



## Review

# Applications of water molecules for analysis of macromolecule properties



Karolina Mitusińska, Agata Raczyńska, Maria Bzówka, Weronika Bagrowska, Artur Góra\*

Tunneling Group, Biotechnology Centre, Silesian University of Technology, Krzywoustego 8, Gliwice, Poland

## ARTICLE INFO

## Article history:

Received 15 October 2019  
Received in revised form 26 January 2020  
Accepted 1 February 2020  
Available online 12 February 2020

## Keywords:

Water molecules  
Water placement  
Macromolecule structure  
Transportation  
Water sites  
Tunnel detection

## ABSTRACT

Water molecules maintain proteins' structures, functions, stabilities and dynamics. They can occupy certain positions or pass quickly via a protein's interior. Regardless of their behaviour, water molecules can be used for the analysis of proteins' structural features and biochemical properties. Here, we present a list of several software programs that use the information provided by water molecules to: i) analyse protein structures and provide the optimal positions of water molecules for protein hydration, ii) identify high-occupancy water sites in order to analyse ligand binding modes, and iii) detect and describe tunnels and cavities. The analysis of water molecules' distribution and trajectories sheds a light on proteins' interactions with small molecules, on the dynamics of tunnels and cavities, on protein composition and also on the functionality, transportation network and location of functionally relevant residues. Finally, the correct placement of water molecules in protein crystal structures can significantly improve the reliability of molecular dynamics simulations.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Contents

1. Introduction	355
2. Software for protein hydration analysis	356
3. Software for water site detection and ligand binding analysis	358
4. Software for tunnel detection and transportation phenomena analysis	360
5. Summary and outlook	362
Funding	363
Conflict of interest	363
CRedit authorship contribution statement	363
References	363

## 1. Introduction

Life began to evolve in an aqueous milieu, and the unique properties of water determine the chemistry of all living organisms. Water is a ubiquitous and essential substance in cells, accounting for about 70% of their mass. It is not only the environment for biological processes, but also an integral part of them [1]. At a macromolecular level, water contributes to biomolecules' formation and their stability, dynamics and functions [2–4]. Water serves as a

reaction reagent or mediates ligand–protein and protein–protein interactions. Water molecules are small enough to penetrate a macromolecule's core, to stabilise its native structure and also to participate in processes occurring in the protein's core [5,6].

X-ray [7] and neutron diffraction [8] crystallography provide an insight into the spatial distribution of water molecules in the vicinity of biomolecular surfaces and confined regions such as active sites, pockets and cavities. Depending on the crystal quality, atomic resolutions can be achieved [9–11]. Protein structures deposited in the Protein Data Bank (PDB) [12] contain an abundance of information, i.e., alternative conformations of amino acid side chains and potential rearrangements of protein compartments. Information

\* Corresponding author.

E-mail address: [a.gora@tunnelinggroup.pl](mailto:a.gora@tunnelinggroup.pl) (A. Góra).

about water molecules' positions is usually incomplete or can be strongly influenced by experimental conditions. Therefore, it is unclear how closely the distribution of crystal water molecules resembles the native conditions of the biomolecule. Nuclear Magnetic Resonance (NMR) is useful for discovering the hydration properties of proteins, especially their dynamics. Unfortunately, this technique cannot provide information about the three-dimensional structure of the hydration sites, and its time scales are shorter by an order of magnitude than the residence times of water molecules [4,13].

The limitations of experimental methods can be overcome by computational techniques. *Ab initio* and DFT (Density Functional Theory) methods can be used for a precise description of a reaction mechanism, including the contribution of water [14,15]. Molecular dynamics (MD) and Monte Carlo (MC) simulations provide a detailed atomic description of a biomolecule and a solvent, along with their dynamics [16,17]. These simulations, however, do not tackle many equilibrium and long-time-scale kinetic properties [18].

The increasing awareness of the significant role of water molecules has given rise to a range of software focused on the analysis of water molecules' behaviour. Recent reviews focused on virtual screening strategies describe several docking software applications that are capable of utilising information related to water molecules [19–21]. This paper provides a review of the available computational methods that employ water molecules for the analysis of macromolecules' properties and structure dynamics. In the first part, we provide an overview of the techniques used for the prediction of water molecules' locations. The following chapter describes the water sites that may participate in ligand binding. Next, water molecules are analysed in terms of ligand transportation and the detection of tunnels and cavities. For all three chapters, a list of software along with their functionality and/or their characteristics and principles is provided. The last chapter comprises conclusions and general remarks, as well as perspectives on the further development of software.

## 2. Software for protein hydration analysis

Water molecules not only maintain the functions of proteins but also stabilise their native structure [13]. The presence of water molecules in proteins' internal cavities is conserved among homologous proteins families, as well as the key residues are [22]. It was shown, by reducing the amount of water during crystallisation [23] or by using mutants of particular proteins [24], that internal water molecules contribute to the structural folding and the stability of enzymes, ion channels, proton pumps and other macromolecules [25–27]. However, as we pointed out above, the experimental results are insufficient and can be inconsistent with each other [28]. The water molecules inside a protein's structure may also be distorted, and their position may depend on the orientation of a particular side chain. They may also be trapped inside a protein's cavity due to a process of large conformational changes.

The residence time of a water molecule buried in an internal cavity or trapped in a narrow cleft depends on its location and connectivity with the bulk solvent [29,30]. Fast minimisation and short equilibrium stages can provide insufficient or inaccurate solvation of the protein interior and can bias the results. Therefore, it is important to fill the internal cavities with water molecules precisely, prior to running MD simulations. Lengthening the minimisation and equilibration procedure can provide sufficient exchange of water molecules between the surroundings and the protein interior; however, it is strongly system-dependent. Application of software developed to place water molecules into a protein's cavities and its surroundings is proposed as an alternative strategy, especially for systems with large interior volumes, homology-modelled proteins or proteins with mutations introduced inside their cores (Table 1).

Three different methods (Fig. 1) have been implemented for the placement of water molecules in a protein's interior: i) based on the docking of water molecules, such as Dowser [31] and WaterDock [32], ii) based on the reference interaction site model (RISM), such as 3D-RISM [33], GAsol [34] and Placevent [35], and iii) based

**Table 1**  
List of available software to predict water molecules' positions and orientation.

Software	Testing set	Accuracy*	Remarks
<b>Docking-based</b>			
Dowser [30]	14 crystal structures of OppA; D- and K-channels of cytochrome c oxidase;	63%	<i>Not available</i>
Dowser+ [31]	Photosystem II	74%	<i>Not available</i>
Dowser++ [39]		85%	Dowser++ standalone <a href="#">link</a>
WaterDock [32]	14 crystal structures of OppA; HIV-1 protease; ribonuclease A; GluR2 ligand binding core; concanavalin A; glutathione S-transferase; carbonic anhydrase	88%	the code is available with the WaterDock 2.0 Pymol plugin: – <a href="#">link</a>
WaterDock 2.0 [47]	14 crystal structures of OppA; HIV-1 protease; GluR2 ligand binding core; bovine pancreatic trypsin; glutathione S-transferase; HSP90; PIM1; series of 184 BRD4-BD1 complexes; androgen receptor; casein kinase II; thrombin; carbonic anhydrase	91%	WaterDock 2.0 standalone <a href="#">link</a> WaterDock 2.0 Pymol plugin <a href="#">link</a>
<b>RISM-based</b>			
3D-RISM [33]	Alanine dipeptide; HIV-1 protease	–	<i>Not available</i>
GAsol [34]	HIV-1 protease; neuraminidase; bovine pancreatic trypsin; series of 184 BRD4-BD1 complexes	94.3%	BSD 3-clause license <a href="#">link</a>
Placevent [35]	HIV-1 protease; rotor ring of F-ATP synthase	water position error –0.5 Å	the code and tutorial <a href="#">link</a>
<b>Similarity-based</b>			
ProBis H2O [37]	Src kinase with bound bosutinib; human programmed death 1 (hPD) with ligand (hPD-L1); DNA Gyrase B; human matrix metalloproteinase (hMMP-1)	–	GUI – PyMOL integrated <a href="#">link</a>
PyWATER [36]	thrombin; trypsin; BPTI; bromodomain-containing protein 4; MHC class I proteins; class A β-lactamases	identified all crystallographic water molecules	GUI – PyMOL integrated <a href="#">link</a>

\*Accuracy was calculated based on the number and quality of identified crystallographic water molecules. The numbers were taken from original publications. Please note that there are some differences in the details of the accuracy measurements. Information about currently unavailable software is in *italics*.

on the assumption that internal water molecules are conserved among similar proteins' structures (PyWATER [36] and ProBiS H2O [37]).

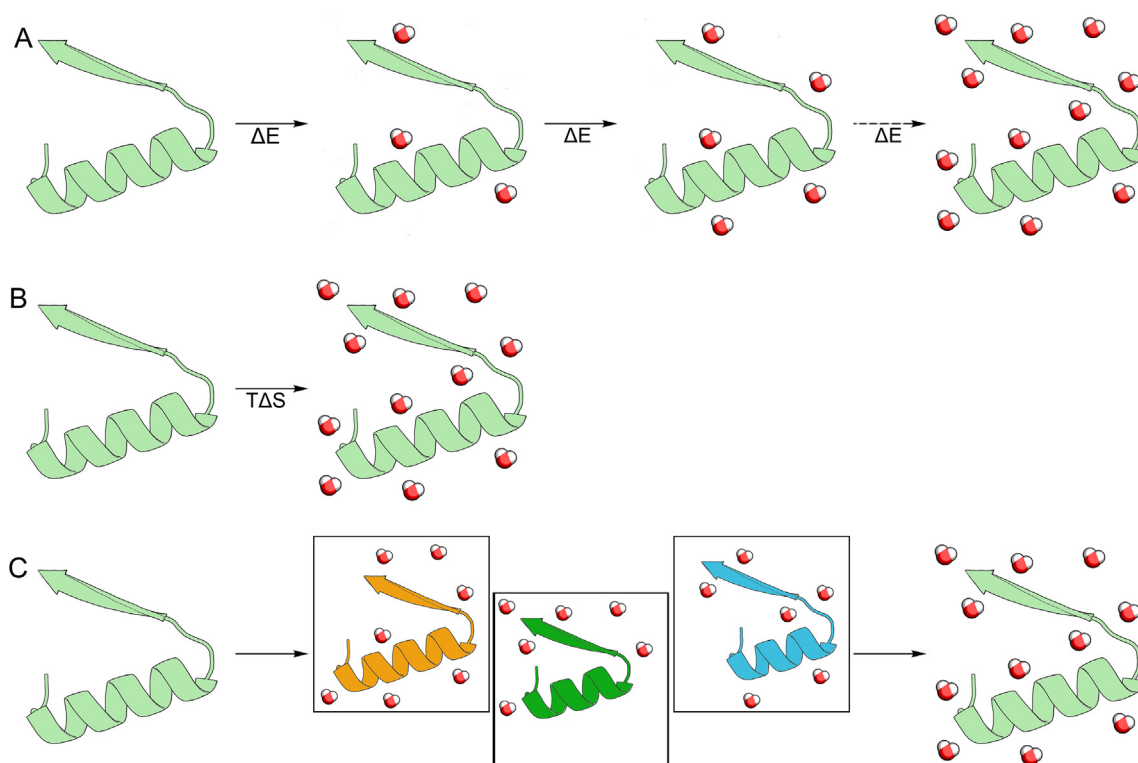
The docking-based methods assume that the protein structure is the target and the water molecule is the ligand. Both Dowser and WaterDock utilise the commonly available docking software - AutoDock Vina [38]. These methods are fast and provide accurate positioning of the water molecules determined by crystallography. The water molecule docking algorithms have improved over new software versions, and for the Dowser software 'generations', the average accuracy of their predictions have increased from 63% in Dowser, to 74% in Dowser+ and up to 85% of the water molecules in Dowser++ [39], when compared to high-resolution crystallographic structures. WaterDock software presented a higher accuracy of crystallographic water molecules' prediction than Dowser ++: it was 88% for the original WaterDock and 91% for WaterDock 2.0; however, it should be kept in mind that there were some differences in the details of the accuracy measurements described in the original publications [39]. Along with their ability to predict water molecules' positions, both WaterDock software releases are also able to determine if water molecules are displaced or ordered. WaterDock 2.0 comes with an easy-to-use PyMOL plugin.

The RISM theory is used for calculating the distribution of solvent molecules around a solute and has its roots in statistical, mechanical integral equation theories (IET) of liquids [34]. Due to the fact that the distribution calculated by 3D-RISM theory is continuous, it is difficult to directly examine specific solvent interactions, especially when they are numerous. 3D-RISM has been successfully used to locate water molecules in proteins as compared to experiment [40] and simulation [41], to calculate hydration free energies [42] and to predict fragment and drug positions [43]. The Placevent algorithm gave an average error for water molecules' positions of about 0.5 Å [35]. The GASol software, in which the 3D-RISM theory was combined with a genetic algorithm and a desirability function, showed the highest accuracy,

with 93.4% of the predicted water molecules within 2 Å from their crystallographic positions [34]. Generally, RISM-based methods for water molecules' prediction are slower than docking-based ones, and the computational time is system-size-dependent. However, they can be more accurate, especially for complex systems (e.g., metalloproteins, proteins equipped with large cavities or in complexes with nucleic acids) [44]. Moreover, it was shown that the RISM theory may break down in larger systems and systematically underestimates the partial molar volume (PMV) of amino acids [13].

As an alternative to the methods based on the physicochemical properties comes a simple similarity-based approach, implemented in PyWATER [36] and ProBiS H2O [37], which both superimpose crystallographic structures similar to the target protein and cluster the positions of conserved water molecules inside the protein cavities. ProBiS H2O is the first software that utilises the ProBiS algorithm [45] to perform local superimposition of the detected conserved water molecules. It also reduces the bias introduced by comparing similar protein structures or structures in different conformations than the query protein. PyWATER searches for similar structures using the PDB database [46]. The accuracy of such an approach strongly depends on the number, similarity and quality of related structures. Generally, ProBiS H2O gives fewer clusters with more tightly packed water molecules in comparison to PyWATER due to the clustering algorithms used (PyWATER uses hierarchical clustering, while ProBiS H2O uses a Python implementation of 3D-DBSCAN (Three Dimensional Density-Based Spatial Clustering of Applications with Noise)). In addition, PyWATER stores information on the degree of conservation of each water molecule cluster with related atom numbers of water oxygen atoms from the superimposed protein structures.

All the tools mentioned above provide relatively fast, intuitive and accurate modelling of the water molecules in low-quality crystal structures and thus provide a more accurate starting point for an MD or MC study. Their usage can be also recommended for *in*



**Fig. 1.** Strategies for placement of water molecules in the protein's interior. (A) Docking-based, (B) RISM-based, and (C) similarity-based strategies.

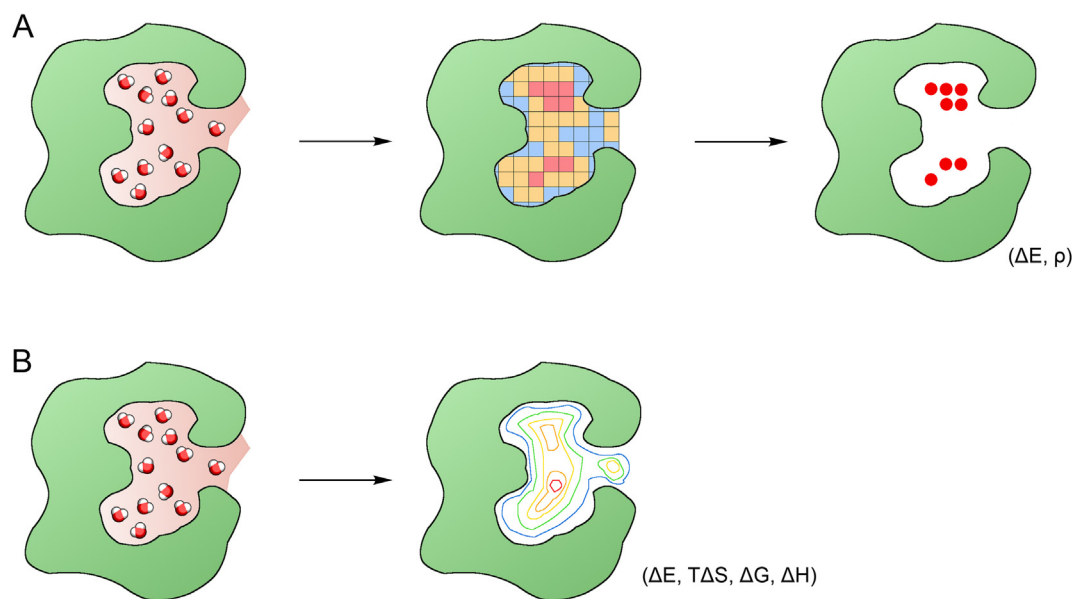
*silico* prepared mutant structures, where substitution of residues enlarges or significantly reshapes internal cavities. However, the user should keep in mind that none of these methods assume flexibility or large conformational changes in the target structure. Also, in the case of preparing a very demanding protein structure, i.e., an unrefined homology model or a sole representative of a particular protein family, the user should be encouraged to avoid similarity-based or docking-based methods and focus on the RISM-based software to properly sample the positions of water molecules.

### 3. Software for water site detection and ligand binding analysis

Studies of protein–ligand interactions are crucial for a better understanding of the mechanisms of biological processes and their regulation [48]. Water-mediated interactions were found in 85% of 392 analysed protein–ligand complexes. Structural and thermodynamic data indicate that water molecules mediate interactions between proteins and ions, substrates, cofactors, inhibitors and other macromolecules [49,50]. Water molecules are placed methodically within the surroundings and inside the protein, and display a particular structure characterised by the presence of hydration or water sites – regions of high-water density. They act as locations that attract water and can be used to describe water behaviour around chemical molecules [7,51,52]. The hydration/dehydration balance is relevant for protein–ligand formation and binding affinity, involving both entropic and enthalpic contributions [53,54]. During the binding process, water molecules can either be displaced or conserved, bridging the protein–ligand interactions in the latter case [55]. The presence of water molecules in protein binding sites may imply different effects on the energy, entropy and enthalpy of the system. Depending on the situation, such effects may be favourable or unfavourable. For example, in a case where water molecules are trapped in a hydrophobic cavity filled by residues that cannot make appropriate hydrogen bonds, the enthalpic contribution is unfavourable. An opposite situation occurs when water molecules are engaged in forming hydrogen bonds to hydrophilic residues, and here the enthalpic effect may be favourable [56]. The displace-

ment of such water molecules can contribute to the binding free energy, impact affinity during ligands' association and govern enthalpy and entropy partitioning, according to the properties of the individual water molecules compared to those in the bulk phase [49]. Developing a ligand with a high binding affinity towards a specific target is one of the most important steps during the entire drug design process. Thus, a lot of effort is focused on the prediction of whether a water site should be displaced and whether this would cause an increase in a ligand's affinity.

Different approaches, both experimentally-based (i.e., location in crystal structures and B-factors) and knowledge-based (i.e., free energy, water's contribution to binding free energy, entropic contribution), have been reported to assess information about water sites [55]. One of the very first experimentally-based software programs, GRID [57], uses a regular array of 'grid points', established throughout and around the protein, to calculate the energetics of water probes inside a macromolecular binding site (Fig. 2a). GRID places a chemical probe and calculates an empirical interaction energy at all grid points [55]. An approach using crystallographic B-factors to determine which water molecules in a protein's structure are likely to be displaced has been implemented in Consolv [58] and WaterScore [59] software. Using geometric criteria can also indicate the positions of water molecules mediating protein–ligand interactions. Such a procedure has been included in the AcquaAlta program [60] for estimating the propensities of ligand hydration. In HINT software [61], the Gibbs free energy of non-covalent interactions is based on van der Waals interactions and partial atomic partition coefficients. A knowledge-based approach has been implemented, e.g., in AQUARIUS [62] or AQUARIUS2 [63] software. The probable positions for hydration sites are predicted based on solvent distributions surrounding particular amino acids derived from the analysis of a protein's structure. However, most of the software applications mentioned above are not currently used or are used very rarely. This is due to the fact that another class of methods, describing the thermodynamic properties of water by analysing data from MD and MC simulations, became very popular and easy to use.



**Fig. 2.** Strategies of analysis of water sites and ligand binding modes. (A) Strategy using a grid to calculate energetics based on water local distribution, and (B) strategy using IFST (Inhomogeneous Fluid Solvation Theory) to assess the role of structural water molecules by calculating their contribution to the thermodynamics of protein solvation. Grid cells (squares at row A) are coloured according to increasing number of water molecules detected in cells (green – low, red – high). Cells with highest occupancy provide information about the energetically preferred position of the water molecules. Calculated isolines (row B) provide information about the same values of the thermodynamic factors calculated for water molecules in protein cavities. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Table 2**

List of software for analysis of ligand binding and drug design with respect to the water molecules in the binding cavity.

Software	Testing set	Functionality	Remarks
<b>Input – a single structure</b>			
AcquaAlta [60]	trypsin; dihydrofolate reductase; thymidine kinase; VEGFR2; glycogen phosphorylase; human phosphodiesterase; beta trypsin; holo-glyceraldehyde 3P dehydrogenase; HSP90; AmpC beta-lactamase; 2CDK2; ACE; COMT; HIV-1 protease; non-nucleoside adenosine deaminase; ACK1; coagulation factor Xa; EGFR	generating of explicit water molecules at the ligand–protein interface; searching for water molecules interacting with generic functional groups of small organic molecules; generating water molecules bridging interactions between ligand and protein considering the hydration propensities of the involved functional groups and aromatic moieties	available on request: <a href="#">link</a>
FLAP [73]	a set of 23 protein kinase structures	target-based pharmacophores; comparison of multiple protein targets; docking ligands into protein targets	commercial, standalone: <a href="#">link</a>
JAWS [77]	neuraminidase; scytalone dehydratase; Major Urinary Protein 1; bovine $\beta$ -lactoglobulin; cyclooxygenase-2	determining the optimal placement of water molecules in a binding site; binding free energy estimation	implemented in a modified version of MCPRO, v. 2.1 [87]
SZMAP [75]	HIV-1 protease; neuraminidase; trypsin; factor Xa; scytalone dehydratase; oligopeptide-binding protein (OppA);	computation of binding free energies and the corresponding thermodynamic components for water molecules in the binding site	commercial <a href="#">link</a>
WaterFLAP [74]	adenosine A <sub>2A</sub> StaR receptor in complex with triazine	generating and scoring water networks for both apo and ligand-complex structures; binding path prediction; lipophilic hot-spot calculation	commercial, standalone: <a href="#">link</a>
WaterScore [59]	<i>cutinase; xylose isomerase; penicillopepsin; galactose/glucose binding protein; proteinase A; rhizopus pepsin; actinidin; DNase I; cholesterol oxidase; RNase A; thermitase; lipid binding protein; Fv fragment of mouse monoclonal antibody D1.3; dihydrofolate reductase</i>	<i>determine conserved water molecule positions; scoring of protein–ligand interactions and determination of ligand binding mode with respect to bound and displaced water molecules</i>	<i><a href="#">link</a> (currently unavailable)</i>
WATGEN [72]	<i>126 protein–peptide binding interface structures</i>	<i>identification of water sites; selection of the ‘best’ water sites for ligand docking; solvation thermodynamics; binding free energy estimation</i>	<i>no information available</i>
WATsite [71]	<i>three different structures of protein–ligand complexes of factor Xa</i>	<i>identification of water sites; free energy estimation</i>	<i><a href="#">link</a> (currently unavailable)</i>
WRAPPA [76]	vinculin binding-site; truncated SNARE complex; potassium channel fragment; human relaxin-3; RNA complexed with Rev peptide; Kv1.3 channel blocker Tc32	identification of water sites, referred to as dehydrons	web server: <a href="#">link</a>
WScore [78]	<i>a set of 542 binding sites within 506 protein–ligand complexes, associated with 22 receptors</i>	<i>predicting the location of water sites; producing a detailed map of the water structure and displacement free energies; ligand docking and scoring</i>	<i>no information available</i>
<b>Input – MD simulations</b>			
AQUA-DUCT [79,80] AquaMMapS [85]	<i>Solanum tuberosum epoxide hydrolase [88] casein kinase 2; A<sub>2A</sub> adenosine receptor</i>	<i>high-density water sites’ and/or co-solvent sites’ identification identification of spatial hot spots within the protein binding site</i>	<i>standalone: <a href="#">link</a> no information available</i>
GIST [68]	Cucurbit[7]uril (CB7); factor Xa	high-density water sites’ identification; map of regions where the solvent interacts favourably with the surface or has unfavourable entropy	implemented in AmberTools
SPAM [84]	HIV-1 protease; hen egg-white lysozyme	qualitative estimation of the thermodynamic profile of water in hydration sites; binding free energy estimation	implemented in AmberTools
SSTMap [81] STOW [70]	Caspase 3 <i>HIV-1 protease–ligand complex; concanavalin A–carbohydrate complexes; cyclophilin A–ligand complexes</i>	identification of water sites <i>computation of contribution of discrete ordered water molecules to the solvation thermodynamics; determine and analyse water sites</i>	<a href="#">link</a> <i>no information available</i>
WATCLUST [69]	AmpC beta-lactamase	determine and analyse water sites	VMD plugin: <a href="#">link</a> the direct transfer of the information to Autodock
Water-swap [82]	neuraminidase in complex with oseltamivir	calculation of binding free energy by water-swap reaction coordinate	part of the Siremol’s Sire application: <a href="#">link</a>
WaterMap [67,89]	streptavidin; Cox-2; antibody DB3; HIV-1 protease	identification of water sites; solvation thermodynamics; entropic and enthalpic contributions to the free energy	commercial, part of the Schrödinger package: <a href="#">link</a>
WatMD [83]	<i>Green Fluorescent Protein; Mannitol 2-Dehydrogenase</i>	<i>identification of water sites</i>	<i>no information available</i>

\*Information about currently unavailable software is in *italics*.

Most of the recently developed tools are based on Inhomogeneous Fluid Solvation Theory (IFST) derived in 1998 by Lazardis [64]. IFST is a statistical mechanical method that calculates free energy differences from short MD or MC simulations by quantifying the effect of a solute acting as a perturbation to bulk water. The solute may be different molecules, such as proteins, peptides or other molecules. One of the major advantages of IFST is that the

system is spatially decomposed to consider the contribution of specific regions to the total solvation free energies. The contributions of each individual water molecule to the enthalpy are calculated by computing the average interaction energies, whereas the contributions to the entropy are calculated from intermolecular correlations. The Gibbs free energy equation can be used to calculate the contribution to the free energy from the enthalpy and

entropy. The result is then compared with the contribution of one water molecule to the free energy of bulk water, again calculated using IFST. The results of IFST depend on the accuracy of the force field and the water model that was used: a detailed discussion can be found elsewhere [65,66]. IFST is implemented in WaterMap [67], GIST [68], WATCLUST [69], STOW [70] and WATsite [71] software. Methods based on IFST are limited to the analysis of high-occupancy hydration sites and therefore omit solvent molecules found in lower density regions [68] (Fig. 2b).

Some of the software listed in Table 2 perform calculations based on static input (WATsite, WATGEN [72], FLAP [73], WaterFLAP [74], SZMAP [75], WRAPPA [76], WaterScore [59], AcquaAlta [60], JAWS [77] and WScore [78]), while some of them rely on data obtained from MC or MD simulations (AQUA-DUCT [79,80], GIST [68], WATCLUST [69], STOW [70], SSTMap [81], WaterMap [67], Water-swap [82], WatMD [83], SPAM [84] and AquaMMapS [85]). In static structure-based software, hydration and water sites are determined by investigating potential binding sites through the placement of water probes. The only exception is WATsite, which conducts MD simulations based on the static input structure. In simulation-based software, the system is simulated with explicit water molecules, which are free to explore the system's space. These water molecules are then clustered in hydration sites, and their thermodynamic properties are calculated. Some of the software only identifies the water sites, without any further information, while some can estimate the binding free energy and the corresponding thermodynamic components for water molecules in the binding sites. In a very recent paper, the authors combined WATsite data with neural networks and deep learning to significantly improve the speed of water site identification and the calculation of the free energy contributions [86]. The authors claim that such an approach will allow the inclusion of solvation components, such as water-mediated interactions or enthalpically stable hydration networks in proximity to the protein–ligand complex, in structure-based ligand design.

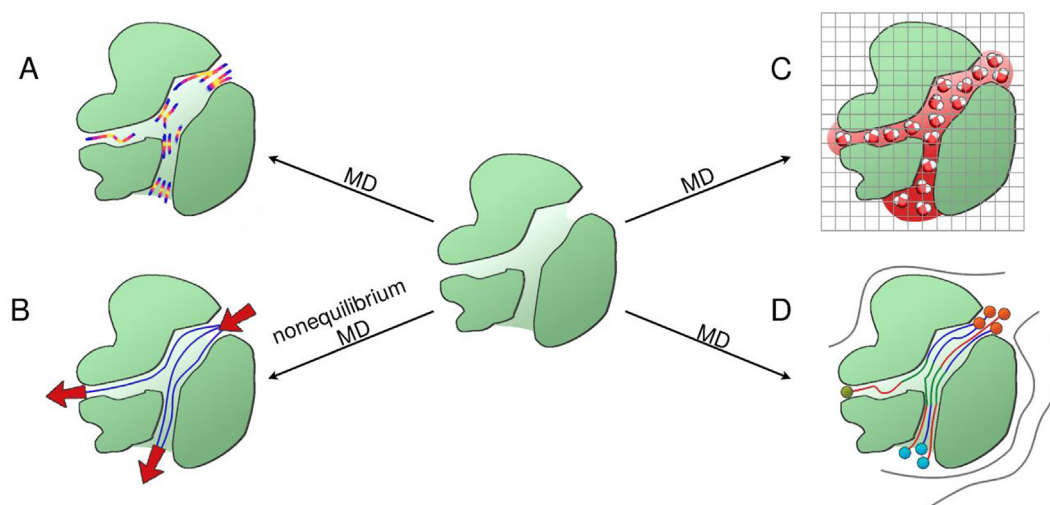
#### 4. Software for tunnel detection and transportation phenomena analysis

The intramolecular voids inside a protein structure, such as cavities, tunnels, channels and pores, are often important for protein functions [90]. While we have already shown the importance of cavities, this section focuses on the function of tunnels and channels. For proteins with a buried active site, tunnels facilitate substrate entry and enable product egress. Tunnels, as well as the whole protein structure, should not be seen as rigid bodies. In fact, a reasonable degree of flexibility is often required to maintain the catalytic reaction. The geometry and amino acid composition of a particular tunnel determine the shape and chemical properties of a potential ligand. Tunnels are also equipped with a much more sophisticated mechanism of small molecule discrimination – gates. Gates are capable of controlling substrate access to the active site, preventing solvent access to particular protein regions and synchronising processes occurring in distant parts of the protein [91]. Tunnels, pores, gates and cavities constitute a dynamic network inside a protein. Therefore, for proper tunnel detection, a single crystal structure of a protein may be insufficient. MD simulations provide a picture of a protein's movements in an aqueous solution. Reasonably long simulations give insights into the dynamics of the tunnel network. Well-defined tunnels allow fast water exchange over a time of about  $10^{-9}$  s, while transient tunnels extend the required time up to  $10^{-3}$  s. In comparison, the exchange time of water molecules at the protein surface with bulk ones is in the sub-nanosecond range [92]. Therefore, from the computational point of view, the lengths of the required molecular

dynamics simulations depend on the studied system. In the case of buried active sites linked with the solvent *via* a network of tunnels, hundreds of nanoseconds are enough to provide good sampling [88,93–96], whereas to observe the exchange of deeply buried water molecules with the bulk solvent up to as much as tens of milliseconds are necessary [30]. Shorter simulations can provide information about the potential pathways of such an exchange and can suggest mutations that can open an alternative tunnel [80]. The second parameter which might influence the required length of simulations is the frequency of gating phenomena. Gates defined by a single amino acid's rotation require shorter experiments than those defined by, e.g., loops or controlled by proteins' breathing motions [91].

The first tunnel detection software used a geometry-based approach to identify 'empty spaces' inside protein structures [90]. The most successful ones, such as Mole 2.0 [97], CAVER 3.0 [98], and CAVER Analyst 1.0 and 2.0 [99,100], are widely used by the scientific community, predominantly to describe tunnels identified in crystal structures. The most successful strategy employs the construction of a Voronoi diagram to detect and describe voids within the macromolecule [101,102]. Using a defined probe radius and internal cavity identification, the software is able to detect tunnels providing access from the selected area to the surroundings. Such a strategy assumes that the tunnels are a summation of connected cavities and is very often used for the analysis of single crystallographic structures. The structural information obtained on such a basis is mostly incomplete, due to tunnels' flexibility. Moreover, using spherical probes for tunnel exploration provides only an approximation of tunnels to tubes with symmetrical diameters and thus prevents analysis of tunnels' asymmetry. It is also difficult to analyse the regulation and direction of the solvent flow, as well as the contribution of tunnels to an enzyme's activity and selectivity. Some of these weaknesses were targeted in 2014 by a non-spherical approach by Benkaidali et al. [103]; however, due to its complex implementation the tool was rarely used. Results provided by geometry-based tunnel detection software were unable to answer questions about solvent flow direction and how tunnels contribute to this.

To analyse the solvent flow direction and tunnels' contribution to this parameter, we need to concentrate on solvent/ligand analysis. Several different methods have been implemented, based on very diverse approaches (Fig. 3). The first attempt was made in 2008 by Bidmon et al. [104], who introduced the Visual Abstractions of Solvent Pathlines method. The pathways of solvent molecules passing through the particular region of interest (so-called ROI) were pre-processed and visualised as Bézier curves. The next attempt at such an analysis was carried out in 2010 by Vasiliev et al. [105]. Their streamline tracing method was applied to photosystem II and was used to visualise water flux in particular regions of the protein. However, the results of the calculations are hard to interpret, and only a few applications of this method can be found in the literature [106]. In 2014, Benson and Pleiss proposed a solvent flux method to study water influx in the *Candida antarctica* lipase B protein cavity from the triglyceride–water interface [107]. They introduced a solvent concentration gradient and the reorientation and rescaling of the velocity vectors of selected water molecules in order to accelerate the influx and increase the probability of rare events in the study. Similarly to widely used strategies (aMD, REMD, SMD and RAMD), it was applied to investigate rare events in a reasonable computing time range (e.g., it overcame the significant energy barriers of slow biophysical events). In contrast to known methods, this technique allowed the flow of multiple molecules, including the selected solvent molecules, to be precisely investigated during a single simulation. Since artificial external forces are introduced to classic MD simulations, one could be concerned about misleading biases and the complicated proto-



**Fig. 3.** Strategies for tunnel detection and description based on water molecule analysis. (A) Streamline tracing, (B) Solvent Flux, (C) *trj\_cavity*, and (D) AQUA-DUCT water tracking approach.

**Table 3**

List of software and methods for tunnel detection and transportation phenomena observation.

Software or method	Applicability System	Functionality	Remarks
AQUA-DUCT [79,80]	<i>Mus musculus</i> epoxide hydrolase <i>D</i> -amino-acid oxidase [111] <i>Pyrococcus furiosus</i> phosphoglucose isomerase [93]	water molecule tracking tunnel and gating residue detection	can be used also for cavity and hot-spot detection Standalone: <a href="#">link</a>
Solvent flux method [107]	Claudin-2 ion channel [112] <i>Candida antarctica</i> lipase B	water molecule tracking and occupancy analysis in the internal cavity ion transportation pathways identification <i>identification of water access pathway; hot-spot identification</i>	<i>based on an artificial gradient. Code not available.</i> visual analysis only; code available on request
Streamline tracing [105]	photosystem II  squalene-hopene cyclase [106]	fibre tracing; tunnel detection; gating residue (access control points) identification changes in water flow after introducing amino acid substitution	visual analysis only; code available on request
<b>trj_cavity</b> [108]*	Der p 2 protein; TM pore; pullulanase  polydicyclopentadiene [114]; herkinorin [115]; glycidoxypropyltrimethoxy silane [116]; mammalian translocator membrane protein [117]; human G-protein coupled receptors [118]; amyloid fibrils [119]; sperm whale myoglobin [120]; 07A metalloprotease [121]; human erythrocyte anion exchanger 1 (Band 3 protein) [122]; amorphous silica [123]; full-length TLR4 dimer [124]; profilin [125]; human serum albumin [126]; laccase [127]; dengue capsid protein (C protein) [128]; horse-radish peroxidase; lactoperoxidase [129]; OmpC–MlaA complex [130–132]; MATE transporter [133] cholesteryl ester transfer protein [134]; acyl carrier proteins [135] mouse myoglobin [109]	generating the trajectory of discovered cavities, quantification of time-dependent cavity volume, solvent presence inside a particular cavity; tunnel detection  cavity analysis  time-dependent cavity analysis	implemented in GROMACS: <a href="#">link</a>
Visual Abstractions of Solvent Pathlines [104]	TEM $\beta$ -lactamase [136]	analysis of the movement of ligands, movements within the cavities and tunnels of proteins <i>identification of the role of water in gating loop flexibility</i>	<i>visual analysis only; code not available</i>
Watergate [113]	haloalkane dehalogenase mutants	visualisation of water molecule trajectories	visual analysis only; code available on request

\*For *trj\_cavity* software only recent applications are presented (2017–2019). Information about currently unavailable software is in *italics*.

col that is dependent on the system. In 2014 another software program emerged as a GROMACS plugin, called *trj\_cavity* [108]. *Trj\_cavity* is capable of cavity and tunnel identification together with time-dependent calculations of their volume and solvent capacity. In the vast majority of research papers, *trj\_cavity* is used only for

the identification of cavities and calculating their volumes and occupancy, while only one study was found where the authors used *trj\_cavity* to actually trace ligands [109]. The existing gap between tools searching for tunnels and pathways, and advanced tools for accelerated water flux investigations was filled in 2017

by AQUA-DUCT [79], an easy-to-use tool facilitating analysis of the behaviour of water (and, if necessary, other solvent molecules) penetrating any selected region in a protein. AQUA-DUCT comprises a *Valve* module, which is capable of tracking water molecules, clustering their trajectories and enabling visualisation in PyMOL [110]. The *Valve* module was used to investigate relatively small proteins, such as D-amino-acid oxidase (DAAO) [111] and *Pyrococcus furiosus* phosphoglucose isomerase [93], as well as ion channels such as claudins [112]. In contrast to the geometry-based approach, water tracking analysis provides information about tunnels' functionality, allows their permeability to be compared and facilitates the detection of the gating residues controlling access to the binding cavity. At the same time, Watergate, a software application for statistical overview of the overall solvent flow, water trajectory clustering, and visualisation was developed [113]. The software programs using water molecules for tunnel detection are listed in Table 3.

Tools based on water molecules as a molecular probe for tunnel detection (listed in Table 3) can provide much more complex information about proteins than simple geometry-based methods. Since they are focused on the information provided by the solvent itself, they also take into account the physicochemical properties of the solute. Such information is useful for examining the effects of the introduced mutation on the solvent flow and thereby the enzyme's activity. By using the pathways of the solvent molecules, the user is able to identify the key residues important for the enzyme's activity and selectivity, and the amino acids that contribute to gating residues and control small ligands' entry/egress. These tools can additionally facilitate the description of cavity shape evolution during simulation time, which can be used for inhibitor design or hot-spot detection for substrate specificity modification, and also the identification of residues distant from the active site which contribute to the activity and selectivity, and thus can be considered as a safe alternative to smart mutant library design. However, it should be kept in mind that to properly sample events such as substrate entrance, product release or the exchange of the trapped solvent molecules with the bulk solvent, the analysed simulations must be of reasonable length and conducted in physiological-like conditions (please see the Summary and outlook section for more details).

## 5. Summary and outlook

The important role of water molecules in structural biology is reflected by a large number of different software programs dedicated to various types of water-molecule-based analysis. Most of the software presented here is focused only on particular aspects of water's presence in a macromolecular structure, such as its contribution to protein stability, ligand binding and drug design, or cavity and tunnel description.

Among all the described software, the role of software in optimising water placement inside proteins' cavities is probably the most underestimated, although placing water molecules is not a trivial task. Three different strategies, RISM theory, the docking of water molecules, and the analysis of conserved water molecules among similar proteins, are used and complement each other. The RISM-based software probably provides the most accurate model; however, it is time-consuming. Docking-based methods are the fastest; however, they may provide biased results for systems, such as metalloproteins, proteins with large cavities and protein-nucleic acids complexes, that are problematic for such software. Both methods can be considered for the prediction of differences in water rearrangement when mutant structures are designed. The third strategy requires a collection of similar structures and may

not be sensitive enough to provide correct predictions when a single mutation occurs. Therefore, depending on the investigated system, different strategies are optimal for water placement and can provide reliable starting points for molecular dynamics studies. As already stated, the water placement method should be considered as a standard approach for homology models or structures with introduced amino acid modifications.

Concerning the role of water in the description of ligand binding, the most commonly used methods are based on IFST. Given a water site, the software can predict how much free energy is gained (or lost) by displacing the water molecules that occupied the potential ligand-binding site. The solvent contribution cannot be neglected, as was shown in several excellent papers [137–139], and therefore continuous progress in both the accuracy and parallel analysis of alternative states is highly desired. Binding enthalpies and entropies of water molecules may also be calculated based on Grid Cell Theory (GCT) [140]. This is a recently developed method for investigating hydration thermodynamics from a molecular dynamics or Monte Carlo standpoint. In the GCT approach, the density, enthalpy, entropy and free energy of water are evaluated for an arbitrary region of space around a system of interest. These parameters refer to the water molecules which enter a particular hydration site of the protein(s) from the bulk concentration. However, this theory has not yet been implemented in known software. The recent AQUA-DUCT version provides an approach combining information on water and co-solvent high-occupancy sites. It can be used for pharmacophore design and suggests directions for future software development. All of the abovementioned methods can provide additional support to drug design and provide more accurate results in comparison to methods neglecting water molecules' contribution.

In the third group of software, water molecules are used as a molecular probe to sample the 'empty spaces' in proteins during molecular dynamics simulations to detect tunnels and cavities. This field is so far monopolised by a widely used geometry-based approach which on the one hand is very simple, but provides rather approximate results. It neglects the tunnels' asymmetry and the physicochemical properties of the tunnel-lining amino acids. Therefore, it is difficult to use such software for analysis of tunnels' functionality. The alternative approaches presented in our review are a most diverse group of software utilising water as a molecular probe. Depending on the implemented algorithms, they provide information about local water flow changes (such as the streamline tracing method), changes in cavity volumes (*trj\_cavity*) or can provide a holistic picture of water flow *via* tunnel networks and an approximation of the energy profiles of particular pathways (AQUA-DUCT). The utilisation of water for cavity and tunnel description removes most of the limitations of the standard approach. It seems that using water molecules as a molecular probe enables more sophisticated analysis of the substrate transportation network provided by tunnels, handling tunnels and cavities together and describing the protein interior in a holistic way as a single entity. However, so far it is hard to provide an estimation of how accurate they are. This problem is caused not only by difficulties in experimental verification of their findings, nor the question of how accurately the hydrophobic cavities can be described, but also due to the lack of benchmarks for the performance inside protein structures of the different water models used in MD simulations.

Protein engineering is one of the most promising, but still largely unexplored, fields of application for software focused on the analysis of water as a molecular probe. So far, most of the examples of such studies are focused on understanding the changes introduced by mutant proteins' construction. However, there are papers showing the potential applicability of the water-based approach



for hot-spot detection [88] and mutant library design [111]. Successful verification of such a strategy can greatly facilitate protein engineering and provide an interesting and easy-to-apply technique. Also, using the information on the water molecules' (or small ligands, or other types of solvent molecules) tracking, the user gains knowledge on a protein's internal architecture, which might be used to develop a successful strategy for further modifications; for instance, to search for more potent inhibitors which will explore previously unused cavities, or to improve the protein's activity and/or selectivity by adjusting the pathways leading to and from the active site.

Since the analysis of water-mediated interactions has become of greater interest, we hope that the number and quality of software programs using water molecules to analyse macromolecules' properties will only increase. However, progress in this promising area cannot be achieved without the further joint efforts of theoreticians and experimentalists. One needs to consider that the majority of the tools described above depend on water models and force fields (e.g., tools based on Inhomogeneous Fluid Solvation Theory or benefiting from MD simulations). Both force fields and water models are being constantly upgraded to provide more accurate descriptions of studied systems. For example, the most recent papers of Huang, et al. provide force fields which can be used for both ordered and disordered proteins [141]. Recent four-point water models have improved the description of its thermodynamic properties; however, water molecules' non-bonded interactions still require validation [142]. Nevertheless, the majority of computational studies employ simple non-polarizable models of water (e.g., TIP3P, SPC/E, TIP4P) and assume that they will describe water molecules in macromolecular surroundings equally well as in the bulk water. Unfortunately, there is no study that can confirm such a presupposition, simply due to the limited access to experimental data providing insight into water's behaviour inside a protein's core. Moreover, even the benchmark analysis of a particular software data's dependency on the used parameters is very limited. As we mentioned above, the comparison of the IFST results obtained with different water models suggests that the quantitative application of IFST to biological systems is strictly model-dependent and has to be carefully analysed. Fortunately, several successful verifications of the findings guided by the software developed to analyse the behaviour and/or properties of water molecules have accelerated research in the field of protein research and each year bring to the scientific community new, optimised, versatile and reliable tools which greatly improve our understanding of nature.

## Funding

This work was funded by the National Science Centre, Poland, grant no DEC-2013/10/E/NZ1/00649 and DEC-2015/18/M/NZ1/00427.

## Conflict of interest

None declared.

## CRediT authorship contribution statement

**Karolina Mitusińska:** Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Agata Raczyńska:** Resources, Data curation, Writing - original draft, Investigation, Visualization. **Maria Bzówka:** Resources, Data curation, Writing - original draft, Investigation, Writing - review & editing. **Weronika Bagrowska:** Resources, Data curation, Writing - original draft, Investigation. **Artur Góra:** Conceptualization, Super-

vision, Funding acquisition, Project administration, Writing - review & editing.

## References

- [1] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. 4th ed. New York: Garland Science; 2002.
- [2] Vallone B, Brunori M. Roles for holes: are cavities in proteins mere packing defects? *Ital J Biochem* 2004;53(1):46–52.
- [3] J. A. Rupley, G. Careri, Protein hydration and function, 1991, 37–172.
- [4] Biedermannová L, Schneider B. Hydration of proteins and nucleic acids: advances in experiment and theory. A review. *Biochim Biophys Acta - Gen Subj* 2016;1860(9):1821–35.
- [5] Mikol V, Papageorgiou C, Borer X. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem* 1995;38(17):3361–7.
- [6] Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Water in protein structure prediction. *Proc Natl Acad Sci U S A* 2004;101(10):3352–7.
- [7] Nakasako M. Water-protein interactions from high-resolution protein crystallography. *Philos Trans R Soc London, Ser B Biol Sci*, 2004;359(1448):1191–206.
- [8] Niimura N, Chatake T, Kurihara K, Maeda M. Hydrogen and hydration in proteins. *Cell Biochem Biophys* 2004;40(3):351–70.
- [9] Savage H, Wlodawer A. Determination of water structure around biomolecules using X-ray and neutron diffraction methods. 1986, 162–183.
- [10] Levitt M, Park BH. Water: now you see it, now you don't. *Structure* 1993;1(4):223–6.
- [11] Carugo O, Bordo D. How many water molecules can be detected by protein crystallography? *Acta Crystallogr D Biol Crystallogr* 1999;55(2):479–83.
- [12] Berman HM. The Protein Data Bank. *Nucleic Acids Res* 2000;28(1):235–42.
- [13] Imai T, Hiraoka R, Kovalenko A, Hirata F. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins Struct Funct Bioinf* 2006;66(4):804–13.
- [14] Hopmann KH, Himo F. Theoretical study of the full reaction mechanism of human soluble epoxide hydrolase. *Chemistry* 2006;12(26):6898–909.
- [15] Sheng X, Lind MES, Himo F. Theoretical study of the reaction mechanism of phenolic acid decarboxylase. *FEBS J* 2015;282(24):4703–13.
- [16] Abrioux C, Coasne B, Maurin G, Henn F, Jeffroy M, Boutin A. Cation behavior in Faujasite zeolites upon water adsorption: a combination of Monte Carlo and molecular dynamics simulations. *J Phys Chem C* 2009;113(24):10696–705.
- [17] Paquet E, Viktor HL. Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: a computational review. *Biomed Res Int* 2015;2015:1–18.
- [18] Levy Y, Onuchic JN. Water and proteins: a love-hate relationship. *Proc Natl Acad Sci U S A* 2004;101(10):3325–6.
- [19] Lavecchia A, Giovanni C. Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem* 2013;20(23):2839–60.
- [20] Sousa SF et al. Protein-ligand docking in the new millennium – A retrospective of 10 years in the field. *Curr Med Chem* 2013;20(18):2296–314.
- [21] Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev* 2017;9(2):91–102.
- [22] Finer-Moore JS, Kossiakoff AA, Hurley JH, Earnest T, Stroud RM. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. *Proteins Struct Funct Genet* 1992;12(3):203–22.
- [23] Nagendra HG, Sukumar N, Vijayan M. Role of water in plasticity, stability, and action of proteins: the crystal structures of lysozyme at very low levels of hydration. *Proteins Struct Funct Genet* 1998;32(2):229–40.
- [24] Takano K, Yamagata Y, Yutani K. Buried water molecules contribute to the conformational stability of a protein. *Protein Eng Des Sel* 2003;16(1):5–9.
- [25] Meyer E. Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. *Protein Sci* 1992;1(12):1543–62.
- [26] Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R. Chemistry of ion coordination and hydration revealed by a K<sup>+</sup> channel-Fab complex at 2.0 Å resolution. *Nature* 2001;414(6859):43–8.
- [27] Tanimoto T, Furutani Y, Kandori H. Structural changes of water in the Schiff base region of bacteriorhodopsin: proposal of a hydration switch model. *Biochemistry* 2003;42(8):2300–6.
- [28] Karplus PA, Faerman C. Ordered water in macromolecular structure. *Curr Opin Struct Biol* 1994;4(5):770–6.
- [29] Otting G, Liepinsh E, Wuthrich K. Protein hydration in aqueous solution. *Science* (80-) 1991;254(5034):974–80.
- [30] Persson F, Halle B. Transient access to the protein interior: simulation versus NMR. *J Am Chem Soc* 2013;135(23):8735–48.
- [31] Morozenko A, Leontyev IV, Stuchebrukhov AA. Dipole moment and binding energy of water in proteins from crystallographic analysis. *J Chem Theory Comput* 2014;10(10):4618–23.
- [32] Ross GA, Morris GM, Biggin PC. Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS ONE* 2012;7(3):e32036.
- [33] Sindhikara DJ, Hirata F. Analysis of biomolecular solvation sites by 3D-RISM theory. *J Phys Chem B* 2013;117(22):6718–23.

- [34] Fusani L, Wall I, Palmer D, Cortes A. Optimal water networks in protein cavities with GAsol and 3D-RISM. *Bioinformatics* 2018;34(11):1947–8.
- [35] Sindhikara DJ, Yoshida N, Hirata F. Placevent: an algorithm for prediction of explicit solvent atom distribution-Application to HIV-1 protease and F-ATP synthase. *J Comput Chem* 2012;33(18):1536–43.
- [36] Patel H, Gruning BA, Gunther S, Merfort I. PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering. *Bioinformatics* 2014;30(20):2978–80.
- [37] Jukić M, Konc J, Gobec S, Janežič D. Identification of conserved water sites in protein structures for drug design. *J Chem Inf Model* 2017;57(12):3094–103.
- [38] 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* 2009;1–18.
- [39] Morozenko A, Stuchebrukhov AA. Dowser++, a new method of hydrating protein structures. *Proteins Struct Funct Bioinf* 2016;84(10):1347–57.
- [40] Imai T, Hiraoka R, Kovalenko A, Hirata F. Water molecules in a protein cavity detected by a statistical–mechanical theory. *J Am Chem Soc* 2005;127(44):15334–5.
- [41] Stumpe MC, Blinov N, Wishart D, Kovalenko A, Pande VS. Calculation of local water densities in biological systems: a comparison of molecular dynamics simulations and the 3D-RISM-KH molecular theory of solvation. *J Phys Chem B* 2011;115(2):319–28.
- [42] Palmer DS, Frolov AI, Ratkova EL, Fedorov MV. Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction. *J Phys: Condens Matter* 2010;22(49):492101.
- [43] Nikolić D, Blinov N, Wishart D, Kovalenko A. 3D-RISM-Dock: A new fragment-based drug design protocol. *J Chem Theory Comput* 2012;8(9):3356–72.
- [44] Truchon J-F, Pettitt BM, Labute P. A cavity corrected 3D-RISM functional for accurate solvation free energies. *J Chem Theory Comput* 2014;10(3):934–41.
- [45] Konc J, Janežič D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* 2010;26(9):1160–8.
- [46] Bernstein FC et al. The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 1977;112(3):535–42.
- [47] Sridhar A, Ross GA, Biggin PC. Waterdock 2.0: water placement prediction for Holo-structures with a pymol plugin. *PLoS ONE* 2017;12(2):e0172743.
- [48] Du X et al. Insights into protein–ligand interactions: mechanisms, models, and methods. *Int J Mol Sci* 2016;17(2):144.
- [49] Schiebel J et al. Intriguing role of water in protein–ligand binding studied by neutron crystallography on trypsin complexes. *Nat Commun* 2018;9(1):3559.
- [50] Lu Y, Wang R, Yang C-Y, Wang S. Analysis of ligand–bound water molecules in high-resolution crystal structures of protein–ligand complexes. *J Chem Inf Model* 2007;47(2):668–75.
- [51] Henchman RH, McCammon JA. Structural and dynamic properties of water around acetylcholinesterase. *Protein Sci* 2009;11(9):2080–90.
- [52] Irwin BWJ, Vukovic S, Payne MC, Huggins DJ. Large-scale study of hydration environments through hydration sites. *J Phys Chem B* 2019;123(19):4220–9.
- [53] Chervenak MC, Toone EJ. A direct measure of the contribution of solvent reorganization to the enthalpy of binding. *J Am Chem Soc* 1994;116(23):10533–9.
- [54] Klebe G, Böhm H-J. Energetic and entropic factors determining binding affinity in protein–ligand complexes. *J Recept Signal Transduct* 1997;17(1–3):459–73.
- [55] de Beer S, Vermeulen N, Oostenbrink C. The role of water molecules in computational drug design. *Curr Top Med Chem* 2010;10(1):55–66.
- [56] Uehara S, Tanaka S. AutoDock-GIST: incorporating thermodynamics of active-site water into scoring function for accurate protein–ligand docking. *Molecules* 2016;21(11):1604.
- [57] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985;28(7):849–57.
- [58] Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J Mol Biol* 1997;265(4):445–64.
- [59] García-Sosa AT, Mancera RL, Dean PM. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein–ligand complexes. *J Mol Model* 2003;9(3):172–82.
- [60] Rossato G, Ernst B, Vedani A, Smieško M. AcquaAlta: a directional approach to the solvation of ligand–protein complexes. *J Chem Inf Model* 2011;51(8):1867–81.
- [61] Eugene Kellogg G, Abraham DJ. Hydrophobicity: is  $\text{LogP}(o/w)$  more than the sum of its parts? *Eur J Med Chem*, 35(7–8), 651–661.
- [62] Pitt WR, Goodfellow JM. Modelling of solvent positions around polar groups in proteins. *Protein Eng Des Sel* 1991;4(5):531–7.
- [63] Pitt WR, Murray-Rust J, Goodfellow JM. AQUARIUS2: knowledge-based modeling of solvent sites around proteins. *J Comput Chem* 1993;14(9):1007–18.
- [64] Lazaridis T. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J Phys Chem B* 1998;102(18):3531–41.
- [65] Huggins DJ, Payne MC. Assessing the accuracy of inhomogeneous fluid solvation theory in predicting hydration free energies of simple solutes. *J Phys Chem B* 2013;117(27):8232–44.
- [66] Hess B, van der Vegt NFA. Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J Phys Chem B* 2006;110(35):17616–26.
- [67] Cappel D, Sherman W, Beuming T. Calculating water thermodynamics in the binding site of proteins – applications of WaterMap to drug discovery. *Curr Top Med Chem* 2017;17(23).
- [68] Ramsey S, Nguyen C, Salomon-Ferrer R, Walker RC, Gilson MK, Kurtzman T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J Comput Chem* 2016;37(21):2029–37.
- [69] López ED et al. WATCLUST: a tool for improving the design of drugs based on protein–water interactions: Fig. 1. *Bioinformatics* 2015;31(22):3697–9.
- [70] Li Z, Lazaridis T. Computing the thermodynamic contributions of interfacial water, 2012, 393–404.
- [71] Hu B, Lill MA. WATsite: hydration site prediction program with PyMOL interface. *J Comput Chem* 2014;35(16):1255–60.
- [72] Bui H-H, Schiewe AJ, Haworth IS. WATGEN: an algorithm for modeling water networks at protein–protein interfaces. *J Comput Chem* 2007;28(14):2241–51.
- [73] Baroni M, Cruciani G, Sciabola S, Perruccio F, Mason JS. 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(2):279–94.
- [74] Mason JS, Bortolato A, Weiss DR, Deflorian F, Tehan B, Marshall FH. High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks. *Silico Pharmacol*. 2013;1(1):23.
- [75] Bayden AS, Moustakas DT, Joseph-McCarthy D, Lamb ML. Evaluating free energies of binding and conservation of crystallographic waters using SZMAP. *J Chem Inf Model* 2015;55(8):1552–65.
- [76] Fraser CM, Fernandez A, Scott LR, WRAPPA: A Screening Tool for Candidate Dehydrogenation, 2011.
- [77] Michel J, Tirado-Rives J, Jorgensen WL. Prediction of the Water Content in Protein Binding Sites. *J Phys Chem B* 2009;113(40):13337–46.
- [78] Murphy RB et al. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand–Receptor Docking. *J Med Chem* May 2016;59(9):4364–84.
- [79] Magdziarz T et al. AQUA-DUCT: a ligands tracking tool. *Bioinformatics* 2017;33(13):2045–6.
- [80] Magdziarz T et al. AQUA-DUCT 1.0: structural and functional analysis of macromolecules from an intramolecular voids perspective. *Bioinformatics* 2019. <https://doi.org/10.1093/bioinformatics/btz946>.
- [81] Haider K, Cruz A, Ramsey S, Gilson MK, Kurtzman T. Solvation structure and thermodynamic mapping (SSTMap): an open-source, flexible package for the analysis of water in molecular dynamics trajectories. *J Chem Theory Comput* 2018;14(1):418–25.
- [82] Woods CJ, Malaisree M, Hannongbua S, Mulholland AJ. A water-swap reaction coordinate for the calculation of absolute protein–ligand binding free energies. *J Chem Phys* 2011;134(5):054114.
- [83] Velez-Vega C, McKay DJJ, Aravamathan V, Pearlstein R, Duca JS. Time-averaged distributions of solute and solvent motions: exploring proton wires of GFP and PfM2DH. *J Chem Inf Model* 2014;54(12):3344–61.
- [84] Cui G, Swails JM, Manas ES. SPAM: a simple approach for profiling bound water molecules. *J Chem Theory Comput* 2013;9(12):5539–49.
- [85] Cuzzolin A, Deganutti G, Salmaso V, Sturlese M, Moro S. AquaMMAPS: an alternative tool to monitor the role of water molecules during protein–ligand association. *ChemMedChem* 2018;13(6):522–31.
- [86] Ghanbarpour A, Mahmoud A, Lill M. On-the-fly prediction of protein hydration densities and free energies using deep learning. 2020.
- [87] Jorgensen WL, Tirado-Rives J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comput Chem* 2005;26(16):1689–700.
- [88] Mitusińska K, Magdziarz T, Bzówka M, Stańczak A, Gora A. Exploring Solanum tuberosum epoxide hydrolase internal architecture by water molecules tracking. *Biomolecules* 2018;8(4):143.
- [89] Yang Y, Lightstone FC, Wong SE. Approaches to efficiently estimate solvation and explicit water energetics in ligand binding: the use of WaterMap. *Expert Opin Drug Disc* 2013;8(3):277–87.
- [90] Brezovsky J, Chovancova E, Gora A, Pavelka A, Biedermannova L, Damborsky J. Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnol Adv* 2013;31(1):38–49.
- [91] Gora A, Brezovsky J, Damborsky J. Gates of enzymes. *Chem Rev* 2013;113(8):5871–923.
- [92] Denisov VP, Halle B. Protein hydration dynamics in aqueous solution. *Faraday Discuss* 1996;103:227.
- [93] Subramanian K et al. Distant Non-obvious mutations influence the activity of a hyperthermophilic *Pyrococcus furiosus* phosphoglucose isomerase. *Biomolecules* 2019;9(6):212.
- [94] Brezovsky J et al. Engineering a de Novo Transport Tunnel. *ACS Catal* 2016;6(11):7597–610.
- [95] Fischer A, Don CG, Smieško M. Molecular dynamics simulations reveal structural differences among allelic variants of membrane-anchored cytochrome P450 2D6. *J Chem Inf Model* 2018;58(9):1962–75.
- [96] Watanabe G, Nakajima D, Hiroshima A, Suzuki H, Yoneda S. Analysis of water channels by molecular dynamics simulation of heterotetrameric sarcosine oxidase. *Biophys. Physicobiol* 2015;12:131–8.

- [97] Sehna D et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J Cheminf* 2013;5(1):39.
- [98] Chovanova E et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* 2012;8(10):e1002708.
- [99] Kozlikova B et al. CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* 2014;30(18):2684–5.
- [100] Jurcik A et al. CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* 2018;34(20):3586–8.
- [101] Kim D-S, Sugihara K. Tunnels and voids in molecules via voronoi diagram. In: 2012 Ninth International Symposium on Voronoi Diagrams in Science and Engineering. p. 138–43.
- [102] Pavelka A, Sebestova E, Kozlikova B, Brezovsky J, Sochor J, Damborsky J. CAVER: algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Trans Comput Biol Bioinf* 2016;13(3):505–17.
- [103] Benkaidali L et al. Computing cavities, channels, pores and pockets in proteins from non-spherical ligands models. *Bioinformatics* 2014;30(6):792–800.
- [104] Bidmon K, Grottel S, Bös F, Pleiss J, Ertl T. Visual abstractions of solvent pathlines near protein cavities. *Comput Graph Forum* 2008;27(3):935–42.
- [105] Vassiliev S, Comte P, Mahboob A, Bruce D. Tracking the flow of water through photosystem II using molecular dynamics and streamline tracing. *Biochemistry* 2010;49(9):1873–81.
- [106] Gustafsson C, Vassiliev S, Kürten C, Syrén P-O, Brinck T. MD simulations reveal complex water paths in squalene-hopene cyclase: tunnel-obstructing mutations increase the flow of water in the active site. *ACS Omega* 2017;2(11):8495–506.
- [107] Benson SP, Pleiss J. Solvent flux method (SFM): a case study of water access to *Candida antarctica* lipase B. *J Chem Theory Comput* 2014;10(11):5206–14.
- [108] Paramo T, East A, Garzón D, Ulmschneider MB, Bond PJ. Efficient characterization of protein cavities within molecular simulation trajectories: trj\_cavity. *J Chem Theory Comput* 2014;10(5):2151–64.
- [109] Chintapalli SV, Anishkin A, Adams SH. Exploring the entry route of palmitic acid and palmitoylcarnitine into myoglobin. *Arch Biochem Biophys* 2018;655:56–66.
- [110] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
- [111] Subramanian K et al. Modulating D-amino acid oxidase (DAAO) substrate specificity through facilitated solvent access. *PLoS ONE* 2018;13(6):e0198990.
- [112] Irudayanathan FJ, Wang X, Wang N, Willsey SR, Seddon IA, Nangia S. Self-assembly simulations of classic claudins—insights into the pore structure, selectivity, and higher order complexes. *J Phys Chem B* 2018;122(30):7463–74.
- [113] Vad V et al. Watergate: visual exploration of water trajectories in protein dynamics. *Vcbm* 2017:33–42.
- [114] Elder RM, Long TR, Bain ED, Lenhart JL, Sirk TW. Mechanics and nanovoid nucleation dynamics: effects of polar functionality in glassy polymer networks. *Soft Matter* 2018;14(44):8895–911.
- [115] Marmolejo-Valencia AF, Martínez-Mayorga K. Allosteric modulation model of the mu opioid receptor by herkinorin, a potent not alkaloidal agonist. *J Comput Aided Mol Des* 2017;31(5):467–82.
- [116] Chowdhury SC, Elder RM, Sirk TW, van Duin ACT, Gillespie JW. Modeling of glycidoxypropyltrimethoxy silane compositions using molecular dynamics simulations. *Comput Mater Sci* 2017;140:82–8.
- [117] Zeng J et al. Structural prediction of the dimeric form of the mammalian translocator membrane protein TSPO: a key target for brain diagnostics. *Int J Mol Sci* 2018;19(9):2588.
- [118] Cao R, Giorgetti A, Bauer A, Neumaier B, Rossetti G, Carloni P. Role of extracellular loops and membrane lipids for ligand recognition in the neuronal adenosine receptor type 2A: an enhanced sampling simulation study. *Molecules* 2018;23(10):2616.
- [119] Choi H, Chang HJ, Lee M, Na S. Characterizing structural stability of amyloid motif fibrils mediated by water molecules. *ChemPhysChem* 2017;18(7):817–27.
- [120] Espinosa YR, Caffarena ER, Grigera JR. The role of hydrophobicity in the cold denaturation of proteins under high pressure: a study on apomyoglobin. *J Chem Phys* 2019;150(7):075102.
- [121] Polêto MD, Alves MP, Ligabue-Braun R, Eller MR, De carvalho AF. Role of structural ions on the dynamics of the *Pseudomonas fluorescens* 07A metalloprotease. *Food Chem., Jul.* 2019;286:309–15.
- [122] Kalli AC, Reithmeier RAF. Interaction of the human erythrocyte Band 3 anion exchanger 1 (AE1, SLC4A1) with lipids and glycophorin A: molecular organization of the Wright (Wr) blood group antigen. *PLoS Comput Biol* 2018;14(7):e1006284.
- [123] Chowdhury SC, Wise EA, Ganesh R, Gillespie JW. Effects of surface crack on the mechanical properties of Silica: a molecular dynamics simulation study. *Eng Fract Mech* 2019;207:99–108.
- [124] Patra MC, Kwon H-K, Batool M, Choi S. Computational insight into the structural organization of full-length toll-like receptor 4 dimer in a model phospholipid bilayer. *Front Immunol* 2018;9.
- [125] Kadirvel P, Anishetty S. Potential role of salt-bridges in the hinge-like movement of apicomplexa specific  $\beta$ -hairpin of Plasmodium and Toxoplasma profilins: A molecular dynamics simulation study. *J Cell Biochem* 2018;119(4):3683–96.
- [126] Tiwari G, Verma CS. Toward understanding the molecular recognition of albumin by p53-activating stapled peptide ATSP-7041. *J Phys Chem B* 2017;121(4):657–70.
- [127] Liu Y et al. Effect of surfactants on the interaction of phenol with laccase: MOLECULAR docking and molecular dynamics simulation studies. *J Hazard Mater* 2018;357:10–8.
- [128] Boon PLS et al. Partial intrinsic disorder governs the dengue capsid protein conformational ensemble. *ACS Chem Biol* 2018;13(6):1621–30.
- [129] Chen M, Zeng G, Xu P, Zhang Y, Jiang D, Zhou S. Understanding enzymatic degradation of single-walled carbon nanotubes triggered by functionalization using molecular dynamics simulation. *Environ Sci Nano* 2017;4(3):720–7.
- [130] Yeow J et al. The architecture of the OmpC–MlaA complex sheds light on the maintenance of outer membrane lipid asymmetry in *Escherichia coli*. *J Biol Chem* 2018;293(29):11325–40.
- [131] Hughes GW et al. Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system. *Nat Microbiol* 2019;4(10):1692–705.
- [132] Yeow J, Chong Z-S, Marzinek J, Bond P. Molecular basis for the maintenance of lipid asymmetry in the outer membrane of *Escherichia coli*. *BioRxiv* 2017. 192500.
- [133] Kraha A, Zachariae U. Insights into the ion-coupling mechanism in the MATE transporter NorM-VC. *Phys Biol* 2017;14(4):045009.
- [134] Revanasiddappa PD, Sankar R, Senapati S. Role of the bound phospholipids in the structural stability of cholesterol ester transfer protein. *J Phys Chem B* 2018;122(15):4239–48.
- [135] Farmer R, Thomas CM, Winn PJ. Structure, function and dynamics in acyl carrier proteins. *PLoS ONE* 2019;14(7):e0219435.
- [136] Bös F, Pleiss J. Multiple molecular dynamics simulations of TEM  $\beta$ -lactamase: dynamics and water binding of the  $\Omega$ -loop. *Biophys J* 2009;97(9):2550–8.
- [137] Darby JF et al. Water networks can determine the affinity of ligand binding to proteins. *J Am Chem Soc* 2019;141(40):15818–26.
- [138] Syrén P-O, Hammer SC, Claasen B, Hauer B. Entropy is key to the formation of pentacyclic terpenoids by enzyme-catalyzed polycyclization. *Angew Chem Int Ed* 2014;53(19):4845–9.
- [139] Chen D et al. Effective lead optimization targeting the displacement of bridging receptor–ligand water molecules. *Phys Chem Chem Phys* 2018;20(37):24399–407.
- [140] Gerogiokas G, Calabro G, Henchman RH, Southey MWY, Law RJ, Michel J. Prediction of small molecule hydration thermodynamics with grid cell theory. *J Chem Theory Comput* 2014;10(1):35–48.
- [141] Huang J et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 2017;14(1):71–3.
- [142] Beauchamp KA, Lin Y-S, Das R, Pande VS. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J Chem Theory Comput* 2012;8(4):1409–14.