Taylor & Francis
Taylor & Francis Group

Expert Opinion on Drug Discovery

ISSN: 1746-0441 (Print) 1746-045X (Online) Journal homepage: www.tandfonline.com/journals/iedc20

Water in drug design: pitfalls and good practices

Balázs Zoltán Zsidó & Csaba Hetényi

To cite this article: Balázs Zoltán Zsidó & Csaba Hetényi (2025) Water in drug design: pitfalls and good practices, Expert Opinion on Drug Discovery, 20:6, 745-764, DOI: 10.1080/17460441.2025.2497912

To link to this article: https://doi.org/10.1080/17460441.2025.2497912

Published online: 27 Apr 2025.

EXPERT OPINION ON DRUG DISCOVERY
2025, VOL. 20, NO. 6, 745–764
https://doi.org/10.1080/17460441.2025.2497912

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Water in drug design: pitfalls and good practices

Balázs Zoltán Zsidó[a] and Csaba Hetényi[a,b]

[a]Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, Pécs, Hungary; [b]National Laboratory for Drug Research and Development, Budapest, Hungary

ABSTRACT

Introduction: Structure-based drug design relies on optimizing drug–target interactions and blocking harmful pathophysiological events at the atomic level. Such events of the human body are modulated by water acting either as a medium or an individual partner in molecular interactions. A precise understanding of the modulatory mechanisms of water is essential for a successful drug design.

Areas covered: The present review discusses different topographical and networking situations that result in radically different roles of water, a root of various pitfalls of drug design. The review surveys good practices for tackling the problems of determining water structure at atomic resolution. Techniques for quantifying the effects of bulk, networking, and individual water molecules on the stability of drug-target complexes are also discussed. The article is based on a literature search using the PubMed, Web of Science, and Google Scholar databases.

Expert opinion: With advances in rapid computational algorithms and a better understanding of the physicochemical machinery of complex formation, theoretical approaches have resulted in elegant and cost-effective tools that fill the knowledge gaps left by the limited experimental methods. Overcoming the technical pitfalls of drug design, water transforms from a frustrating challenge into a handy tool for fine-tuning drug–target interactions.

ARTICLE HISTORY
Received 13 February 2025
Accepted 22 April 2025

KEYWORDS
Water; drug design; network; ligand; target; free energy; enthalpy; structure

1. Introduction

The engineering of potent ligands requires precise knowledge of the role of water in structure-based drug discovery [1–5]. Experimental techniques supply spatial coordinates of the water structure of the targets and their complexes with ligands that are indispensable for molecular engineering. However, determining such atomic resolution data [6] is still challenging and partly originates from water molecules' enormous networking and mobility (the root of these challenges are further analyzed in Section 2). They tend to form a hydrogen-bonded network around the solute molecules, and the disruption or re-organization of this hydration network also influences the overall strength of target-ligand binding [1,6–9]. The importance of water networks is now well-known for drug designers and numerous theoretical methods were released in the past decades to complement the missing structural data (Section 3). However, even small changes like adding a methyl group to a series of congeneric ligands can result in an unexpectedly large change in the binding affinity due to the rearrangement of the adjacent water networks [10]. Such large, ligand-induced variability of water networks still imposes a heavy challenge on experimental and theoretical techniques.

Computational docking can be considered a poster child for the above-mentioned water-related problems of ligand design. It has long been introduced as a major technique for calculating target–ligand complex structures [11–13]. Indeed, docking methods allow a fast selection of top candidates in high throughput screening campaigns on large ligand libraries [12,14,15]. However, the precision of computational docking is often deteriorated by the lack of an appropriate water model. Usually, only simplified, implicit water models are applied [16–18] that obviously cannot account for atomic interactions (H-bonds) of water molecules. This limitation can be annoying if the hydration structure becomes unpredictable due to the diverse interaction profile of the large variety of possible ligands binding to the target pocket. During the binding process, the ligand competes with the surrounding water molecules (and ions) for binding sites on the target (Figure 1(a)).

The water molecules can be classified into three main functional categories according to their roles in ligand binding.

(1) Conserved water molecules stay in the binding pockets of the target or the ligand and either form bridges [9] between hydrophilic groups of the partners (Figure 1(a)), or fill their sub-pockets. Conserved waters have thermodynamic/kinetic stability ('happy' with their position [19,20]) and become an essential constituent of the forming target–ligand interface (Figure 1(b)) [1].
(2) Displaceable waters have lower positional stability (unhappy) on the target/ligand surface. They are ready to leave their unfavorable (mostly hydrophobic) [21] sub-pockets and are expelled by the ligand during the competition (red arrow, Figure 1(a)).

CONTACT Csaba Hetényi hetenyi.csaba@pte.hu Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, Szigeti út 12, Pécs 7624, Hungary

Article highlights

- Incorporating water molecules into drug–target interactions is one of the greatest challenges in drug design.
- Experimental structure determination often results in incomplete water structures, and the (de)stabilizing role of individual water molecules cannot be measured instrumentally.
- Theoretical methods provide good answers for 'aqueous challenges' in structure determination and predicting drug-target complex stability.
- The application of molecular dynamics simulations and fast quantum mechanics-based scoring schemes will further increase the accuracy of the calculation of drug binding affinity.
- The further development of physicochemical methods and their clever combination with machine learning will likely increase the success of drug design projects in the next decades.
- By overcoming the technical pitfalls of drug design, water transforms from a frustrating challenge into a handy tool in fine-tuning drug–target interactions.

(3) Incoming water molecules that join the target–ligand complex (blue arrow, Figure 1(a)) during the binding process from the bulk (Figure 1(b)). They act as stabilizing staples bridging between the target and the ligand at the edges of the interface.

Every ligand impacts the local water structure and can affect the static/dynamic properties of the surrounding water molecules. Thus, the above categorization of waters is ligand-dependent. Anyway, the prediction of the actual function would be rather hopeless if explicit water molecules were not applied in the calculations. However, the incorporation of explicit water molecules in docking calculations is still not trivial, as their function largely depends on the actual ligand [22] (Section 3).

While there are established experimental methods for the determination of thermodynamic [23–28] and kinetic [25,26,29] binding parameters, it is still impossible to measure the effect of individual water molecules on the stability of drug-target complexes [30] experimentally. Changes in water structure can either increase [31] or decrease [32] ligand binding affinity or change the partitioning of enthalpic and entropic contributions. (In the latter case, mutual compensating effects can result in the binding affinity remaining unaffected.) Therefore, the determination of the influence of individual water on the thermodynamic signature of ligand binding poses a significant challenge and an untapped potential at the same time [33]. Among theoretical methods, physicochemical 'first principle' approaches have shown excellent results in predicting ligand binding affinity [3,27,28] and in

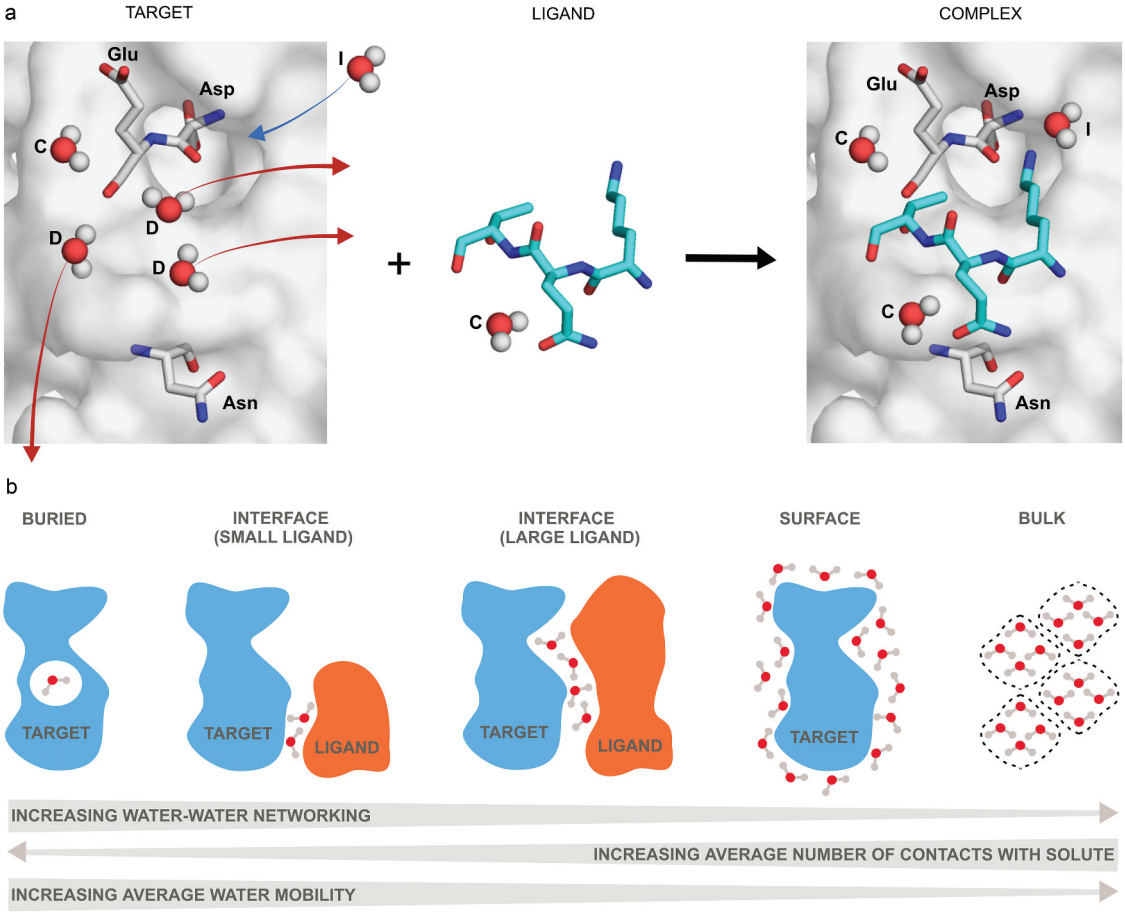

Figure 1. Functional classification and topologies of water molecules. (a) Functional classification of water molecules in a typical ligand binding situation. The binding of a ligand (teal sticks) is shown to a hydrophilic binding pocket (gray surface) with hydrophilic amino acid side chains highlighted as sticks. Hydrating water molecules are represented by red and white spheres. Red and blue arrows indicate leaving (displaced, D), and incoming, I water molecules, respectively. A conserved, C water molecule remains on the ligand and another on the target during binding. (b) Water topologies involved in target-ligand binding.

the thermodynamic contributions of individual water molecules [34–36], as well. Developing emerging technologies based on previous knowledge and artificial intelligence (AI) has also addressed the challenges of the structure and stability of drug-target complexes [37–41]. Section 4 provides further details on practices available for stability predictions. To support this work, PubMed, Web of Science, and Google Scholar database searches were performed until February of 2025, prioritizing recent developments in the field. Non-recent, but fundamental papers were also included.

2. Topography, networking, and models of bulk water

Many pitfalls and challenges of drug design are rooted in the molecular properties of water. The large difference in electronegativity of oxygen and hydrogen atoms, the two lone electron pairs of oxygen, and the angular (C_{2v}) symmetry of the molecule determine the large dipole moment [42,43] (polarity) and the maximal four hydrogen bonding (H-bonding) possibilities of water. The small size is responsible for its high mobility as a solvent (and also as a binding partner). With these unique physicochemical characteristics, water is a molecular Jolly Joker as it can stabilize or destabilize target-ligand complexes depending on its topographic and networking situation.

2.1. Topography

The (de)stabilizing effects of water can happen at different levels of networking depending on the actual location, and the number of contacts with the solutes (target and ligand). Due to the local topography, buried, interfacial, surface, and bulk waters can be distinguished (Figure 1(b)) [1]. On average, a buried or an interfacial water molecule tends to have the largest contact surface with the solutes and the lowest mobility. It mostly plays the role of an autonomous (individual) partner of the solutes and is important in the stabilization of either the solute structure [44] or the target-ligand complex [1,9,35]. With an increase in the size of the ligand and the interface, there is a higher level of networking between water molecules. A water molecule residing on the target (or ligand) surface has relatively few contacts with the solute, and they are also involved in large, loose cage-like networks [45] with each other. At the end of the scale, bulk waters organize in highly dynamic networks of grape-like clusters [42,43], forming the mobile medium surrounding the target-ligand complexes (Figure 1(b)).

2.2. Networking

Individual water molecules residing in tiny interfaces (Figure 1(b)) show low networking and mobility, and various practices are available for their structural determination (Section 3). The thermodynamic and kinetic characterization of their role in the (de)stabilization of target-small ligand complexes has also been well-studied, and a detailed account is given in Section 4. However, the precise description of the complex networking of water molecules in an extended target-large ligand interface or on a target surface (Figure 1(b)) is still a great challenge in drug design. Even displacement (Figure 1(a)) of a single networking water molecule by a ligand can disrupt or rearrange a series of water–water or water–target interactions [1], with an overall effect on kinetic and thermodynamic stability. The role of a single water molecule in an interaction network can be quantitatively analyzed using graph theoretical approaches [46]. Furthermore, information on hydration networks can be extracted from computational simulations and used for ligand (hit) selection and design. For example, in the topological water network fragment screening (TWN-FS) method, the shape and geometry similarity of the fragments to TWNs is used for hit selection [47].

A 'divide and conquer' analysis can split large water networks into sub-nets and help uncover their role in the target–ligand interactions. Water molecules can be considered as nodes of a network graph, and depending on their interconnectivity and mobility, they can be assigned to static or dynamic sub-nets of the full hydration network [9]. Static sub-nets are mostly formed by thermodynamically or kinetically stable water molecules trapped at the bottom of deep binding pockets interacting with hydrophilic residues. Static sub-nets are generally small and often conserved during ligand binding. Their waters can be formally considered part of the target contributing to complex stability [9,48]. A hydration network analysis [49] by MobyWat [50] showed that the increased amount of static water sub-nets corresponds to elevated dissociation barriers and a decrease in ligand dissociation rates (higher binding affinity to the oligosaccharide ligands) in the case of a mutant of *E. coli* bacteriophage HK620 (HK620TSP) [49] if compared to the wild-type protein. The results of the study were confirmed by experimental kinetic analyses, as well.

In dynamic sub-nets, the water–water interactions dominate (instead of the solute–water interactions), and have loose connectivity quantified by the low density of edges [9,51,52]. Dynamic sub-nets can be found around hydrophobic residues where water molecules are turned away from the residue and form loose cage structures with each other [35,53]. The transformation of the dynamic sub-net of these cage structures during ligand binding is responsible for the stabilizing hydrophobic effects [54–56]. Dynamic sub-nets are also formed at the shallow regions of the binding interface that cannot provide a topological [57] trap or at surface regions of the target that continuously communicate with the bulk. They are important for recognizing various ligands [58] and affect binding affinity [59]. As the dynamic sub-nets are located between the static sub-nets and the bulk, they also mediate the migration of water molecules into the interface, a key step in the mechanism of ligand dissociation [8].

In particular, large water networks play an important role in designing protein–protein interaction (PPI) inhibitors (Figure 2(a)). The PPI inhibitors have to compete with the binding of the protein partners and appropriately modulate the interfacial hydration network between them. Due to the large interface areas, the PPI water networks are extended (Figure 2(a)) and very heterogeneous, including static regions at the interaction hot spots [20,63], and dynamic sub-nets at the shallow interface regions. A better adaptation of an inhibitor to a favorable PPI water

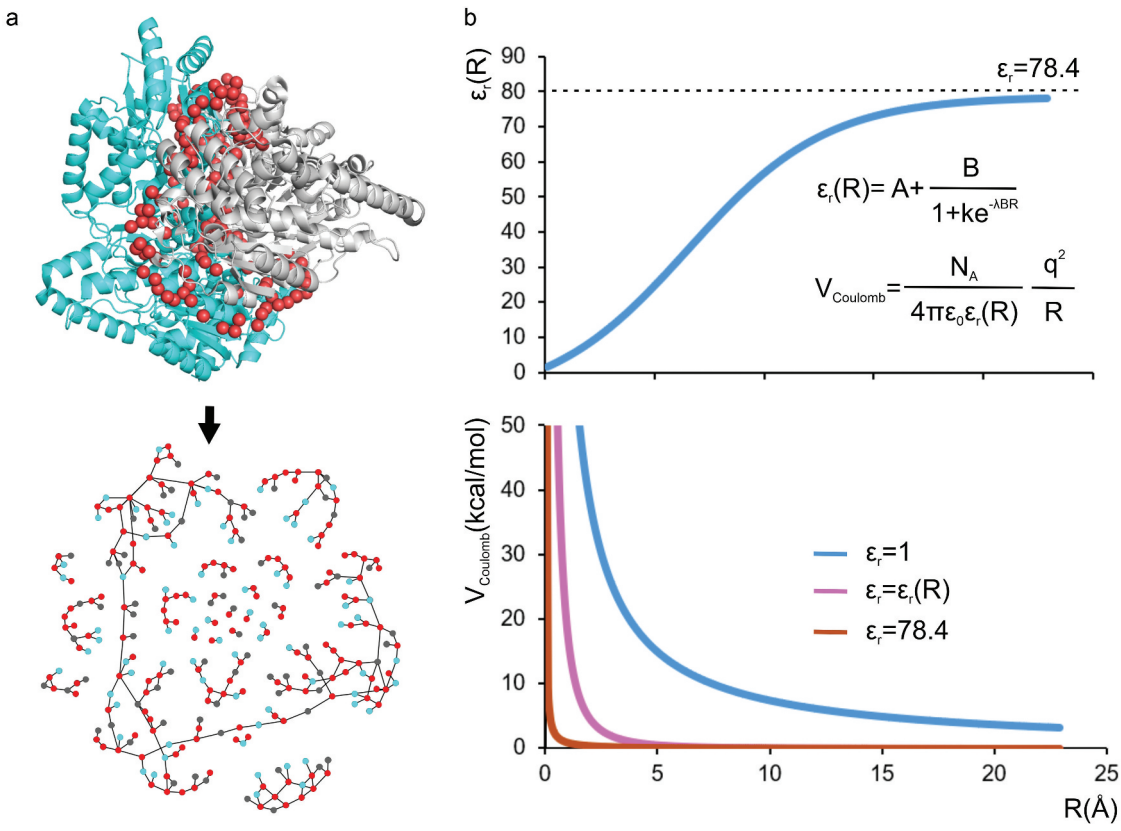

Figure 2. The role of hydration networks in protein–protein interactions and the shielding effect of water. (a) Protein–protein interaction as seen in the structure of *Escherichia coli* transketolase homodimer (PDB code 6tj9 [60]). The two interacting protein chains are shown in gray and teal cartoons (top). Interface water oxygen atoms are shown as red spheres. The network graph of the water molecules was produced from the structure by MobyWat NetDraw mode [50] and visualized with the Gephi [61] software. The colors of the network nodes follow the coloring of the structure. (b) The shielding effect of bulk water is shown on the example of an oxygen-oxygen atom pair. The plot of a distance-dependent relative permittivity ε_r(R) function [62] (top), where R is the distance between two neighboring atoms. The dotted line represents water as a bulk medium with $\varepsilon_r = 78.4$. The corresponding (repulsive) Coulomb potential (V) was calculated for two neighboring oxygen atoms (bottom). A, B, λ and k are constants [62], ε_0 is the vacuum permittivity. In the Coulomb potential q is the partial charge of an oxygen atom, N_A is the Avogadro number. V was also calculated for constant ε_rs of the vacuum (1, blue) and the bulk (78.4, red). The intermediate (purple) V curve with an ε_r(R) allows realistic modeling of the change of shielding effect of water on the electrostatic interactions as it provides a transition between the vacuum and the bulk curves in the region of ca. 1–2<R < 10 Å important for water-water H-bonds, and ligand binding. That is, after the covalent bond region (ca. R = 1–2 Å) there is still a repulsion that drops reaching the cut-off (ca. R = 10 Å), due to the increase of the shielding effect of bulk waters.

network increases its potency by orders of magnitude [20]. The interface region of PPIs is hydrophobic or contains hydrophobic patches with polar interactions and water molecules present [64]. The displacement/replacement of these water molecules by PPI inhibitors has important thermodynamic consequences [64], and the release of water molecules from hydrophobic interface regions is labeled as one of the most important driving forces of PPIs [21].

2.3. Models of bulk water

Bulk waters represent the largest topographical group with the largest networks and highest mobility on the right end of the scale of Figure 1(b). In the liquid phase, water molecules form extensive but weak (very dynamic) sub-nets in/between their loose grape-like clusters [42,43]. They supply incoming (individual) water molecules during ligand binding (Figure 1(a)), and the medium of target–ligand interactions. Their collective effects are usually accounted for by implicit or explicit solvent models instead of vacuum simulations, whose limitations were recognized early on [65]. Water as a Jolly Joker medium has different effects on the interaction of apolar and polar sites [66]. For apolar sites, the hydrophobic effect was described above. Polar interactions are effectively shielded (dampened) by water molecules located along the electric force field of the surrounding molecules, following their large dipole moment. This effect can be (roughly) accounted for in the Coulomb potential using distance-dependent relative permittivity [ε_r(R)] functions [62,67,68] that have a sigmoidal increase of ε_r (Figure 2(b)) with an increase in the distance of the interacting charges (R in Figure 2(b)). Accordingly, the repulsive electrostatic interaction calculated for the example of two neighboring oxygen atoms (Figure 2(b)) scales between vacuum ($\varepsilon_r = 1$) and bulk (ε_r = 78.4) with a continuous transition provided by the ε_r(R) function and converges to zero together with the bulk ($\varepsilon_r = 78.4$) curve. This is especially important in the target-ligand contact region (at an R between the covalent distance of ca. 1–2 Å and the electrostatic interaction cutoff of ca. 10 Å). Thus, ε_r(R) functions have been implemented and proved efficient in fast docking programs [69].

Molecular mechanics (MM) level calculations of target–ligand interactions apply implicit or explicit water models for the bulk. Implicit water models of continuum solvation methods, like the molecular mechanics Poisson-Boltzmann [70–72]

(MM/PBSA) and generalized Born surface area [73–75] (MM/GBSA) do not consider explicit water molecules, just a continuous isotropic medium [76]. In such models, the solute is represented as a cavity in the water continuum, and its charge density is restricted by the boundaries of the cavity [76]. When water plays a critical role in the interaction of the solutes (target and ligand), the implicit solvent models have limitations [77] for the calculation of the hydrophobic effects [54–56] and the description of water-solute networks [78]. Explicit solvent models offer a much more realistic approach as they work with the known geometry and partial charge distribution of water molecules at the atomic level. The flexibility of water molecules is often kept restricted, point charges and pairwise additive interactions are used [79] in the applications of the popular transferable intermolecular potential (TIP [80]), and simple point charge (SPC [81]) models to various problems where water molecules play a key role [82–84]. The explicit solvent models are still under massive development, including flexibility and polarizability [77,85,86]. Implicit solvent models have been implemented in quantum mechanics (QM) approaches [87], and there is increasing use of explicit waters and hybrid models in current target-ligand calculations (Section 4.2).

3. Patching the potholes of structure determination

3.1. Problems of experiments

The quick and accurate determination of target–ligand complex structures is inevitable for structure-based drug design [88]. More than half of the experimental structures deposited in the Protein Data Bank (PDB) [89] contain at least one water molecule [2], usually found at interface regions (Figure 1(b)) [90]. However, the state-of-the-art experimental structure determination methods still have several limitations, as outlined in the next paragraphs. Despite the limitations, it must be emphasized that experimental measurements remain the ultimate source of data for validating theoretical predictions. In particular, knowledge-based computational methods (next Section) heavily rely on large experimental databases [91]. Several thousands of crystallographic water positions (on proteins) were used to validate quick [30] and computationally demanding theoretical methods [50], as well.

Cryo-electron microscopy provided fascinating results in determining large macromolecular systems [92,93]. However, its performance is still limited in smaller details, like measuring ligands and water molecules on the protein surface [59,94]. Cryo-electron microscopic maps with a resolution of < 2.0 Å are not routinely produced [95] at a large scale, necessitated by drug design. This resolution is often necessary for smaller details, like ligands and water molecules.

To this day, the main experimental source of protein structures is X-ray crystallography [96,97]. Notably, X-ray crystallographic experiments provide static snapshots of biological systems and (except in a few cases [98]) lack information about their dynamics [99]. The number of human proteins is limited, of which only a fraction is druggable [100]. Many human druggable proteins are membrane-bound, with limited solubility [101]. Apart from membrane-bound proteins, the expression, crystallization, and purification of proteins are challenging [102]. Furthermore, a crystallized protein is not in the native form present in the matrix of the living body. Therefore, the crystallization process might result in artifacts (structural features not typical of native proteins). For membrane-bound proteins, the artifacts are also caused by detergents (required for extracting the protein from its original cellular environment) that necessitate the development of new methods to detect artifacts [103]. The measurement of the crystallized protein requires high energy radiation and a low (cryogenic) temperature, leading to further artifacts [104] mainly regarding waters, especially those involved in the conformational change of highly hydrated amino acid side chains. Fitting atoms and amino acids into X-ray crystallographic electron density maps also have a natural uncertainty [105]. It might necessitate the manual intervention of a researcher, bringing in the human factor. The development of automation in high-throughput X-ray crystallography is promising especially in fragment-based lead discovery [106,107]. However, it is important to notice that the crystallization of many protein-ligand complexes requires tremendous efforts to set the appropriate experimental environment.

Still, X-ray crystallography is the most popular method for the experimental determination of hydrated structures. However, it also has some limitations for structure determination of hydration [1,50,57,104,108]. The main problem with the experimental determination of water molecules lies in their great mobility and high degrees of freedom. In smaller systems, fitting water molecules into electron density maps is more accurate than in larger systems [109], where setting and controlling thousands of water molecules is a tremendous work and the 'human factor' can play a role. The lonely oxygen atom of water molecules has a small peak in electron density compared to the atoms of surrounding amino acids, and the small scattering of water hydrogen atoms further complicates structure determination [109] and increases the uncertainty of the resulting water structure. In protein crystal packing, the interface regions might amount to up to 30–40% of the whole protein surface [110], and the water molecules therein might be an artifact of crystallization. Neutron diffraction experiments [111,112] improve the experimental determination of hydration structure over X-ray crystallography [113]. However, only a few laboratories can perform such experiments, and it may be difficult to find sufficient beam time as the flux of a neutron source is by orders of magnitude less powerful if compared with a synchrotron source [114].

Finally, nuclear magnetic resonance spectroscopy (NMR) can be mentioned because it gives a good picture of the atomic dynamics of smaller proteins or peptides and shows advances in hydration dynamics determination [115]. However, the number of water molecules determined by NMR is still low [2]. Besides the monetary and technical drawbacks of the above methods, there are fundamental challenges that experiments cannot overcome without the help of computational methods described in the next Section. On the other hand, computer methods can hardly be developed without experimental data.

3.2. Practices offered by computational methods

Like in other fields of drug design [116], computational methods can complement experimental methods for a better understanding of the water structure [2]. These methods are

based on the first principles of physical chemistry or the available experimental knowledge, including recent artificial intelligence (AI) techniques [117–119]. The hydration structure can be predicted for the separate target and ligand partners alone or in combination with the docking simulations. The performance of physicochemical methods was discussed previously [2] and their success rate (the match with the experimental structure) can reach 80% or 90% for the prediction of water positions on the whole target surface or in the target–ligand interface, respectively [2]. A brief list of available practices updated with knowledge-based techniques and docking protocols is provided in Table 1.

3.2.1. Physico-chemical methods

Physico-chemical methods are based on molecular mechanics (MM) force fields that allow the prediction of water structure and energy contributions. The methods are based on grid (docking) calculations or MD simulations [5]. Grid-based approaches generally account for solute–water interactions and do not consider water–water H-bonds in the hydration shell. They often use a precalculated energy grid around the target that allows fast prediction of the water positions. For example, HydraMap v.2 [144] is a grid-based method that calculates statistical potentials for protein-water and ligand–water interactions, then places the binding pocket of a protein in a grid box, and generates water probes on the grid points. The potential of mean force energies of the water probes are calculated, and the probes matching a threshold are kept as hydration sites. The method also distinguishes between three functional (displaced, conserved, and shifted) types of water molecules based on ligand proximity. WaterDock [123,124] docks water molecules into the target using AutoDock Vina, and clustering the sites of the best scores results in the predicted water sites. The method was validated on a carefully assembled set of consensus water positions from high-resolution crystal structures, neutron diffraction data, and MD simulations. WaterDock performs independent docking runs, water–water interactions are not considered, and water site locations are not optimized to form H-bonds with each other.

Molecular dynamics (MD) simulations also utilize MM force fields and can be performed with explicit water models mentioned in Section 2.3. Besides solute–water interactions, they account for water–water H-bonds and can describe the dynamics of water networks. For example, WaterMap [120] is commercial software with various applications [145]. WaterMap calculates water positions from a short MD simulation followed by an occupancy-based clustering of hydration sites submitted to a thermodynamic analysis (described in Section 4). Open-source programs like MobyWat [50] are also available to extract hydration structure from MD trajectories, generating mobility and occupancy values for each water molecule. MobyWat was written in standard C and can be installed on any platform. It is compatible with any MD program, including the popular open-source MD program package GROMACS [146], and can read binary trajectories and PDB files. Besides extracting hydration structures, MobyWat has a Validation sub-mode to compare results to experiments, and a NetDraw mode to map the complex water-mediated network of solutes.

Physico-chemical methods offer good performance (success rates) [2] and the above applications of explicit water models can handle water–water contacts and internal changes in the hydration structure. However, MM-based approaches are limited by the common use of force fields that do not consider the polarizability of the electron systems. Due to the large dipole moment of water, the involvement of polarizable force fields is an important direction for improving the MM-based methods in calculating the hydration structure. A polarizable force field and a grand canonical nonequilibrium candidate Monte-Carlo method were used to sample cavity hydration in proteins [147]. Water molecules were exchanged between a virtual reservoir and a buried cavity of the protein trypsin. This approach eliminated the kinetic barrier present in such exchanges during MD simulations. The polarizable force field resulted in good water placement accuracy in the unique microenvironment of buried cavities [147]. SZMAP [36] is based on a combination of explicit and implicit solvent models (Section 2). A grid representation of the solute is applied, and then the explicit water probe is scored.

3.2.2. Knowledge-based methods

Knowledge-based methods are trained on available experimental structural data. WATGEN [133] is a tool for predicting PPI hydration structure, a major problem of PPI inhibitor design (Section 2). The method is based on experimental observations of amino acid solvation. Water sites are distributed around the solute's H-bond acceptor and donor centers. Sites within a meaningful distance of the solutes are kept if they do not collide with any solute atom. Then, the water sites are scored based on the number of interacting solute polar atoms around them (number of H-bonds formed) [133]. The water-scoring approach of WaterFlap [132] also relies on the number of H-bonds formed by a water molecule [148,149]. WaterFlap uses a grid representation instead of a spherical one, and the energetic contours of the grid are based on molecular interaction fields [132]. Probes are fitted into protein cavities and favorable interactions are searched. The increasing number of favorable interactions corresponds to an increased probability of ligand (or water) binding to the target. WarPP [30], a similar, geometry-based approach, was validated on a large set of experimental data. Polar solute atoms with unsaturated H-bonding capacities or geometrically non-ideal H-bonds (that could be replaced by solvating water molecules) are selected. The information on favorable H-bonding geometries was derived from experiments. These sites (if not occupied) were then classified as potential water positions, and explicit water molecules were positioned accordingly. The water molecules were then refined to achieve the most H-bonds possible. A hybrid approach, Fold-X [131] combines an empirical MM force field and experimental data from X-ray crystallographic structures. The position of a water molecule and interacting protein atoms are extracted from X-ray crystallographic structures and superimposed onto the query structure. Then, water probes are placed into these positions, and clashes are removed. A score is calculated for the water probes based on an empirical MM force field; those with a favorable calculated score are kept. The statistical method 3D-RISM [130] calculates the probability of the presence of a solvent molecule in an area surrounding the solute, leading to a solvent shell covering the solute. 3D-RISM

Table 1. Water structure: prediction methods and implementations in computational docking.

Method name	Features	Availability	Year of publication
Physico-chemical methods			
WaterMap [120]	An all-atom MD-based method with explicit water to predict hydration sites and their thermodynamic contributions.	Standalone program, commercial	2008
HyPred [121]	An all-atom MD-based method with explicit water for predicting the solvent density near the surface of a protein.	Server unavailable, free for academic use	2010
STOW [122]	Applies inhomogeneous fluid solvation theory to calculate the thermodynamic contributions of interfacial water molecules to ligand binding.	Server unavailable	2012
WaterDock [123,124]	Fast and accurately validated method that docks water molecules to the target using AutoDock Vina.	Standalone program, open source	2012, 2017
SPAM [125]	An all-atom MD-based method with explicit water to predict hydration sites at the water-protein interface and compute a local free energy score based on the distribution of interaction energies between water and the environment at a specific site.	Standalone program, free for academic use	2013
WATsite [126]	An all-atom MD-based method with explicit water model to calculate the thermodynamic contributions of interfacial water molecules to ligand binding.	Server unavailable, free for academic use	2014
MobyWat [50]	An all-atom MD-based method with explicit water for predicting water molecules on the surface of targets and at the interface of target-ligand complexes.	Standalone program, open source	2015
WATCLUST [127]	An all-atom MD-based method with explicit water for predicting water molecules on the surface of targets.	Standalone program, free for academic use	2015
SZMAP [36]	Places water probes on grid points and calculates their free energy of binding compared to the same probe with removed charges and van der Waals terms.	Standalone program, commercially available	2015
Knowledge-based methods			
Auto-SOL [128]	Provides surface and interface hydration for proteins based on H-bond directionality.	Standalone program, free for academic use	1991
AQUARIUS [129]	Based on an analysis of experimental solvent positions around polar and charged amino acid residues.	Server unavailable	1991
3D-RISM [130]	Calculates thermodynamic properties and solvent structure around a solute based on a distribution function.	Standalone program, free for academic use	1999
Fold-X [131]	Predicts interface and surface waters based on a continuum solvation model and its own force field.	Standalone program, free for academic use	2005
WaterFlap [132]	Places interface and surface water molecules based on a grid representation with energetic contours calculated by molecular interaction fields.	Standalone program, commercially available	2007
WATGEN [133]	Places interface and surface water molecules based on experimental observations of amino acid hydration.	Website unavailable	2007
WarPP [30]	Places surface and interface water molecules based on interaction geometries derived from experimental data.	Standalone program, free for academic use	2018
Splash'Em [134]	Places water molecules into nucleic acids based on experimental data and a gradient descent refinement step.	Standalone program, free for academic use	2019
Knowledge-based (AI) methods			
Accutar [40]	Places surface and interface water molecules based on a scoring function that probes proteins for missing water molecules derived from water-water distances.	Server, free for academic use	2021
Galaxy-CNN [38]	Places surface and interface water molecules based on a water score map derived from experimental water positions.	Server and also standalone program, open source	2022
HydraProt [37]	Applies a coarse representation of the protein to quickly place water candidates and evaluates them based on confidence scores.	Standalone program, open source	2024
Implementations in computational docking			
FlexX [135]	Precalculates favorable water sites and spherical waters can be switched on during docking.	Standalone program, free for academic use	1999
SLIDE [136]	Applies another program Consolv to predict water molecules that are likely to be displaced, the remaining waters are kept and can be displaced during docking at the cost of a penalty.	Standalone program, free for academic use	2000
AutoDock [137]	Calculates multiple energy grids during docking with or without water molecules and 'choose' between them keeping the one with the most favorable score.	Standalone program, open source	2002
Glide [138]	Water molecules are docked into the binding site and scored by empirical scoring terms. These can be kept or removed by the user, the remaining waters are minimized with the ligand in the binding site.	Standalone program, commercially available	2004
GOLD [139]	Water molecules are allowed to rotate and their energetic contributions are scored during docking, displacement of water molecules is rewarded by entropic gain.	Standalone program, commercially available	2005
FITTED [140,141]	Either uses crystallographic water molecules or places water probes on grid points, optimizes them, and keeps the energetically most favorable ones.	Standalone program, free for academic use	2007
DOCK 3.5.54 [142]	Water molecules are treated as flexible parts of the target, and they are switched off (removed) if their contribution to ligand binding is unfavorable.	Standalone program, open source	2008
HADDOCK [143]	The target is hydrated before docking using MD simulations, and water molecules are displaced by the ligand binding based on experimental data.	Server, free for academic use	2016
HydroDock [108]	Performs dry docking and hydrates the binding site of the apo target using MD simulations and MobyWat. Assembles and refines candidate complex structures and selects representative hydrated binding modes.	A protocol combining open-source programs	2021

performed weaker than its physical-chemical peers in a comparative study [150].

The use of artificial intelligence (AI) for scientific predictions has been widely investigated in the past decades following the pioneering works of John von Neumann [151] and other founding computer scientists. In 2024, the Nobel Prize in Chemistry was awarded partly for protein structure predictions performed by AlphaFold [152], an AI-based software. AlphaFold was developed to predict 3D structures of proteins on a large scale to assist target-based drug design. From

amino acid sequences and a large training set of known 3D protein structures, the method predicts 3D structures of proteins of previously unknown structures with deep learning. However, currently, the handling of non-amino acid entities, such as water molecules, is not routinely implemented in the program [153]. Thus, further AI-based tools were developed to predict water positions in protein structures. For example, HydraProt [37] starts from a coarse representation of the protein, sampling possible water positions. The final prediction is based on multi-layer perception, and a refinement step is performed to minimize sterically clashing water positions. The performance of three deep learning-based programs was also evaluated on the test set of 600 proteins (the number of experimental waters in this set was not specified). HydraProt, Accutar [40] and Galaxy-CNN [38] were compared. The success rates of accurately predicted water molecules within 1 Å distance from an experimental water position in the first hydration shell were 76, 53, and 68 %, respectively [37]. Additionally, the prediction of interface water molecules was performed on a test set of 151 complexes (with an unknown number of waters). The highest success rate (79% within 1 Å) was achieved by HydraProt [37]. Knowledge-based prediction of hydration sites is also available for DNA, and a recent method applying the probability of water density distribution of hydrated building blocks successfully predicted 40% of experimental waters within 0.5 Å [154].

3.2.3. Hydration in computational docking

The programs surveyed in the previous Sections calculate the water structure for an apo (target) or a holo (complex) structure where the solute structure is known. However, at the beginning of drug design projects, the structures of the target-ligand complexes are not known for all possible candidate ligands. Fast computational docking techniques can provide complex structures even at large numbers required by high throughput screening (HTS) campaigns. However, their accuracy is limited by the simplistic (implicit) water models (Section 2) discussed in the Introduction. For example, Cov_DOX [155] is a covalent docking program that uses implicit solvent models COSMO and SMD. The program achieved better structural matches with experiments than other benchmark programs [155]. However, the authors reported problematic cases, when the scoring function of their program prioritized target–ligand interactions over target–water interactions that are not regarded in their calculations.

Thus, including explicit water molecules would theoretically help drug designers to avoid such pitfalls, and various solutions have been suggested (Table 1). However, it remains an open question if computational docking should start with placing the ligand or the waters into the binding pocket (a 'chicken-and-egg problem') [2,25]. The most common strategies treat the waters as a static part of the target [26] or as a flexible part of the ligand [27], in some cases allowing their displacement/movement from their original position on the target [28,29]. If experiments cannot capture all available water molecules, then certain theoretical methods can predict hydration structure before docking [30,31]. In DOCK 3.5.54 [142], the displacement of water molecules is optionally considered. The off (displaced) and on (conserved) states of water molecules are based on whether the calculated interaction energy of the ligand with the target (and flexible waters as parts of the target) improves by keeping them (on) or displacing them (off). In GOLD [139], a penalty term is introduced for the displacement of water molecules. Similarly, SLIDE [136] penalizes the displacement of favorably bound water molecules in the ligand score. AutoDock [137] can combine several experimental structures with water molecules into the same grid by a weighted average approach. In a FlexX [135] protocol, water particles are added before docking, and their displacement occurs during docking. FITTED [140,141] uses a similar approach, and the displacement of water molecules depends on their distance from the docked ligand. Glide [138] adds water molecules after ligand docking. Besides the above-mentioned serial approaches, a parallel combination of fast docking and MD-based hydration is also possible. The HydroDock protocol [108] uses the void-free (cavity-free) hydrated structure of the target as predicted by MobyWat [50]. A simultaneous docking step is performed on the dry target to calculate the dry target-ligand complex. Then, the hydrated target and the dry target–ligand complex are combined, and the overlapping water molecules are removed. The hydrated complex is minimized to accurately set the orientations of the hydrogen atoms of the water molecules. A short MD simulation is performed with an explicit solvent model to generate a set of ligand-binding modes on the target, driving the ligands and waters to the real (experimental) binding modes [108]. The closest to average ligand-binding mode is finally selected from all the MD frames as the representative binding mode [25].

4. Fine-tuning the thermodynamic and kinetic stability of ligand binding

Water contributes to the thermodynamic and kinetic stabilization of the target-ligand complexes, acting as a medium, network, or individual molecule (Section 2). The stabilization effect of conserved or displaced (Figure 1(a)) individual water molecules increases binding affinity measured by a change in the thermodynamic functions. The net effect is expressed as the Gibbs free energy change ($\Delta G_{b,lig}$) of the ligand-binding reaction (Equation. 1, for 1:1 stoichiometry, non-covalent complex formed), that is a sum of enthalpic ($\Delta H_{b,lig}$) and entropic ($\Delta S_{b,lig}$) contributions (Eq. 2, where T is the absolute temperature) and, therefore, both types of contributions have to be considered in the stabilization of target-ligand complexes.

$$Target(H_2O)_T + Ligand(H_2O)_L = Target : Ligand(H_2O)_{T+L-w} + wH_2O \quad (1)$$

$$\Delta G_{b,lig} = \Delta H_{b,lig} - T\Delta S_{b,lig} \quad (2)$$

Conserved (Figure 1(a)) water molecules bridging between the target and ligand partners usually have a favorable contribution to $\Delta H_{b,lig}$ [156] via H-bonds formed with both partners [157]. Useful practices of molecular engineering are shown on well-known drug targets in Figure 3. The incorporation of a key water position (as an isosteric oxygen atom) into the ligand results in a more potent and/or selective compound (Figure 3(f)). Accordingly, the ligand is extended

by a hydrophilic group isosteric to the key interface water molecule. In other cases, the new ligands can form additional interactions with the target by displacing the key water molecules (Figure 3(c,d)). Water molecules can also fill empty spaces, and sometimes they are trapped in an energetically unfavorable hydrophobic environment in the binding pocket [35]. The displacement of such 'unhappy' waters (Figure 1(a)) by a ligand has a favorable contribution to binding affinity via the release of waters into the bulk (desolvation, Figure 3(c)) and a corresponding increase of $\Delta S_{b,lig}$ [10]. In this case, the entropic gain can outweigh the (enthalpic) desolvation penalty [162]. The enthalpic contribution of the new target–ligand interactions is more pronounced in the example of Figure 3(d). The design of blockers against influenza A viral ion channel is another good example, where the hydrophobic interior of the channel is not well druggable by small molecules [163], and an elongated bulky adamantane derivative capable of displacing all water molecules from the channel achieved a good potency against virus replication [164].

The number of good practices is increasing for calculating the contributions of the n-th individual water molecule to the solvation thermodynamics ($\Delta G_{b,wat}$, $\Delta H_{b,wat}$, and $\Delta S_{b,wat}$) of the solute (target/ligand) partners (Equation. 3).

$$Target(H_2O)_{n-1} + H_2O = Target(H_2O)_n (n = 1, 2, \ldots) \quad (3)$$

Such approaches are especially applicable for congeneric ligand series to determine key hydration sites that should be targeted during drug design [35,144,165]. In the cases of small, static hydration networks (Section 2), the classification and use of individual water positions can support the design of a potent ligand. However, the thermodynamic effects of individual water molecules are not necessarily additive if they participate in a hydration network. For example, the simultaneous displacement of all water molecules results in a smaller $\Delta G_{b,lig}$ than the sum of $\Delta G_{b,wat}$s of individually replaced water molecules [166]. Thus, the perturbation of local water sub-nets should also be considered during energy calculations, not just the displacement of individual water molecules [166].

Besides equilibrium thermodynamics, the reaction kinetics of drug binding is an equally important factor of pharmacological activity [167,168]. The association rate constant (k_{on}) and the dissociation rate constant (k_{off}) are two deterministic parameters of binding kinetics. The residence time (τ) of a ligand (or a water) molecule on the surface of the target can be expressed as $\tau = 1/k_{off}$ [167]. Waters can form H-bonds with hydrophilic residues (Figure 1(a)) and with each other [35]. The displacement of water molecules by the ligand (dehydration) requires breaking the H-bonds and crossing the corresponding energy barrier determined by $\Delta G_{b,wat}$ of the interaction of the target and the individual water molecule (Equation. 3) under consideration. However, a correlation cannot be observed between τ_{wat} and $\Delta G_{b,wat}$, and τ_{wat} is rather determined by the surrounding topography [8,57] and the overall dynamics of the protein. For example, a deep pocket on the target surface can trap water at the bottom and prevent the cooperative exchange of the trapped water molecules with others in the bulk. This may lead to a long τ_{wat} of the trapped waters even if they have a weak interaction (small negative $\Delta G_{b,wat}$) with the target [8,57].

In addition to the above target (de)hydration, water may directly modulate ligand-binding kinetics [168]. It can attack (and disrupt) H-bonding between the target and the ligand leading to dissociation, which is the reverse of the association process described in Figure 1(a). Local target topography can largely affect the outcome of the association/dissociation processes [8]. In the case of narrow binding pockets, there is not enough space for the simultaneous binding of water and a ligand. Ligand association (Figure 1(a)) starts with breaking target-water H-bonds, leading to a thermodynamic penalty, which is only asynchronously compensated later by the binding of the ligand [8]. The thermodynamic penalty corresponds to an increased kinetic stability of the water molecules shielded by the narrow target pocket [8]. Similarly, ligand dissociation is also slower from a narrow binding pocket [8] that topologically shields the target-ligand H-bonds and protects them from the intruder water molecules. The rearrangement of water–water interactions in networks or the bulk can also contribute to the kinetic stability of target-ligand binding. The disruption of a single layer of water shell between the target and ligand poses a kinetic barrier to ligand binding [169]. On the other hand, the displacement of disordered water molecules from a poorly hydrated binding pocket is favorable [170]. The mobility (τ_{wat}) in water networks also affects the thermodynamic stability of large protein complexes. For example, in a complex of ATRX ADD histone reader and the trimethylated histone H3K9me3, a static sub-net of water molecules formed around the charged ligand side chain [9]. The static sub-net shields target-ligand H-bonds from the bulk solvent, leading to increased thermodynamic stability of the complex [9]. As a result, the dissociation constant of the ATRX ADD-H3K9me3 (trimethylated) complex is 7 times lower than that of ATRX ADD-H3K9me0 (non-methylated) [171].

Despite the importance of individual water molecules or their networks in ligand binding, experimental techniques are not available for directly measuring their thermodynamic/kinetic effects. At the same time, instrumental techniques of net thermodynamics [1,27,33,172], and kinetics [173] of target-ligand binding are available and indispensable for the (indirect) validation of computation methods that calculate individual contributions of water molecules or their networks. Similar to the case of structural calculations (Section 3.2), the computational methods are based on physicochemical or knowledge-based algorithms (Table 2) described in the next paragraphs. The physicochemical approaches can work either at the molecular mechanics (MM) or quantum mechanics (QM) levels and apply implicit, explicit, and more recently hybrid water models that combine the advantages of the other two.

4.1. MM-level methods

Besides the structural performance, individual water molecules also influence the performance of predicting the thermodynamic stability (scoring) of target-ligand complexes in docking calculations (Introduction). While there are routinely used (implicit) methods [188] for estimating bulk (de)

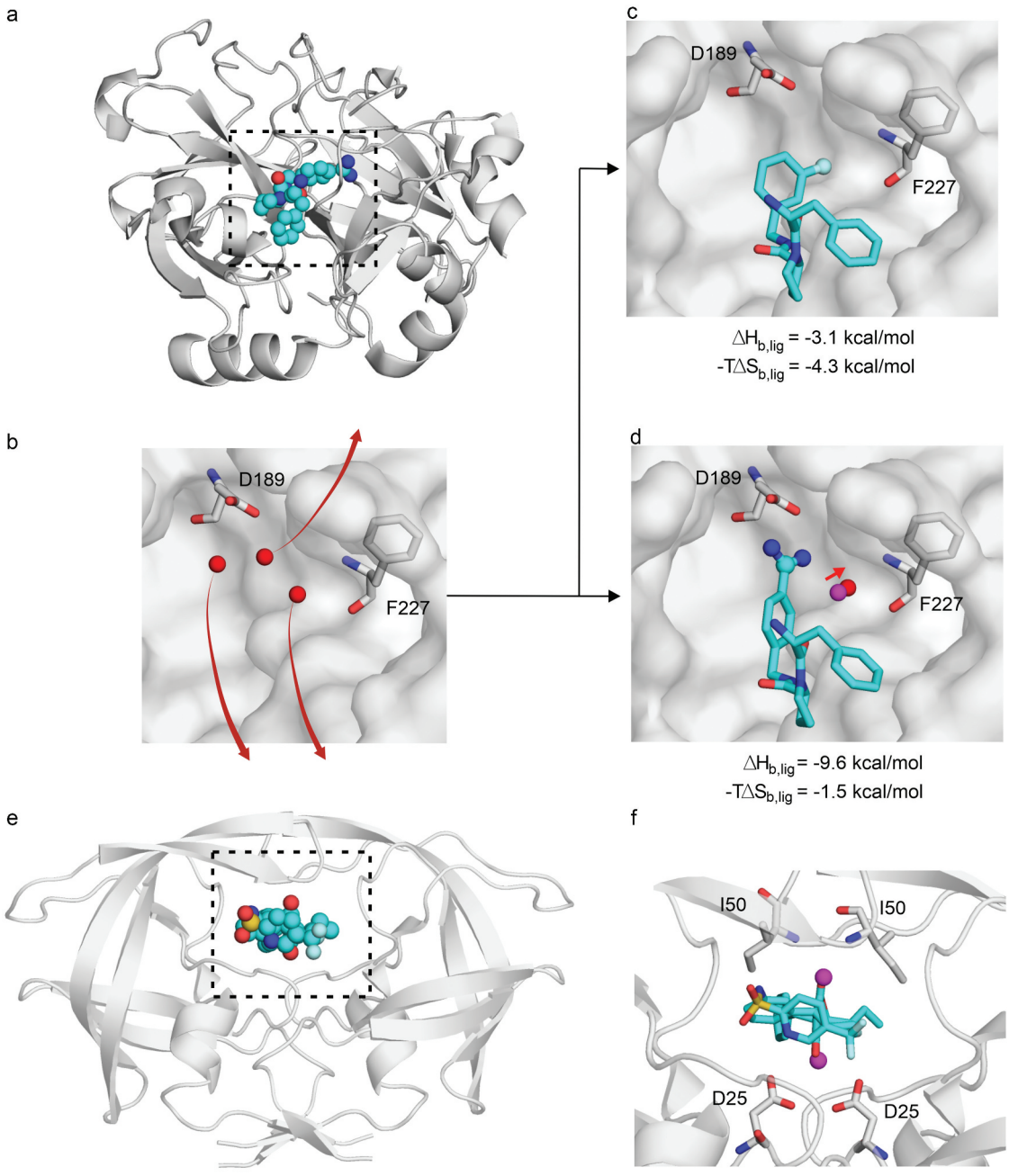

Figure 3. The effect of individual water molecules on ligand binding thermodynamics. (a) The inhibitor-bound (teal spheres) structure of thrombin (gray cartoon, PDB code 2zda [158]). The black dotted rectangle borders the binding pocket featured in (b)–(d). (b) The binding pocket of apo thrombin (PDB code 3d49, grey surface) with a highlight on three important waters (red spheres) and two interacting amino acids (grey sticks). Red arrows indicate that the waters will be displaced or replaced by inhibitors. (c) Inhibitor 1a (teal sticks) contains a chlorine atom (sphere) and displaces all three water molecules, resulting in an unfavorable enthalpic contribution, a relatively small $\Delta H_{b,lig}$, and a larger entropic gain [159]. (d) Inhibitor 4 (teal sticks) with a guanidino group (spheres) displaces two water molecules and forms a strong salt bridge with D189, yielding a considerably more favorable $\Delta H_{b,lig}$ than inhibitor 1a. The third water (of the three apo waters) is not displaced (smaller entropic gain) just shifted from the purple (apo) to the red (holo) position. (e) The inhibitor-bound (teal spheres) structure of the human immunodeficiency virus (HIV-1) protease enzyme (gray cartoon, PDB code 3spk [160]). The black dotted rectangle borders the active site featured in f. f) an HIV-1 protease (gray cartoon) inhibitor, tipranavir (teal sticks) bound to the active site of the enzyme. The positions of two water molecules in the apoenzyme (PDB code 1g6l [161]) are shown as purple spheres. The isosteric oxygen atoms of the inhibitor replace both water molecules upon binding and form favorable interactions with residues D25 and I50.

solvation effects (Section 2) based on the molecular surface area, the fast estimation of the $\Delta G_{b,wat}$ of a certain water molecule or the corresponding hydration site of the target surface remains a challenge. The use of molecular volume (instead of the surface) proved to be effective [7] for the estimation of $\Delta G_{b,wat}$ for a target atom (hydration site). This approach of Stouten et al. uses a fragmental volume of the atom and a set of atomic solvation parameters for six different atom types. The solvation term was implemented in the popular docking tool AutoDock [69,137,189] and widely applied, yielding realistic target-ligand complexes during the fast docking calculations without an explicit water model [7].

While implicit water models provide fast structural solutions, explicit water models are appropriate for calculating

Table 2. Calculation of thermodynamics of target-water and target–ligand binding.

Method name	Features	Calculated quantity	Water model	Year of publication
Molecular mechanics level				
3D-RISM [130]	Produces an average water distribution around the solute for calculation of thermodynamic parameters.	$\Delta G_{b,wat}$	Explicit	1999
Barillari et al [34]	Double decoupling method with replica exchange thermodynamic integration in Monte-Carlo simulations.	$\Delta G_{b,wat}$	Explicit	2007
WaterMap [120]	Hydration sites are produced by MD simulations, and an IST analysis is performed.	$\Delta G_{b,wat}$	Explicit	2008
SZMAP [36]	Explicit water probe surrounded by implicit water model.	$\Delta G_{b,wat}$, $\Delta S_{b,wat}$	Hybrid	2015
Quantum mechanics level				
Nikitina et al. [174]	Semi-empirical, PM3. Predicted waters.	$\Delta H_{b,lig}$	Explicit	2004
Raha et al. [175]	Semi-empirical PM3 interaction energy combined with other MM terms.	$\Delta G_{b,lig}$, $\Delta H_{b,lig}$	Implicit	2005
Nikitina et al. [176]	Semi-empirical, PM3. Predicted waters and COSMO.	$\Delta H_{b,lig}$	Hybrid	2006
Fanfrlik et al. [177]	Semi-empirical PM6-DH2-M-C level enthalpy, combined with rigid rotor harmonic oscillator entropy, and ligand desolvation calculation. COSMO and SMD.	$\Delta G_{b,lig}$, $\Delta H_{b,lig}$	Implicit	2010
Dobes et al. [178]	Semi-empirical PM6-DH2-M-C level enthalpy combined with rigid rotor harmonic oscillator entropy, and ligand desolvation calculation. COSMO and SMD.	$\Delta G_{b,lig}$, $\Delta H_{b,lig}$	Implicit	2011
Gonzalez et al. [179]	Semi-empirical PM7.	$\Delta H_{b,lig}$	Explicit (crystallographic)	2017
Ehrlich et al. [180]	HF-3c electronic energy, DFTB3-3D thermostatic correction with C-RS solvation free energy calculation. Crystallographic waters + COSMO-RS.	$\Delta G_{b,lig}$	Hybrid	2017
Hylsová et al. [181]	DFT-D3 and PM6-D3X4-C levels in combination.	$\Delta G_{b,lig}$	Explicit (crystallographic and predicted)	2017
Pecina et al. [182]	Two levels of QM calculations to calculate interaction energy: semi-empirical PM7 and DFTB3-D3H4X.	$\Delta G_{b,lig}$	Explicit (crystallographic)	2018
Horváth et al. [183]	Cuts out the target-ligand binding interface from the system and performs semi-empirical PM7 level calculations with the MOZYME linear scaling. Predicted waters + COSMO.	$\Delta H_{b,lig}$	Hybrid	2019
QMH-L [53]	Semi-empirical 1SCF PM7 level calculation on MM-level preoptimized complexes. A descriptor of elemental composition. Predicted waters + COSMO.	$\Delta G_{b,lig}$ $\Delta H_{b,lig}$	Hybrid	2023
DOX_BDW [184]	Semi-empirical PM7 and DFT level. Predicted waters + COSMO.	$\Delta G_{b,lig}$	Hybrid	2023
Szél et al. [35]	Semi-empirical 1SCF PM7 level calculation on MM-level preoptimized complexes. Predicted waters + COSMO.	$\Delta H_{b,wat}$	Hybrid	2024
Lindahl et al. [185]	DFT single point energies, def2-TZVP basis set, ΩB97X-D functional for dispersion correction. SMD.	$\Delta G_{b,lig}$	Implicit	2024
SQM2.20 [186]	Semi-empirical PM6-D3H4X and MM-level preoptimized complexes. COSMO2 + some crystallographic waters.	$\Delta G_{b,lig}$	Implicit	2024
Knowledge-based (AI)				
Δ_{vina}XGB [187]	Calculates Vina scores for water molecules.	$pK_{d,lig}$	Crystallographic waters	2019
DeepWATsite [39]	Uses WATsite to calculate thermodynamics of waters at grid points.	$\Delta H_{b,lig}$	Explicit	2020
ECIFGraph:HM-Holo-Apo [41]	Trained on experimental data with an RMSD-based weighting.	$\Delta G_{b,lig}$	Explicit	2024

thermodynamic and kinetic properties [190,191] that are often combined with MD. For example, the inhomogeneous fluid solvation theory (IST [192]) calculates the solvation free energy based on the solute–solvent interaction and reorganization in solvent–solvent interactions. The theory is based on fixing the solute in space, and then a solvent density fluctuation is observed around it. Additionally, calculations on the displacement of surface water molecules and the rearrangement of water networks upon ligand binding are also possible with IST [18]. WaterMap [120] applies IST and MD simulations, predicting hydration sites and their corresponding thermodynamic properties in binding pockets. After a short MD simulation, an occupancy-based clustering of hydration sites is performed and submitted to an IST analysis resulting $\Delta G_{b,wat}$. Applications of WaterMap [120] are surveyed in a recent review [145], along with some limitations: flexible-binding sites and buried hydration sites might not be mapped adequately, considerable computational costs, water displacement energy calculations require a reference state, and electrostatic contributions are neglected during calculations. More advanced techniques must be involved for calculating $\Delta G_{b,lig}$ as WaterMap does not work as a standalone scoring function [145,193]. The grid inhomogeneous solvation theory (GIST) [194] is a modification of IST, and treats the solute and its surrounding solvent in a three-dimensional grid. GIST accounts for hydration sites separately and considers individual water positions and the corresponding binding thermodynamics [194]. While GIST calculations are accurate for the first hydration layer (static sub-nets) around the solute, they are less successful for loosely bound hydration layers [195] (dynamic sub-nets). The solute is restrained to only one conformation [196], and therefore, its flexibility is neglected unless multiple calculations are performed [196].

Replica exchange thermodynamic integration and Monte-Carlo simulations with a double decoupling method were applied for the calculation of $\Delta G_{b,wat}$ of individual water molecules binding to apoprotein structures by Barillari et al. [34]. The method performs a functional classification (Figure 1(a)) of water molecules. The distribution plots (Figure 4) show a remarkable shift of the $\Delta G_{b,wat}$ of conserved waters toward the negative values if compared with the more positive replaced ones. SZMAP [36], a grid-based method places an explicit water molecule at grid points in the binding site and represents the rest of the solvent with a Poisson-Boltzmann implicit solvent model. The difference in binding thermodynamics between the explicit water and either a vacuum or an indifferent probe is calculated in the next step. It was shown that conserved and displaced (Figure 1(a)) water molecules can be distinguished based on their $\Delta S_{b,wat}$ [36]. The methods of Barillari et al. and SZMAP are useful for designing ligands that can displace thermodynamically unfavorable water molecules from binding sites. Other grid-based methods including WaterFlap [132], 3D-RISM [130], HydraMap v.2 [144] are also available for the calculation of $\Delta G_{b,wat}$. HydraMap calculations are based on statistical potentials of protein–water interactions derived from experimental protein structures.

Computational calculation of binding kinetics data is challenging, but it is an emerging approach in drug design [197]. Kinetics parameters such as k_{on} and k_{off} were calculated using explicit solvent MD simulations for cyclodextrin host-guest complexes with reasonable agreement with experimental data [198], and a distinction could be made between entropy and enthalpy-driven [198] binding events. The τ was calculated using τ-random acceleration molecular dynamics with explicit solvent [199] for a series of heat shock protein (HSP90α) inhibitors and a good correlation ($r^2 = 0.86$) with experimental τ data was achieved for 78% of the ligands [199].

4.2. QM-level methods

The neglect of electronic effects is a major limitation [200] of common MM-based approaches described in the previous section. For example, polarization is rarely considered in thermodynamic/kinetic stability, which is disturbing considering that the large dipole moment of water predestinates its involvement in such effects. Quantum mechanics (QM) offers a more accurate description of the electronic effects of molecular interactions. The polarizability of the electron system at water molecules [35,53] and even exotic π stacking interactions of heterocyclic ring systems [201] are handled better by QM than MM methods. While the structural optimization of target-ligand complexes with explicit hydration is still time-consuming purely at the QM level, the calculation of target-ligand binding thermodynamics (scoring) is possible using QM [28,202]. As a hybrid solution, QM/MM methods have been developed and applied, and their limitations are well-described [203,204].

Recently, there has been an increase in fast-scoring techniques that use a (fragmented) [183] structural single point as input. Open-source semi-empirical QM programs like Mopac [205] can be applied very efficiently [206] for such calculations

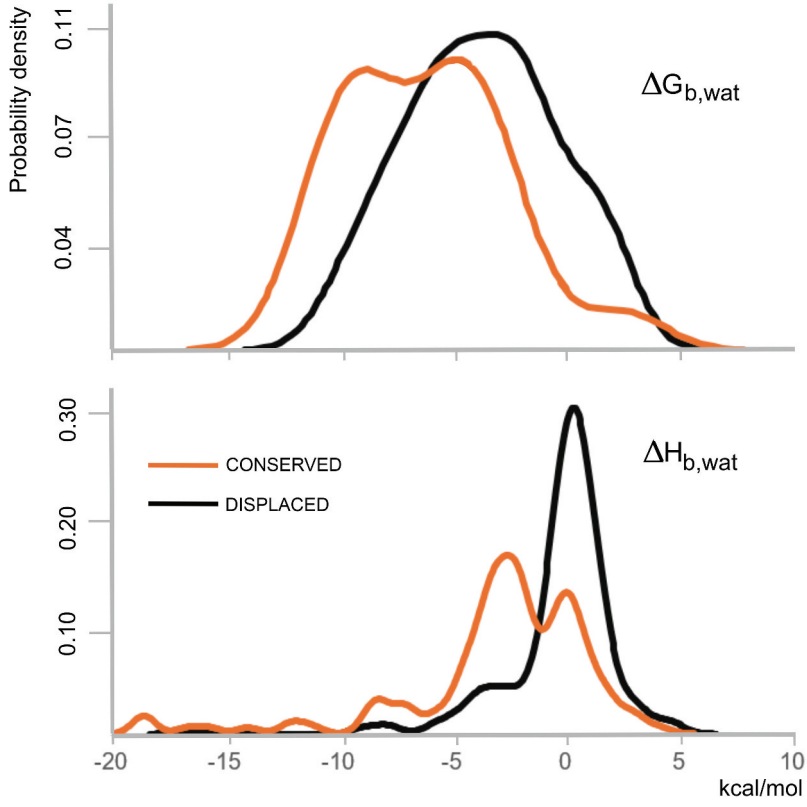

Figure 4. The distribution plots of $\Delta G_{b,wat}$ and $\Delta H_{b,wat}$ values of conserved and displaced water molecules. The $\Delta G_{b,wat}$ curves were re-drawn from an MM-based study [34]. The $\Delta H_{b,wat}$ curves were re-drawn using the data based on QM calculations [35]. Adapted from [34] and [35] with permission of the American Chemical Society.

in combination with MM/MD programs [146] that provide the structural input. For example, QMH-L [53] uses an MM-level optimized structure as an input complemented with a hybrid solvation model of MobyWat [50] predicted explicit water molecules and COSMO [207] to calculate $\Delta H_{b,lig}$ and $\Delta G_{b,lig}$. The calculations were performed at the semi-empirical PM7 QM level, and a descriptor was introduced accounting for the entropic contributions. On 43 test systems, the QMH-L approach predicted binding affinity with a root mean squared error of 0.94 kcal/mol and an r^2 of 0.6 [53]. Mopac-based single-point calculations were also introduced to calculate $\Delta H_{b,wat}$, and subsequent classification (Figure 1(a)) of water molecules in ligand binding [35]. An enthalpic threshold was specified to distinguish between conserved and displaced water molecules in binding pockets of important druggable targets. The conserved water molecules of the binding pocket of the human immunodeficiency virus (HIV-1) protease, and the ion channel of influenza A virus were identified in agreement with experiments [35]. Interestingly, the distribution of the QM-calculated $\Delta H_{b,wat}$ values [35] (Figure 4) is very similar to that of the MM-calculated $\Delta G_{b,wat}$ values of the previous study [34] discussed above. The similarity in the shift between the probability density curves of conserved and displaced waters may indicate that $\Delta H_{b,wat}$ is a determining factor of classification between the two groups (Figure 4).

The ligand is usually not involved in the hydration thermodynamics (Equation 3) calculators ($\Delta H_{b,wat}$ and $\Delta G_{b,wat}$ in Table 2). However, some of these calculators aim to predict the role of individual waters in target-ligand binding taking only the unliganded target structure as an input. This is obvious that the binding of different ligands results in different hydration networks in the target–ligand interface seriously limiting the applicability of such target-based methods in decision-making for an individual water molecule. For an accurate decision, the involvement of the ligand and an MD simulation with all three partners (target, ligand, waters) would be necessary to solve this chicken-and-egg problem. The refinements can be performed for the best candidates. However, without such refinements, it is more realistic to emphasize that available target-based methods listed in Table 2 offer statistical help for decision-making, in this sense, they can provide a probability that the water molecule belongs to conserved or displaced classes as it was concluded in the above studies [34,35] (Figure 4).

Another method, SQM2.20 [186] combines a gas–phase interaction energy, solvation free energy, and entropic components, like the conformational change of the ligand, handling the interaction interface at the semi-empirical PM6 QM level. However, the method considers explicit water molecules only for some systems. The SQM2.20 method achieved an r^2 of 0.69 [186] on 164 systems, similar to the above method QMH-L. Thus, an r^2 of 0.6–0.7 is realistic for such approaches. DOX_BDW [184] is another QM-based method distinguishing displaced/conserved water molecules in ligand binding affinity prediction. The method decomposes the binding interface into layers where different levels of theory are applied. Explicit interface water molecules (predicted with a grid-based approach) were involved in the calculations. Scoring with DOX_BDW achieved a correlation of a Spearman r of 0.66–0.85 with experimental affinity data, for different test sets [184]. Lindahl and Friedman applied [185] single-point density functional theory (DFT) QM calculations and an energy decomposition analysis to investigate the impact of protein chain selection on calculated $\Delta G_{b,lig}$ if multiple chains (models) are present in a PDB structure. The authors conclude that due to the deviations of experimental structures, multiple protein chains should be considered in calculating $\Delta G_{b,lig}$ [185].

4.3. Knowledge-based approaches

Comprehensive reviews of Kairys et al [28] and by Harren et al [208] are available on the recent progress of knowledge-based (machine learning, ML) scoring functions predicting ligand-binding affinity. For example, Δ_{vina}XGB [187] incorporates explicit water molecules in an ML scoring function. In the training set of the scoring function, protein-ligand complexes with target-bound water molecules were involved. Experimental water molecules within 3.5 Å of the target and with a negative Vina score [209] were kept in the training set of 3257 complexes. The validation set consisted of 315 complexes. The water-related scoring features were the number of bridging waters and Vina scores between water-target and water-ligand. On the CASF-2016 benchmark set the method performed with a Pearson r of 0.796 and an error of 1.32 in pK_d units [187]. DeepWATsite [39] is based on a convolutional neural network that uses protein, ligand density, and hydration as inputs. These density data are then distributed in a 3D grid filling the binding site of the target. The MD-based WATsite [126] predictions produced the hydration data. The thermodynamic data generated by WATsite were included in the grid points. The ECIFGraph … etc. method [41] relies on water networks in apo and holo-structures constructed with HydraMap [150]. A graph is formed by protein, ligand, and water atoms as nodes, the nodes are connected by edges representing covalent or hydrogen bonds. The difference between the holo and apo graphs is evaluated by a graph transformer module. A combined ML/QM approach with an implicit Poisson-Boltzmann water model [210] showed a yet modest predictive performance of binding affinity when compared to benchmark methods [202]. The modest performance might be explained by the critical point-based approach in electron density to represent the target–ligand interactions or the lack of the calculation of the entropic contribution of the ligand [202].

5. Conclusions

The optimization of target–ligand interactions is a central question of drug design, and for this purpose, a precise understanding of the coupled modulatory mechanisms of water is essential. The various local topographies of the binding partners correspond to different levels of networking and mobility in the hydration spheres. This results in rather controversial roles of water that can stabilize or destabilize the target–ligand interactions. In addition, water acts as a surrounding medium, making complex formation possible, but simultaneously acts as an individual partner mediating binding

between protein and ligand. It continuously exchanges and migrates between the bulk and the solute (surface or interface) regions.

The present review discusses why the above features have rendered water a problematic factor in structure-based drug design. The problems of experimental determination of water structure were shortly sketched. We enumerated the physicochemical and knowledge-based (AI) methods offering practical answers to the structural challenges. A set of good practices was collected that can quantify the contribution of individual water molecules to the thermodynamic stability of the target-ligand complexes. We gave an overview of the use of water molecules in computational docking and recent scoring methods for calculating target-ligand complex stability at MM and QM levels.

6. Expert opinion

Structural biochemists and drug designers have long considered water a problematic factor. Indeed, it is rather difficult to account for the effects of water due to its small size, large mobility in bulk, and other unique physicochemical characteristics that render this molecule so important on our planet. The particular problems with water can be further specified for the case of target-ligand binding, the heart of structure-based drug design, as follows.

(1) Explanation of water effects very often requires both thermodynamic and kinetic approaches also accounting for the topography [57] (local surroundings) of a water molecule.
(2) The topographical issues become even more complex for large water networks (on the target surfaces or in PPI interfaces) that are often very dynamic and difficult to handle (Section 2). The calculation of the effects of individual waters in small, static sub-nets is relatively well-described (Section 4).
(3) The competition between the ligand and water molecules (Figure 1(a)) imposes a 'chicken-and-egg' challenge and limits the applicability of computational docking (Section 3) and thermodynamic classification (Section 4) methods.
(4) The limitations of experimental structure determination techniques (Section 2).
(5) Lack of experimental methods for measuring the effects of individual water molecules on the stability of target-ligand complexes (Section 4).

With advances in rapid computational algorithms and a better understanding of the physicochemical machinery of complex formation, theoretical approaches have resulted in elegant and cost-effective tools and tricks that address the above challenges. For example, it can be foreseen how conserved water molecules can bridge or fill void spaces in the target–ligand interface, and it is even possible to incorporate the water into a new, improved ligand in the form of an isosteric functional group (Section 4). Thus, in the hands of a skilled (and lucky) drug designer familiar with the (combination of) good practices presented in the previous Sections, water can easily become a Jolly Joker.

Despite the amazing developments described in the present review, adding water to target-ligand binding remains one of the greatest challenges affecting various fields of drug design [211]. Although physicochemical methods offer reliable hydration structures in many cases, they are also limited by the lack of modeling electronic contributions to H-bonds, as commonly used MM force fields rely on simple Lennard-Jones and Coulomb terms. Hopefully, this limitation will be overcome by polarizable force fields and better water models [212–215]. Further improvement of accuracy can be expected, especially in the case of MD-based methods applying explicit water models. While AI-based emerging technologies have much to offer, serious concerns were raised about their applicability in drug design [91,216] and computational docking [217]. In the case of ligand docking and water structure prediction, the weakest point of AI-based technologies is probably the training set. The chemical universe [218] offers enormous libraries of different ligand molecules (fragments) for HTS campaigns that would require teaching an astronomical number of target–ligand interaction situations, which is impossible to undertake (because of a lack of appropriate training data). Similarly, the number of different target/ligand–water interaction situations is also tremendous. Thus, the presence of the ligand would impose an intractable challenge for AI, and a combination of physicochemical approaches with AI would become necessary. However, the AI-based calculation of the water structure of common protein or DNA targets may be a realistic alternative for further improvements. Furthermore, the uncertainty of ligand and water positions is a rather annoying feature of many experimental structures, and it will persistently hamper the assembly of proper structural training sets for AI. Notably, physicochemical (first principle) methods do not have such limitations and can be used even to correct structural errors in experimental or AI-derived structures [219–221].

Besides the expected further improvement of the emerging machine learning techniques, physicochemical methods remain indispensable for calculating water-mediated interactions. The target-based predicting of the role of water in target-ligand binding without the ligand is a chicken-and-egg problem most precisely answered using MD simulations in which all three partners (target, ligand, and waters) are present. Fast QM-based scoring schemes will increase the accuracy of the calculation of drug binding affinity. The further development of such physicochemical methods and their combination with machine learning [222] will likely increase the success rate of drug design projects in the next decades. By overcoming the technical pitfalls of drug design, water transforms from a frustrating challenge into a handy tool in fine-tuning drug–target interactions.

Acknowledgments

The authors thank Digital Government Development and Project Management Ltd. for providing the authors with access to the Komondor HPC facility based in Hungary.

Funding

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The authors are also supported by the National Research, Development and Innovation Office (PharmaLab, RRF-2.3.1-21-2022-00015) and Digital Government Development and Project Management Ltd.

Declaration of interest

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosure

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Zsidó BZ, Hetényi C. The role of water in ligand binding. Curr Opin Struct Biol. 2021;67. doi: 10.1016/j.sbi.2020.08.002
2. Zsidó BZ, Bayarsaikhan B, Börzsei R, et al. The advances and limitations of the determination and applications of water structure in molecular engineering. Int J Mol Sci. 2023;24:11784. doi: 10.3390/ijms241411784
3. Bodnarchuk MS. Water, water, everywhere… It's time to stop and think. Drug Discov Today. 2016;21(7):1139–1146. doi: 10.1016/j.drudis.2016.05.009
4. Ladbury JE. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. Chem Biol. 1996;3(12):973–980. doi: 10.1016/S1074-5521(96)90164-7
5. Bucher D, Stouten P, Triballeau N. Shedding light on important waters for drug design: simulations versus grid-based methods. J Chem Inf Model. 2018;58:692–699. doi: 10.1021/acs.jcim.7b00642
6. Homans S. Water, water everywhere — except where it matters? *drug discov Today*. Drug Discov Today. 2007;12(13–14):534–539. doi: 10.1016/j.drudis.2007.05.004
7. Stouten PFW, Frömmel C, Nakamura H, et al. An effective solvation term based on atomic occupancies for use in protein simulations. Mol Simul. 1993;10(2–6):97–120. doi: 10.1080/08927029308022161
8. Schmidtke P, Luque FJ, Murray JB, et al. Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design. J Am Chem Soc. 2011;133:18903–18910. doi: 10.1021/ja207494u
9. Jeszenoi N, Bálint M, Horváth I, et al. Exploration of interfacial hydration networks of target–ligand complexes. J Chem Inf Model. 2016;56(1):148–158. doi: 10.1021/acs.jcim.5b00638
10. Krimmer SG, Betz M, Heine A, et al. Methyl, ethyl, Propyl, butyl: futile but not for water, as the correlation of structure and thermodynamic signature shows in a congeneric series of thermolysin inhibitors. Chem Med chem. 2014;9:833–846. doi: 10.1002/cmdc.201400013
11. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. Nature. 2023;616:673–685. doi: 10.1038/s41586-023-05905-z
12. Potlitz F, Link A, Schulig L. Advances in the discovery of new chemotypes through ultra-large library docking. Expert Opin Drug Discov. 2023;18:303–313. doi: 10.1080/17460441.2023.2171984
13. Paggi JM, Pandit A, Dror RO. The art and science of molecular docking. Annu Rev Biochem. 2024;93:389–410. doi: 10.1146/annurev-biochem-030222-120000
14. Wéber E, Ábrányi-Balogh P, Nagymihály B, et al. Target-templated construction of functional proteomimetics using photo-foldamer libraries. Angew Chem. 2025;137. doi: 10.1002/ange.202410435
15. Mihalovits LM, Szalai TV, Bajusz D, et al. Exploring chemical spaces in the billion range: Is docking a computational alternative to DNA-Encoded libraries? J Chem Inf Model. 2024;64:8963–8979. doi: 10.1021/acs.jcim.4c00803
16. Sulimov VB, Kutov DC, Sulimov AV. Advances in docking. Curr Med Chem. 2020;26:7555–7580. doi: 10.2174/0929867325666180904115000
17. Katkova EV, Onufriev AV, Aguilar B, et al. Accuracy comparison of several common implicit solvent models and their implementations in the context of protein-ligand binding. J Mol Graph Model. 2017;72:70–80. doi: 10.1016/j.jmgm.2016.12.011
18. Balius TE, Fischer M, Stein RM, et al. Testing inhomogeneous solvation theory in structure-based ligand discovery. In: Proceedings of the National Academy of Sciences; 2017. Vol. 114. doi: 10.1073/pnas.1703287114
19. Lukac I, Wyatt PG, Gilbert IH, et al. Ligand binding: evaluating the contribution of the water molecules network using the fragment molecular orbital method. J Comput Aided Mol Des. 2021;35:1025–1036. doi: 10.1007/s10822-021-00416-3
20. Ratkova EL, Dawidowski M, Napolitano V, et al. Water envelope has a critical impact on the design of protein–protein interaction inhibitors. Chem Commun. 2020;56(31):4360–4363. doi: 10.1039/C9CC07714F
21. Huggins DJ, Marsh M, Payne MC. Thermodynamic properties of water molecules at a protein–protein interaction surface. J Chem Theory Comput. 2011;7:3514–3522. doi: 10.1021/ct200465z
22. Wong SE, Lightstone FC. Accounting for water molecules in drug design. Expert Opin Drug Discov. 2011;6:65–74. doi: 10.1517/17460441.2011.534452
23. Geschwindner S, Ulander J, Johansson P. Ligand binding thermodynamics in drug discovery: still a hot tip? J Med Chem. 2015;58:6321–6335. doi: 10.1021/jm501511f
24. Baranauskienė L, Petrikaitė V, Matulienė J, et al. Titration calorimetry standards and the precision of isothermal titration calorimetry data. Int J Mol Sci. 2009;10:2752–2762. doi: 10.3390/ijms10062752
25. Liu W, Jiang J, Lin Y, et al. Insight into thermodynamic and kinetic profiles in small-molecule optimization. J Med Chem. 2022;65:10809–10847. doi: 10.1021/acs.jmedchem.2c00682
26. Yu Z, van Veldhoven JPD, Louvel J, et al. Structure–affinity relationships (SARs) and structure–kinetics relationships (SKRs) of K_v 11.1 blockers. J Med Chem. 2015;58(15):5916–5929. doi: 10.1021/acs.jmedchem.5b00518
27. Kairys V, Baranauskiene L, Kazlauskiene M, et al. Binding affinity in drug design: experimental and computational techniques. Expert Opin Drug Discov. 2019;14:755–768. doi: 10.1080/17460441.2019.1623202
28. Kairys V, Baranauskiene L, Kazlauskiene M, et al. Recent advances in computational and experimental protein-ligand affinity determination techniques. Expert Opin Drug Discov. 2024;19(6):649–670. doi: 10.1080/17460441.2024.2349169
•• **An up-to-date review, and a good starting point to learn the advances in computational and experimental protein-ligand affinity determination techniques.**
29. Linkuvienė V, Talibov VO, Danielson UH, et al. Introduction of intrinsic kinetics of protein–ligand interactions and their implications for drug design. J Med Chem. 2018;61:2292–2302. doi: 10.1021/acs.jmedchem.7b01408
30. Nittinger E, Flachsenberg F, Bietz S, et al. Placement of water molecules in protein structures: from large-scale evaluations to single-case examples. J Chem Inf Model. 2018;58:1625–1637. doi: 10.1021/acs.jcim.8b00271
31. Liu C, Wrobleski ST, Lin J, et al. 5-cyanopyrimidine derivatives as a novel class of potent, selective, and orally active inhibitors of P38α MAP kinase. J Med Chem. 2005;48:6261–6270. doi: 10.1021/jm0503594
32. Wissner A, Berger DM, Boschelli DH, et al. 4-anilino-6,7-dialkoxyquinoline-3-carbonitrile inhibitors of epidermal growth factor receptor kinase and their bioisosteric relationship to the 4-anilino-

6,7-dialkoxyquinazoline inhibitors. J Med Chem. 2000;43:3244–3256. doi: 10.1021/jm000206a
33. Geschwindner S, Ulander J. The current impact of water thermodynamics for small-molecule drug discovery. Expert Opin Drug Discov. 2019;14:1221–1225. doi: 10.1080/17460441.2019.1664468
34. Barillari C, Taylor J, Viner R, et al. Classification of water molecules in protein binding sites. J Am Chem Soc. 2007;129(9):2577–2587. doi: 10.1021/ja066980q
•• **A statistical distinction between (individual) conserved and displaced water molecules based on their calculation of binding free energy values.**
35. Szél V, Zsidó BZ, Hetényi C. Enthalpic classification of water molecules in target–ligand binding. J Chem Inf Model. 2024;64:6583–6595. doi: 10.1021/acs.jcim.4c00794
36. Bayden AS, Moustakas DT, Joseph-McCarthy D, et al. Evaluating free energies of binding and conservation of crystallographic waters using SZMAP. J Chem Inf Model. 2015;55:1552–1565. doi: 10.1021/ci500746d
37. Zamanos A, Ioannakis G, Emiris IZ. HydraProt: a new deep learning tool for fast and accurate prediction of water molecule positions for protein structures. J Chem Inf Model. 2024;64:2594–2611. doi: 10.1021/acs.jcim.3c01559
38. Park S, Seok C. GalaxyWater-CNN: prediction of water positions on the protein structure by a 3D-Convolutional neural network. J Chem Inf Model. 2022;62:3157–3168. doi: 10.1021/acs.jcim.2c00306
39. Mahmoud AH, Masters MR, Yang Y, et al. Elucidating the multiple roles of hydration for accurate protein-ligand binding prediction via deep learning. Commun Chem. 2020;3:19. doi: 10.1038/s42004-020-0261-x
40. Huang P, Xing H, Zou X, et al. Accurate prediction of hydration sites of proteins using energy model with atom embedding. Front Mol Biosci. 2021;8. doi: 10.3389/fmolb.2021.756075
41. Qu X, Dong L, Luo D, et al. Water network-augmented two-state model for protein–ligand binding affinity prediction. J Chem Inf Model. 2024;64:2263–2274. doi: 10.1021/acs.jcim.3c00567
42. Nilsson A, Pettersson LGM. Perspective on the structure of liquid water. Chem Phys. 2011;389(1–3):1–34. doi: 10.1016/j.chemphys.2011.07.021
43. Némethy G, Scheraga HA. Structure of water and hydrophobic bonding in proteins. I. A model for the thermodynamic properties of liquid water. J Chem Phys. 1962;36:3382–3400. doi: 10.1063/1.1732472
44. Carugo O. Statistical survey of the buried waters in the protein data bank. Amino Acids. 2016;48:193–202. doi: 10.1007/s00726-015-2064-4
45. Chen X, Weber I, Harrison RW. Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures. J Phys Chem B. 2008;112:12073–12080. doi: 10.1021/jp802795a
46. Brysbaert G, Blossey R, Lensink MF. The inclusion of water molecules in residue interaction networks identifies additional central residues. Front Mol Biosci. 2018;5. doi: 10.3389/fmolb.2018.00088
47. Yoon HR, Park GJ, Balupuri A, et al. TWN-FS method: a novel fragment screening method for drug discovery. Comput Struct Biotechnol J. 2023;21:4683–4696. doi: 10.1016/j.csbj.2023.09.037
48. Schiebel J, Gaspari R, Wulsdorf T, et al. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. Nat Commun. 2018;9:3559. doi: 10.1038/s41467-018-05769-2
49. Kunstmann S, Gohlke U, Broeker NK, et al. Solvent networks tune thermodynamics of oligosaccharide complex formation in an extended protein binding site. J Am Chem Soc. 2018;140:10447–10455. doi: 10.1021/jacs.8b03719
50. Jeszenoi N, Horváth I, Bálint M, et al. Mobility-based prediction of hydration structures of protein surfaces. Bioinformatics. 2015;31:1959–1965. doi: 10.1093/bioinformatics/btv093
51. Hulovatyy Y, Chen H, Milenković T. Exploring the structure and function of temporal networks with dynamic graphlets. Bioinformatics. 2015;31:i171–i180. doi: 10.1093/bioinformatics/btv227
52. Kim H, Anderson R. Temporal node centrality in complex networks. Phys Rev E. 2012;85:026107. doi: 10.1103/PhysRevE.85.026107
53. Szél V, Zsidó BZ, Jeszenői N, et al. Target–ligand binding affinity from single point enthalpy calculation and elemental composition. Phys Chem Chem Phys. 2023;25(46):31714–31725. doi: 10.1039/D3CP04483A
54. Chandler D. Interfaces and the driving force of hydrophobic assembly. Nature. 2005;437:640–647. doi: 10.1038/nature04162
55. Southall NT, Dill KA, Haymet ADJ. A view of the hydrophobic effect. J Phys Chem B. 2002;106:521–533. doi: 10.1021/jp015514e
56. Cheng Y-K, Rossky PJ. Surface topography dependence of biomolecular hydrophobic hydration. Nature. 1998;392:696–699. doi: 10.1038/33653
57. Halle B, Daniel RM, Finney JL, et al. Protein hydration dynamics in solution: a critical survey. Philos Trans R Soc Lond B Biol Sci. 2004;359(1448):1207–1224. doi: 10.1098/rstb.2004.1499
•• **A fundamental paper on protein hydration dynamics discussing (among others) mobility and topography differences between bulk and surface-bound water molecules. A very comprehensive digest of experimental results.**
58. Venkatakrishnan AJ, Ma AK, Fonseca R, et al. Diverse GPCRs exhibit conserved water networks for stabilization and activation. Proc Natl Acad Sci U S A. 2019;116:3288–3293. doi: 10.1073/pnas.1809251116
59. Darby JF, Hopkins AP, Shimizu S, et al. Water networks can determine the affinity of ligand binding to proteins. J Am Chem Soc. 2019;141:15818–15826. doi: 10.1021/jacs.9b06275
60. Rabe von Pappenheim F, Aldeghi M, Shome B, et al. Structural basis for antibiotic action of the B1 antivitamin 2′-Methoxy-Thiamine. Nat Chem Biol. 2020;16:1237–1245. doi: 10.1038/s41589-020-0628-4
61. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Web and Social Media; 2009. Vol. 3. p. 361–362. doi: 10.1609/icwsm.v3i1.13937
62. Mehler EL, Solmajer T. Electrostatic effects in proteins: comparison of dielectric and charge models. Protein Eng Des Sel. 1991;4(8):903–910. doi: 10.1093/protein/4.8.903
63. Stachowski TR, Vanarotti M, Seetharaman J, et al. Water networks repopulate protein–ligand interfaces with temperature. Angew Chem Int Ed. 2022;61(31). doi: 10.1002/anie.202112919
64. Scott DE, Bayly AR, Abell C, et al. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. Nat Rev Drug Discov. 2016;15:533–550. doi: 10.1038/nrd.2016.29
65. Levitt M. Protein folding by restrained energy minimization and molecular dynamics. J Mol Biol. 1983;170:723–764. doi: 10.1016/S0022-2836(83)80129-6
66. Monroe J, Barry M, DeStefano A, et al. Water structure and properties at hydrophilic and hydrophobic surfaces. Annu Rev Chem Biomol Eng. 2020;11:523–557. doi: 10.1146/annurev-chembioeng-120919-114657
67. Ramstein J, Lavery R. Energetic coupling between DNA bending and base pair opening. Proc Natl Acad Sci USA. 1988;85(19):7231–7235. doi: 10.1073/pnas.85.19.7231
68. Hingerty BE, Ritchie RH, Ferrell TL, et al. Dielectric effects in biopolymers: the theory of ionic saturation revisited. Biopolymers. 1985;24:427–439. doi: 10.1002/bip.360240302
69. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785–2791. doi: 10.1002/jcc.21256
70. Lamm G. The Poisson–Boltzmann Equation. Rev Comput Chem. 2003;19:147–365.
71. Fogolari F, Brigo A, Molinari H. The Poisson–boltzmann equation for biomolecular electrostatics: a tool for structural biology. J Mol Recognit. 2002;15(6):377–392. doi: 10.1002/jmr.577
72. Feig M, Onufriev A, Lee MS, et al. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem. 2004;25:265–284. doi: 10.1002/jcc.10378
73. Born M. Volumen Und Hydratationswärme Der Ionen. Z für Phys. 1920;1(1):45–48. doi: 10.1007/BF01881023

74. Qiu D, Shenkin PS, Hollinger FP, et al. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate born radii. J Phys Chem A. 1997;101:3005–3014. doi: 10.1021/jp961992r
75. Still WC, Tempczyk A, Hawley RC, et al. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc. 1990;112:6127–6129. doi: 10.1021/ja00172a038
76. Skyner RE, McDonagh JL, Groom CR, et al. A review of methods for the calculation of solution free energies and the modelling of systems in solution. Phys Chem Chem Phys. 2015;17 (9):6174–6191. doi: 10.1039/C5CP00288E
77. Onufriev AV, Izadi S. Water models for biomolecular simulations. WIREs Comput Mol Sci. 2018;8(2). doi: 10.1002/wcms.1347
78. Levy Y, Onuchic JN. Water mediation in protein folding and molecular recognition. Annu Rev Biophys Biomol Struct. 2006;35 (1):389–415. doi: 10.1146/annurev.biophys.35.040405.102134
79. Bernal JD, Fowler RH. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. J Chem Phys. 1933;1:515–548. doi: 10.1063/1.1749327
80. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc. 1981;103:335–340. doi: 10.1021/ja00392a016
81. Berendsen HJC, Grigera RJ, Straatsma PT. The missing term in effective pair potentials. J Phys Chem. 1987;91(24):6269–6271. doi: 10.1021/j100308a038
82. Fu C, Xiao Y, Zhou X, et al. Insight into binding of endogenous neurosteroid ligands to the sigma-1 receptor. Nat Commun. 2024;15:5619. doi: 10.1038/s41467-024-49894-7
83. Bálint M, Jeszenoi N, Horváth I, et al. Systematic exploration of multiple drug binding sites. J Cheminform. 2017;9. doi: 10.1186/s13321-017-0255-6
84. Aguayo-Ortiz R, Dominguez L. Unveiling the possible oryzalin-binding site in the α-tubulin of *toxoplasma gondii*. ACS Omega. 2022;7:18434–18442. doi: 10.1021/acsomega.2c00729
85. van der Spoel D, van Maaren PJ, Berendsen HJC. A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. J Chem Phys. 1998;108(24):10220–10230. doi: 10.1063/1.476482

•• **A fundamental paper on the application of explicit water models.**

86. Kiss PT, Baranyai A. Sources of the deficiencies in the popular SPC/E and TIP3P models of water. J Chem Phys. 2011;134. doi: 10.1063/1.3548869
87. Guerard JJ, Arey JS. Critical evaluation of implicit solvent models for predicting aqueous oxidation potentials of neutral organic compounds. J Chem Theory Comput. 2013;9:5046–5058. doi: 10.1021/ct4004433
88. Merino F, Raunser S. Electron cryo-microscopy as a tool for structure-based drug development. Angew Chem Int Ed. 2017;56 (11):2846–2860. doi: 10.1002/anie.201608432
89. Berman HM, Battistuz T, Bhat TN, et al. The protein data bank. Acta Crystallogr D Biol Crystallogr. 2002;58:899–907. doi: 10.1107/S0907444902003451
90. de Beer S, Vermeulen N, Oostenbrink C. The role of water molecules in computational drug design. Curr Top Med Chem. 2010;10:55–66. doi: 10.2174/156802610790232288
91. Moore PB, Hendrickson WA, Henderson R, et al. The protein-folding problem: not yet solved. Science 2022;375:507–507. doi: 10.1126/science.abn9422
92. Adrian M, Dubochet J, Lepault J, et al. Cryo-electron microscopy of viruses. Nature. 1984;308:32–36. doi: 10.1038/308032a0
93. Frank J, Zhu J, Penczek P, et al. A model of protein synthesis based on cryo-electron microscopy of the E. Coli ribosome. Nature. 1995;376:441–444. doi: 10.1038/376441a0
94. Pintilie G, Zhang K, Su Z, et al. Measurement of atom resolvability in cryo-EM maps with Q-Scores. Nat Methods. 2020;17:328–334. doi: 10.1038/s41592-020-0731-1
95. Chari A, Stark H. Prospects and limitations of high-resolution single-particle cryo-electron microscopy. Annu Rev Biophys. 2023;52:391–411. doi: 10.1146/annurev-biophys-111622-091300
96. Shoemaker SC, Ando N. X-Rays in the cryo-electron microscopy era: structural biology's dynamic future. Biochemistry. 2018;57:277–285. doi: 10.1021/acs.biochem.7b01031
97. Weichenberger CX, Afonine PV, Kantardjieff K, et al. The solvent component of macromolecular crystals. Acta Crystallogr D Biol Crystallogr. 2015;71:1023–1038. doi: 10.1107/S1399004715006045
98. Wolff AM, Nango E, Young ID, et al. Mapping protein dynamics at high spatial resolution with temperature-jump X-Ray crystallography. Nat Chem. 2023;15:1549–1558. doi: 10.1038/s41557-023-01329-4
99. Biedermannová L, Schneider B. Hydration of proteins and nucleic acids: advances in experiment and theory. A review. Biochim et Biophys Acta (BBA) - Gener Subj. 2016;1860:1821–1835. doi: 10.1016/j.bbagen.2016.05.036
100. Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov. 2002;1:727–730. doi: 10.1038/nrd892
101. Kermani AA. A guide to membrane protein X-ray crystallography. FEBS J. 2021;288:5788–5804. doi: 10.1111/febs.15676
102. Niedzialkowska E, Gasiorowska O, Handing KB, et al. Protein purification and crystallization artifacts: the tale usually not told. Protein Sci. 2016;25:720–733. doi: 10.1002/pro.2861
103. Tsuchiya Y. Discrimination between biological interfaces and crystal-packing contacts. Adv Appl Bioinf Chem. 2008;2008:99. doi: 10.2147/AABC.S4255
104. Halle B. Biomolecular cryocrystallography: structural changes during flash-cooling. Proc Natl Acad Sci USA. 2004;101(14):4793–4798. doi: 10.1073/pnas.0308315101
105. DePristo MA, de Bakker PIW, Blundell TL. Heterogeneity and inaccuracy in protein structures solved by X-Ray crystallography. Structure. 2004;12(5):831–838. doi: 10.1016/j.str.2004.02.031
106. Blundell TL, Jhoti H, Abell C. High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov. 2002;1:45–54. doi: 10.1038/nrd706
107. Renaud J-P, Chari A, Ciferri C, et al. Cryo-EM in drug discovery: achievements, limitations and prospects. Nat Rev Drug Discov. 2018;17:471–492. doi: 10.1038/nrd.2018.77
108. Zsidó BZ, Börzsei R, Szél V, et al. Determination of ligand binding modes in hydrated viral ion channels to foster drug design and repositioning. J Chem Inf Model. 2021;61:4011–4022. doi: 10.1021/acs.jcim.1c00488
109. Finney JL, Eley DD, Richards RE, et al. The organization and function of water in protein crystals. Phil Trans R Soc Lond B Biol Sci. 1977;278:3–32. doi: 10.1098/rstb.1977.0029
110. Islam SA, Weaver DL. Molecular interactions in protein crystals: solvent accessible surface and stability. Proteins Struct Function Bioinf. 1990;8:1–5. doi: 10.1002/prot.340080103
111. Kossiakoff AA, Sintchak MD, Shpungin J, et al. Analysis of solvent structure in proteins using neutron D $_2$ OðH $_2$ O solvent maps: pattern of primary and secondary hydration of trypsin. Proteins Struct Function Bioinf. 1992;12:223–236. doi: 10.1002/prot.340120303
112. Chatake T, Fujiwara S. A technique for determining the deuterium/hydrogen contrast map in neutron macromolecular crystallography. Acta Crystallogr D Struct Biol. 2016;72:71–82. doi: 10.1107/S2059798315021269
113. Levitt M, Park BH. Water: Now you see it, Now you don't. Structure. 1993;1(4):223–226. doi: 10.1016/0969-2126(93)90011-5
114. Kono F, Kurihara K, Tamada T. Current status of neutron crystallography in structural biology. Biophys Physicobiol. 2022;19: e190009. doi: 10.2142/biophysico.bppb-v19.0009
115. Armstrong BD, Han S. Overhauser dynamic nuclear polarization to study local water dynamics. J Am Chem Soc. 2009;131:4641–4647. doi: 10.1021/ja809259q
116. Sabe VT, Ntombela T, Jhamba LA, et al. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: a review. Eur J Med Chem. 2021;224:113705. doi: 10.1016/j.ejmech.2021.113705
117. Tropsha A, Isayev O, Varnek A, et al. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nat Rev Drug Discov. 2024;23:141–155. doi: 10.1038/s41573-023-00832-0

118. Mullowney MW, Duncan KR, Elsayed SS, et al. Artificial intelligence for natural product drug discovery. Nat Rev Drug Discov. 2023;22(11):895–916. doi: 10.1038/s41573-023-00774-7
119. Nagra NS, van der Veken L, Stanzl E, et al. The company landscape for artificial intelligence in large-molecule drug discovery. Nat Rev Drug Discov. 2023;22:949–950. doi: 10.1038/d41573-023-00139-0
120. Abel R, Young T, Farid R, et al. Role of the active-site solvent in the thermodynamics of factor xa ligand binding. J Am Chem Soc. 2008;130:2817–2831. doi: 10.1021/ja0771033
121. Virtanen JJ, Makowski L, Sosnick TR, et al. Modeling the hydration layer around proteins: HyPred. Biophys J. 2010;99:1611–1619. doi: 10.1016/j.bpj.2010.06.027
122. Li Z, Lazaridis T. Computing the thermodynamic contributions of interfacial water. Methods Mol Biol. 2012;819:393–404.
123. Ross GA, Morris GM, Biggin PC, et al. Rapid and accurate prediction and scoring of water molecules in protein binding sites. PLOS ONE. 2012;7(3):e32036. doi: 10.1371/journal.pone.0032036e32036
• **WaterDock is a correctly documented and validated method that predicts and scores water molecules in protein binding sites.**
124. Sridhar A, Ross GA, Biggin PC, et al. Waterdock 2.0: water placement prediction for holo-structures with a Pymol Plugin. PLOS ONE. 2017;12(2):e0172743. doi: 10.1371/journal.pone.0172743
125. Cui G, Swails JM, Manas ES. SPAM: a simple approach for profiling bound water molecules. J Chem Theory Comput. 2013;9:5539–5549. doi: 10.1021/ct400711g
126. Hu B, Lill MA. Watsite: hydration site prediction program with PyMOL interface. J Comput Chem. 2014;35:1255–1260. doi: 10.1002/jcc.23616
127. López ED, Arcon JP, Gauto DF, et al. WATCLUST: a tool for improving the design of drugs based on protein-water interactions. Bioinformatics. 2015;31:3697–3699. doi: 10.1093/bioinformatics/btv411
128. Vedani A, Huhta DW. Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. J Am Chem Soc. 1991;113:5860–5862. doi: 10.1021/ja00015a049
129. Pitt WR, Goodfellow JM. Modelling of solvent positions around polar groups in proteins. Protein Eng Des Sel. 1991;4(5):531–537. doi: 10.1093/protein/4.5.531
130. Kovalenko A, Hirata F. Self-consistent description of a metal–water interface by the Kohn–sham density functional theory and the three-dimensional reference interaction site model. J Chem Phys. 1999;110:10095–10112. doi: 10.1063/1.478883
131. Schymkowitz JWH, Rousseau F, Martins IC, et al. Prediction of water and metal binding sites and their affinities by using the fold-X force field. Proc Natl Acad Sci USA. 2005;102(29):10147–10152. doi: 10.1073/pnas.0501980102
132. Baroni M, Cruciani G, Sciabola S, et al. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): theory and application. J Chem Inf Model. 2007;47:279–294. doi: 10.1021/ci600253e
133. Bui H-H, Schiewe AJ, Haworth IS. WATGEN: an algorithm for modeling water networks at protein–protein interfaces. J Comput Chem. 2007;28(14):2241–2251. doi: 10.1002/jcc.20751
134. Wei W, Luo J, Waldispühl J, et al. Predicting positions of bridging water molecules in nucleic acid–ligand complexes. J Chem Inf Model. 2019;59:2941–2951. doi: 10.1021/acs.jcim.9b00163
135. Rarey M, Kramer B, Lengauer T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. Proteins Struct Function Genet. 1999;34:17–28. doi: 10.1002/(SICI)1097-0134(19990101)34:1<17:AID-PROT3>3.0.CO;2-1
136. Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. In: Klebe G, editor. Virtual screening: an alternative or complement to high throughput screening? Vol. 20. Dordrecht: Springer; 2000. p. 171–190.
137. Österberg F, Morris GM, Sanner MF, et al. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins Struct Function Bioinf. 2002;46:34–40. doi: 10.1002/prot.10028
138. Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47:1739–1749. doi: 10.1021/jm0306430
139. Verdonk ML, Chessari G, Cole JC, et al. Modeling water molecules in protein–ligand docking using GOLD. J Med Chem. 2005;48(20):6504–6515. doi: 10.1021/jm050543p
140. Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. J Chem Inf Model. 2007;47:435–449. doi: 10.1021/ci6002637
141. Therrien E, Weill N, Tomberg A, et al. Docking ligands into flexible and solvated macromolecules. 7. Impact of protein flexibility and water molecules on docking-based virtual screening accuracy. J Chem Inf Model. 2014;54(11):3198–3210. doi: 10.1021/ci500299h
• **Fitted is a user-friendly software that can be applied for covalent and non-covalent docking. It takes care of target flexibility and explicit waters.**
142. Huang N, Shoichet BK. Exploiting ordered waters in molecular docking. J Med Chem. 2008;51:4862–4865. doi: 10.1021/jm8006239
143. van Zundert GCP, Rodrigues JPGLM, Trellet M, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol. 2016;428:720–725. doi: 10.1016/j.jmb.2015.09.014
144. Li Y, Zhang Z, Wang R. HydraMap v.2: prediction of hydration sites and desolvation energy with refined statistical potentials. J Chem Inf Model. 2023;63:4749–4761. doi: 10.1021/acs.jcim.3c00408
145. Kaczor AA, Zięba A, Matosiuk D. The application of WaterMap-guided structure-based virtual screening in novel drug discovery. Expert Opin Drug Discov. 2024;19:73–83. doi: 10.1080/17460441.2023.2267015
146. van der Spoel D, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701–1718. doi: 10.1002/jcc.20291
•• **The citation classic open-source GROMACS remains an indispensable tool for MD calculations with explicit water models.**
147. Deng J, Cui Q. Efficient sampling of cavity hydration in proteins with nonequilibrium grand canonical monte carlo and polarizable force fields. J Chem Theory Comput. 2024;20:1897–1911. doi: 10.1021/acs.jctc.4c00013
148. Nittinger E, Gibbons P, Eigenbrot C, et al. Water molecules in protein–ligand interfaces. Evaluation of software tools and SAR comparison. J Comput Aided Mol Des. 2019;33:307–330. doi: 10.1007/s10822-019-00187-y
149. Szalai TV, Bajusz D, Börzsei R, et al. Effect of water networks on ligand binding: computational predictions vs experiments. J Chem Inf Model. 2024;64:8980–8998. doi: 10.1021/acs.jcim.4c01291
150. Li Y, Gao Y, Holloway MK, et al. Prediction of the favorable hydration sites in a protein binding pocket and its application to scoring function formulation. J Chem Inf Model. 2020;60:4359–4375. doi: 10.1021/acs.jcim.9b00619
151. von Neumann J. The computer and the Brain;. New Haven & London: Yale University Press; 1958.
152. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–589. doi: 10.1038/s41586-021-03819-2
153. Dahlström KM, Salminen TA. Apprehensions and emerging solutions in ML-Based protein structure prediction. Curr Opin Struct Biol. 2024;86:102819. doi: 10.1016/j.sbi.2024.102819
•• **A well-balanced review of current machine learning methodologies of protein structure prediction with relevant remarks on the limitations of ligand binding and water structure.**
154. Biedermannová L, Černý J, Malý M, et al. Knowledge-based prediction of DNA hydration using hydrated dinucleotides as building blocks. Acta Crystallogr D Struct Biol. 2022;78:1032–1045. doi: 10.1107/S2059798322006234
155. Wei L, Chen Y, Liu J, et al. Cov_dox: a method for structure prediction of covalent protein–ligand bindings. J Med Chem. 2022;65:5528–5538. doi: 10.1021/acs.jmedchem.1c02007
156. Komarov IV, Bugrov VA, Cherednychenko A, et al. Insights into modeling approaches in chemistry: assessing ligand-protein binding thermodynamics based on rigid-flexible model molecules. Chem Rec. 2024;24. doi: 10.1002/tcr.202300276

157. Chen D, Li Y, Zhao M, et al. Effective lead optimization targeting the displacement of bridging receptor–ligand water molecules. Phys Chem Chem Phys. 2018;20(37):24399–24407. doi: 10.1039/C8CP04118K
158. Baum B, Muley L, Heine A, et al. Think twice: understanding the high potency of Bis(Phenyl)Methane inhibitors of thrombin. J Mol Biol. 2009;391:552–564. doi: 10.1016/j.jmb.2009.06.016
159. Baum B, Mohamed M, Zayed M, et al. More than a simple lipophilic contact: a detailed thermodynamic analysis of nonbasic residues in the S1 pocket of thrombin. J Mol Biol. 2009;390:56–69. doi: 10.1016/j.jmb.2009.04.051
160. Wang Y, Liu Z, Brunzelle JS, et al. The higher barrier of darunavir and tipranavir resistance for HIV-1 protease. Biochem Biophys Res Commun. 2011;412:737–742. doi: 10.1016/j.bbrc.2011.08.045
161. Pillai B, Kannan KK, Hosur MV. 1.9 Å X-Ray study shows closed flap conformation in crystals of tethered HIV-1 PR. Proteins: Struct Function Genet. 2001;43:57–64. doi: 10.1002/1097-0134(20010401)43:1<57:AID-PROT1017>3.0.CO;2-D
162. Zhang D, Meng Q, Guo F. Incorporating water molecules into highly accurate binding affinity prediction for proteins and ligands. Int J Mol Sci. 2024;25:12676. doi: 10.3390/ijms252312676
163. Thomaston JL, Polizzi NF, Konstantinidi A, et al. Inhibitors of the M2 proton channel engage and disrupt transmembrane networks of hydrogen-bonded waters. J Am Chem Soc. 2018;140:15219–15226. doi: 10.1021/jacs.8b06741
164. Wang J, Ma C, Fiorin G, et al. Molecular dynamics simulation directed rational design of inhibitors targeting drug-resistant mutants of influenza a virus M2. J Am Chem Soc. 2011;133:12834–12841. doi: 10.1021/ja204969m
165. Morningstar-Kywi N, Wang K, Asbell TR, et al. Prediction of water distributions and displacement at protein–ligand interfaces. J Chem Inf Model. 2022;62(6):1489–1497. doi: 10.1021/acs.jcim.1c01266
166. Barros EP, Ries B, Champion C, et al. Accounting for solvation correlation effects on the thermodynamics of water networks in protein cavities. J Chem Inf Model. 2023;63:1794–1805. doi: 10.1021/acs.jcim.2c01610
167. Copeland RA. The drug–target residence time model: a 10-year retrospective. Nat Rev Drug Discov. 2016;15:87–95. doi: 10.1038/nrd.2015.18
168. Schuetz DA, de Witte WEA, Wong YC, et al. Kinetics for drug discovery: an industry-driven effort to target drug residence time. Drug Discov Today. 2017;22(6):896–911. doi: 10.1016/j.drudis.2017.02.002
169. Shan Y, Kim ET, Eastwood MP, et al. How does a drug molecule find its target binding site? J Am Chem Soc. 2011;133:9181–9183. doi: 10.1021/ja202726y
170. Englert L, Biela A, Zayed M, et al. Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: binding of phosphonamidate to the S1′-pocket of thermolysin. Biochim et Biophys Acta (BBA) - Gener Subj. 2010;1800:1192–1202. doi: 10.1016/j.bbagen.2010.06.009
171. Iwase S, Xiang B, Ghosh S, et al. ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. Nat Struct Mol Biol. 2011;18:769–776. doi: 10.1038/nsmb.2062
172. Nittinger E, Schneider N, Lange G, et al. Evidence of water molecules—A statistical evaluation of water molecules based on electron density. J Chem Inf Model. 2015;55:771–783. doi: 10.1021/ci500662d
173. Crabtree MD, Shammas SL. Stopped-flow kinetic techniques for studying binding reactions of intrinsically disordered proteins. Methods Enzymol. 2018;611:423–457.
174. Nikitina E, Sulimov V, Zayets V, et al. Semiempirical calculations of binding enthalpy for protein–ligand complexes. Int J Quantum Chem. 2004;97(2):747–763. doi: 10.1002/qua.10778
175. Raha K, Merz KM. Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein–ligand complexes. J Med Chem. 2005;48:4558–4575. doi: 10.1021/jm048973n
176. Nikitina E, Sulimov V, Grigoriev F, et al. Mixed implicit/explicit solvation models in quantum mechanical calculations of binding enthalpy for protein–ligand complexes. Int J Quantum Chem. 2006;106(8):1943–1963. doi: 10.1002/qua.20943
177. Fanfrlík J, Bronowska AK, Řezáč J, et al. A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-Bonding: HIV-1 protease with 22 ligands. J Phys Chem B. 2010;114:12666–12678. doi: 10.1021/jp1032965
178. Dobeš P, Fanfrlík J, Řezáč J, et al. Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. J Comput Aided Mol Des. 2011;25:223–235. doi: 10.1007/s10822-011-9413-5
179. González R, Suárez CF, Bohórquez HJ, et al. Semi-empirical quantum evaluation of peptide – MHC class II binding. Chem Phys Lett. 2017;668:29–34. doi: 10.1016/j.cplett.2016.12.015
180. Ehrlich S, Göller AH, Grimme S. Towards full quantum-mechanics-based protein–ligand binding affinities. Chemphyschem. 2017;18:898–905. doi: 10.1002/cphc.201700082
181. Hylsová M, Carbain B, Fanfrlík J, et al. Explicit treatment of active-site waters enhances quantum mechanical/implicit solvent scoring: inhibition of CDK2 by new Pyrazolo[1,5-a]pyrimidines. Eur J Med Chem. 2017;126:1118–1128. doi: 10.1016/j.ejmech.2016.12.023
182. Pecina A, Brynda J, Vrzal L, et al. Ranking power of the SQM/COSMO scoring function on carbonic anhydrase II–inhibitor complexes. Chemphyschem. 2018;19(7):873–879. doi: 10.1002/cphc.201701104
183. Horváth I, Jeszenői N, Bálint M, et al. A fragmenting protocol with explicit hydration for calculation of binding enthalpies of target-ligand complexes at a quantum mechanical level. Int J Mol Sci. 2019;20:20. doi 10.3390/ijms20184384
184. Liu J, Wan J, Ren Y, et al. DOX_BDW: incorporating solvation and desolvation effects of cavity water into nonfitting protein–ligand binding affinity prediction. J Chem Inf Model. 2023;63:4850–4863. doi: 10.1021/acs.jcim.3c00776
185. Lindahl E, Friedman R. Exploring the impact of protein chain selection in binding energy calculations with DFT. Chemphyschem. 2024;25(24). doi: 10.1002/cphc.202400119
186. Pecina A, Fanfrlík J, Lepšík M, et al. SQM2.20: semiempirical quantum-mechanical scoring function yields DFT-Quality protein–ligand binding affinity predictions in minutes. Nat Commun. 2024;15:1127. doi: 10.1038/s41467-024-45431-8
187. Lu J, Hou X, Wang C, et al. Incorporating explicit water molecules and ligand conformation stability in machine-learning scoring functions. J Chem Inf Model. 2019;59:4540–4549. doi: 10.1021/acs.jcim.9b00645
188. Kleinjung J, Fraternali F. Design and application of implicit solvent models in biomolecular simulations. Curr Opin Struct Biol. 2014;25:126–134. doi: 10.1016/j.sbi.2014.04.003
189. Uehara S, Tanaka S. AutoDock-GIST: incorporating thermodynamics of active-site water into scoring function for accurate protein-ligand docking. Molecules. 2016;21:1604. doi: 10.3390/molecules21111604
190. van der Spoel D, Zhang J, Zhang H. Quantitative predictions from molecular simulations using explicit or implicit interactions. WIREs Comput Mol Sci. 2022;12(1). doi: 10.1002/wcms.1560
•• **A state-of-the-art review of classical simulations including explicit and implicit water models for generating quantitative predictions.**
191. Zhang J, Zhang H, Wu T, et al. Comparison of implicit and explicit solvent models for the calculation of solvation free energy in organic solvents. J Chem Theory Comput. 2017;13:1034–1043. doi: 10.1021/acs.jctc.7b00169
192. Lazaridis T. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. J Phys Chem B. 1998;102:3531–3541. doi: 10.1021/jp9723574
193. Matter H, Güssregen S. Characterizing hydration sites in protein-ligand complexes towards the design of novel ligands.

Bioorg Med Chem Lett. 2018;28:2343–2352. doi: 10.1016/j.bmcl.2018.05.061
194. Nguyen CN, Kurtzman Young T, Gilson MK. Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor Cucurbit[7]Uril. J Chem Phys. 2012;137:044101. doi: 10.1063/1.4733951
195. Hüfner-Wulsdorf T, Klebe G. Protein–ligand complex solvation thermodynamics: development, parameterization, and testing of GIST-Based solvent functionals. J Chem Inf Model. 2020;60:1409–1423. doi: 10.1021/acs.jcim.9b01109
196. Loeffler JR, Schauperl M, Liedl KR. Hydration of aromatic heterocycles as an adversary of π-stacking. J Chem Inf Model. 2019;59:4209–4219. doi: 10.1021/acs.jcim.9b00395
197. Decherchi S, Cavalli A. Thermodynamics and kinetics of drug-target binding by molecular simulation. Chem Rev. 2020;120:12788–12833. doi: 10.1021/acs.chemrev.0c00534
198. Tang Z, Chang CA. Binding thermodynamics and kinetics calculations using chemical host and guest: a comprehensive picture of molecular recognition. J Chem Theory Comput. 2018;14(1):303–318. doi: 10.1021/acs.jctc.7b00899
199. Kokh DB, Amaral M, Bomke J, et al. Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. J Chem Theory Comput. 2018;14:3859–3869. doi: 10.1021/acs.jctc.8b00230
200. Kar RK. Benefits of hybrid QM/MM over traditional classical mechanics in pharmaceutical systems. Drug Discov Today. 2023;28:103374. doi: 10.1016/j.drudis.2022.103374
201. Loeffler JR, Fernández-Quintero ML, Schauperl M, et al. STACKED – solvation theory of aromatic complexes as key for estimating drug binding. J Chem Inf Model. 2020;60:2304–2313. doi: 10.1021/acs.jcim.9b01165
202. Ginex T, Vázquez J, Estarellas C, et al. Quantum mechanical-based strategies in drug discovery: finding the pace to new challenges in drug design. Curr Opin Struct Biol. 2024;87:102870. doi: 10.1016/j.sbi.2024.102870
•• **A useful review of the present QM-level strategies in drug discovery discussing prospects of QM-tailored force fields and machine learning.**
203. Söderhjelm P, Aquilante F, Ryde U. Calculation of protein–ligand interaction energies by a fragmentation approach combining high-level quantum chemistry with classical many-body effects. J Phys Chem B. 2009;113:11085–11094. doi: 10.1021/jp810551h
204. Klähn M, Braun-Sand S, Rosta E, et al. On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. J Phys Chem B. 2005;109:15645–15650. doi: 10.1021/jp0521757
205. Řezáč J Stewart JJP. How well do semiempirical QM methods describe the structure of proteins? J Chem Phys. 2023;158(4):044118.
•• **The first modern open-source release of the evergreen semi-empirical QM software MOPAC is available for drug designers.**
206. Řezáč J, Hobza P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. J Chem Theory Comput. 2012;8:141–151. doi: 10.1021/ct200751e
207. Klamt A, Schüürmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans. 1993;2(5):799–805. doi: 10.1039/P29930000799
208. Harren T, Gutermuth T, Grebner C, et al. Revolutionizing peptide-based drug discovery: advances in the post-AlphaFold era. Wiley Interdiscip Rev Comput Mol Sci. 2024;14(1). doi: 10.1002/wcms.1693
209. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31:455–461. doi: 10.1002/jcc.21334
210. Isert C, Atz K, Riniker S, et al. Exploring protein–ligand binding affinity prediction with electron density-based geometric deep learning. RSC Adv. 2024;14(7):4492–4502. doi: 10.1039/D3RA08650J
211. Morales-Herrejón G, Mendoza-Figueroa HL, Martínez-Archundía M, et al. The importance of structural water in HDAC8 for correct binding pose applied for drug design of anticancer molecules. Anticancer Agents Med Chem. 2024;24:1109–1125. doi: 10.2174/0118715206299644240523054454
212. Lambros E, Paesani F. How good are polarizable and flexible models for water: insights from a many-body perspective. J Chem Phys. 2020;153. doi: 10.1063/5.0017590
213. Xiong Y, Izadi S, Onufriev AV. Fast polarizable water model for atomistic simulations. J Chem Theory Comput. 2022;18:6324–6333. doi: 10.1021/acs.jctc.2c00378
214. Jing Z, Liu C, Cheng SY, et al. Polarizable force fields for biomolecular simulations: recent advances and applications. Annu Rev Biophys. 2019;48:371–394. doi: 10.1146/annurev-biophys-070317-033349
215. Lemkul JA, Huang J, Roux B, et al. An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications. Chem Rev. 2016;116:4983–5013. doi: 10.1021/acs.chemrev.5b00505
216. Hann MM, Keserű GM. The continuing importance of chemical intuition for the medicinal chemist in the era of artificial intelligence. Expert Opin Drug Discov. 2025;20(2):137–140. doi: 10.1080/17460441.2025.2450785
217. Buttenschoen M, Morris GM, Deane CM. PoseBusters: AI-Based docking methods fail to generate physically valid poses or generalise to novel sequences. Chem Sci. 2024;15:3130–3139. doi: 10.1039/D3SC04185A
218. Cherkasov A. The 'big bang' of the chemical universe. Nat Chem Biol. 2023;19:667–668. doi: 10.1038/s41589-022-01233-x
219. Zhang Y, Vass M, Shi D, et al. Benchmarking refined and unrefined AlphaFold2 structures for hit discovery. J Chem Inf Model. 2023;63:1656–1667. doi: 10.1021/acs.jcim.2c01219
220. Gu S, Yang Y, Zhao Y, et al. Evaluation of AlphaFold2 structures for hit identification across multiple scenarios. J Chem Inf Model. 2024;64(9):3630–3639. doi: 10.1021/acs.jcim.3c01976
221. Yang Z, Zeng X, Zhao Y, et al. AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduct Target Ther. 2023;8:115. doi: 10.1038/s41392-023-01381-z
222. Valsson Í, Warren MT, Deane CM, et al. Narrowing the gap between machine learning scoring functions and free energy perturbation using augmented data. Commun Chem. 2025;8(1):41. doi: 10.1038/s42004-025-01428-y
• **An interesting approach to fill the philosophical gap between machine learning and physicochemical approaches.**