking program.[85,117] However, it appears clear that the rigorous inclusion of water in docking is highly conducive to performance, as not including waters or even not allowing them to be displaced has been shown to have very negative effects on docking accuracy (see Table 4).

6 Free energy methods

A number of methods in this field have been developed to calculate the binding free energy of individual water molecules — a thermodynamic observable which it is not yet possible to measure directly experimentally. It is often thought that the binding free energy, related to the stability of a given water site, can be used to predict whether a site should be conserved or displaced, though this is not always the case.[118] These methods vary in terms of the approximations made in the interest of speed. Some methods employ rigorous statistical mechanics to calculate binding free energies, whereas others calculate the binding enthalpy and entropy separately, with the latter often requiring some approximation. It should be noted that a number of these methods implicitly include the ability to predict water binding locations, as well as the free energies. The methods discussed in this section are summarised in Table 5.

Widely considered to be the most rigorous method in calculating binding free energies of water molecules,[136] double decoupling is often used to provide reference results in the development of more rapid calculations. Double decoupling calculations involve the gradual, alchemical decoupling of a water molecule's interactions from both a protein and bulk solvent environment, from which the binding free energy can be extracted.[59,118–123] This typically requires the use of constraints (or restraints) which ensure that the decoupled water is not replaced by a water from bulk, and the calculated free energy must be corrected for the impact of this.[119] Grand canonical Monte Carlo (GCMC) simulations simulate a statistical ensemble which allows water molecules to be added to and removed from a binding site (according to the chemical potential of the simulation), in a theoretically rigorous manner.[124–129] These simulations can be performed at multiple chemical potential values, to calculate the binding free energy of all water molecules in the binding site, thereby capturing cooperative effects.[126,127]

As discussed above, many methods in this area seek to calculate the binding enthalpy and entropy of water sites separately, as these values determine whether displacement will stabilise or destabilise a complex. The change in enthalpy is often closely approximated as the difference in interaction energy between the protein environment and bulk solvent, whereas the entropy change is often calculated in a more approximate fashion, which varies by method. SPAM, using results from molecular dynamics (MD) simulations, evaluates the distributions of interaction energies in protein and water to evaluate a free energy and extracts the entropic contribution by subtracting the enthalpy.[133] The WATsite method is also used to process MD simulations, with the enthalpy calculated as above and the entropy approximated using the translational and rotational probability distributions of each water.[94] The SZMAP package[137] calculates the free energy, enthalpy and entropy of a given location by sampling a set of water orientations at that site and then constructing partition functions and probability distributions.[134] Grid cell theory (GCT) has been developed to process MD results to assign water binding enthalpy and entropy values to regions of a grid, with the entropy calculated using the forces and torques on each molecule.[130]

There are a large collection of methods and programs, similar to those above, which calculate the enthalpy and entropy separately, using inhomogeneous fluid solvation theory (IFST).[138,139] A full description of IFST is beyond the scope of this review, but the core principle is that translational and orientational correlation functions are used to approximate the entropy of protein-bound water, relative to that of bulk. This theory can be applied to water sites, as in WaterMap,[113–115] STOW,[140] WATCLUST[141] and SSTMap,[142] or the thermodynamics can be resolved onto a grid, referred to as GIST[111] and implemented in SSTMap.[142] It should be noted that many implementations of IFST neglect the impact of water-water correlations on the entropy — the GIST approach has been extended to account for this.[143]

It should be noted that not all approximate methods in this area follow the same approach to binding free energy calculations of water molecules. The JAWS method is a simulation technique in which waters can be partially decoupled from their environment to varying extents during a simulation, depending on what is most stable, via Monte Carlo (MC) sampling.[131] The probability of each water being fully interacting or fully decoupled can be used to approximate the binding free energy, provided that both extremes are sampled during the simulation. Setny published an approach which samples the interaction energy of a water probe at points on a lattice (similar to earlier work with Zacharias[67]), to approximate the excess chemical potential at each point, whilst the minima can be used to locate water binding sites, the chemical potentials calculated can be used to estimate binding free energies for each site.[132]

Whilst not strictly a free energy method, Zia *et al.* published an approach to estimate the stabilities of water binding sites, using MD simulation.[135] This approach involves applying a desolvating bias potential to a binding site, which is gradually increased through the simulation, with the persistence of each water molecule assumed to be proportional to their binding strength.

Whilst many of the methods discussed in previous sections of this review are generally very quick, a number of those discussed in this section are significantly slower, and would not be suitable for screening large compound libraries (unless a very large amount of computing power were available). This is often due to the need to carry out a long simulation of the structure of interest, rather than analysing snapshots of a structure. Although many of the methods in this section apply significant approximations in the free energy calculations, a number of these are able to produce results at least correlated with those from double decoupling. However, there are two core issues from which

Table 5 Brief descriptions of the free energy methods, along with their reported performances and the number of structures used in their testing (the number in parentheses indicates the estimated number of proteins covered).

Method	Description	Validation	Structures
Double Decoupling [59,118–123]	Rigorous calculation of water binding free energy by decoupling a single water molecule from both protein and solvent.	Binding free energies associated with predetermined water locations are correlated with the probability of conservation. [118]	35 (6)
GCMC [124–129]	Monte Carlo simulations in the grand canonical ensemble allow water molecules to be inserted and deleted from a region of interest.	All consensus crystallographic water sites are identified (within 2.0 Å). [126] Binding free energies of water networks match those from double decoupling. [127]	20 (5)
GCT [130]	Molecular dynamics simulation data is analysed to resolve thermodynamic properties onto a grid, using grid cell theory.	Water binding free energies show good agreement with those from double decoupling.	3 (3)
GIST [111]	Inhomogeneous fluid solvation theory is used to analyse a molecular dynamics simulation, with the results resolved onto a grid.	Modelling of congeneric ligands with water binding free energies produced similar results to those previously reported.	28 (1)
JAWS [131]	Water molecules are assigned a parameter which allows them to switch partially on/off during a Monte Carlo simulation.	Water locations qualitatively agree well with the crystallographic sites. Water binding free energies qualitatively match those from double decoupling.	6 (5)
Setny [132]	A semi-explicit water model is used to sample a grid around the binding site, then an iterative process places waters at the most favourable locations.	29/33 water sites with negative binding free energies are correctly predicted (within 2.0 Å). Free energy values are correlated with double decoupling results.	8 (8)
SPAM [133]	The distribution of interaction energies between a water site and its environment are compared with those in bulk solvent to calculate thermodynamic properties.	Results obtained are consistent with inhomogeneous fluid solvation theory, but not double decoupling.	14 (2)
SZMAP [134]	A semi-explicit water model is used to calculate interaction energies for a number of water orientations at a given site, from which thermodynamic properties are calculated.	Water binding free energies calculated are correlated with those from double decoupling.	35 (6)
WaterMap [113–115]	Water molecules from a molecular dynamics simulation are clustered and analysed with inhomogeneous fluid solvation theory.	Water locations are qualitatively consistent with prior knowledge. Water binding free energies were used to model differences between congeneric ligands for a single target.	33 (5)
WATsite [94]	Water molecules from a molecular dynamics simulation are analysed using their interaction energies and rotational/translational distributions.	Water positions qualitatively matched known sites. Water thermodynamics were self-consistent and agreed with known behaviour.	5 (3)
Zia *et al.* [135]	A desolvating bias potential is added to the system in order to determine the resistance of water sites to the bias applied.	The persistence of active site water molecules in the presence of the bias was consistent with prior knowledge of these sites.	4 (1)

a significant number of these methods suffer; the first is the neglect of the influence of water molecules on each other,[41] in either using a single water probe (with or without a continuum bulk solvent) or by analysing the sites independently. The second issue is the common use of very short molecular dynamics simulations (sometimes only several nanoseconds) to generate configurations of the hydration structure. MD simulations can take tens of nanoseconds (or longer) to show converged water sites,[144,145] and buried water sites can exchange on a timescale of nanoseconds to milliseconds, depending on the kinetics of the system.[146] Results from short simulations are therefore susceptible to bias by the starting water configuration. This issue could be alleviated somewhat by using much longer simulations, but the slow exchange of waters may require an inaccessibly long simulation to show convergence. A better solution to this issue might involve the use of Monte Carlo techniques to accelerate the exchange of buried waters with bulk, such as GCMC [124–129] or the method proposed by Ben-Shalom *et al.* [58]

It should also be noted that the majority of these methods employ water models which have been parameterised to reproduce bulk water behaviour,[147] and these models may not be appropriately polarised for water molecules at the protein interface, given the dependence of the water dipole on its environment [13]. This is also true for other methods which model water-protein interactions, such as many interaction-based site prediction and ligand docking methods. Additionally, as mentioned in section 2, the results from methods which resolve water structural and thermodynamic information onto grids (GCT and GIST) could be more difficult to interpret, but do offer some practical advantages — both AutoDock and DOCK have incorporated GIST results to improve docking performance.[92,96,97]

7 Challenges and limitations

7.1 Definition of success

With a variety of testing and validation procedures for water predictions, it is difficult for a potential user to know which method is most suitable for their requirements. In part, this issue arises from the different ways in which authors measure the success of

their approach, as can be seen from the respective method sections and also Tables 1 - 5. This is exacerbated by the fact that no large and challenging dataset for testing interfacial water predictions has been widely adopted by the community. As previously mentioned, it is difficult to validate the characterisation of water molecules, so this section focuses on the validation of location predictions.

Fig. 3 shows a plot of the accuracy of water binding site predictions against the distance threshold used to define a correct prediction, for the methods that clearly quote values for these parameters in their publications (where specifics are not given, the results are not included). It can be seen that whilst there is some variation between the methods, the accuracy is in part linked to how strictly an accurate prediction is defined. If a larger distance cut-off is used, then there is greater leniency in the classification of a water molecule as being predicted correctly. Additionally, some methods use different criteria to determine which crystallographic sites are considered in the accuracy test, such as those which exhibit a particular number or type of interactions, for example. These factors, combined with the fact that most methods are tested on very different sets of structures, makes a direct comparison problematic.

Furthermore, the rigid receptor effect, which is well recognised in the field of docking,[85] has a significant effect on the accuracy of water binding site predictions. Methods which do not allow flexibility in the environment when placing a ligand (or water molecules) tend to achieve higher success rates, as the environment is often pre-organised for a successful prediction. In those methods which allow full motion of the protein, the structure can diverge from the crystal structure, thereby apparently reducing the accuracy by making the comparison difficult.[148] Of the methods discussed in this review, the treatment of the flexibility of the ligand-protein environment varies significantly, and likely has a large impact on the reported performances.

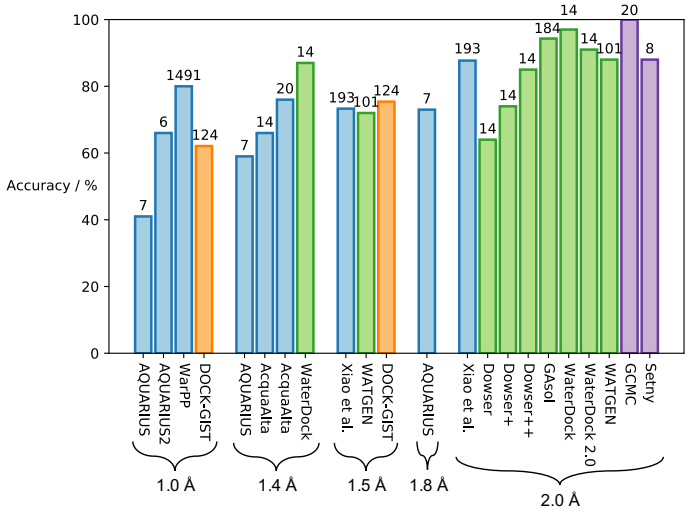

Fig. 3 Bar chart showing the percentage of crystallographic water sites correctly predicted by different methods, where this has been reported. The bars are grouped by the distance threshold which was used to define a successful prediction. The number above each bar indicates the number of crystal structures used in testing. It should also be noted that many of these methods filter the water sites which are considered, in alphabetical order, these are briefly summarised as follows; AcquaAlta: binding site waters; AQUARIUS: no apparent filtering; DOCK-GIST: binding site waters; Dowser: binding site waters; GAsol: binding site waters; GCMC: consensus waters in the binding site; Setny: buried water sites with negative binding free energies; WarPP: binding site waters with at least two hydrogen bonds to protein or ligand; WaterDock: binding site waters; WATGEN: bridging waters; Xiao *et al.*: binding site waters.

It should be noted that the validation of methods that seek to predict the effects of water on ligand binding affinities requires high quality binding affinity data alongside the crystallographic data, to first ensure that there is a difference in water binding and then to quantify this effect. Additionally, many other differences between two ligands will be subsumed into the binding free energy, so the contribution of the water conservation/displacement cannot be fully isolated. The quality of binding affinity data is difficult to ascertain and quantitatively assess on a large scale, and as such, this is not discussed in this section.

7.2 Data used in testing

Related to the theme of method testing is the quality of the structural data used in both the parameterisation of a method and the validation of the results obtained. We have assembled a list of all of the Protein Data Bank (PDB)[28,29] codes reported in test sets for the methods discussed in this review (where they are available), to assess the quality of the crystallographic data typically used. Of the 6650 PDB codes obtained, 6622 of these were returned from the PDB (the others have since been removed from the PDB), for which both the resolutions and release years were extracted. Looking at the extent to which structures are recycled reveals that the vast majority of the structures considered (5167) have been used by only a single method, however there are also a significant number (496), which have been used with 3 or methods, and 15 structures which have been used in 7 or more different validations. These data, along with the distributions of the resolutions and release years of the structures are plotted in Fig. 4. The resolutions are mostly distributed between 0.5 and 3.5 Å, with 95.6 % of the values lying between 1.0 and 2.5 Å. This would indicate that many good quality structures have been used in testing, although there are some cases of particularly high or low quality structures being used. The distribution of the PDB release years for the structures used is rather broad, and includes a significant number of older structures. This observation is in part because there will be a lag between the release of a structure and its use, and also there is a tendency to reuse structures tested by others, resulting in some structures persisting for a number of years.

Fig. 4 shows that many of the structures which tend to be reused (the darker points) have resolutions better than around 2.5 Å — it is likely that they were selected for testing based on this. However, it is of some concern that a significant portion of these structures were released prior to the widespread used of cryogenic crystallography (in the mid- to late 1990s),[149] as this has been found to have a significant impact on the identification of water molecules.[150,151] This is in part due to the fact that

these structures have been used by older methods, when these older structures would have been new, but some of these recycled structures (such as 1F0R and 1HPX) have been used to assess water placement in very recent years.

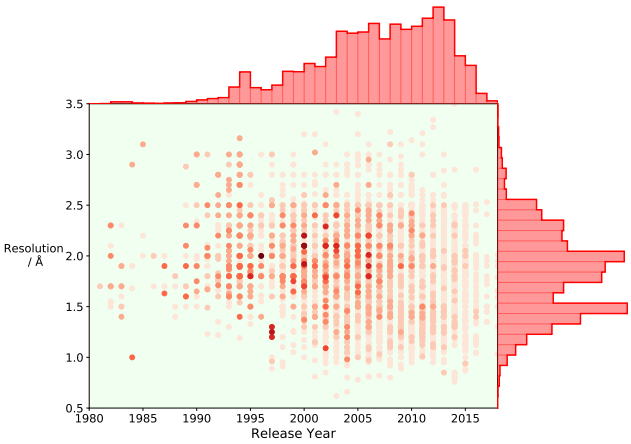

Fig. 4 Scatter plot showing the release year and resolution of the 6622 PDB entries considered. The darkness of the points corresponds to the number of methods which have used a given PDB code, with the darkest points corresponding to 10 methods, and the lightest to 1. Histograms of the resolutions and release years are included alongside the axes.

To estimate the number of proteins covered by these crystal structures, FASTA sequences were extracted for each of the chains presented in each of these structures, and a similarity matrix based on sequence identity was calculated using the BLAST program.[152,153] This similarity matrix was used to cluster these chains such that no two protein chains in the same cluster had a similarity of less than 75 %. From this, any chain clusters which are only ever present in a protein structure together were assumed to be part of a heteromultimeric protein and their clusters merged (additionally, four other clusters of protein chains were manually identified as all being thrombin, and two as HIV protease, given their high populations). This analysis led to the estimation that the 6622 PDB codes contain 1852 distinct proteins (this is likely an overestimate, given the strict threshold used for clustering). Fig. 5 shows ten of the proteins used by the largest number of methods, with HIV protease having been used the most, likely due to it being well known as a case where water displacement is associated with a large increase in affinity.[14] It was also considered how the structures are distributed across the proteins represented, with Fig. 6 showing the numbers of structures for the ten most represented proteins. Two points are of interest here: first, modern drug targets do not appear to be well represented, and second, these ten proteins cover 23.5 % of the 6622 structures considered (the 50 most populated cover 44.1 % of the structures). This observation indicates that some of the methods may suffer from undetected biases, related to the apparent biases in the validation data.

The data presented in this section highlight the need for a consistent, diverse, well-curated dataset of high-quality and relevant

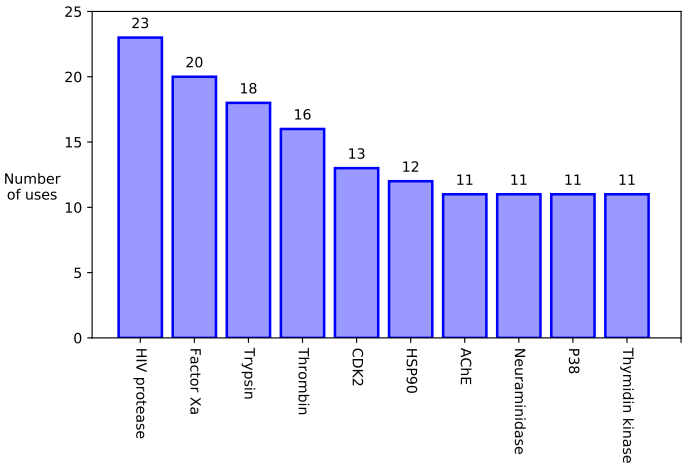

Fig. 5 Bar chart showing the number of times that the ten most frequently used proteins have been used by different methods. CDK2: cyclin-dependent kinase 2, HSP90: heat shock protein 90, AChE: acetylcholinesterase.

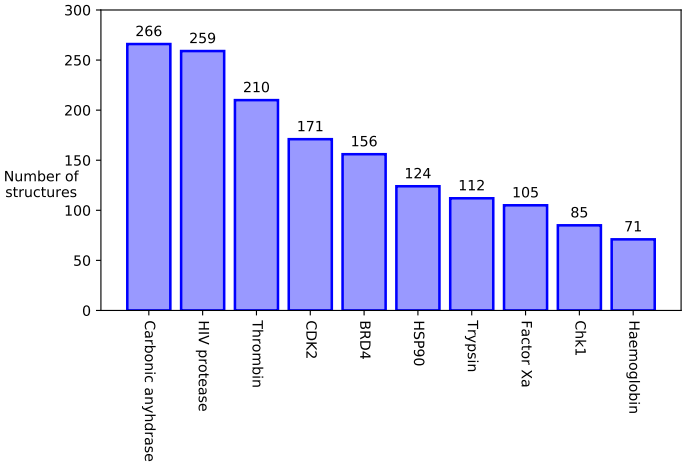

Fig. 6 Bar chart showing the estimated number of structures for the ten most common proteins contained in the 6622 PDB codes considered. CDK2: cyclin-dependent kinase 2, BRD4: bromodomain-containing protein 4, HSP90: heat shock protein 90.

structures, to provide a common reference for comparison of different methods. It is clear that no single dataset has been widely adopted, and many of those which have do not necessarily represent drug targets. Additionally, the size of a dataset is very important, as different classes of methods are often tested on datasets of different sizes, owing to the speed differences between them. Therefore, a common set should be large enough to be statistically meaningful, but not practically infeasible for the testing of the more time-consuming methods, which typically employ simulation-based techniques.

7.3 Application to compound design

As discussed in depth in this review, many different approaches have been developed for the prediction of protein-bound water molecules with respect to both their location and properties.

Whilst the relative successes have been assessed, both quantitatively and qualitatively, in very different ways, it is clear that these aspects can be predicted rather well at varying degrees of computational cost, and it does appear that the field is approaching maturity in this regard.[154] However, the challenge remains largely unsolved as to how computational modelling can directly guide the displacement and/or conservation of binding site water molecules.

Some free energy methods have attempted to predict changes in ligand affinity based on water binding free energies,[115,155] showing good correlation across a range of results for congeneric ligands. This area has been heavily investigated with ligand docking methods, with varying degrees of success, and does not appear to have been attempted using knowledge-based modelling. It has been observed that relative ligand binding free energy calculations can be very negatively impacted if a displaced water molecule is not eliminated to bulk within the timescale of the simulation.[156] This problem can be solved using either GCMC[125,129,157] or the method proposed by Ben-Shalom *et al.*[58,158] alongside the free energy perturbation, allowing much more rapid binding and unbinding of water molecules from the binding site. These advances have so far been found to improve the accuracy of free energy calculations and reduce their sensitivity to the initial water configuration of the binding site.[129,157,158]

8 Conclusions

Here we have provided a comprehensive review and description of many of the methods available for the computational study of water molecules in computer-aided drug design. Given that the broad range of computational methods available are able to capture the dynamics of molecular systems, probe systems at the atomic level and readily assess compounds which may be synthetically challenging, along with the well recognised role that water can play in drug binding, there is the potential for these approaches to have a significant impact on drug design and development.

The introductory section of this review outlined three questions which medicinal chemists will typically ask themselves with respect to water molecules within a protein structure. The first of these, pertaining to the locations of water molecules, can now be answered by a number of methods, offering differing balances of the trade-off between time required and rigour of the results obtained. However, a thorough and fair comparison of the accuracies of different water prediction methods would require a reliable, high-quality, curated dataset of relevant complexes containing water-mediated interactions, with a single measure of accuracy, to provide a consistent frame of reference for testing, in order to directly compare the many approaches available. Whilst a number of popular datasets already exist,[43,114,118,159–164] this will likely require a new set, including more recent experimental data. However, it is also important that a common frame of reference contains diverse structures, and that these are pharmaceutically relevant.

The second of these questions, relating to the relative importance of water molecules in ligand optimation, is much more difficult to answer. Many methods have been developed to assess water sites found within a protein structure, from the knowledge-based classification methods, to the more rigorous (and computationally expensive) free energy methods. The former have primarily been developed to predict the likelihood of a water being conserved or displaced from a structure upon ligand binding, based on a knowledge base of crystal structures. A number of these methods have demonstrated good performance when applied to pairs of *apo/holo*-structures; however, it is important to note that the training and testing of such methods can be biased by the data used, and may not be applicable to classes of ligands different from those used in training the models. The free energy methods are typically assessed in terms of their ability to reproduce results from double decoupling calculations, given that these are widely accepted to be a gold standard[136], as water binding free energies cannot be directly measured experimentally. Although a quantitative assessment of the performance of these methods is not trivial, it appears that they are indeed able to offer insight which is of use in drug design, particularly the methods which are able to predict the binding entropy of water sites, which offer an indication of the free energy which might be released when the water site is displaced.[18]

However, the third of these questions, and the most important, given the ultimate aim of these approaches — regarding how compound design should be guided based on information relating to water molecules — cannot yet be unambiguously answered. Docking protocols have been developed to account for water molecules during ligand pose and affinity prediction, with mixed degrees of success, and these methods also inherit the difficulties and limitations typically associated with docking.[85] Many other methods have also been applied prospectively in live drug design,[20,41,165,166] though it is difficult to determine the contribution of the water analysis method to the design strategy, and it has been suggested that many compound modifications suggested would likely have been proposed by medicinal chemists using other structure-based design techniques.[41] It is important that more diverse methods are developed with the aim of forming more direct links between water molecules and ligand structure-activity relationships, to maximise utility in drug discovery and compound development. However, an increase in the amount of effort expended on predicting water-related changes in ligand binding affinity would require a very large amount of high quality experimental data, of both ligand affinities and corresponding protein-ligand crystal structures. This would be necessary to both maximise the confidence in the differences in hydration and binding affinity, to provide a consistent and reliable dataset for the testing of various methods and also for the training of knowledge-based models.

Finally, we must reiterate the importance of reliable and consistent datasets for the purpose of quantifying and comparing the accuracy of these methods. Whilst prospective testing on new therapeutic compounds is often proposed as a superior alternative to retrospective validation, in this context, the performance of different water prediction/analysis methods cannot be inferred from the number of pharmaceutical insights already offered, as it is difficult to isolate the contribution of the water analysis method. First, some methods are more widely used in the pharmaceutical

industry, often owing to their being well established over time. Additionally, prospective validations typically focus on positive results, as negative results can be difficult to confirm. Therefore, for the purpose of comparison, retrospective validation would be a more consistent and fair test. The field would certainly benefit from a blind, community-wide challenge involving the prediction of water binding locations in unpublished crystal structures.

In summary, the broad field of applying computational methods and programs to the study of water molecules in macromolecular biological structures is well developed. Water molecule locations can be predicted and characterised with good accuracy, via a number of different methods. As discussed in this review, future work should focus on the development and use of reliable and consistent test sets, and the development of methods to form rigorous links between binding site water molecules and drug affinity. These developments will be of great use to the pharmaceutical industry in expediting computer-aided drug design.

Conflicts of interest

RDT is an employee of UCB, and research conducted by MLS, HEBM and JWE has been partly funded by UCB.

Acknowledgements

The authors thank the EPSRC for funding. MLS is supported by the EPSRC-funded CDT in Next Generation Computational Modelling, under grant EP/L015382/1. HEBM was supported by the EPSRC-funded CDT in Theory and Modelling in Chemical Sciences, under grant EP/L015722/1 during this work. The authors thank Dr Philip Williamson (Biological Sciences, University of Southampton) for helpful discussions regarding the use of NMR in this field.

Notes and references

1 M. Hendlich, A. Bergner, J. Günther and G. Klebe, *J. Mol. Biol.*, 2003, **326**, 607–620.

2 J. Günther, A. Bergner, M. Hendlich and G. Klebe, *J. Mol. Biol.*, 2003, **326**, 621–636.

3 Y. Lu, R. Wang, C.-Y. Yang and S. Wang, *J. Chem. Inf. Model.*, 2007, **47**, 668–675.

4 O. Carugo, *Amino Acids*, 2016, **48**, 193–202.

5 J. M. Chen, S. L. Xu, Z. Wawrzak, G. S. Basarab and D. B. Jordan, *Biochemistry*, 1998, **37**, 17735–17744.

6 C. Liu, S. T. Wrobleski, J. Lin, G. Ahmed, A. Metzger, J. Wityak, K. M. Gillooly, D. J. Shuster, K. W. McIntyre, S. Pitt, D. R. Shen, R. F. Zhang, H. Zhang, A. M. Doweyko, D. Diller, I. Henderson, J. C. Barrish, J. H. Dodd, G. L. Schieven and K. Leftheris, *J. Med. Chem.*, 2005, **48**, 6261–6270.

7 A. Wissner, D. M. Berger, D. H. Boschelli, M. B. Floyd, L. M. Greenberger, B. C. Gruber, B. D. Johnson, N. Mamuya, R. Nilakantan, M. F. Reich, R. Shen, H. R. Tsou, E. Upeslacis, Y. F. Wang, B. Wu, F. Ye and N. Zhang, *J. Med. Chem.*, 2000, **43**, 3244–3256.

8 N. N. Nasief, H. Tan, J. Kong and D. Hangauer, *J. Med. Chem.*, 2012, **55**, 8283–8302.

9 M. Kim and A. E. Cho, *Sci. Rep.*, 2016, **6**, 36807.

10 M. Aldeghi, G. A. Ross, M. J. Bodkin, J. W. Essex, S. Knapp and P. C. Biggin, *Commun. Chem.*, 2018, **1**, 19.

11 J. R. H. Tame, S. H. Sleigh, A. J. Wilkinson and J. E. Ladbury, *Nat. Struct. Mol. Biol.*, 1996, **3**, 998.

12 J. D. Dunitz, *Science*, 1994, **264**, 670–670.

13 M. Maurer and C. Oostenbrink, *J. Mol. Recognit.*, 2019, **32**, 32.

14 P. Y. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. Rayner and Y. N. Wong, *Science*, 1994, **263**, 380–384.

15 J. E. Chrencik, A. Patny, I. K. Leung, B. Korniski, T. L. Emmons, T. Hall, R. A. Weinberg, J. A. Gormley, J. M. Williams, J. E. Day, J. L. Hirsch, J. R. Kiefer, J. W. Leone, H. D. Fischer, C. D. Sommers, H.-C. Huang, E. J. Jacobsen, R. E. Tenbrink, A. G. Tomasselli and T. E. Benson, *J. Mol. Biol.*, 2010, **400**, 413–433.

16 J. Le Roux, C. Leriche, P. Chamiot-Clerc, J. Feutrill, F. Halley, D. Papin, N. Derimay, C. Mugler, C. Grépin and L. Schio, *Bioorg. Med. Chem. Lett.*, 2016, **26**, 454–459.

17 A. Bortolato, B. G. Tehan, M. S. Bodnarchuk, J. W. Essex and J. S. Mason, *J. Chem. Inf. Model.*, 2013, **53**, 1700–1713.

18 A. Joncour, N. Desroy, C. Housseman, X. Bock, N. Bienvenu, L. Cherel, V. Labeguere, C. Peixoto, D. Annoot, L. Lepissier, J. Heiermann, W. J. Hengeveld, G. Pilzak, A. Monjardet, E. Wakselman, V. Roncoroni, S. Le Tallec, R. Galien, C. David, N. Vandervoort, T. Christophe, K. Conrath, M. Jans, A. Wohlkonig, S. Soror, J. Steyaert, R. Touitou, D. Fleury, L. Vercheval, P. Mollat, N. Triballeau, E. van der Aar, R. Brys and B. Heckmann, *J. Med. Chem.*, 2017, **60**, 7371–7392.

19 J. G. Kettle, R. Anjum, E. Barry, D. Bhavsar, C. Brown, S. Boyd, A. Campbell, K. Goldberg, M. Grondine, S. Guichard, C. J. Hardy, T. Hunt, R. D. O. Jones, X. Li, O. Moleva, D. Ogg, R. C. Overman, M. J. Packer, S. Pearson, M. Schimpl, W. Shao, A. Smith, J. M. Smith, D. Stead, S. Stokes, M. Tucker and Y. Ye, *J. Med. Chem.*, 2018, **61**, 8797–8810.

20 M. S. Bodnarchuk, M. J. Packer and A. Haywood, *ACS Med. Chem. Lett.*, 2020, **11**, 77–82.

21 A. M. Davis, S. J. Teague and G. J. Kleywegt, *Angew. Chem. Int. Ed.*, 2003, **42**, 2718–2736.

22 A. McPherson, *Methods*, 2004, **34**, 254–265.

23 G. J. Kleywegt, *Acta Crystallogr. D*, 2000, **56**, 249–265.

24 D. H. Ohlendorf, *Acta Crystallogr. D*, 1994, **50**, 808–812.

25 B. A. Fields, H. H. Bartsch, H. D. Bartunik, F. Cordes, J. M. Guss and H. C. Freeman, *Acta Crystallogr. D*, 1994, **50**, 709–730.

26 D. A. A. Myles, *Curr. Opin. Struct. Biol.*, 2006, **16**, 630–637.

27 W. B. O'Dell, A. M. Bodenheimer and F. Meilleur, *Arch. Biochem. Biophys.*, 2016, **602**, 48–60.

28 *Protein Data Bank*, `www.rcsb.org`.

29 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat,

H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
30 G. Otting, E. Liepinsh and K. Wüthrich, *Science*, 1991, **254**, 974–980.
31 G. Otting, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1997, **31**, 259–285.
32 B. Halle, *Philos. Trans. R. Soc. B*, 2004, **359**, 1207–1223; discussion 1223–1224, 1323–1328.
33 K. Modig, E. Liepinsh, G. Otting and B. Halle, *J. Am. Chem. Soc.*, 2004, **126**, 102–114.
34 N. V. Nucci, M. S. Pometun and A. J. Wand, *Nat. Struct. Mol. Biol.*, 2011, **18**, 245–249.
35 J. M. Gruschus and J. A. Ferretti, *J. Biomol. NMR*, 2001, **20**, 111–126.
36 B. Halle, *J. Chem. Phys.*, 2003, **119**, 12372–12385.
37 B. D. Armstrong and S. Han, *J. Am. Chem. Soc.*, 2009, **131**, 4641–4647.
38 I. Kaminker, R. Barnes and S. Han, *Methods in Enzymology*, Elsevier, 2015, vol. 564, pp. 457–483.
39 L. Biedermannová and B. Schneider, *Biochim. Biophys. Acta*, 2016, **1860**, 1821–1835.
40 T. S. G. Olsson, M. A. Williams, W. R. Pitt and J. E. Ladbury, *J. Mol. Biol.*, 2008, **384**, 1002–1017.
41 A. P. Graves, I. D. Wall, C. M. Edge, J. M. Woolven, G. Cui, A. Le Gall, X. Hong, K. Raha and E. S. Manas, *Curr. Top. Med. Chem.*, 2017, **17**, 2599–2616.
42 D. Bucher, P. Stouten and N. Triballeau, *J. Chem. Inf. Model.*, 2018, **58**, 692–699.
43 G. Rossato, B. Ernst, A. Vedani and M. Smieško, *J. Chem. Inf. Model.*, 2011, **51**, 1867–1881.
44 W. R. Pitt and J. M. Goodfellow, *Protein Eng. Des. Sel.*, 1991, **4**, 531–537.
45 W. R. Pitt, J. Murray-Rust and J. M. Goodfellow, *J. Comput. Chem.*, 1993, **14**, 1007–1018.
46 S. V. Rakhmanov and V. J. Makeev, *BMC Struct. Biol.*, 2007, **7**, 19.
47 M. L. Verdonk, J. C. Cole and R. Taylor, *J. Mol. Biol.*, 1999, **289**, 1093–1108.
48 M. L. Verdonk, J. C. Cole, P. Watson, V. Gillet and P. Willett, *J. Mol. Biol.*, 2001, **307**, 841–859.
49 E. Nittinger, F. Flachsenberg, S. Bietz, G. Lange, R. Klein and M. Rarey, *J. Chem. Inf. Model.*, 2018, 1625–1637.
50 M. Zheng, Y. Li, B. Xiong, H. Jiang and J. Shen, *J. Comput. Chem.*, 2013, **34**, 583–592.
51 W. Xiao, Z. He, M. Sun, S. Li and H. Li, *J. Chem. Inf. Model.*, 2017, **57**, 1517–1528.
52 S. Miyazawa and R. L. Jernigan, *Macromolecules*, 1985, **18**, 534–552.
53 I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor and M. L. Verdonk, *J. Comput. Aided Mol. Des.*, 1997, **11**, 525–537.
54 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. B*, 2016, **72**, 171–179.
55 D. Beglov and B. Roux, *J. Phys. Chem. B*, 1997, **101**, 7821–7826.
56 T. Imai, R. Hiraoka, A. Kovalenko and F. Hirata, *J. Am. Chem. Soc.*, 2005, **127**, 15334–15335.
57 T. Imai, R. Hiraoka, A. Kovalenko and F. Hirata, *Proteins*, 2007, **66**, 804–813.
58 I. Y. Ben-Shalom, C. Lin, T. Kurtzman, R. C. Walker and M. K. Gilson, *J. Chem. Theory Comput.*, 2019, **15**, 2684–2691.
59 L. Zhang and J. Hermans, *Proteins*, 1996, **24**, 433–438.
60 A. Morozenko, I. V. Leontyev and A. A. Stuchebrukhov, *J. Chem. Theory Comput.*, 2014, **10**, 4618–4623.
61 A. Morozenko and A. A. Stuchebrukhov, *Proteins*, 2016, **84**, 1347–1357.
62 J. W. H. Schymkowitz, F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher and L. Serrano, *Proc. Natl. Acad. Sci.*, 2005, **102**, 10147–10152.
63 L. Fusani, I. Wall, D. Palmer and A. Cortes, *Bioinformatics*, 2018, **34**, 1947–1948.
64 P. J. Goodford, *J. Med. Chem.*, 1985, **28**, 849–857.
65 A. Miranker and M. Karplus, *Proteins*, 1991, **11**, 29–34.
66 D. J. Sindhikara, N. Yoshida and F. Hirata, *J. Comput. Chem.*, 2012, **33**, 1536–1543.
67 P. Setny and M. Zacharias, *J. Phys. Chem. B*, 2010, **114**, 8667–8675.
68 G. A. Ross, G. M. Morris and P. C. Biggin, *PLoS ONE*, 2012, **7**, e32036.
69 A. Sridhar, G. A. Ross and P. C. Biggin, *PLOS ONE*, 2017, **12**, e0172743.
70 S. Chakrabarty and A. Warshel, *Proteins: Struct., Funct., Bioinf.*, 2013, **81**, 93–106.
71 H. Yoon, V. Kolev and A. Warshel, *J. Phys. Chem. B*, 2017, **121**, 9358–9365.
72 H.-H. Bui, A. J. Schiewe and I. S. Haworth, *J. Comput. Chem.*, 2007, **28**, 2241–2251.
73 P. S. Nerenberg and T. Head-Gordon, *Curr. Opin. Struct. Biol.*, 2018, **49**, 129–138.
74 T. D. Bergazin, I. Y. Ben-Shalom, N. M. Lim, S. C. Gill, M. K. Gilson and D. L. Mobley, *J. Comput. Aided Mol. Des.*, 2021, **35**, 167–177.
75 A. Amadasi, F. Spyrakis, P. Cozzini, D. J. Abraham, G. E. Kellogg and A. Mozzarelli, *J. Mol. Biol.*, 2006, **358**, 289–309.
76 A. Amadasi, J. A. Surface, F. Spyrakis, P. Cozzini, A. Mozzarelli and G. E. Kellogg, *J. Med. Chem.*, 2008, **51**, 1063–1067.
77 C. A. Bottoms, T. A. White and J. J. Tanner, *Proteins: Struct., Funct., Bioinf.*, 2006, **64**, 404–421.
78 M. L. Raymer, P. C. Sanschagrin, W. F. Punch, S. Venkataraman, E. D. Goodman and L. A. Kuhn, *J. Mol. Biol.*, 1997, **265**, 445–464.
79 M. Jukič, J. Konc, S. Gobec and D. Janežič, *J. Chem. Inf. Model.*, 2017, **57**, 3094–3103.
80 H. Patel, B. A. Gruning, S. Gunther and I. Merfort, *Bioinformatics*, 2014, **30**, 2978–2980.

81 P. C. Sanschagrin and L. A. Kuhn, *Protein Sci.*, 1998, **7**, 2054–2064.

82 A. T. García-Sosa, R. L. Mancera and P. M. Dean, *J. Mol. Model.*, 2003, **9**, 172–182.

83 J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano, *Nucleic Acids Res.*, 2005, **33**, W382–388.

84 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.

85 R. D. Taylor, P. J. Jewsbury and J. W. Essex, *J. Comput. Aided Mol. Des.*, 2002, **16**, 151–166.

86 G. E. Kellogg, S. F. Semus and D. J. Abraham, *J. Comput. Aided Mol. Des.*, 1991, **5**, 545–552.

87 D. J. Abraham and G. E. Kellogg, *J. Comput. Aided Mol. Des.*, 1994, **8**, 41–49.

88 G. E. Kellogg and D. J. Abraham, *Eur. J. Med. Chem.*, 2000, **35**, 651–661.

89 D. L. Chen and G. E. Kellogg, *J. Comput. Aided Mol. Des.*, 2005, **19**, 69–82.

90 L. A. Kuhn, C. A. Swanson, M. E. Pique, J. A. Tainer and E. D. Getzoff, *Proteins*, 1995, **23**, 536–547.

91 F. Österberg, G. M. Morris, M. F. Sanner, A. J. Olson and D. S. Goodsell, *Proteins: Struct., Funct., Bioinf.*, 2002, **46**, 34–40.

92 S. Uehara and S. Tanaka, *Molecules*, 2016, **21**, 1604.

93 A. H. Mahmoud, M. R. Masters, Y. Yang and M. A. Lill, *Commun. Chem.*, 2020, **3**, 1–13.

94 B. Hu and M. A. Lill, *J. Comput. Chem.*, 2014, **35**, 1255–1260.

95 N. Huang and B. K. Shoichet, *J. Med. Chem.*, 2008, **51**, 4862–4865.

96 H. Sun, L. Zhao, S. Peng and N. Huang, *Proteins: Struct., Funct., Bioinf.*, 2014, **82**, 1765–1776.

97 T. E. Balius, M. Fischer, R. M. Stein, T. B. Adler, C. N. Nguyen, A. Cruz, M. K. Gilson, T. Kurtzman and B. K. Shoichet, *Proc. Natl. Acad. Sci.*, 2017, **114**, E6839–E6846.

98 C. R. Corbeil, P. Englebienne and N. Moitessier, *J. Chem. Inf. Model.*, 2007, **47**, 435–449.

99 M. Rarey, B. Kramer and T. Lengauer, *Proteins: Struct., Funct., Bioinf.*, 1999, **34**, 17–28.

100 *Glide*, 2017.

101 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.

102 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2006, **49**, 6177–6196.

103 M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor and R. Taylor, *J. Med. Chem.*, 2005, **48**, 6504–6515.

104 M. A. Lie, R. Thomsen, C. N. S. Pedersen, B. Schiøtt and M. H. Christensen, *J. Chem. Inf. Model.*, 2011, **51**, 909–917.

105 G. Lemmon and J. Meiler, *PLoS ONE*, 2013, **8**, e67536.

106 V. Schnecke and L. A. Kuhn, *Perspect. Drug Discov.*, 2000, **20**, 171–190.

107 R. B. Murphy, M. P. Repasky, J. R. Greenwood, I. Tubert-Brohman, S. Jerome, R. Annabhimoju, N. A. Boyles, C. D. Schmitz, R. Abel, R. Farid and R. A. Friesner, *J. Med. Chem.*, 2016, **59**, 4364–4384.

108 M. Rarey, B. Kramer, T. Lengauer and G. Klebe, *J. Mol. Biol.*, 1996, **261**, 470–489.

109 M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins*, 2003, **52**, 609–623.

110 D. M. Lorber and B. K. Shoichet, *Protein Sci.*, 1998, **7**, 938–950.

111 C. N. Nguyen, T. Kurtzman Young and M. K. Gilson, *J. Chem. Phys.*, 2012, **137**, 044101.

112 D. S. Goodsell and A. J. Olson, *Proteins*, 1990, **8**, 195–202.

113 *WaterMap*, 2017.

114 T. Young, R. Abel, B. Kim, B. J. Berne and R. A. Friesner, *Proc. Natl. Acad. Sci.*, 2007, **104**, 808–813.

115 R. Abel, T. Young, R. Farid, B. J. Berne and R. A. Friesner, *J. Am. Chem. Soc.*, 2008, **130**, 2817–2831.

116 J. Meiler and D. Baker, *Proteins*, 2006, **65**, 538–548.

117 J. C. Cole, C. W. Murray, J. W. M. Nissink, R. D. Taylor and R. Taylor, *Proteins: Struct., Funct., Bioinf.*, 2005, **60**, 325–332.

118 C. Barillari, J. Taylor, R. Viner and J. W. Essex, *J. Am. Chem. Soc.*, 2007, **129**, 2577–2587.

119 M. K. Gilson, J. A. Given, B. L. Bush and J. A. McCammon, *Biophys. J.*, 1997, **72**, 1047–1069.

120 W. L. Jorgensen, J. K. Buckner, S. Boudon and J. Tirado-Rives, *J. Chem. Phys.*, 1988, **89**, 3742–3746.

121 B. Roux, M. Nina, R. Pomès and J. C. Smith, *Biophys. J.*, 1996, **71**, 670–681.

122 D. Hamelberg and J. A. McCammon, *J. Am. Chem. Soc.*, 2004, **126**, 7683–7689.

123 L. R. Olano and S. W. Rick, *J. Am. Chem. Soc.*, 2004, **126**, 7991–8000.

124 H.-J. Woo, A. R. Dinner and B. Roux, *J. Chem. Phys.*, 2004, **121**, 6392–6400.

125 Y. Deng and B. Roux, *J. Chem. Phys.*, 2008, **128**, 115103.

126 G. A. Ross, M. S. Bodnarchuk and J. W. Essex, *J. Am. Chem. Soc.*, 2015, **137**, 14930–14943.

127 G. A. Ross, H. E. Bruce Macdonald, C. Cave-Ayland, A. I. Cabedo Martinez and J. W. Essex, *J. Chem. Theory Comput.*, 2017, **13**, 6373–6381.

128 J. Wahl and M. Smieško, *J. Chem. Inf. Model.*, 2019, **59**, 754–765.

129 H. E. Bruce Macdonald, C. Cave-Ayland, G. A. Ross and J. W. Essex, *J. Chem. Theory Comput.*, 2018, **14**, 6586–6597.

130 G. Gerogiokas, M. W. Y. Southey, M. P. Mazanetz, A. Hefeitz, M. Bodkin, R. J. Law and J. Michel, *Phys. Chem. Chem. Phys.*, 2015, **17**, 8416–8426.

131 J. Michel, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2009, **113**, 13337–13346.

132 P. Setny, *J. Chem. Theory Comput.*, 2015, **11**, 5961–5972.

133 G. Cui, J. M. Swails and E. S. Manas, *J. Chem. Theory Com-*

put., 2013, **9**, 5539–5549.

134 A. S. Bayden, D. T. Moustakas, D. Joseph-McCarthy and M. L. Lamb, *J. Chem. Inf. Model.*, 2015, **55**, 1552–1565.

135 S. R. Zia, R. Gaspari, S. Decherchi and W. Rocchia, *J. Chem. Theory Comput.*, 2016, **12**, 6049–6061.

136 M. S. Bodnarchuk, R. Viner, J. Michel and J. W. Essex, *J. Chem. Inf. Model.*, 2014, **54**, 1623–1633.

137 *SZMAP*, 2017, `http://www.eyesopen.com`.

138 T. Lazaridis, *J. Phys. Chem. B*, 1998, **102**, 3531–3541.

139 T. Lazaridis, *J. Phys. Chem. B*, 1998, **102**, 3542–3550.

140 Z. Li and T. Lazaridis, *Computational Drug Discovery and Design*, Springer, New York, NY, 2012, pp. 393–404.

141 E. D. López, J. P. Arcon, D. F. Gauto, A. A. Petruk, C. P. Modenutti, V. G. Dumas, M. A. Marti and A. G. Turjanski, *Bioinformatics*, 2015, **31**, 3697–3699.

142 K. Haider, A. Cruz, S. Ramsey, M. K. Gilson and T. Kurtzman, *J. Chem. Theory Comput.*, 2018, **14**, 418–425.

143 C. N. Nguyen, T. Kurtzman and M. K. Gilson, *J. Chem. Theory Comput.*, 2016, **12**, 414–429.

144 L. Defelipe, J. Arcon, C. Modenutti, M. Marti, A. Turjanski and X. Barril, *Molecules*, 2018, **23**, 3269.

145 M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu and A. E. Garcia, *Chem. Rev.*, 2016, **116**, 7673–7697.

146 D. Laage, T. Elsaesser and J. T. Hynes, *Chem. Rev.*, 2017, **117**, 10694–10725.

147 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

148 M. E. Wall, G. Calabró, C. I. Bayly, D. L. Mobley and G. L. Warren, *J. Am. Chem. Soc.*, 2019, **141**, 4711–4720.

149 E. F. Garman and T. R. Schneider, *J. Appl. Crystallogr.*, 1997, **30**, 211–237.

150 M. Nakasako, *Philos. Trans. R. Soc. B*, 2004, **359**, 1191–1206.

151 M. Gnesi and O. Carugo, *J. Appl. Crystallogr.*, 2017, **50**, 96–101.

152 S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403–410.

153 C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, *BMC Bioinformatics*, 2009, **10**, 421.

154 M. S. Bodnarchuk, *Drug Discov. Today*, 2016, **21**, 1139–1146.

155 C. N. Nguyen, A. Cruz, M. K. Gilson and T. Kurtzman, *J. Chem. Theory Comput.*, 2014, **10**, 2769–2780.

156 J. Michel, J. Tirado-Rives and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2009, **131**, 15403–15411.

157 G. A. Ross, E. Russell, Y. Deng, C. Lu, E. D. Harder, R. Abel and L. Wang, *J. Chem. Theory Comput.*, 2020, **16**, 6061–6076.

158 I. Y. Ben-Shalom, Z. Lin, B. K. Radak, C. Lin, W. Sherman and M. K. Gilson, *J. Chem. Theory Comput.*, 2020, **16**, 7883–7894.

159 M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson and C. W. Murray, *J. Med. Chem.*, 2007, **50**, 726–741.

160 M. L. Verdonk, P. N. Mortenson, R. J. Hall, M. J. Hartshorn and C. W. Murray, *J. Chem. Inf. Model.*, 2008, **48**, 2214–2225.

161 J. B. Dunbar, R. D. Smith, C.-Y. Yang, P. M.-U. Ung, K. W. Lexa, N. A. Khazanov, J. A. Stuckey, S. Wang and H. A. Carlson, *J. Chem. Inf. Model.*, 2011, **51**, 2036–2046.

162 J. B. Dunbar, R. D. Smith, K. L. Damm-Ganamet, A. Ahmed, E. X. Esposito, J. Delproposto, K. Chinnaswamy, Y.-N. Kang, G. Kubish, J. E. Gestwicki, J. A. Stuckey and H. A. Carlson, *J. Chem. Inf. Model.*, 2013, **53**, 1842–1852.

163 N. Huang, B. K. Shoichet and J. J. Irwin, *J. Med. Chem.*, 2006, **49**, 6789–6801.

164 J. Wahl and M. Smieško, *ChemMedChem*, 2018, **13**, 1325–1335.

165 D. Cappel, W. Sherman and T. Beuming, *Curr. Top. Med. Chem.*, 2017, **17**, 2586–2598.

166 X. Hu, I. Maffucci and A. Contini, *Curr. Med. Chem.*, 2019, **26**, 7598–7622.