Bioorganic & Medicinal Chemistry Letters xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Digest

Characterizing hydration sites in protein-ligand complexes towards the design of novel ligands

Hans Matter*, Stefan Güssregen

Sanofi-Aventis Deutschland GmbH, Integrated Drug Discovery (IDD), Synthetic Molecular Design, Building G838, Industriepark Höchst, 65926 Frankfurt am Main, Germany

ARTICLE INFO

Keywords:
Structure-based design
Hydration sites
WaterMap
3D RISM
Computer-aided drug design

ABSTRACT

Water is an essential part of protein binding sites and mediates interactions to ligands. Its displacement by ligand parts affects the free binding energy of resulting protein-ligand complexes. Therefore the characterization of solvation properties is important for design. Of particular interest is the propensity of localized water to be favorably displaced by a ligand. This review discusses two popular computational approaches addressing these questions, namely WaterMap based on statistical mechanics analysis of MD simulations and 3D RISM based on integral equation theory of liquids. The theoretical background and recent applications in structure-based design will be presented.

Introduction

Water molecules are essential parts of protein sites and mediate interactions to ligands, as revealed by X-ray crystallography plus further experimental and theoretical evidence.[1–5] Displacement of water molecules by ligands is possible and affects the free energy of binding (ΔG) to proteins.[5,6] The impact of water on binding energetics has been analyzed in detail, in particular on the thermodynamic signature in terms of entropy and enthalpy contributions.[5,7] An early assessment for the entropic cost of freezing water was provided by Dunitz from hydrated inorganic crystals.[8] He estimated the free energy cost between 0 and 2 kcal/mol at 300 K for transferring a water molecule from liquid to the protein, while the higher value provides an upper limit for tightly bound water interacting with polar protein moieties.[8] In contrast, most other water at protein surfaces and hydrophobic pockets was estimated to cost only a fraction of this value.[8]

Therefore the characterization of solvation properties including position and thermodynamic profile is important for design. X-ray structures and other techniques can unveil the structure of water networks in protein-ligand complexes, although most solvent molecules are not sufficiently resolved. Understanding the effect of desolvation on binding energy requires an assessment of the thermodynamic properties of solvent contacting the protein and mediating interactions to ligands.[9]

There are classical examples like aspartyl proteases illustrating the impact of water in structure-based design. For example, the identification of non-peptidic HIV-1 protease inhibitors by replacement of structurally conserved water was demonstrated for a series of cyclic ureas.[10] There is one structurally conserved water molecule linking two aspartic acid residues in the catalytic site of aspartyl proteases, which is displaced by almost all cyclic and non-cyclic ligands in HIV-1 protease and other proteases like Renin and BACE (β-secretase). Moreover, the design of a cyclic urea series by Lam et al.[10] targeted a second structural water molecule in the HIV-1 protease X-ray structure linking an earlier inhibitor to the glycine-rich P-strands of the HIV-1 protease dimer. This water molecule interacts with carbonyl oxygens of residues Ile50 and Ile50′ and donates two hydrogen bonds to carbonyl oxygens of this earlier inhibitor.

Water-mediated interactions have been experimentally characterized for some proteins at remarkably accuracy.[11,12] A neutron crystal structure of a family 4-type carbohydrate-binding module complexed with a xyloglucan revealed atomic details of protein-carbohydrate hydrogen-bonds and water-mediated interactions.[11] This accurate knowledge of water positions has a significant influence on docking and design, as discussed below.[13,14]

Understanding the thermodynamic characteristics of water might impact structure-based design, as protein-ligand interactions are modulated by solvent features next to the cavity.[15] Therefore it is important to estimate the propensity of conserved water to be replaced by ligands. The main question here is which water molecule can be favorably displaced. The energetic consequences associated with this are highly relevant for ligand pose and affinity ranking. However, energetic estimations can only be provided by computation, as thermodynamic data for individual water molecules cannot be obtained due to the macroscopic nature of this property.

* Corresponding author.
E-mail address: Hans.Matter@sanofi.com (H. Matter).

https://doi.org/10.1016/j.bmcl.2018.05.061
Received 29 March 2018; Received in revised form 28 May 2018; Accepted 30 May 2018

An adequate treatment of solvation effects is difficult in computational chemistry.[16,17] Solvation models ranging from continuum (implicit) to explicit models based on atomistic treatment of solvent and quantum-mechanical treatments are available.[18] Many approaches have been developed to characterize hydration sites in proteins, including empirical scoring functions like HYDE[19] containing a desolvation term, hybrid continuum/explicit solvent approaches like SZMAP[20], grid-based molecular interaction fields like WaterFLAP[21] and statistical mechanics models like 3D RISM.[22–26]

The computational costs for extensive free energy methods has prompted for the development of protocols, which are less computationally expensive, but allow integrating structural water and hydrogen-bond directionality. In this article we discuss two approaches for modelling water in drug design with an emphasis on accuracy balanced with computational costs. The first of these approaches is WaterMap, which is popular for this task already. The second approach is 3D RISM, which has a long history in modelling different molecules in solution, but only recently gained interest for modelling water in proteins due to its attractive properties. After introduction into their theoretical background, we will focus on recent drug design applications of both methods.

Analysis of hydration sites using WaterMap

WaterMap is based on molecular dynamics (MD) simulations to describe thermodynamic properties of water interacting with proteins.[27,28] Therefore solvent contributions to protein-ligand affinity become accessible. In particular, the free energy balance after displacing solvent from a binding site can be estimated. Often, addition of a hydrophobic substituent to a hydrophobic pocket can have a favorable effect on affinity, which is related to a gain in free energy of binding after releasing solvent molecules into bulk solvent.[27]

The computation starts with an MD simulation of the protein or complex followed by clustering of solvent positions. MD simulations are typically performed using the OPLS3 force field[29] and the program Desmond[30–32] for relatively short simulation times of 2–10 ns. Typically the protein is solvated in a water box extending 10 Å in all directions, followed by relaxation using minimization and heating steps. Finally, positional restraints are applied to protein atoms (5 kcal*mol^{-1}*Å^{-2}) to balance realistic protein thermal mobility and convergence for hydration site energies with simulation time. Therefore, results are relevant only for a single conformation.

Clustering of solvent then produces high occupancy maps of "hydration sites" (see Fig. 1). The average interaction energy between each water in such a site and the system is computed and the excess entropy for those is estimated using inhomogeneous solvation theory.[33,34] This formalism computes the free energy cost of transferring a solvent molecule from a hydration site to bulk water. In this model, bulk water corresponds to zero entropy, while excess entropy is estimated for water in hydration sites. To enhance sampling of extremely buried protein hydration sites that are difficult to sample by short MD simulations, a grand canonical Monte-Carlo (MC) step was added before MD.[35]

The analysis of energetic terms for distinct protein hydration sites allows estimating their entropy, enthalpy and free energy.[36] Energetic values are compared to bulk solvent reference values to estimate the enthalpic cost and excess entropy to transfer water from a hydration site next to the protein to bulk solvent. This forms the basis for the classification of hydration sites into favorable and unfavorable sites relative to bulk solvent.

While MD-based analyses approaches like WaterMap focus on properties of individual water molecules during simulation, it is also possible to discretize the inhomogeneous-solvation-theory-based MD estimates of solvation energy and entropy on regular 3D grids, as has been described by Gilson et al. in their GIST approach (grid inhomogeneous solvation theory).[37,38]

Regions of energetically unfavorable water molecules are termed "hot spots".[39] It is assumed that displacement of water molecules from these unfavorable sites by hydrophobic ligand parts favorably contributes to the binding free energy, which is the case for hydration sites in hydrophobic protein environments, where water molecules cannot form tight hydrogen bonds. In contrast, replacement of water by a polar ligand moiety can be considered for hydration sites showing strong interactions to the protein, but with significantly higher entropy due to its location in a more hydrophobic protein environment. A new polar ligand part can improve binding free energy, if the same interaction to the protein can be formed with similar geometry and without entropic penalty. The water release into the bulk might provide a net positive effect for affinity. In contrast, favorable hydration sites are characterized by potential water molecules forming tight H-bonds to a polar environment. A replacement of water situated in these sites should be avoided, while bridging interactions to the protein via conserved solvent molecules in sites could be considered, if geometrically possible. In any case, replacement or interaction with a polar ligand must consider preferred geometrical arrangements for hydrogen-bonding.

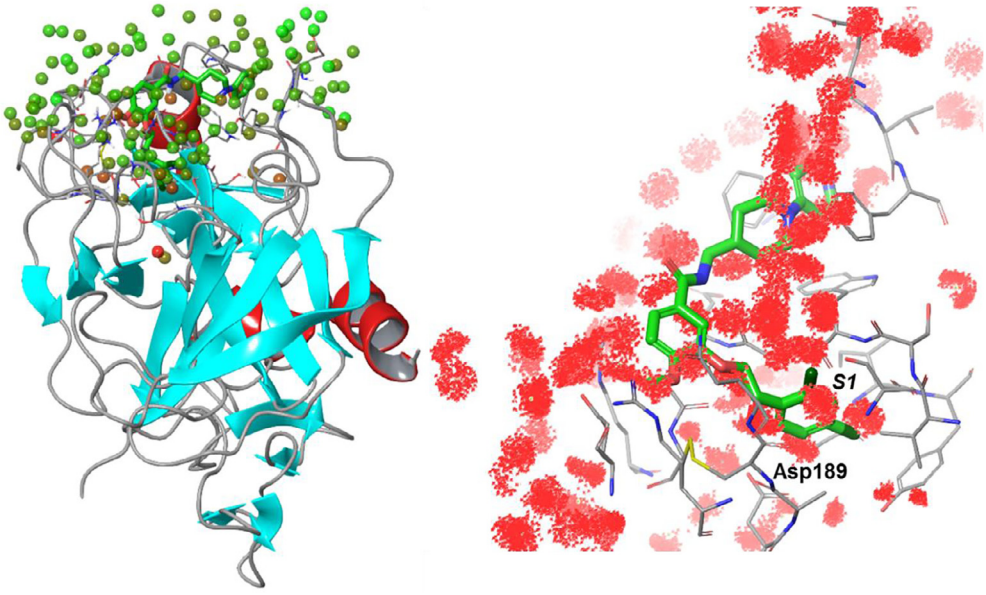

Fig. 1. WaterMap hydration sites for the X-ray structure of a 3-oxybenzamide factor Xa inhibitor (PDB 2BMG, 2.70 Å resolution) from the analysis of a 10 ns MD simulation without ligand (Desmond, OPLS3 force field, Schrödinger version 2017-2). Left: Average hydration sites colored by free energy of binding (ΔG); green spheres indicate favorable sites with low ΔG compared to bulk solvent, red spheres indicate unfavorable sites with high ΔG. Right: View into the binding site with sampled water oxygen atom positions (red atoms) during 10 ns MD as input for clustering and hydration site analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)