

## Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 J. Phys.: Condens. Matter 28 344007

(<http://iopscience.iop.org/0953-8984/28/34/344007>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

### Download details:

IP Address: 163.1.203.121

This content was downloaded on 30/08/2016 at 11:26

Please note that [terms and conditions apply](#).

You may also be interested in:

[Applications of large-scale density functional theory in biology](#)

Daniel J Cole and Nicholas D M Hine

[Water based on a molecular model behaves like a hard-sphere solvent for a nonpolar solute when the reference interaction site model and related theories are employed](#)

Tomohiko Hayashi, Hiraku Oshima, Yuichi Harano et al.

[Designing molecular complexes using free-energy derivatives from liquid-state integral equation theory](#)

Florian Mrugalla and Stefan M Kast

[Protein–protein interactions: organization, cooperativity and mapping](#)

Ozlem Keskin, Buyong Ma, Kristina Rogale et al.

[All-atom simulations of crowding effects on ubiquitin dynamics](#)

Luciano A Abriata, Enrico Spiga and Matteo Dal Peraro

[Iron in biology: a structural viewpoint](#)

Peter F Lindley

# Exploring the role of water in molecular recognition: predicting protein ligandability using a combinatorial search of surface hydration sites

Sinisa Vukovic<sup>1</sup>, Paul E Brennan<sup>2</sup> and David J Huggins<sup>1,3</sup>

<sup>1</sup> Department of Physics, Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge, CB3 0HE, UK

<sup>2</sup> Structural Genomics Consortium and Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, NDM Research Building, Roosevelt Drive, Oxford, OX3 7FZ, UK

<sup>3</sup> Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

E-mail: [djh210@cam.ac.uk](mailto:djh210@cam.ac.uk) (DJH) and [sv375@cam.ac.uk](mailto:sv375@cam.ac.uk) (SV)

Received 29 January 2016, revised 4 April 2016

Accepted for publication 12 April 2016

Published 1 July 2016



## Abstract

The interaction between any two biological molecules must compete with their interaction with water molecules. This makes water the most important molecule in medicine, as it controls the interactions of every therapeutic with its target. A small molecule binding to a protein is able to recognize a unique binding site on a protein by displacing bound water molecules from specific hydration sites. Quantifying the interactions of these water molecules allows us to estimate the potential of the protein to bind a small molecule. This is referred to as ligandability. In the study, we describe a method to predict ligandability by performing a search of all possible combinations of hydration sites on protein surfaces. We predict ligandability as the summed binding free energy for each of the constituent hydration sites, computed using inhomogeneous fluid solvation theory. We compared the predicted ligandability with the maximum observed binding affinity for 20 proteins in the human bromodomain family. Based on this comparison, it was determined that effective inhibitors have been developed for the majority of bromodomains, in the range from 10 to 100 nM. However, we predict that more potent inhibitors can be developed for the bromodomains BPTF and BRD7 with relative ease, but that further efforts to develop inhibitors for ATAD2 will be extremely challenging. We have also made predictions for the 14 bromodomains with no reported small molecule  $K_d$  values by isothermal titration calorimetry. The calculations predict that PBRM1(1) will be a challenging target, while others such as TAF1L(2), PBRM1(4) and TAF1(2), should be highly ligandable. As an outcome of this work, we assembled a database of experimental maximal  $K_d$  that can serve as a community resource assisting medicinal chemistry efforts focused on BRDs. Effective prediction of ligandability would be a very useful tool in the drug discovery process.

Keywords: hydration, proteins, ligandability, drug discovery, bromodomains, water, molecular recognition

Online supplementary data available from [stacks.iop.org/JPhysCM/28/344007/mmedia](http://stacks.iop.org/JPhysCM/28/344007/mmedia)

(Some figures may appear in colour only in the online journal)

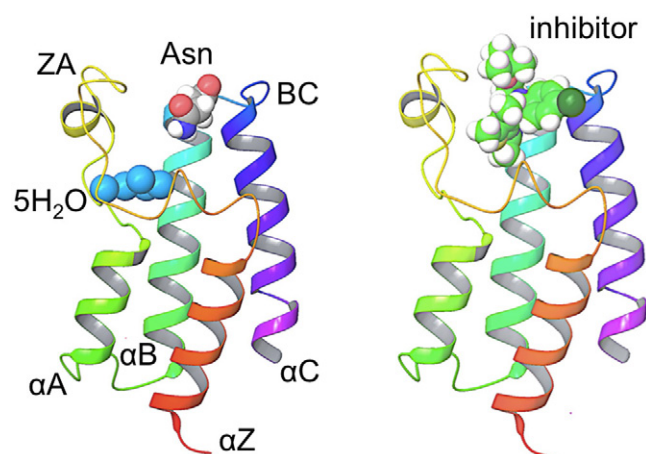


Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## Introduction

Water molecules at protein surfaces can be weakly or strongly bound. Therefore, the binding of a small molecule to a particular region of the surface is affected by the existing network of hydration sites in the absence of the small molecule. Quantifying the propensity of a protein to bind a small molecule provides an estimate of ligandability in drug design. The druggability of a protein target is defined as the relative ease or difficulty of developing a small molecule that can effectively modulate the protein's activity *in vivo* [1]. However, there are many complex pharmacodynamics and pharmacokinetic factors influencing the ability of a small molecule to be effective as a drug [2, 3]. For these reasons predicting druggability is an extremely difficult task [4, 5]. However, a good indicator of protein druggability is protein ligandability. The ligandability of a protein is defined as the relative ease or difficulty of developing a small molecule that can inhibit the protein *in vitro* [6]. Predicting ligandability is somewhat more straightforward and, since it is necessary for druggability (though not sufficient), it is of great interest in drug discovery. For this reason, computational prediction of ligandability is a useful goal. In many cases, ligandability is phrased as the maximal affinity of a ligand for the given target [7–9]. There are a number of computational methods to predict ligandability including MAP<sub>POD</sub> [9], SiteMap [10], fpocket [11], DLID [12], DrugFEATURE [13], LIGSITE [14], Q-SiteFinder [15], i-SITE [16], DrugPred [17], JEDI [18]. All of these methods apply large-scale data analysis to structural, and physicochemical data that is empirically derived using a training sets for ligands, proteins, and protein–ligand complexes.

Methods such as FTMap take a different approach, namely to predict ligandability without any dependence on training set data [19, 20]. FTMap tests different molecular fragments against the entire surface of the protein for locations with the strongest interactions. Here we report a similar approach that uses water molecule instead of fragments. The basis for this approach is the idea that the displacement of water molecules from a protein surface is a key contributor to protein–ligand binding affinity [21–29]. A set of strongly bound waters yields a high desolvation penalty and makes it difficult to bind a ligand with high affinity. The task of predicting ligandability is then the task of identifying the combination of water molecules on the surface of the protein that is the easiest to displace. Whilst quantifying the displaceability of individual water molecules has been reported previously using inhomogeneous fluid solvation theory (IFST) [23, 30–34], quantifying the displaceability of combinations of water molecules has not been reported. This is a challenging combinatorial problem due to the large number of hydration sites surrounding a protein. Approximation methods cannot be applied in this case because selectivity is based on the preference of one single binding site over all other possible sites, such that all possible combinations must be evaluated. Moreover, proteins are dynamic systems and constant changes in their structure are reflected in responsive changes of the solvent molecules on the surface.



**Figure 1.** The bromodomain binding pocket (left) is formed from the conserved fold of helices  $\alpha Z$ ,  $\alpha A$ ,  $\alpha B$  and  $\alpha C$  as well as the connecting loops ZA, AB and BC. It is bounded at the base by five strongly-conserved water molecules and at the top by a conserved asparagine residue. Small molecule inhibitors such as (+)JQ1 (PDB ID 4FLP) bind in this pocket (right).

In this work, we describe a new computational tool to quantify the binding free energy of combinations of water molecules from protein surfaces. We apply the method to the bromodomains (BRDs) family of proteins. BRDs represent a good test case due to the large number of available crystal structures, the limited structural flexibility, and experimental data on small-molecule binding. BRDs are very well structurally characterized. They were identified for the first time in the *brahma* gene of *Drosophila melanogaster* [35] and classified as epigenetic readers of covalent post-translational modifications of chromatin [36, 37]. In particular, the great majority of BRDs are capable of selectively recognizing  $\epsilon$ -N-acetylated lysine residues (Kac) in histone tails [38–40]. BRD possess a structurally conserved interaction module about 110 amino acid long characterized by a left-handed four-helix bundle (referred to as the BRD fold:  $\alpha Z$ ,  $\alpha A$ ,  $\alpha B$  and  $\alpha C$ ) [41] shown in figure 1. The binding site for acetylated-lysine is a hydrophobic pocket made by helices and inter-helical  $\alpha Z$ - $\alpha A$  and  $\alpha B$ - $\alpha C$  loops (ZA and BC loops, respectively). Within the binding pocket lies a highly-conserved asparagine residue at the beginning of BC loop (at the top of the pocket) and four or five conserved water molecules (at the bottom of the pocket) that participate in recognition of Kac [22]. In the last two decades, the binding motif between BRDs and Kac histone has been established [42, 43] and extensively reviewed [44–46]. Based on this protein-protein interaction motif, many inhibitors have been developed [47–50]. Knowing the structure of a protein is very useful in the process of drug discovery and The Structural Genomics Consortium reveals a new BRD structure almost monthly [51].

There are 61 BRDs found in the human proteome appearing within 41 different proteins [43] that, based on their structure, can be divided in eight distinctive subfamilies [52]. Inhibitors have been developed for each of the subfamilies, from those deemed druggable such as BRD4 [53, 54] to those deemed difficult to drug such as BAZ2B [55], including inhibitors with mixed BRD pharmacology [56]. However most BRDs,

excluding BRD4, have not yet been studied extensively and thus their biological role and potential for ligandability is not fully understood. Thus, we apply our method to the BRDs in order to predict their ligandability. Because extensive lead-optimization campaigns have not been prosecuted against many of the BRDs, and in some cases no inhibitors have been reported, we can make prospective predictions on the relative ease or difficulty of developing potent small molecule inhibitors.

We developed a suite of programs to quantify the binding free-energy of combinations of hydration sites from protein surfaces through a combinatorial search. We process information from molecular dynamics (MD) simulation and convert the locations and interaction energies of explicit water molecules into their statistical representation over time, referred to as hydration sites [30]. We then identify which combination of hydration sites is most weakly bound by summing the binding energies of hydration sites in each combination. Once determined, such a hot spot (the location and predicted ligandability) on the surface of one protein can be compared with hot spots on the other proteins. In this work we present details of the procedure for generating individual hydration site and assessing combinations of hydration sites as well as comparing the predicted and experimentally available results for BRDs. We have also compiled a list of the maximal reported affinities for each BRD, reported as  $K_d$  values measured using isothermal calorimetry (ITC), and discuss the potential for improvement in each case as well as the potential of BRD family members with no reported data. We hope that this tool will prove useful, quite early in the drug discovery process, in distinguishing proteins that are ligandable from those where developing a potent small molecule inhibitor will be challenging.

It is important to state that analysis of hydration sites around a protein only looks at part of the story for small molecule binding. Obviously, one must also consider direct interaction of a small molecule with the protein and the ligand in solution. In a hydrophobic pocket, solvent molecules can be displaced from the surface of the protein with little cost and the corresponding non-polar interactions formed with the ligand are of modest size. However, strongly bound water molecules near polar protein groups are much harder to displace, but the corresponding polar interactions formed with the ligand allow much greater energetic compensation. This work does not consider these compensation effects, but relies on the relationship between surface hydration sites and ligand displaceability established previously [30].

## Methods

We identify the location of the combination of hydration sites with the weakest binding affinity with the ligand-binding hotspot and the summed binding free energy of the hydration sites with the potential for ligandability [30]. The procedure of locating such hotspots on the surface of a protein is performed in two steps. Firstly, hydration sites around the protein are generated from MD simulation. Secondly, combinations of these hydration sites are compared to identify the

combination with the weakest total binding free energy. The method of generating hydration sites and calculating binding free energies using IFST has been reported before [30, 57–59]. However, the evaluation of all possible combinations of hydration is reported here for the first time. Details of both steps of the method are given below. In terms of computer time, the first stage of hydration site analysis can be completed within 24 h on a 16 CPU machine and the second stage of assessing all combinations can be completed within a few minutes or a few days on a 4 CPU machine, depending on the number of hydration sites in total and the number of hydration sites per combination. We apply the method to the BRD family. The list of BRDs whose structures have been processed are given in tables 1 and 2.

### Protein structure preparation

To begin, downloaded protein structures from the Protein Databank [60] were prepared for analysis. We selected only apo protein structures for this analysis. We did not select crystal structures in complex with small molecules, because we are interested in applying the method to potential drug targets lacking existing inhibitors. Downloaded structures were initially processed by Schrodinger's Preparation Wizard [61] by changing seleno-methionines to methionines, adding side chains when they were missing, checking for orientations of the asparagine, glutamine and histidine residues, checked for protonation state of all ionizable residues and removing all metal, buffer and solvent molecules. Proteins were solvated by SOLVATE<sup>4</sup> and we used the TIP4P-2005 water model [62]. Overall neutrality of the system was preserved by adding chloride and sodium ions placed at least 10.0 Å from the protein.

### Molecular dynamics simulations

Prepared structures were then subjected to MD simulation in water. Simulations were performed using NAMD version 2.9 [63]. Force field parameters were from the CHARMM27 force field [64, 65] with heavy-atom restraints of 1.0 kcal mol<sup>-1</sup> Å<sup>-2</sup> for the protein and ions. Equilibration was performed for 1.0 ns in an NPT ensemble at 300 K using Langevin temperature control [66]. It was checked that the system was equilibrated by verifying that energy fluctuations were stable. MD simulations were performed using a time step of 2.0 fs, saving coordinates every 200 steps (400 fs), for a total run of 3.6 ns. Throughout the equilibration and dynamics runs, electrostatic interactions were modeled with a uniform dielectric and a dielectric constant of 1.0. Van der Waals interactions were truncated at 11 Å with switching from 9 Å. The electrostatic model was the particle mesh Ewald method [67]. The entire systems, protein plus solvent, was treated using rhombic dodecahedral periodic boundary conditions which extended at least 10 Å from the protein surface.

<sup>4</sup> SOLVATE is written by Helmut Grubmüller Theoretical Biophysics Group Institut für Medizinische Optik Ludwig-Maximilians-Universität München München Germany. Copyright © 1996–2013. All rights reserved.



### Hydration site identification

Since water molecules tend to cluster at protein surfaces and concentrate at distinct locations [57, 69], IFST is usefully applied to small subvolumes referred to as hydration sites [31–34, 70–74]. Hydration sites are then individual spherical regions with a high number density of water. Individual hydration sites are identified by clustering water molecules from the MD simulation. In this case, the clustering procedure was applied to a radius of 16 Å around the BRD binding site with a hydration site radius of 1.2 Å. Hydration sites were determined iteratively, starting with the water molecule in the region of highest water density. No hydration site was selected that was 2.4 Å within of an already selected hydration site. Detailed procedures for clustering have been described previously [30, 57].

### IFST calculations

Trajectories from the simulation were then analyzed using IFST to obtain the of hydration site free energies. IFST calculations are performed on simulation data from equilibrium MD trajectories and thus the chemical potential of water is constant throughout the solution. The free energy change calculated by IFST is the solvation free energy of the solute. An IFST calculation on the whole system volume yields the total solvation free energy. An IFST calculation on a hydration site yields the contribution of water molecules in that hydration site to the solvation free energy. In the context of IFST, the end states are a system of bulk water with  $N$  solvent particles and a solution with  $N$  solvent particles. Thus, the calculations can be considered to capture the perturbing effect of the solute on the water molecules [58, 59, 68]. The binding free energy of a hydration site ( $\Delta G_{\text{IFST}}$ ) is defined as the difference in free energy between the water molecules in the presence of the protein and the same number of water molecules in pure bulk solution. Unlike alchemical free-energy calculations such as double decoupling, IFST calculations never consider a state with an artificial cavity.  $\Delta G_{\text{IFST}}$  is calculated from the energetic ( $\Delta E_{\text{IFST}}$ ) and entropic ( $\Delta S_{\text{IFST}}$ ) contributions to free energy:

$$\Delta G_{\text{IFST}} = \Delta E_{\text{IFST}} - T \Delta S_{\text{IFST}} \quad (1)$$

$\Delta E_{\text{IFST}}$  is the contribution of water molecules in a hydration site to the total hydration energy and can be calculated from average solute–water ( $E_{\text{sw}}$ ), water–water ( $E_{\text{ww}}$ ), and bulk solution ( $E_{\text{bulk}}$ ) energies:

$$\Delta E_{\text{IFST}} = E_{\text{sw}} + E_{\text{ww}} - E_{\text{bulk}} \quad (2)$$

$\Delta S_{\text{IFST}}$  is the contribution of water molecules in a hydration site to the total hydration entropy. The entropy is based on a multiparticle expansion of the molecular distribution functions [75, 76]. IFST is based on the two-particle approximation and thus the entropy expansion is truncated at the two-particle level [58]. We use a truncation of Wallace’s expansion for the entropy in the canonical ensemble [77].  $\Delta S_{\text{IFST}}$  can then be calculated from solute–water ( $S_{\text{sw}}$ ), water–water ( $S_{\text{ww}}$ ), and bulk solution ( $S_{\text{bulk}}$ ) entropy terms:

$$\Delta S_{\text{IFST}} = S_{\text{sw}} + S_{\text{ww}} - S_{\text{bulk}} \quad (3)$$

It is important to note that the excess energy and the excess entropy of solution depend on the ensemble, but that the ensemble dependent terms cancel in the expression for the solute chemical potential. Thus, the chemical potential is equal for insertion at constant volume and constant pressure [78, 79]. We have shown previously that  $S_{\text{ww}}$  is negligible for ordered water molecules surrounding a protein and thus we set this term to zero [80]. We exploit a  $k$ -nearest neighbours approach to calculate  $S_{\text{sw}}$  using the first nearest neighbour ( $k = 1$ ) in all cases [29, 72, 80]:

$$S_{\text{sw}} = nR \left\{ \frac{1}{nF} \sum_{i=1}^F \sum_{j=1}^n \ln \left[ \frac{\pi d_{\text{total}}^6 nF}{48 V_i} \right] + \gamma \right\} \quad (4)$$

$$d_{\text{total}} = \sqrt{d_{\text{total}}^2 + d_{\text{orient}}^2} \quad (5)$$

$$d_{\text{trans}} = \sqrt{(x_{ij} - x_{kl})^2 + (y_{ij} - y_{kl})^2 + (z_{ij} - z_{kl})^2} \quad (6)$$

$$d_{\text{orient}} = 2 \times \text{acos}(|\mathbf{q}_{ij} \cdot \mathbf{q}_{kl}|) \quad (7)$$

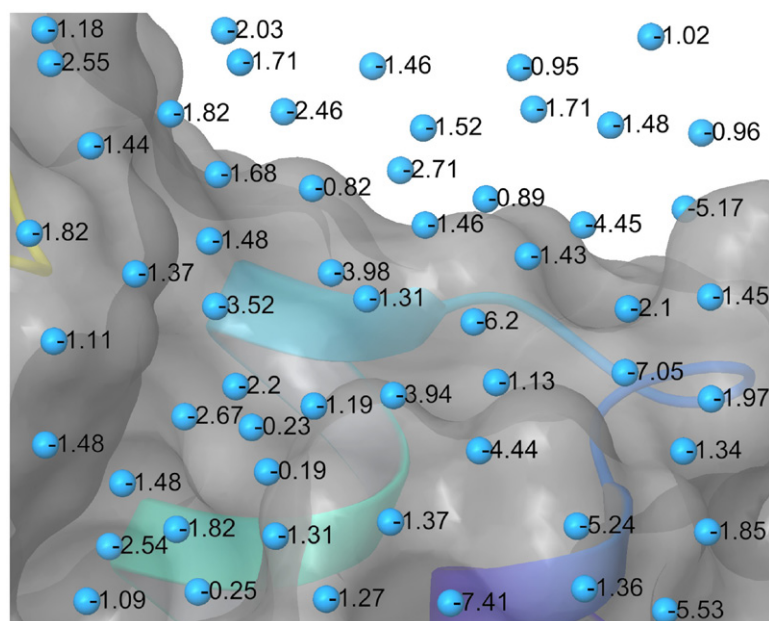
$R$  is the gas constant,  $F$  is the number of frames sampled,  $V_i$  is the volume of the system in frame  $i$ , and  $\gamma$  is Euler’s constant, which corrects for the asymptotic bias. The nearest neighbour distance in translational space ( $d_{\text{trans}}$ ) between two water molecules is the Euclidean norm between the Cartesian coordinates of water molecule  $j$  in frame  $i$  and its nearest neighbour water molecule  $k$  in frame  $l$ . The nearest neighbour distance in orientational space ( $d_{\text{orient}}$ ) between two water molecules is the distance between the rotations required to bring the two orientations to the same reference orientation. The correct distance metric for the rotation group is twice the geodesic distance on the unit sphere. This quaternion representations of the rotations for water molecule  $j$  in frame  $i$  and its nearest neighbour water molecule  $k$  in frame  $l$  are denoted by  $\mathbf{q}_{ij}$  and  $\mathbf{q}_{kl}$ . The total solute–water distance ( $d_{\text{total}}$ ) is estimated by combining the translational and orientational distances.

For an individual hydration site,  $\Delta G_{\text{IFST}}$  is a predominantly negative number between 0 and  $-15.0 \text{ kcal mol}^{-1}$  [80] and tends to zero with increasing distance from the protein (e.g. for water molecules 7–8 Å from the protein) where interactions become more like those in bulk solvent. For the TIP4P-2005 water model  $\Delta E_{\text{bulk}}$  is  $-11.57 \text{ kcal mol}^{-1}$  and  $\Delta S_{\text{bulk}}$  is  $15.5097 \text{ cal K}^{-1} \text{ mol}^{-1}$  [80]. Hydration sites free energies at the surface of CECR2 are shown in figure 2.

For each protein, 9000 snapshots from molecular dynamics runs were processed in order to obtain  $\Delta G_{\text{IFST}}$  for the hydration sites.

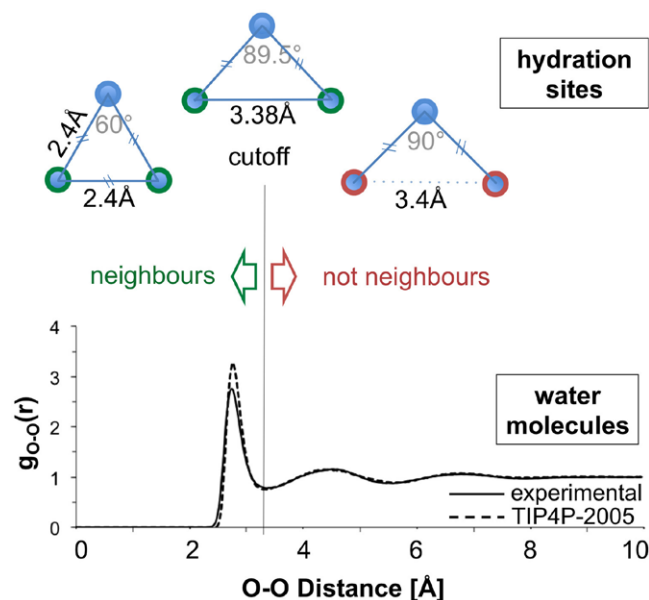
### Generation of combinations of hydration sites

Combinations of interest are those that include hydration sites on the surface of the protein. Therefore, the cut off distance was applied at 3.6 Å from the surface of the protein (using heavy atom to heavy atom distance) and was extended to 4.1 Å when the hydration site was over a hydrophobic residue. The reason for the extended cut off was that hydration sites are found at a further distance from the protein for hydrophobic residues.



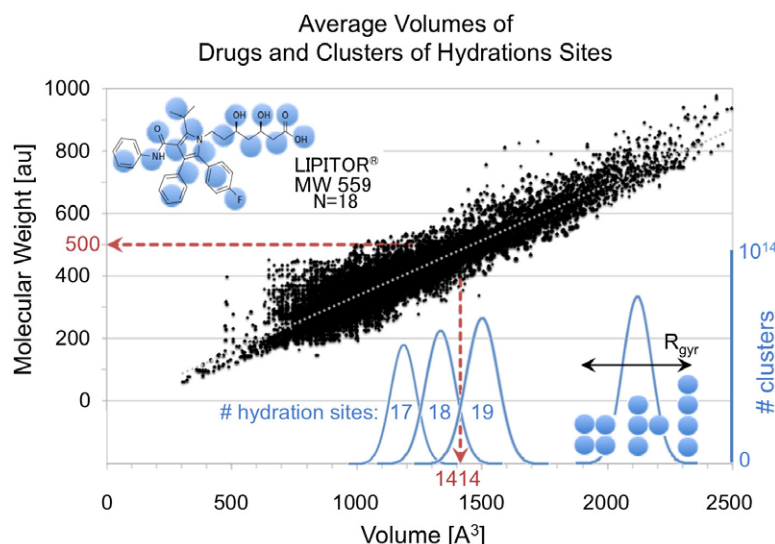
**Figure 2.** Values of  $\Delta G_{\text{HST}}$  in  $\text{kcal mol}^{-1}$  for hydration sites above the surface of the binding site of CECR2 (PDB ID 3NXB).

Choosing combinations of hydration sites is in principle very simple (i.e. take combinations of hydration sites like those in figure 2). However, the technical implementation of generating, evaluating, and storing such combinations is a challenging problem. Small proteins like BRDs have a surface covered by about 2000 hydration sites (by the definition above). If one is to take all possible combinations of combinations of hydration sites then the number of combinations is in excess of  $10^{40}$ , which is not possible to process. This number can be significantly smaller if one only considers combinations that enclose hydration sites within a contiguous volume (since a single small molecule inhibitor must be contiguous). In order to avoid generating all possible combinations and then checking if each combination of hydration sites can be enclosed by a contiguous volume, a different approach was taken. Prior to generation of combinations, every hydration site was evaluated for the number of its neighboring hydration sites. The cut off distance for two hydration sites to be considered neighbors was taken to be 3.38 Å. This distance is a reflection of the hydration site being a statistical representation of water molecules. In figure 3 it is shown that this distance is the cut off for O–O distance in the first solvation shell for TIP4P-2005 water model and that at this distance two hydration sites cannot be considered neighbors since the third one can be placed in-between them (which happens because hydration sites are never closer than 2.4 Å from each other). The consequence of this algorithm was that the resulting combinations were constituted of hydration sites that each had at least one neighbor (which guaranteed that a combination constituted a contiguous volume). For BRDs, the total number of such combinations did not exceed  $10^{14}$ . The number is still large yet cannot be smaller because in the search for the hot spot it is important not to miss a single combination. The reason for such exhaustiveness is to mimic proteins' ability to have only one hot spot for the substrate among all possible spots on its surface. The method must then identify this site.

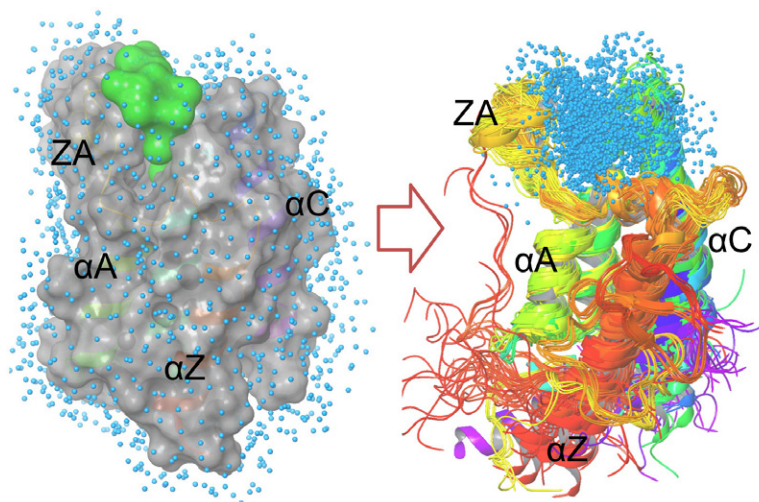


**Figure 3.** The distance cutoff between two neighboring hydration sites is 3.38 Å. Two hydration sites cannot be considered neighbors if there is the third hydration site between them.

One can perform the search for combinations containing any number of hydration sites. Combinations of 4–6 hydration sites would be more appropriate for screening of fragments, while combinations of 30–40 hydration sites would correspond to small peptides. BRDs were first investigated using combinations from 4 to 20 hydration sites. To identify an appropriate number of hydration sites for binding of a drug-like molecule, we calculated the volumes of hydration site combinations of increasing size and compared them with the volume of a molecule of molecular weight 500 au (figure 4). The linear correlation between the molecular weight and the volume of a



**Figure 4.** A plot of the molecular weight against the volume enclosed by the solvent accessible surface area for 250 000 structures from the BindingDB [81]. Molecules are shown as black dots. An approximation of the distribution of volumes available to combinations of hydration sites of a given size is shown in blue (for 17, 18 and 19 hydration sites).



**Figure 5.** The predicted hot spot (green surface enclosing a combination of 18 hydration sites) for CECR2 from PDB ID 3NXB (left) and superimposed hot spots (blue; 18 hydration sites per combination) for all 165 chains from 34 BRD structures (right).

drug ( $R^2 = 0.85$ ) yielded an average volume of  $1414 \text{ Å}^3$  for a molecule of molecular weight 500 au. The number of hydration sites per combination that resulted in the volume closest to that size was for a combination with 18 hydration sites. Thus, we use combinations of 18 hydration sites for the final analysis reported here, since we estimate that this is the number of water molecules that a molecule of molecular weight 500 au would displace on average. It is interesting to note that a linear correlation between molecular weight and surface accessible solvent area has been reported previously [9].

The binding free energy of every combination ( $\Delta G_{\text{sum}}$ ) was calculated as the sum of free energies for the 18 constituent hydration sites.  $\Delta G_{\text{sum}}$  for a set of  $N$  hydration sites is the contribution of water molecules in these  $N$  hydration site to the hydration free energy of the protein. The binding

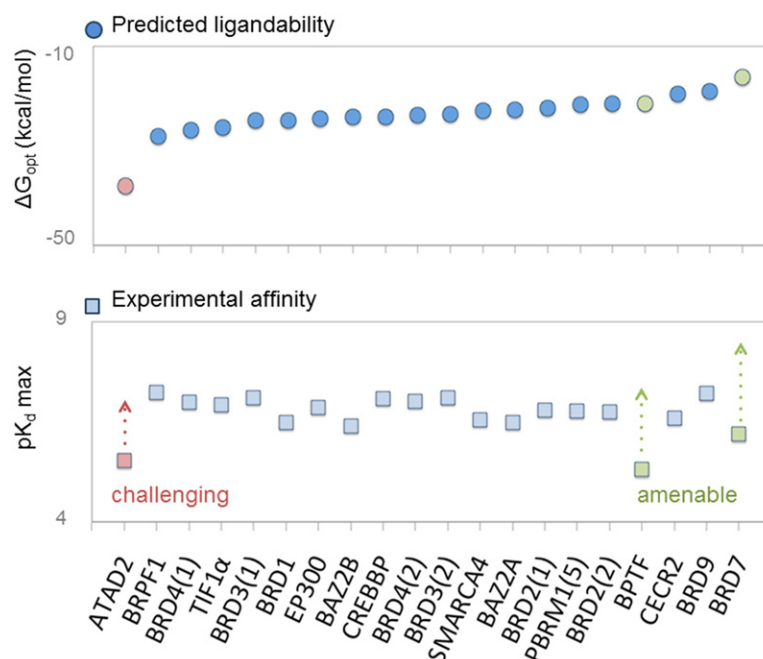
free energy of the optimal combination ( $\Delta G_{\text{opt}}$ ) is then the maximal value of  $\Delta G_{\text{sum}}$  amongst all combinations.

$$\Delta G_{\text{sum}} = \sum_{n=1}^{18} \Delta G_{\text{IFST}}(n) \quad (8)$$

$$\Delta G_{\text{opt}} = \max(\Delta G_{\text{sum}}) \quad (9)$$

## Results

For each of 34 BRDs whose structures we studied (165 protein chains), we enumerated all possible combinations of 18 contiguous hydration sites. In each case, the combination with the weakest binding energy is referred to as the hot spot. Thus, every protein has a hot spot, but the question is how ligandable



**Figure 6.** Comparison between predicted ligandability calculated as  $\Delta G_{\text{opt}}$  (kcal mol<sup>-1</sup>) for combinations of size  $N = 18$  (circles) and the maximum observed  $pK_d$  values measured by ITC (squares).

**Table 1.** The values of  $\Delta G_{\text{opt}}$  for combinations of 18 hydration sites and the best values for  $K_d$  measured by ITC.

BRD	$\Delta G_{\text{opt}} (N = 18)$ (kcal mol <sup>-1</sup> )	Max $pK_d$	Max $K_d$ (ITC) (nM)	Ligand MW	$K_d$ ref.
ATAD2	-38.05	5.82	1500	350.40	[83]
BRPF1	-28.06	7.89	13	380.45	[84]
BRD4(1)	-26.83	7.59	26	423.17	[85]
TIF1 $\alpha$	-26.35	7.51	31	643.00	[86]
BRD3(1)	-24.91	7.72	19	423.17	[85]
BRD1	-24.87	6.97	108	383.42	[87]
EP300	-24.53	7.42	38	508.20	[88]
BAZ2B	-24.22	6.87	136	371.00	[89]
CREBBP	-24.16	7.68	21	508.20	[88]
BRD4(2)	-23.93	7.60	25	414.50	[90]
BRD3(2)	-23.77	7.72	19	423.17	[85]
SMARCA4	-22.94	7.05	89	321.40	[91]
BAZ2A	-22.78	6.96	109	357.42	[92]
BRD2(1)	-22.44	7.34	46	423.17	[85]
PBRM1(5)	-21.66	7.32	48	321.40	[91]
BRD2(2)	-21.53	7.28	52	423.17	[85]
BPTF	-21.50	5.55	2800	187.00	[93]
CECR2	-19.63	7.10	80	495.70	[94]
BRD9	-19.07	7.85	14	353.40	[95]
BRD7	-16.25	6.62	239	353.40	[95]

is it. As an illustration, the hot spot for the BRD of CECR2 is shown in figure 5 (green surface in the left-hand panel). Importantly, a correspondence between the locations of the predicted hot spots (blue spheres in the right-hand panel of figure 5), and the actual Kac binding sites (shown in figure 1) was observed for all 165 structures. This observation provides a basic qualitative validation of the approach.

We calculated  $\Delta G_{\text{opt}}$  for each of the BRDs and compared this with the maximum reported  $pK_d$  value obtained by ITC [48, 82]. This data is summarized in table 1.

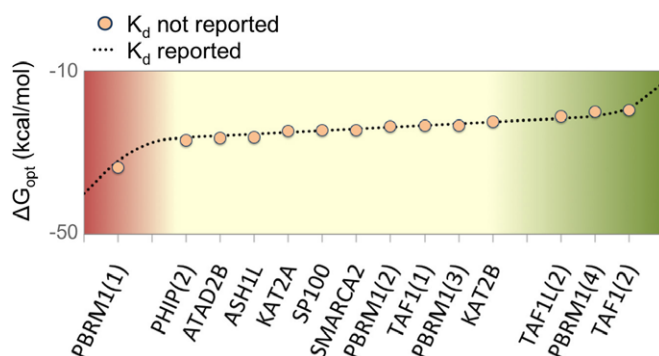
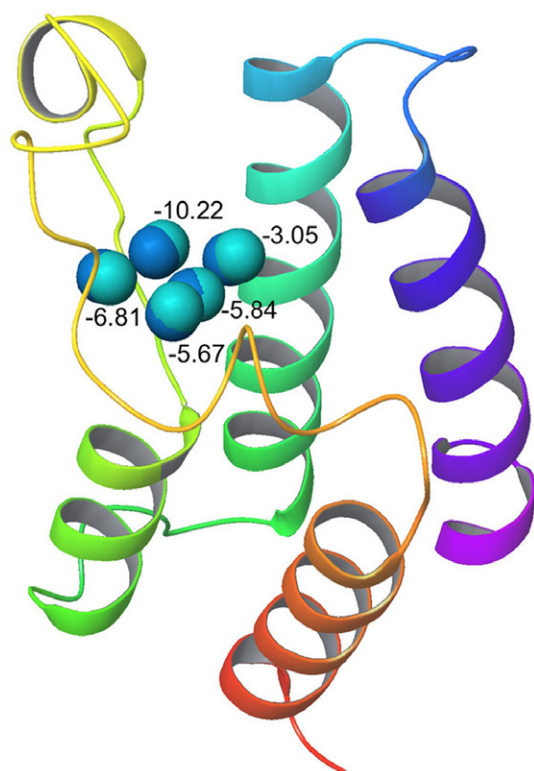
A graphical representation of the data can be seen in figure 6.

The majority of BRDs are predicted to have similar ligandability and correspondingly most BRDs have experimental  $K_d$  values in the range of 10–100 nM. However, BPTF and



**Table 2.** The best values for  $\Delta G_{\text{opt}}$  of combinations of 18 hydration sites for BRDs whose values for  $K_d$  measured by ITC were not reported.

BRD	$\Delta G_{\text{opt}} (N = 18)$ (kcal mol <sup>-1</sup> )	BRD	$\Delta G_{\text{opt}} (N = 18)$ (kcal mol <sup>-1</sup> )
PBRM1(1)	-33.66	PBRM1(2)	-23.65
PHIP(2)	-26.93	TAF1(1)	-23.40
ATAD2B	-26.49	PBRM1(3)	-23.29
ASH1L	-26.17	KAT2B	-22.50
KAT2A	-24.73	TAF1L(2)	-21.10
SP100	-24.57	PBRM1(4)	-19.87
SMARCA2	-24.56	TAF1(2)	-19.60

**Figure 7.** Predictions for ligandability of BRDs with no  $K_d$  values measured by ITC (orange circles) and the correspondence for BRDs whose  $K_d$  values are known (dotted line).**Figure 8.** The crystal structure for CECR2 (PDB ID 3NXB), showing five crystallographic water molecules (dark blue) overlapping with predicted hydration sites (light blue). Hydration sites are labeled by their  $\Delta G_{\text{IFST}}$  values.

BRD7 (green circles and squares) are different, having experimental  $K_d$  values greater than 100 nM. Based on the calculated  $\Delta G_{\text{opt}}$ , the prediction is that more potent inhibitors could be developed for these BRDs without significant difficulty. In particular, BRD7 is predicted to be a highly ligandable target. Conversely, ATAD2 (red circle and square) is predicted to have a relatively low ligandability and the prediction is that significant effort will be required to develop inhibitors approaching 10 nM. As a product of this work, we offer a dataset of maximal  $pK_d$  values, hoping it will serve as a community resource in drug discovery efforts focused on BRDs ([www.tcm.phy.cam.ac.uk/~sv375/BRDs/table/Kd\\_for\\_BRDs.html](http://www.tcm.phy.cam.ac.uk/~sv375/BRDs/table/Kd_for_BRDs.html)). We also considered 14 BRDs that have protein structures deposited in the PDB but do not have reported  $K_d$  values from ITC data. This data is summarized in table 2.

A graphical representation of the data can be seen in figure 7.

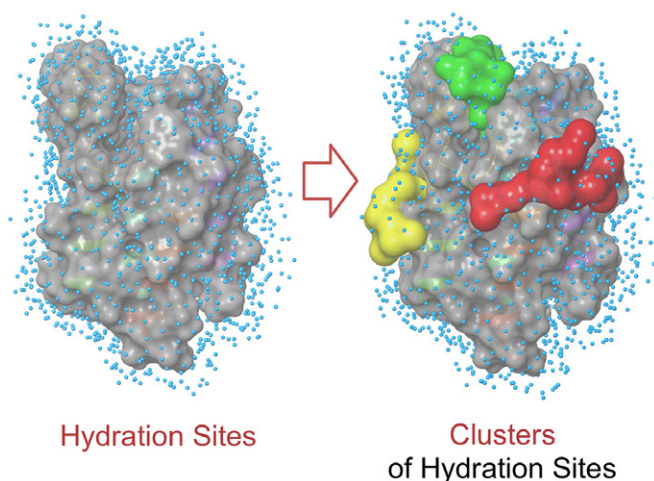
The trend observed for BRDs that do have reported  $K_d$  values from ITC data (figure 7 dotted line) suggests that identifying inhibitors with  $K_d$  in the range of 10–100 nM will not be a significant challenge for most other BRDs. However, four of these BRDs are notably different. PBRM1(1) is predicted to be a relatively difficult target (figure 7 red region), whereas TAF1L(2), PBRM1(4), and TAF1(2) are predicted to be a relatively ligandable targets (figure 7 green region). In addition to predicting ligandability, we also considered the five ‘conserved’ water molecules at the bottom of the Kac binding pocket (figure 1). Our analysis placed hydration sites that reproduced the positions of these conserved water molecules for each BRD. The  $\Delta G_{\text{IFST}}$  values are also strongly negative, suggesting that they are tightly bound. An example is shown for CECR2 in figure 8. The search for hot spots tends not to favor selection of these hydration sites.

## Discussion

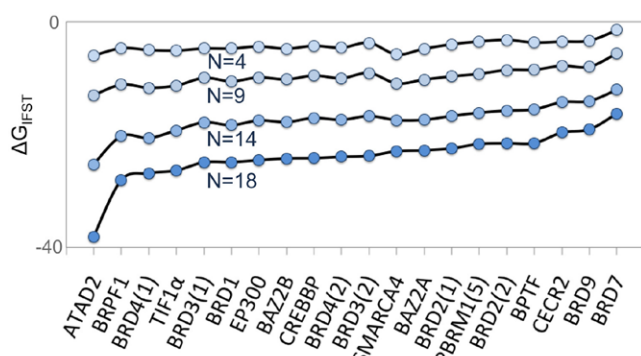
The primary technical challenge in this work is the enumeration of the trillions of viable combinations of hydration sites to identify the most weakly bound. This is illustrated for CECR2 in figure 9. The most weakly bound combination (shown in green) is located in the binding site, while the most strongly bound (red) is outside the binding site. This is the case for all 165 BRDs structures tested and is an encouraging result.

It is also important to discuss the assumptions, advantages, and limitations of the method. We chose combinations of 18 hydration sites because the volume of such combinations corresponds to a volume of a drug with molecular weight of 500 (figure 4). However, we studied whether the number of hydration sites influences the predicted ligandability. Trends observed with other number of hydration sites have also been computed and are presented in figure 10.

We observe that, in general, trends in relative ligandability (values of  $\Delta G_{\text{opt}}$ ) are almost unchanged as the size of the combination changes. We observe that hot spots tend to be built around a small group of hydration sites that are very

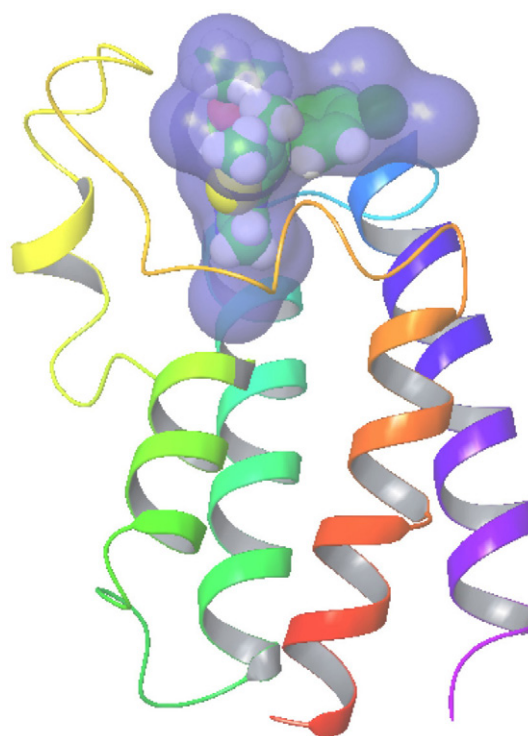


**Figure 9.** Combinations of hydration sites on CECR2 (PDB ID 3NXXB) with the lowest (green), intermediate (yellow) and highest (red)  $\Delta G_{\text{opt}}$  values. The most weakly bound combination (the hot spot) is in the Kac binding site.



**Figure 10.**  $\Delta G_{\text{opt}}$  for BRDs as the number of hydration sites changes (number of hydration sites  $N = 4, 9, 14, 18$ ).

weakly bound. For example, ATAD2 always has the lowest and BRD7 always has the highest  $\Delta G_{\text{opt}}$  value. However the relative ordering does change for smaller combinations ( $N < 10$ ). The interpretation of this observation is that success in fragment screening might not always correspond to ligandability for a 500 MW inhibitor. The method described here can detect such scenarios and inform decision making in fragment-based drug design. It can also inform decision making in the hit-to-lead and lead-optimization stages of drug design. Even if a protein is ligandable and worth pursuing as a therapeutic target, major challenges remain in identifying an inhibitor *in vitro* and then optimizing it for use *in vivo*. An additional advantage of the method described here is that the results present a template for molecular design, suggesting the position, shape and characteristics of the eventual inhibitor. This information embedded in the results of the method is very useful in approaching the first of these challenges. An example is shown for BRDT in figure 11, with an inhibitor overlaying the most weakly bound combination of hydration sites. The position of the hot spot can also be used to predict when allosteric inhibition might be effective, while the 3D shape of the hot spot can be used to deduce vector poses for de novo drug design.



**Figure 11.** Volume overlap for the predicted hot spot (purple) and the (+)-JQ1 ligand in BRDT (PDB ID 4FLP).

Whilst the BRDs are adequate for testing a proof of concept, the applicability of the method to more diverse structures of protein families is yet to be established. BRDs are relatively new therapeutic target and most of those tested did not undergo systematic lead optimization campaigns. Thus, the currently available maximum  $K_d$  values are not reflective of the true potential. Testing this method against the well-studied set of protein targets [9] would therefore be a logical progression, and is the subject of our current work. It will also be important to study the compensating effect of ligand-protein interactions noted above, as this was not included in the current analysis. In addition, it would be interesting to quantitatively assess the two-particle approximation used in IFST, to develop more accurate models. However, these are useful goals, because knowing the potential of a protein to be ligandable helps focus drug discovery efforts on targets that can be inhibited effectively thus saving time, effort and money.

## Conclusion

We have presented a new method capable of generating and evaluating all possible combinations of hydration sites on the surface of a protein. Based on the identification and quantification of the most weakly bound combination of hydration sites, we predicted the ligandability of a set of BRDs. The majority of BRDs are predicted to have similar ligandability. This is in line with the maximal observed  $K_d$  values determined by ITC for BRDs, which are mostly in the 10–100 nM range. The best target for binding a small

molecule was predicted to be BRD9 and the worst ATAD2. This is in good agreement with the maximal observed  $K_d$  values determined by ITC for BRD9 (14 nM) and ATAD2 (1500 nM). Based on the maximal observed  $K_d$  values, it was concluded that the majority of BRDs have good inhibitors already developed. There are two notable cases where the maximal observed  $K_d$  values are lower than might be expected based on their predicted ligandability. Thus we predict that inhibitors for BPTF and BRD7 in the 10–100 nM range will be developed without significant difficulty. We also predicted the ligandability of the 14 BRDs with protein structures available in the PDB but with no  $K_d$  values determined by ITC. As with the majority of other BRDs, we expect that it will not be particularly challenging to develop inhibitors in the 10–100 nM range. However, we predict that significant effort will be required to develop inhibitors of PBRM1(1). Conversely, TAF1L(2), PBRM1(4) and TAF1(2) are expected to be highly ligandable targets. As a final result, we have assembled an experimental dataset of maximal binding affinities for the BRDs, which we hope will serve as a community resource to assist in medicinal chemistry efforts against this target family ([www.tcm.phy.cam.ac.uk/~sv375/BRDs/table/Kd\\_for\\_BRDs.html](http://www.tcm.phy.cam.ac.uk/~sv375/BRDs/table/Kd_for_BRDs.html)).

## Acknowledgments

We would like to thank S Knapp and O Fedorov for helpful discussions and the Cambridge High Performance Computing Service for technical support. Work in the DJH laboratory is supported by the Medical Research Council under grant ML/L007266/1. All calculations were performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service ([www.hpc.cam.ac.uk/](http://www.hpc.cam.ac.uk/)) provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England, and were funded by the EPSRC under grants EP/F032773/1 and EP/J017639/1.

## References

- [1] Hopkins L A and Groom C R 2002 The druggable genome *Nat. Rev. Drug Discovery* **1** 727–30
- [2] Hajduk P J, Huth J R and Tse C 2005 Predicting protein druggability *Drug Discovery Today* **10** 1675–82
- [3] Overington J P, Al-Lazikani B and Hopkins A L 2006 How many drug targets are there? *Nat. Rev. Drug Discovery* **5** 993–6
- [4] Brown S P and Muchmore S W and Hajduk P J 2009 Healthy skepticism: assessing realistic model performance *Drug Discovery Today* **14** 420–7
- [5] Fauman E B, Rai B K and Huang E S 2011 Structure-based druggability assessment—identifying suitable targets for small molecule therapeutics *Curr. Opin. Chem. Biol.* **15** 463–8
- [6] Edfeldt F N B and Folmer R H A and Breze A L 2011 Fragment screening to predict druggability (ligandability) and lead discovery success *Drug Discovery Today* **16** 284–7
- [7] Kuntz I D and Chen K, Sharp K A and Kollman P A 1999 The maximal affinity of ligands *Proc. Natl Acad. Sci. USA* **96** 9997–10002
- [8] Surade S and Blundell T L 2012 Structural biology and drug discovery of difficult targets: the limits of ligandability *Chem. Biol.* **19** 42–50
- [9] Cheng C A, Coleman G R, Smyth T K, Cao Q, Soullard P, Caffrey R D, Salzberg C A and Huang S E 2007 Structure-based maximal affinity model predicts small-molecule druggability *Nat. Biotechnol.* **25** 71–5
- [10] Halgren T A 2009 Identifying and characterizing binding sites and assessing drugability *J. Chem. Inf. Model.* **49** 377–89
- [11] Schmidtke P and Barril X 2010 Understanding and predicting druggability. A high-throughput method for detection of drug binding sites *J. Med. Chem.* **53** 5858–67
- [12] Sheridan R P, Maiorov V N, Holloway M K, Cornell W D and Gao Y D 2010 Drug-like density: a method of quantifying the ‘bindability’ of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank *J. Chem. Inf. Model.* **50** 2029–40
- [13] Liu T and Altman R B 2014 Identifying druggable targets by protein microenvironments matching: application to transcription factors *CPT Pharmacometrics Syst. Pharmacol.* **3** 1–9
- [14] Hendlich M, Rippmann F and Barnickel G 1997 LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins *J. Mol. Graph. Model.* **15** 359–63
- [15] Laurie A T and Jackson R M 2005 Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites *Bioinformatics* **21** 1908–16
- [16] Morita M, Namakura S and Shimizu K 2008 Highly accurate method for ligand-binding site prediction in unbound state (apo) protein structures *Proteins* **73** 468–79
- [17] Krasowski A, Muthas D, Sarkar A, Schmitt S and Brenk R 2011 DrugPred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set *J. Chem. Inf. Model.* **51** 2829–42
- [18] Cuchillo R, Pinto-Gil K and Michel J A 2015 Collective variable for the rapid exploration of protein druggability *J. Chem. Theory Comput.* **11** 1292–307
- [19] Brenke R, Kozakov D, Chuang G Y, Beglov D, Hall D, Landon M R, Mattos C and Vajda S 2009 Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques *Bioinformatics* **25** 621–7
- [20] Kozakov D, Hall D R, Napoleon R L, Yueh C, Whitty A and Vajda S 2015 New frontiers in druggability *J. Med. Chem.* **58** 9063–88
- [21] Lemieux R 1996 How water provides the impetus for molecular recognition in aqueous solution *Acc. Chem. Res.* **29** 373–80
- [22] Owen D J, Ornaghi P, Yang J-C, Lowe N, Evans P R, Ballario P, Neuhaus D, Filetici P and Travers A A 2000 The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p *EMBO J.* **19** 6141–9
- [23] Li Z and Lazaridis T 2003 Thermodynamic contributions of the ordered water molecule in HIV-1 protease *J. Am. Chem. Soc.* **125** 6636–7
- [24] Abel R, Wang L, Friesner R A and Berne B J 2010 A Displaced-solvent functional analysis of model hydrophobic enclosures *J. Chem. Theory Comput.* **6** 2924–34
- [25] Hummer G 2010 Under water’s influence *Nat. Chem.* **2** 906–7
- [26] Beuming T, Che Y, Abel R, Shanmugasundaram V and Sherman W 2011 Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization *Proteins* **80** 871–83



- [27] Graziano G 2012 Molecular driving forces of the pocket-ligand hydrophobic association *Chem. Phys. Lett.* **533** 95–9
- [28] Mondal J, Friesner R A and Berne B J 2014 Role of desolvation in thermodynamics and kinetics of ligand binding to a kinase *J. Chem. Theory Comput.* **10** 5996–705
- [29] Nguyen C N, Cruz A, Gilson M K and Kurtzman T 2014 Thermodynamics of water in an enzyme active site: grid-based hydration analysis of coagulation factor Xa *J. Chem. Theory Comput.* **10** 2769–80
- [30] Haider K and Huggins D J 2013 Combining solvent thermodynamic profiles with functionality maps of the Hsp90 binding site to predict the displacement of water molecules *J. Chem. Inf. Model.* **53** 2571–86
- [31] Li Z and Lazaridis T 2006 Thermodynamics of buried water clusters at a protein–ligand binding interface *J. Phys. Chem. B* **110** 1464–75
- [32] Young T, Abel R, Kim B, Berne B J and Friesner R A 2007 Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding *Proc. Natl Acad. Sci. USA* **104** 808–13
- [33] Robinson D D, Sherman W and Farid R 2010 Understanding kinase selectivity through energetic analysis of binding site waters *ChemMedChem* **5** 618–27
- [34] Huggins D J, Marsh M and Payne M C 2011 Thermodynamic properties of water molecules at a protein–protein interaction surface *J. Chem. Theory Comput.* **7** 3514–22
- [35] Tamkun J W, Deuring R, Scott M P, Kissinger M, Pattatucci A M and Kaufman T C 1992 Brahma: a regulator of drosophila homeotic genes structurally related to the yeast transcriptional factor SNF2/SWI2 *J. Med. Chem.* **68** 561–72
- [36] Jacobson R H, Ladurner A G, King D S and Tjian R 2000 Structure and function of a human TAF<sub>II</sub>250 bromodomain module *Science* **288** 1422–5
- [37] Fischle W, Wang Y M and Allis C D 2003 Histone and chromatin cross-talk *Curr. Opin. Cell Biol.* **15** 1172–83
- [38] Allfrey V G, Faulkner R and Mirsky A E 1964 Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis *Proc. Natl Acad. Sci.* **51** 786–94
- [39] Gershey E L, Vidali G and Allfrey V G 1968 Chemical studies of histone acetylation *J. Biol. Chem.* **243** 5018–22
- [40] Dhalluin C, Carlson J E, Zeng L, He C, Aggarwal A K and Zhou M-M 1999 Structure and ligand of a histone acetyltransferase bromodomain *Nature* **399** 491–6
- [41] Sanchez R and Zhou M-M 2009 The role of human bromodomains in chromatin biology and gene transcription *Curr. Opin. Drug Discovery Dev.* **12** 659–65
- [42] Moriniere J et al 2009 Cooperative binding of two acetylation marks on a histone tail by a single bromodomain *Nature* **461** 664–8
- [43] Filippakopoulos P et al 2012 Histone recognition and large-scale structural analysis of the human bromodomain family *Cell* **149** 214–31
- [44] Ruthenburg A J, Li H, Patel D J and Allis C D 2007 Multivalent engagement of chromatin modifications by linked binding molecules *Nat. Rev. Mol. Cell Biol.* **8** 983–94
- [45] Mujtaba S, Zeng L, Zhou M-M 2007 Structure and acetyl-lysine recognition of the bromodomain *Oncogene* **26** 5521–7
- [46] Marchand J-R and Caffisch A 2015 Binding mode of acetylated histone to bromodomains: variations on common motif *ChemMedChem* **10** 1327–33
- [47] Muller S, Filippakopoulos P and Knapp S 2011 Bromodomains as therapeutic targets *Expert Rev. Mol. Med.* **13** e29
- [48] Filippakopoulos P and Knapp S 2014 Targeting bromodomains: epigenetic readers of lysine acetylation *Nat. Rev. Drug Discovery* **13** 337–56
- [49] Arrowsmith C H, Bountra C, Fish P V, Lee K and Schapira M 2012 Epigenetic protein families: a new frontier for drug discovery *Nat. Rev. Drug Discovery* **11** 384–400
- [50] Brand M, Measures A M, Wilson B G, Cortopassi W A, Alexander R, Höss M, Hewings D S, Rooney T P C, Paton R S and Conway S J 2015 Small molecule inhibitors of bromodomain-acetyl-lysine interactions *ACS Chem. Biol.* **10** 22–39
- [51] SGC, Structural Genomics Consortium [www.thesgc.org](http://www.thesgc.org) (accessed 29 June 2015)
- [52] Vidler L R, Brown N, Knapp S and Hoelder S 2012 Druggability analysis and structural classification of bromodomain Acetyl-lysine binding sites *J. Med. Chem.* **55** 7346–59
- [53] Miyoshi S, Ooike S, Iwata K, Hikawa H and Sugahara K (Mitsubishi Tanabe Pharma Co. JP). Antitumor Agent. WO/2009/084693 9 July 2009
- [54] Filippakopoulos P et al 2010 Selective inhibition of BET bromodomains *Nature* **468** 1067–73
- [55] Ferguson F M et al 2013 Targeting low-druggability bromodomains: fragment based screening and inhibitor design against the BAZ2B bromodomain *J. Med. Chem.* **56** 10183–7
- [56] Fedorov O et al 2014 [124] Tirazolo [43-a] phthalazines: inhibitors of diverse bromodomains *J. Med. Chem.* **57** 462–76
- [57] Huggins D J 2012 Application of inhomogeneous fluid solvation theory to model the distribution and thermodynamics of water molecules around biomolecules *Phys. Chem. Chem. Phys.* **14** 15105–17
- [58] Lazaridis T 1998 Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory *J. Phys. Chem. B* **102** 3531–41
- [59] Lazaridis T 1998 Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids *J. Phys. Chem. B* **102** 3542–50
- [60] Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 The protein data bank *Nucleic Acids Res.* **28** 235–42
- [61] Sastry G M, Adzhigirey M, Day T, Annabhimoju R, Sherman W 2013 Protein and ligand preparation: parameters protocols and influence on virtual screening enrichments *J. Comput. Aided Mol. Des.* **27** 221–34
- [62] Abascal J L F and Vega C 2005 A general purpose model for the condensed phases of water: TIP4P/2005 *J. Chem. Phys.* **123** 234505
- [63] Phillips J C, Braun R and Schulten K 2005 Scalable molecular dynamics with NAMD *J. Comput. Chem.* **26** 1781–802
- [64] MacKerell A D Jr et al 1998 All-atom empirical potential for molecular modeling and dynamics studies of proteins *J. Phys. Chem. B* **102** 3586–616
- [65] MacKerell A D Jr, Feig M and Brooks C L 3rd 2004 Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations *J. Comput. Chem.* **25** 1400–15
- [66] Feller S E, Zhang Y H and Brooks B R 1995 Constant-pressure molecular dynamics simulation: the Langevin piston method *J. Chem. Phys.* **103** 4613–21
- [67] Essmann U, Perera L, Berkowitz M L, Darden T, Lee H and Pedersen L G 1995 A smooth particle mesh Ewald method *J. Chem. Phys.* **103** 8577–93
- [68] Morita T and Hiroike K 1961 A new approach to the theory of classical fluids III *Prog. Theor. Phys.* **25** 537–78



- [69] Yang J, Zhang L, Wang L and Zhong D 2012 Femtosecond conical intersection Dynamics of Tryptophan in proteins and validation of slowdown of hydration layer dynamics *J. Am. Chem. Soc.* **134** 16460–3
- [70] Lazaridis T 2000 Solvent reorganization energy and entropy in hydrophobic hydration *J. Phys. Chem. B* **104** 4964–79
- [71] Young T, Abel R, Friesner R A and Berne B J Methods of calculating differences of binding affinities between congeneric pairs of ligands by way of a displaced solvent functional *US Patent* 7756674 (13 July 2010)
- [72] Nguyen C N, Young T K and Gilson M K 2012 Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril *J. Chem. Phys.* **137** 044101
- [73] Raman E P and MacKerell A D Jr 2013 Rapid estimation of hydration thermodynamics of macromolecular regions *J. Chem. Phys.* **139** 055105
- [74] Huggins D J and Payne M C 2013 Assessing the accuracy of inhomogeneous fluid solvation theory in predicting hydration free energies of simple solutes *J. Phys. Chem. B* **117** 8232–44
- [75] Nettleton R E and Green M S 1958 Expression in terms of molecular distribution functions for the entropy density in an infinite system *J. Chem. Phys.* **29** 1365–70
- [76] Baranyai A and Evans D J 1989 Direct entropy calculation from computer simulation of liquids *Phys. Rev. A* **40** 3817–22
- [77] Wallace D C 1987 On the role of density fluctuations in the entropy of a fluid *J. Chem. Phys.* **87** 2282–4
- [78] Matubayasi N, Reed L H and Levy R M 1994 Thermodynamics of the hydration shell. 1. Excess energy of a hydrophobic solute *J. Phys. Chem.* **98** 10640–9
- [79] Matubayasi N, Gallicchio E and Levy R M 1998 On the local and nonlocal components of solvation thermodynamics and their relation to solvation shell models *J. Phys. Chem.* **109** 4864–72
- [80] Huggins D J 2015 Quantifying the entropy of binding for water molecules in protein cavities by computing correlations *Biophys. J.* **108** 928–36
- [81] Chen X, Liu M and Gilson M K 2001 Binding DB: a web-accessible molecular recognition database *J. Comb. Chem.* **4** 719–25
- [82] Meslamani J, Steven S G, Sanchez R and Zhou M-M 2014 ChEpiMod: a knowledgebase for chemical modulators of epigenome reader domains *Bioinformatics* **30** 1481–3
- [83] Demont E H *et al* 2015 Fragment-based discovery of low-micromolar ATAD2 bromodomain inhibitors *J. Med. Chem.* **58** 5649–73
- [84] SGC, Structural Genomics Consortium [www.thesgc.org/chemical-probes/PFI-4](http://www.thesgc.org/chemical-probes/PFI-4) (accessed 29 June 2015)
- [85] Chung C-W *et al* 2011 Discovery and characterization of small molecule inhibitors of the BET family bromodomains *J. Med. Chem.* **54** 3827–38
- [86] Palmer W S *et al* 2016 Structure-guided design of IACS-9571 a selective high-affinity dual TRIM24-BRPF1 bromodomain inhibitor *J. Med. Chem.* **59** 1440–54
- [87] SGC, Structural Genomics Consortium [www.thesgc.org/chemical-probes/N1-57](http://www.thesgc.org/chemical-probes/N1-57) (accessed 29 June 2015)
- [88] Hay D A *et al* 2014 Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains *J. Am. Chem. Soc.* **136** 9308–19
- [89] Chen P *et al* 2016 Discovery and characterization of GSK2801 a selective chemical probe for the bromodomains BAZ2A and BAZ2B *J. Med. Chem.* **59** 1410–24
- [90] Zhang G *et al* 2012 Down-regulation of NF- $\kappa$ B transcriptional activity in HIV-associated kidney disease by BRD4 inhibition *J. Biol. Chem.* **287** 28840–51
- [91] SGC, Structural Genomics Consortium [www.thesgc.org/chemical-probes/PFI-3](http://www.thesgc.org/chemical-probes/PFI-3) (accessed 29 June 2015)
- [92] Drouin L *et al* 2015 Structure enabled design of BAZ2-ICR a chemical probe targeting the bromodomains of BAZ2A and BAZ2B *J. Med. Chem.* **58** 2553–9
- [93] Urlick A K, Hawk L. M L, Cassel M K, Mishra N K, Liu S, Adhikari N, Zhang W, dos Santos C O, Hall J L and Pomerantz W C 2015 Dual screening of BPTF and Brd4 using protein-observed fluorine NMR uncovers new bromodomain probe molecules *ACS Chem. Biol.* **10** 2246–56
- [94] SGC, Structural Genomics Consortium [www.thesgc.org/chemical-probes/NVS-1](http://www.thesgc.org/chemical-probes/NVS-1) (accessed 29 June 2015)
- [95] SGC, Structural Genomics Consortium [www.thesgc.org/chemical-probes/BI-9564](http://www.thesgc.org/chemical-probes/BI-9564) (accessed 29 June 2015)