Straightforward incorporation of multiple ligand types into molecular dynamics simulations for efficient binding site detection and characterization

Yaw Sing Tan,[*,†] *Chandra S. Verma*[†,‡,§]

† Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, Singapore 138671

‡ Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

§ School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

ABSTRACT

Binding site identification and characterization is an important initial step in structure-based drug design. To account for the effects of protein flexibility and solvation, several cosolvent molecular dynamics (MD) simulation methods that incorporate small organic molecules into the protein's solvent box to probe for binding sites have been developed. However, most of these

methods are highly inefficient as they allow for the use of only one probe type at a time, which means that multiple sets of simulations have to be performed to map different types of binding sites. The high probe concentrations used in some of these methods also necessitate the use of artificial repulsive forces to prevent the probes from aggregating. Here, we present multiple-ligand-mapping MD (mLMMD), a method that incorporates multiple types of probes for simultaneous and efficient mapping of different types of binding sites without the need for introduction of artificial forces that may cause unintended mapping artifacts. We validate the method on a diverse set of 10 proteins, and show that the mLMMD probes are able to reliably identify hydrophobic, hydrogen-bonding, charged, and cryptic binding sites in all the test cases. Our results also highlight the potential utility of mLMMD to virtual screening and rational drug design.

1. INTRODUCTION

The success of structure-based drug design hinges on knowledge of the locations of ligand binding sites on the target protein. This information may be obtained experimentally using either NMR-based[1] or X-ray crystallographic methods,[2] which tend to be laborious, expensive and time-consuming to perform, and so there is a high demand for computational methods to complement experimental efforts. A variety of computational algorithms that base their predictions on available three-dimensional protein structures have been developed to rapidly identify binding pockets.[3,4] However, they largely neglect the effects of protein flexibility and aqueous solvation, thus compromising their accuracy. The detection of cryptic binding sites, which only reveal themselves upon ligand binding and are otherwise occluded,[5] is particularly

challenging because they are only exposed by conformational changes of the protein. Individual water molecules are known to play an important role in protein–ligand binding,[6] but most computational methods either consider water implicitly or model it as a continuum. These limitations are partially overcome by pocket detection algorithms that have been specifically adapted to analyze protein conformational ensembles generated by molecular dynamics (MD) simulations.[7-9]

MD allows for an atomistic description of both protein and water, thus accounting for protein flexibility and solvation effects. MD-based methods which incorporate small molecular fragments that are commonly found as substructures in drug molecules into the protein's solvent box provide a compelling alternative to conventional pocket detection algorithms. These fragments interact dynamically with the protein surface and compete with water molecules for binding.[10-19] They not only act as probes to map the locations of binding sites on the protein surface, but also can identify sites that favour certain specific interactions. These probe-based MD simulations are commonly referred to as solvent-mapping or cosolvent simulations because they are analogous to the multiple solvent crystal structure method,[2] and high concentrations of probe molecules are usually used. However, since these probe molecules actually map out where ligands would bind to proteins, we prefer here to refer to them as ligands rather than solvents.

The majority of probe-based MD methods allow for the use of only one type of ligand in one simulation run, which means that multiple sets of simulations need to be performed to obtain affinity maps for different functional groups. One of these methods, MixMD, was recently adapted to incorporate multiple types of ligands, but it did not work as well as the single-ligand implementation.[20] Furthermore, it was not clear how the binding site information obtained with different ligands in MixMD could be used for structure-based drug design. To date, three probe-

based MD methods that utilize a cocktail of ligands for simultaneous mapping have been developed.[13,21,22] They allow affinity maps for multiple functional groups to be obtained from a single set of simulations, thus increasing efficiency. Another benefit of multiple-ligand simulations is that they can identify displaceable water sites as representative ligands and water molecules are allowed to compete with one another for binding to the protein.[21,23] Of the three known multiple-ligand methods, only SILCS allows for the use of hydrophobic ligands (benzene and propane) in the probe mixture.[21] Hydrophobic probes are highly desirable as they reduce the overall solvent polarity, thus encouraging the opening and enlargement of hydrophobic pockets that may otherwise remain undetected in simulations of the protein in pure water (Figure 1). In fact, benzene has been found to be the most effective probe for detecting cryptic pockets,[24] which tend to be hydrophobic in nature.[25-29]

Ligand-mapping molecular dynamics (LMMD)[14,15,30] and SWISH[24] are two single-ligand MD methods that employ hydrophobic probes for pocket detection. Similar to the original Tier-I implementation of SILCS (1M benzene and 1 M propane),[11] SWISH (1 M benzene) uses a high concentration of hydrophobic probes for pocket detection. However, the hydrophobic probes are prone to aggregating, which hampers proper sampling of the protein surface, and could induce protein denaturation.[31] Consequently, both methods require the use of artificial repulsive potentials and harmonic restraints to prevent probe aggregation and protein denaturation, respectively. SILCS Tier II requires the additional implementation of repulsive potentials between all charged probes to minimize the formation of ion aggregates, as moderately high concentrations of charged probes (0.25 M methylammonium and 0.25 M acetate) are used. However, these artificial forces could create artifacts during mapping, and their compatibility with the protein and small molecule force fields is unknown. The interligand repulsive potentials

could affect the proper mapping of large pockets that are able to accommodate more than one hydrophobic or charged ligand, while the protein restraints may hinder the opening of certain cryptic pockets and their conformational sampling.[11] In contrast, LMMD avoids the use of these additional artificial forces by employing a low hydrophobic probe concentration (0.15–0.2 M) to prevent ligand aggregation and protein unfolding. In fact, it is the first probe-based MD method that was developed to use hydrophobic probes without additional artificial forces and to identify cryptic pockets.[14] The usefulness of LMMD in revealing cryptic pockets, enhancing the conformational sampling of proteins, and identifying novel ligand binding modes has been demonstrated in several studies,[14,15,30] most notably one in which a novel binding site was predicted by LMMD and subsequently confirmed to be functional in biophysical binding assays and X-ray crystal structures.[32] LMMD has been successfully employed in various studies to identify druggable cryptic pockets,[33] discover novel protein–protein interaction inhibitors,[34,35] and understand protein–peptide[36] and protein–RNA binding.[37] The success of the technique in predicting binding sites certainly suggests that LMMD holds great promise as a tool for drug design. However, only hydrophobic probes (benzene and chlorobenzene) have been validated for use in LMMD thus far, thus limiting it to the detection of hydrophobic and halogen binding sites.

Here, we describe the development of mLMMD (multiple-ligand-mapping molecular dynamics), which is a multiple-ligand version of LMMD that allows for the incorporation of different ligand types to simultaneously probe for hydrophobic, hydrogen-bonding and charged binding sites. The multiple-ligand implementation of LMMD that we have developed addresses a number of limitations of existing multiple-ligand MD methods. Unlike SILCS, mLMMD avoids the use of interligand repulsive potentials and protein restraints that may cause sampling artifacts, as we discuss later. Propane, which is used as a hydrophobic probe along with benzene

in SILCS, was deemed redundant as our previous work had shown that benzene alone is able to map the interactions of both aromatic and aliphatic amino acid side chains.[30] An alternative multiple-ligand method that was developed by Bakan et al.[13] lacks hydrophobic probes like benzene, which may limit its ability to detect hydrophobic cryptic binding sites.[24] Developed primarily for druggability assessment, Bakan et al.'s method identifies the locations of binding sites, but does not characterize their binding preferences. In contrast, mLMMD is able to predict both the location of binding sites and their favored types of interactions. Such information may be used to guide drug design. Furthermore, the ability of existing multiple-ligand MD methods to interrogate protein–protein interfaces, which are challenging to map because of their flat and shallow nature, has not been rigorously validated so far. In this study, we extensively validated the mLMMD method on a diverse set of protein systems that includes enzymes, a receptor protein, and protein–protein interactions (PPIs).
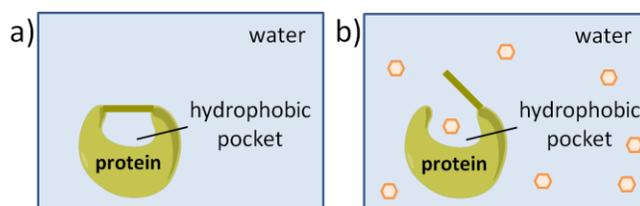

Figure 1. Detection of hydrophobic binding pockets in MD simulations. (a) Hydrophobic pockets tend to stay closed in standard MD simulations. (b) Hydrophobic ligands (orange) reduce solvent polarity, thus facilitating the opening of hydrophobic pockets in LMMD simulations.

2. METHODS

2.1. Probe set. A minimal probe set was selected to represent common functional groups found in drug molecules. Benzene represents hydrophobic ligands and serves as probes for hydrophobic binding sites. For the rest of the probes, only a single methyl group is attached to the active functional group to prevent them from being dragged into nonpolar environments, thus improving probe specificity. Methanol and acetaldehyde represent hydrogen bond donors and hydrogen bond acceptors, respectively, and serve as probes for hydrogen-bonding binding sites. Methylammonium and acetate represent positively charged ligands and negatively charged ligands, respectively, and probe for charged binding sites.

2.2. System setup. The set of proteins used for validation includes four enzymes from different families (p38 mitogen-activated protein kinase [p38 MAPK], *Escherichia coli* AmpC β-lactamase [AmpC], *Lactobacillus casei* dihydrofolate reductase [DHFR], and factor Xa), one receptor protein (estrogen receptor α [ERα]), and five proteins involved in PPIs (BIR3 domain of X-linked inhibitor of apoptosis protein [XIAP], interleukin-2 [IL-2], polo-box domain [PBD] of polo-like kinase 1 [PLK1], von Hippel-Lindau disease tumor suppressor [VHL], and WD repeat-containing protein 5 [WDR5]). This ensures that the method is rigorously tested against a wide variety of binding sites with different shapes, sizes and polarities. Each protein has multiple ligand-bound structures deposited in the Protein Data Bank (PDB)[38] and contains at least three different types of interaction sites. Some of these proteins were selected (p38 MAPK, AmpC, DHFR, factor Xa, and ERα) because they have been used as test proteins by other probe-based MD methods, thus allowing for comparison of results. The ability of mLMMD to map cryptic binding sites using the probe mixture was evaluated on three proteins with known cryptic binding pockets, p38 MAPK, IL-2, and PLK1.

The following PDB structures were used as starting structures for the LMMD simulations: 3HL5[39] (XIAP BIR3), 1M47[40] (IL-2), 3FVH[41] (PLK1 PBD), 1LM8[42] (VHL), 6E1Z[43] (WDR5), 3ERT[44] (inactive ERα), 1OUY[45] (p38 MAPK), 2HDS[46] (AmpC), 3DFR[47] (DHFR), and 1G2L[48] (factor Xa). For IL-2, PLK1 PBD, and p38 MAPK, structures with the cryptic pockets closed were specifically chosen. All crystallographic ligands and structurally non-essential ions were removed, while crystallographic water molecules were retained. The ModLoop web server was used to model missing residues (1-5, 75-76, and 99-102) in IL-2 and revert mutated residues in the DHFR structure to wild type.[49] Acetyl and N-methyl groups were used to cap the N- and C-termini, respectively, if the first or last residue of the protein chain was missing. PDB2PQR[50] was used to determine the protonation states of residues and add missing hydrogen atoms.

The LEaP module in the AMBER 14[51] package was used to solvate each protein with TIP3P water molecules in a periodic truncated octahedron box such that its walls were at least 10 Å away from the protein, followed by charge neutralization with sodium or chloride ions. The appropriate number of probe molecules to be added was determined from the average volume of the solvent box obtained from a 50-ns MD simulation of the apo protein only. Ten different random distributions of the probes around the protein were then generated using Packmol, followed by solvation with TIP3P water molecules using LEaP with a closeness value of 0.8 Å to give a final probe composition of approximately 0.2 M benzene, 0.4 M methanol, 0.4 acetaldehyde, 0.1 M methylammonium, and 0.1 M acetate.

2.3. MD simulations. Energy minimizations and MD simulations were carried out by the PMEMD module of AMBER 14,[51] using the ff14SB[52] force field for proteins, the generalized AMBER force field (GAFF)[53] for probe molecules, the cationic dummy atom model[54] for the zinc ion in XIAP BIR3, and the compromised (CM) set of divalent metal ion parameters[55] for the

8

calcium ion in factor Xa. Atomic charges for the probe molecules (Tables S1-S5) were derived using the R.E.D. Server,[56] which fits restrained electrostatic potential (RESP) charges[57] to a molecular electrostatic potential (MEP) computed by the Gaussian 09 program[58] at the HF/6-31G* theory level. The SHAKE algorithm[59] was used to constrain all bonds involving hydrogen atoms, allowing for a time step of 2 fs. A cutoff distance of 9 Å was implemented for nonbonded interactions, while the particle mesh Ewald method[60] was used to treat long-range electrostatic interactions under periodic boundary conditions. Energy minimization was performed for 500 steps with the steepest descent algorithm, followed by another 500 steps with the conjugate gradient algorithm. The system was then heated gradually to 300 K over 50 ps at constant volume before equilibration at a constant pressure of 1 atm for another 50 ps. During the energy minimization and equilibration steps, weak harmonic positional restraints with a force constant of 2.0 kcal mol^{-1} Å^{-2} were imposed on all non-hydrogen atoms of the protein. The system was allowed to further equilibrate with no restraints for 2 ns before proceeding to the production run, which was carried out for 50 ns at 300 K and 1 atm. The Langevin thermostat[61] was used to maintain the temperature with a collision frequency of 2 ps^{-1}. Pressure was maintained by the Berendsen barostat[62] with a pressure relaxation time of 2 ps.

An MD simulation of just the probe molecules in aqueous solution was also performed. A cubic box containing 40 benzenes, 80 methanols, 80 acetaldehydes, 20 methylammoniums, 20 acetates, and 10000 waters was generated by Packmol. This probe mixture contains 0.2045 M benzene, 0.4090 M methanol, 0.4090 M acetaldehyde, 0.1023 M methylammonium, and 0.1023 M acetate at 300 K and 1 atm, which is similar to that used in the LMMD simulations. The probe mixture was simulated according to the steps outlined above. Radial distribution functions (RDF)

for the probe molecules were calculated with a bin size of 0.05 Å, based on the last 10 ns of the production run.

2.4. MD trajectory analysis. The 10 individual LMMD simulation runs for each protein were combined into a single trajectory for analysis. Atoms from the trajectory snapshots saved at 10-ps intervals were binned into 1 Å × 1 Å × 1 Å grid cells. By counting the number of times a defined probe atom appears in each grid cell, five types of ligand occupancy maps were generated: hydrophobic (benzene carbons), hydrogen bond donor (methanol polar hydrogen), hydrogen bond acceptor (acetaldehyde oxygen), positive charge (methylammonium polar hydrogens), and negative charge (acetate oxygens). The oxygen of methanol was not considered for the hydrogen bond acceptor occupancy map so as to remove false positives that occur when methanol is dragged by its hydroxyl hydrogen to an interaction site with hydrogen bond acceptors but no nearby hydrogen bond donors. The converse is less likely to happen as acetaldehyde, with one fewer hydrogen-bonding atom, incurs a lower desolvation penalty than methanol, and therefore will be preferentially bound at hydrogen bond acceptor sites. Based on the inverse Boltzmann relationship, the binding free energy associated with a probe atom at voxel i is given by the following equation:

$$\Delta G_i \; = \; -RT \, ln \frac{N_i}{N_0}$$

(1)

where R is the gas constant, T is the absolute temperature, and N_i/N_0 is the ratio of the observed occupancy to the expected occupancy. Since the molar concentrations of the probe molecules are fixed, the expected occupancy of a voxel by a probe atom can be obtained using the equation:

$$N_0 = \frac{n_{atoms} \times n_{frames} \times c_p \times N_A}{10^{27}}$$

(2)

where n_{atoms} is the number of defined probe atoms per molecule (e.g. n_{atoms} = 6 for benzene, 1 for methanol, 3 for methylammonium), n_{frames} is the total number of trajectory frames being analyzed, c_p is the molar concentration of the probe (mol/L), N_A is the Avogadro constant $(6.02214076 \times 10^{23}$ mol$^{-1})$, and 10^{27} is the number of grid cells in 1 L of the simulation box. It should be noted that the free energies obtained by equation 1 are approximate and not physically meaningful because the formula assumes that the atoms behave as independent entities.[13,63] Here, we only use the free energies to quantify our cutoff values for visualization of the ligand occupancy maps. Hydrophobic occupancy maps were visualized at -1.5 kcal/mol, hydrogen bond donor and acceptor occupancy maps at -1.75 kcal/mol, and positive and negative charge occupancy maps at -1.9 kcal/mol. These free energy cutoff values were found to be the best compromise between filtering out spurious binding sites and keeping the true positives. The equivalent isovalue cutoffs (N_i) for benzene, methanol, acetaldehyde, methylammonium, and acetate densities are in the ratio 6:3:3:3:2. This ratio is maintained for the ranking of binding sites. To compare the overlap of the ligand occupancy maps with known ligands, experimental protein complex structures were aligned using PyMOL[64] to the average protein structure sampled during the LMMD simulations.

3. RESULTS AND DISCUSSION

A few changes were made to the original implementation of LMMD[14,30] to facilitate the detection of different types of binding sites. Firstly, in addition to benzene, methanol, acetaldehyde, methylammonium, and acetate were included as probe molecules. As the number and type of probes have increased, simulation length was accordingly increased from 5 ns to 50 ns for better sampling. Ligand-induced denaturation of the proteins[31] was not observed during these longer LMMD simulations (Figure S1). Root mean square deviation (RMSD) of the Cα atoms remained below 3 Å in almost all simulations, with the exception of a few runs for VHL. This is not unexpected and does not indicate unfolding as large relative motions of the loosely coupled VHL α and β domains are also observed in the simulation of the apo protein only. Lastly, the criterion for visualization of the probe atom densities was standardized for all proteins. Previously, an arbitrary isovalue cutoff of three times the highest isovalue at which benzene carbons were detected in the bulk solvent, which varies for different systems, was used. Here, the isovalues are converted into free energies, which are then used to specify the cutoff for visualization of the probe atom densities. Benzene carbon occupancy maps were visualized at a cutoff of -1.5 kcal/mol, while methanol polar hydrogens and acetaldehyde oxygen occupancy maps were visualized at -1.75 kcal/mol, and methylammonium polar hydrogens and acetate oxygen occupancy maps were visualized at -1.9 kcal/mol.

3.1. Optimizing the probe mixture composition. One enzyme and one protein involved in PPIs were initially selected from the protein test set to optimize the composition of the probe mixture. XIAP BIR3 and p38 MAPK were chosen as each of them has four different types of interaction sites. In addition, p38 MAPK has a cryptic pocket that could be used to evaluate whether benzene retains its ability to detect cryptic binding sites in the presence of other types of probe molecules. Benzene concentration was fixed at 0.2 M, while the concentrations of

hydrogen-bonding and charged probes were varied from 0.2 M to 0.6 M and from 0.1 M to 0.2 M, respectively. Concentrations of charged probes were deliberately kept low to prevent oppositely charged probes from forming ion aggregates. Various permutations of these probe concentrations were used to map XIAP BIR3 and p38 MAPK (Table S6). Only the probe mixture consisting of 0.2 M benzene, 0.4 M methanol, 0.4 M acetaldehyde, 0.1 M methylammonium, and 0.1 M acetate was able to reproduce all known binding interactions in both proteins. The results of these simulations are described in detail below. LMMD simulations employing this optimal probe mixture were then performed on the rest of the proteins in the test set.

The miscibility of this optimal probe mixture was also evaluated by calculating the RDF for individual probe atoms in an MD simulation of the probe mixture in aqueous solution. This has been used to assess appropriate solvent behaviour in the single-ligand implementations of LMMD[30] and MixMD.[65] If the RDF converges to 1 at long distances, it means the probe molecules and water are evenly mixed. However, if the RDF converges to values larger than 1, probe aggregation is indicated. There was no evidence of probe aggregation in the optimized probe–water mixture, as all the probe–probe RDFs converged to 1 at the van der Waals cutoff of 9 Å (Figure S2). This shows that the probe concentrations used are low enough to prevent probe aggregation.
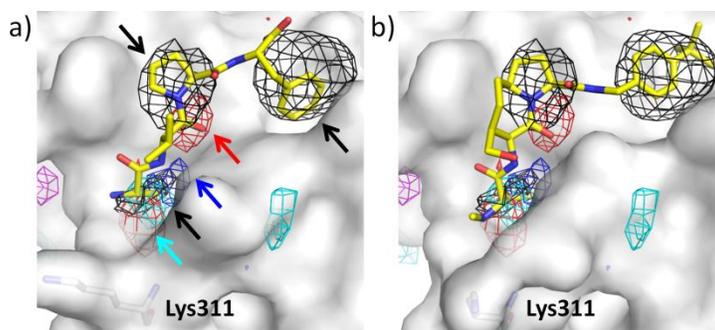
3.2. XIAP BIR3.

Figure 2. Ligand occupancy maps of XIAP BIR3 superimposed on the PDB structures (a) 2OPZ and (b) 6EY2. Hydrophobic, hydrogen bond donor, hydrogen bond acceptor, positive charge and negative charge maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

XIAP is an antiapoptotic protein that interacts with Smac-DIABLO and caspase-9. They contain a conserved tetrapeptide motif that is responsible for binding to the BIR3 domain of XIAP. The N-terminal alanine residue of the peptide is crucial to the binding, as both the positively charged N-terminus and nonpolar alanine side chain engage in interactions with the XIAP BIR3 surface. Both interactions were mapped in the LMMD simulations, as indicated by Figure 2, which shows good overlap of the hydrophobic and positive charge ligand maps with the aliphatic moieties and positively charged amine of the tetrapeptide and a peptidomimetic.

Important hydrogen bonding interactions of the peptide backbone that are mimicked by small-molecule XIAP antagonists were also identified by the hydrogen-bonding probes. Spurious hydrogen bond acceptor densities that overlap with adjacent hydrogen bond donor densities were generated when methanol oxygen was included in the analysis (Figure S3). This observation supports our decision to use methanol to find hydrogen bond donating sites only. Two other

hydrophobic interactions involving proline and phenylalanine of the peptide were reproduced by the benzene probes (Figure 2a). Besides reproducing these known interactions, the simulations also indicate favourable hydrogen bond acceptor densities close to where the N-terminal alanine residue binds. Acetaldehyde probes were observed to hydrogen bond with the backbone nitrogen of Lys311 in this region during the simulations. This interaction has not been utilized by any known ligand of XIAP BIR3. Since the positive charge and hydrogen bond acceptor densities do not overlap that much, it may be possible to incorporate both features into new inhibitors for enhanced binding to XIAP.
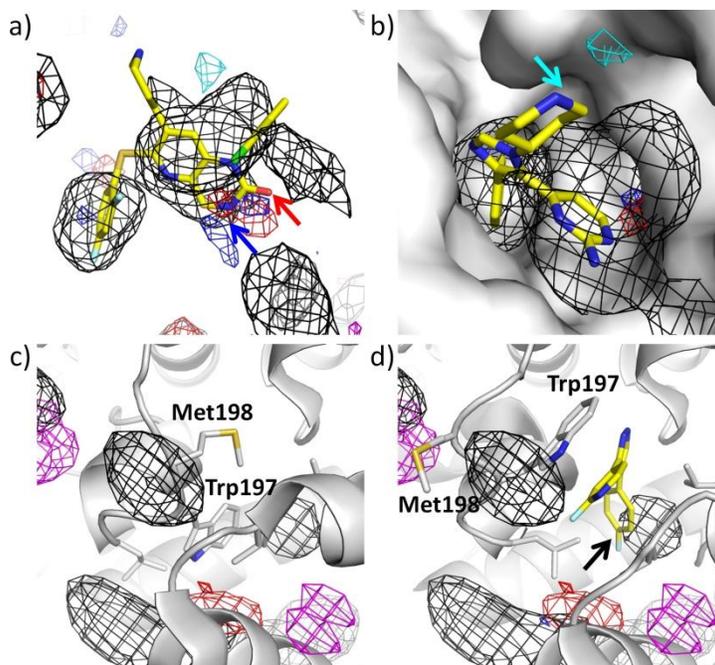
3.3. p38 MAPK.

Figure 3. Ligand occupancy maps of p38 MAPK superimposed on the PDB structures (a) 1OUY, showing the ligand at the ATP binding site, (b) 1LB7, (c) 1OUY, showing the closed cryptic pocket, and (d) 3HL7, showing the open cryptic pocket. Hydrophobic, hydrogen bond donor, hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as

black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

The majority of known p38 MAPK ligands bind to the ATP binding site. These ligands contain several aromatic and nonpolar moieties whose hydrophobic interactions with the protein are well reproduced by the benzene probes (Figure 3a, b). The hydrogen bond donor and acceptor densities also match the locations of donor and acceptor atoms found in the aromatic rings of the inhibitors (Figure 3a). A favorable location for the placement of a positive charge, which is occupied by a positively charged secondary amine of a known ligand (Figure 3b), was revealed by the positive charge ligand map. A tiny hydrogen bond donor density that disappears at a slightly lower free energy cutoff of -1.94 kcal/mol was observed within the difluorophenyl ring of the ligand in Figure 3a. It is not surprising that this rather weak interaction is not exploited by any of the p38 MAPK ATP-competitive ligands, as the hydrophobic interaction is much stronger in this region, as shown by the presence of benzene densities at very high isovalues.

Some ligands have been found to bind at a cryptic allosteric site that is occluded by Trp197 and Met198 in most p38 MAPK structures (Figure 3c). This pocket is exposed by the protein backbone movement of residues 195-198. Different probe-based MD methods have varying degrees of success in detecting this cryptic pocket, with Bakan et al.'s method[13] being successful and MDMix[10] being unsuccessful. It was not mentioned in the SILCS paper[21] whether this cryptic pocket was detected. To test the ability of our method to identify cryptic binding sites, we started the LMMD simulations of p38 MAPK using a structure with the cryptic pocket closed. The pocket was successfully detected, as the hydrophobic ligand map generated overlaps with the aromatic ring of a ligand bound within the pocket (Figure 3d).
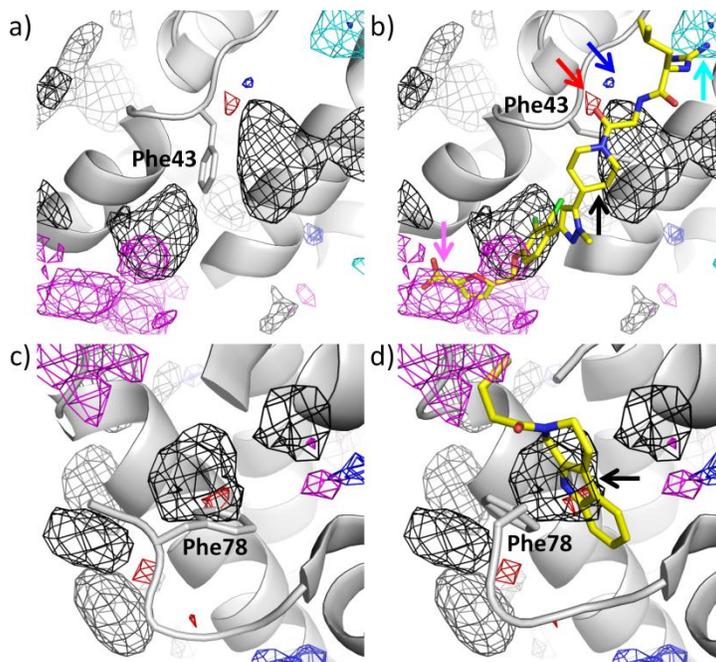
3.4. IL-2.

Figure 4. Ligand occupancy maps of IL-2 superimposed on the PDB structures (a) 1M47, showing the first closed cryptic pocket, (b) 1PY2, (c) 1M47, showing the second closed cryptic pocket, and (d) 1NBP. Hydrophobic, hydrogen bond donor (shown at a cutoff of -1.5 kcal/mol), hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

IL-2 is a cytokine protein that binds to the IL-2 receptor to induce T-cell proliferation. It has not one, but two cryptic pockets that are hidden by phenylalanine residues and are not observed in structures of the apo IL-2 protein and the IL-2–IL-2R complex.[29,40,66] Previous studies have shown that SILCS and LMMD simulations are effective at revealing these pockets.[15,31] Here, mLMMD is shown to be similarly effective, with prominent benzene densities overlapping with ligands bound at the two cryptic binding sites (Figure 4b,d). The hydrogen bond donor and

acceptor maps recapitulated hydrogen bonding interactions made by IL-2 ligands (Figure 3b). The hydrogen bond donor interactions appear to be rather weak though, as they are absent at the standard cutoff of -1.75 kcal/mol, and become prominent only at -1.5 kcal/mol. Positive and negative charge interactions at either end of the main binding site are also successfully reproduced by the charged probes (Figure 3b).
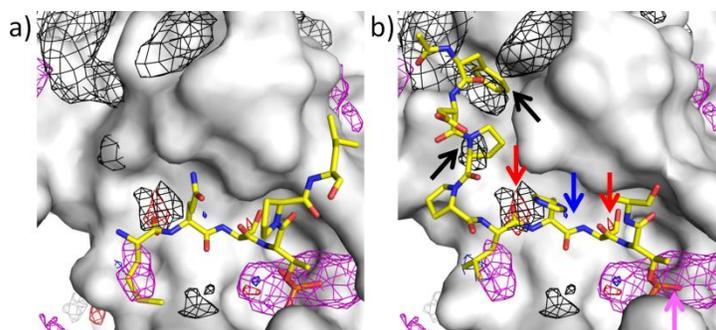
3.5. PLK1.

Figure 5. Ligand occupancy maps of PLK1 superimposed on the PDB structures (a) 1Q4K and (b) 3P37. Hydrophobic, hydrogen bond donor (shown at a cutoff of -1.5 kcal/mol), hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

The PBD of PLK1 helps in the subcellular localization of PLK1 by binding to serine- or threonine-phosphorylated sequences on target proteins. The phosphate binding site is highly polar. A cryptic hydrophobic binding site close to the phosphate binding site has also been identified in two independent X-ray crystallographic studies.[27,67] Crystal structures have shown that side-chain movements of two tyrosine residues allow the pocket to accommodate hydrophobic moieties from several ligands. The pocket is closed when ligands are not bound to it

(Figure 5a). In fact, LMMD was originally developed to sample the open conformation of this pocket in MD simulations and to design a ligand to specifically target its semi-open state.[14]

Similar to what was observed for IL-2, the inclusion of additional probes did not affect the ability of the benzene probes to reveal the cryptic hydrophobic pocket in PLK1 PBD (Figure 5b). They also reproduced the hydrophobic interactions of proline from the peptide ligand in a region between the phosphate and cryptic pockets, which was similarly indicated as a hydrophobic binding site by previous LMMD simulations.[14] These results obtained for IL-2 and PLK1 PBD suggest that mLMMD is at least as effective as single-ligand LMMD in identifying hydrophobic binding sites. The PBD phosphate binding site was revealed by the presence of prominent negative charge densities. In addition, two hydrogen bond acceptor regions identified by red arrows shown in Figure 5b match the location of hydrogen bond accepting backbone oxygen atoms. However, as in IL-2, the hydrogen bond donor interaction made by the peptide backbone seems to be weaker than the hydrogen bond acceptor interactions, with the hydrogen bond donor densities only appearing at a cutoff of -1.5 kcal/mol.
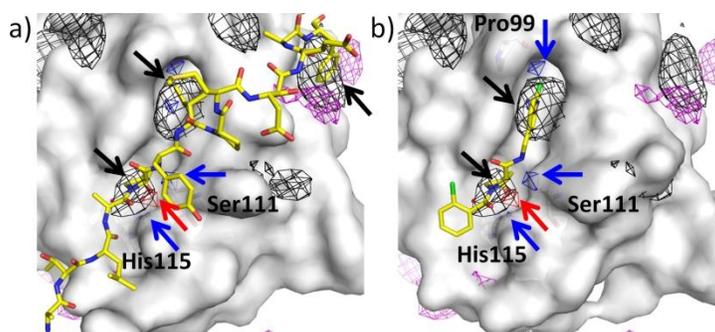
3.6. VHL.

Figure 6. Ligand occupancy maps of VHL superimposed on the PDB structures (a) 6BVB and (b) 4B95. Hydrophobic, hydrogen bond donor (shown at a cutoff of -1.5 kcal/mol), hydrogen

bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

VHL is the substrate-binding component of a multi-subunit E3 ubiquitin ligase that also includes elongins B and C, cullin-2, and RBX1. It binds to proteins that contain a hydroxylated proline residue, such as HIF-1α and HIF-2α. The hydroxylproline is critical to the interaction of substrate proteins with VHL, with the partially buried pyrrolidine ring making hydrophobic contacts and the 4-hydroxyl group forming hydrogen bonds with His115 and Ser111 within the highly complementary hydroxylpropline binding pocket. All three interactions were successfully mapped in the LMMD simulations, although hydrogen bond donor densities only appear at a cutoff of -1.5 kcal/mol (Figure 6). The hydrogen bond donor map also identifies a nearby hydrogen bond donor site that is exploited by both peptide and small-molecule ligands of VHL. Two additional hydrophobic binding sites upstream of the hydroxylproline binding site were indicated by the hydrophobic ligand map. These are occupied by isoleucine and phenylalanine residues (Figure 6a). There is a hydrogen bond donor site close to the central hydrophobic site that has not been exploited by any small-molecule VHL ligand, except for a para-chlorophenyl-bearing compound that appears to form a halogen bond with the backbone oxygen of Pro99 in this region (Figure 6b). Aromatic halogens contain a σ-hole[68] that is positively charged, allowing them to behave like hydrogen bond donors and interact with hydrogen bond acceptors. Modification of existing small-molecule VHL inhibitors to include a hydrogen bond donating functional group in this location could lead to a significant boost in binding affinity.
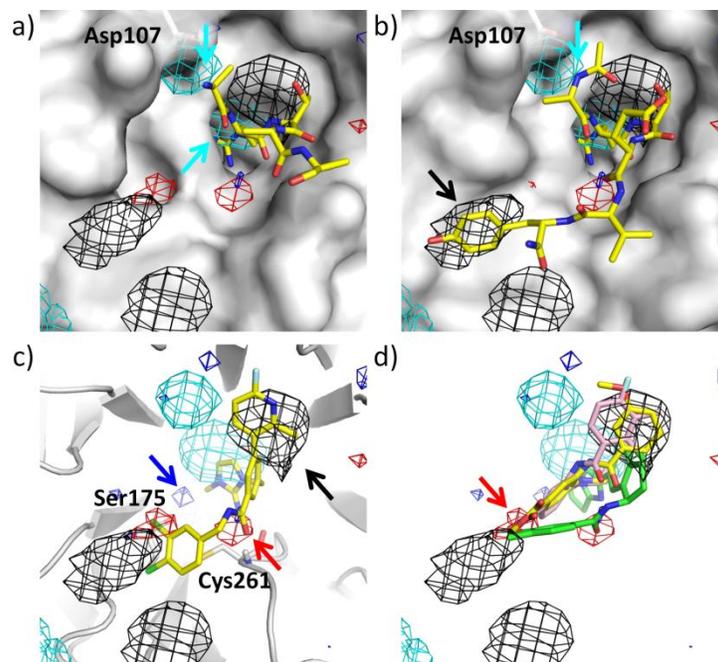
3.7. WDR5.

Figure 7. Ligand occupancy maps of WDR5 superimposed on the PDB structures (a) 4O45, (b) 5SXM, (c) 6E23, (d) 3UR4 (yellow), 4IA9 (pink) and 6DAS (green). Hydrophobic, hydrogen bond donor, hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

WDR5 is a core subunit of multi-subunit histone methyltransferases. It contains a deep pocket called the WIN site that binds to a conserved arginine-containing WIN motif of the methyltransferase subunit. This binding site was identified by the positive charge ligand map, which shows prominent densities within the pocket (Figure 7a). Another positive charge site just outside the WIN site was indicated by the positive charge ligand map, which captures the salt bridge interaction between the positively charged N-terminus of the peptide with Asp107 of WDR5 (Figure 7a, b). Besides these charge-charge interactions, the LMMD probes also recapitulated hydrophobic interactions made by peptide ligands with a shallow hydrophobic

binding site on the WDR5 protein surface (Figure 7b). Another hydrophobic binding site just outside the WIN site that is utilized for binding by small-molecule ligands was similarly detected by the benzene probes (Figure 7c).

Two hydrogen bond acceptor regions identified by red arrows in Figure 7c and 7d match the location of acceptor atoms in several small-molecule ligands. One of them involves the displacement of a crystallographic water molecule that mediates hydrogen bonding between the backbone nitrogen of Cys261 in WDR5 and the backbone oxygen of peptide ligands (Figure 7c). This shows that the displacement of this water molecule is energetically favorable, and also suggests that mLMMD can be used to predict the displaceability of crystallographic water molecules, as has been demonstrated for other probe-based MD methods.[23,63] A deeply buried hydrogen bond donor site that is not utilized by existing WDR5 inhibitors is found inside the WIN site, beneath the positive charge densities. The hydrogen bond donor map predicts a hydrogen bond interaction with the backbone oxygen of Ser175 (Figure 7c). Given the non-overlapping nature of the hydrogen bond donor and positive charge densities in this region, it is possible for both functionalities to be incorporated into the structure of a WDR5 inhibitor for enhanced binding affinity.
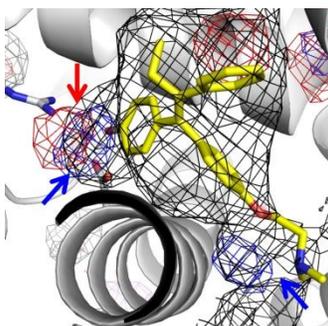
3.8. ERα.

Figure 8. Ligand occupancy maps of ERα superimposed on the PDB structure 3ERT. Hydrophobic, hydrogen bond donor, and hydrogen bond acceptor ligand maps are shown as black, blue, and red meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

ERα is a nuclear receptor protein that binds to its agonist estradiol at its C-terminal ligand binding domain. This induces a conformational change that promotes homodimerization and the recruitment of coactivator proteins. The hydrophobic ligand map generated by the LMMD simulations shows excellent overlap with the nonpolar moieties of an ERα antagonist (Figure 8). The hydroxyl group of the antagonist found at the bottom of the pocket makes important hydrogen bonding contacts with an arginine and a glutamate residue, both of which were identified by the hydrogen bond donor and acceptor maps. The hydrogen bond donor map also identified a favorable interaction site outside the pocket, which is usually occupied by a positively charged amine group. However, the positive charge densities here completely disappear at a cutoff of -1.6 kcal/mol, as opposed to -2.18 kcal/mol for the hydrogen bond donor densities. This suggests that a hydrogen bond interaction is more energetically favourable than a charge-charge interaction at this binding site, and current antagonists could improve their binding affinity by having their amine group modified to a hydrogen bond donating group. An apparently spurious hydrogen bond acceptor site with no hydrogen donors nearby was detected near the entrance of the pocket. Visualization of the simulations shows acetaldehyde probes being drawn by their methyl groups and stacking onto benzene probes that are already bound to the pocket. This observation highlights the need to corroborate identified binding sites by checking for the complimentarity of their immediate environment.
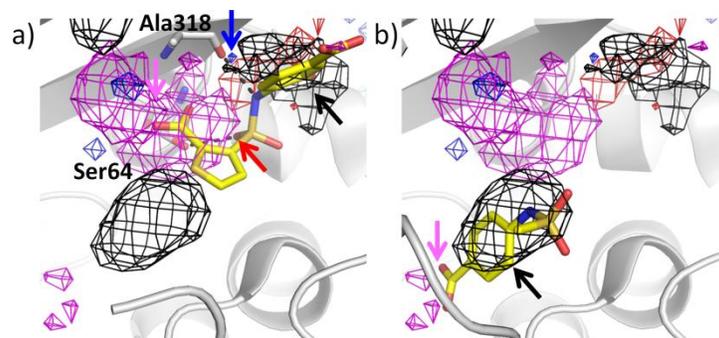
3.9. AmpC.

Figure 9. Ligand occupancy maps of AmpC superimposed on the PDB structures (a) 1XGJ and (b) 2HDS. Hydrophobic, hydrogen bond donor, hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

AmpC binds to a β-lactam ring at its active site and catalyzes its hydrolysis. There are two negative charge sites: one is the oxyanion hole at the active site and the other is found in the adjacent tunnel carboxylate site. Both interactions are reproduced in the negative charge ligand map (Figure 9). There is good overlap of the benzene densities with the aromatic rings of the ligands in these two sites. In addition, the active site is occupied by hydrogen bond donor and acceptor densities, which reflect the hydrogen bonding interactions formed with the backbone oxygen of Ala318 and side chain of Ser64 (Figure 9a). AmpC was also used as a test protein by SILCS, which identified similar ligand features in the active site.[69] However, it is not known if the neighboring tunnel carboxylate site was also detected by SILCS.
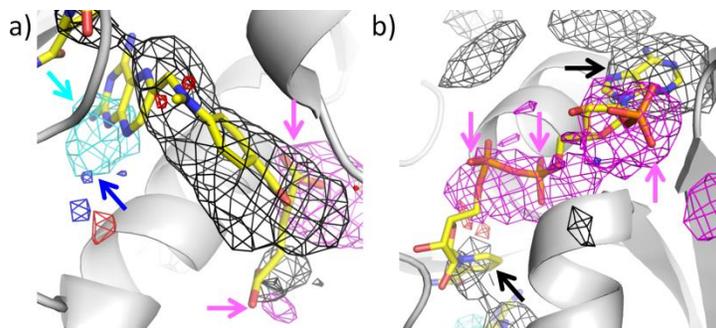
3.10. DHFR.

Figure 10. Ligand occupancy maps of DHFR superimposed on the PDB structure 3DFR, showing the (a) active site and (b) NADPH binding site. Hydrophobic, hydrogen bond donor, hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

DHFR is an enzyme that catalyzes the reduction of dihydrofolate to tetrahydrofolate, which is used for the de novo synthesis of purines, thymidylate, and certain amino acids. The cofactor NADPH is used as the electron donor. We evaluated the ability of mLMMD to detect both the active site and NADPH binding site by removing both the ligand and cofactor from the starting protein structure. A positive charge site, which is exploited by DHFR inhibitors (e.g. methotrexate and trimethoprim) that bear a positive charge in their heterocyclic rings for binding,[70,71] was identified at the bottom of the active site (Figure 10a). Hydrogen bond donor densities were also observed in this region but they are less intense, suggesting that charge-charge interactions are preferred. This could explain the high potency of methotrexate and trimethoprim as competitive inhibitors of the natural substrate dihydrofolate, which is uncharged at its heterocyclic ring. Two negative charge sites are identified outside the active site, which coincide with the locations of both carboxylate groups in methotrexate (Figure 10a). In contrast, only the α-carboxylate site was identified as a pharmacophore feature by another multiple-ligand

MD method, SILCS.[69] However, a low grid free energy cutoff was used to visualize negative acceptor densities in that work, and it is not known if they were present in the γ-carboxylate site at a higher cutoff. Since the oxygens of the two carboxylate groups are only 5.2–5.6 Å from each other, which is well within the interligand repulsive potential cutoff of 8 Å, an acetate probe that occupies the higher-affinity α-carboxylate site can potentially repel other acetate probes from the nearby lower-affinity γ-carboxylate site. Further study needs to be carried out to evaluate the effect of repulsive potentials between similarly charged probes on the mapping of adjacent charged binding sites with fairly different affinities.

The NADPH binding site is also well mapped by the LMMD probes, with prominent benzene densities overlapping with the adenine and nicotinamide moieties, and negative charge densities matching the positions of NADPH's three phosphate groups (Figure 10b).
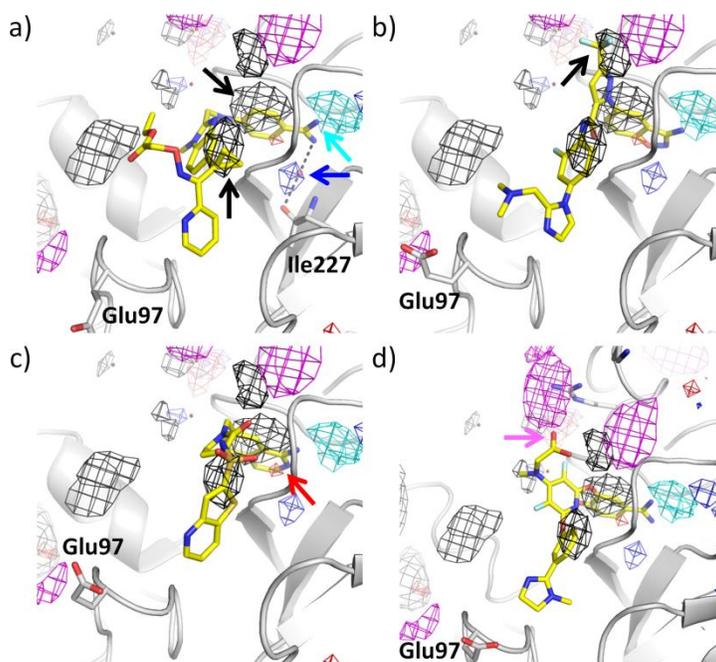
3.11. Factor Xa.

Figure 11. Ligand occupancy maps of factor Xa superimposed on the PDB structures (a) 1G2L, (b) 1Z6E, (c) 1F0R, and (d) 1FJS. Hydrophobic, hydrogen bond donor, hydrogen bond acceptor, positive charge and negative charge ligand maps are shown as black, blue, red, cyan, and magenta meshes, respectively. The correspondingly colored arrows indicate ligand interactions that were mapped by the LMMD probes.

Factor X is a serine protease of the coagulation cascade. Its activated form, factor Xa, cleaves prothrombin to produce thrombin. The attractive electrostatic interaction between the benzamidine group of factor Xa ligands and an aspartate residue deep in the active site, or the S1 subpocket, is mapped by the positive charge ligand map (Figure 11a). Three distinct hydrophobic binding sites were identified by the benzene probes, two of which (S1 and S3) are exploited by the ligand shown in Figure 11a. The third hydrophobic site (S2) is exploited by a trifluoro group of the ligand shown in Figure 11b.

Deep within the S1 subpocket, as indicated by the blue arrow in Figure 11a, lies a putative hydrogen bond donor site that is not occupied by any known ligand. It overlaps with the position of a crystallographic water that is found in the unbound form of factor Xa[72] and many of its liganded structures, and which forms hydrogen bonds with the backbone oxygen of Ile227 and amidine group of several ligands. The presence of hydrogen bond donor densities in this region suggests that the displacement of this water molecule is energetically favorable, and gains in binding affinity could be made by placing a hydrogen bond donating functional group in this region. A hydrogen bond acceptor site adjacent to this hydrogen bond donor site coincides with the position of a heterocyclic nitrogen in a ligand (Figure 11c), which forms hydrogen bonds with two crystallographic waters, including the aforementioned water molecule.

27

Finally, the negative charge ligand map identifies a negative charge site just outside the active site, which is exploited by some ligands to form charge-charge interactions with nearby lysine and arginine residues (Figure 11d). This negative charge site was not identified by SILCS in an earlier work.[21] Another discrepancy was found near the S4 pocket, where a positive charge site was identified by SILCS,[21] but not by mLMMD at the default free energy cutoff of –1.9 kcal/mol. We only saw positive charge densities appearing at –1.6 kcal/mol, which suggests that charge-charge interactions formed here are relatively weak. Our observation is supported by structures of factor Xa complexes, in which the positively charged amine group of ligands in this region is separated by at least 4.7 Å, and up to more than 10 Å, from the carboxylate group of the nearest acidic residue Glu97 (Figure 11). These differences in the identification of charged ligand sites could be caused by the presence of repulsive potentials between charged probes in SILCS and their absence in mLMMD. Indeed, we identified another negative charge site right next to the one we highlighted in Figure 11d, which was not mapped by SILCS.

3.12. Varying the free energy/isovalue cutoff

The default free energy cutoff values that we have recommended for visualization of the ligand occupancy maps are suitable for revealing most functional binding sites. However, this sometimes leads to the generation of ligand densities across significant regions of the protein surface, and it is not immediately obvious where the most important binding sites are. This issue is most acute for proteins whose binding sites are unknown. To test the ability of mLMMD to identify biologically relevant binding sites in such targets, we ranked the binding sites identified in all of our test proteins based on the free energy or isovalue at which the associated ligand densities disappear. In other words, the ligand densities at the top-ranked binding site disappear at the lowest energy or the highest isovalue.

We found that at least one of the top two sites is a known binding site in all the test proteins, with the exception of factor Xa, for which the third-ranked site is a known binding site (Figure 12). Most notably, a known binding site is identified as the top-ranked site in six of the proteins. This shows that high isovalues can be used with reasonably high confidence to visualize biologically relevant binding sites in proteins that are under-characterized structurally.

For structurally well-characterized protein targets, the identification of new and cryptic binding sites is of greater interest. However, when several new interaction sites are identified within the vicinity of the main binding site, it may be difficult to decide which one of them is the most relevant for ligand design. In these cases, the isovalue cutoff can be increased to rank the free energy contribution of each of these auxiliary interaction sites. Those interaction sites that retain ligand densities at higher isovalues are likely to be more important and should be prioritized for binding by new ligands.

We would like to emphasize that the cutoff is a flexible value that can be changed according to the purpose. The recommended default free energy cutoffs are useful to identify cryptic binding sites and additional interaction sites in proteins that are structurally well-characterized. Higher isovalue cutoffs may be used for new protein targets with limited structural data and whose binding sites are unknown and uncharacterized. When resources and budget are limited, higher isovalue cutoff values may also be used to select the most important auxiliary binding sites for exploitation by new ligands.
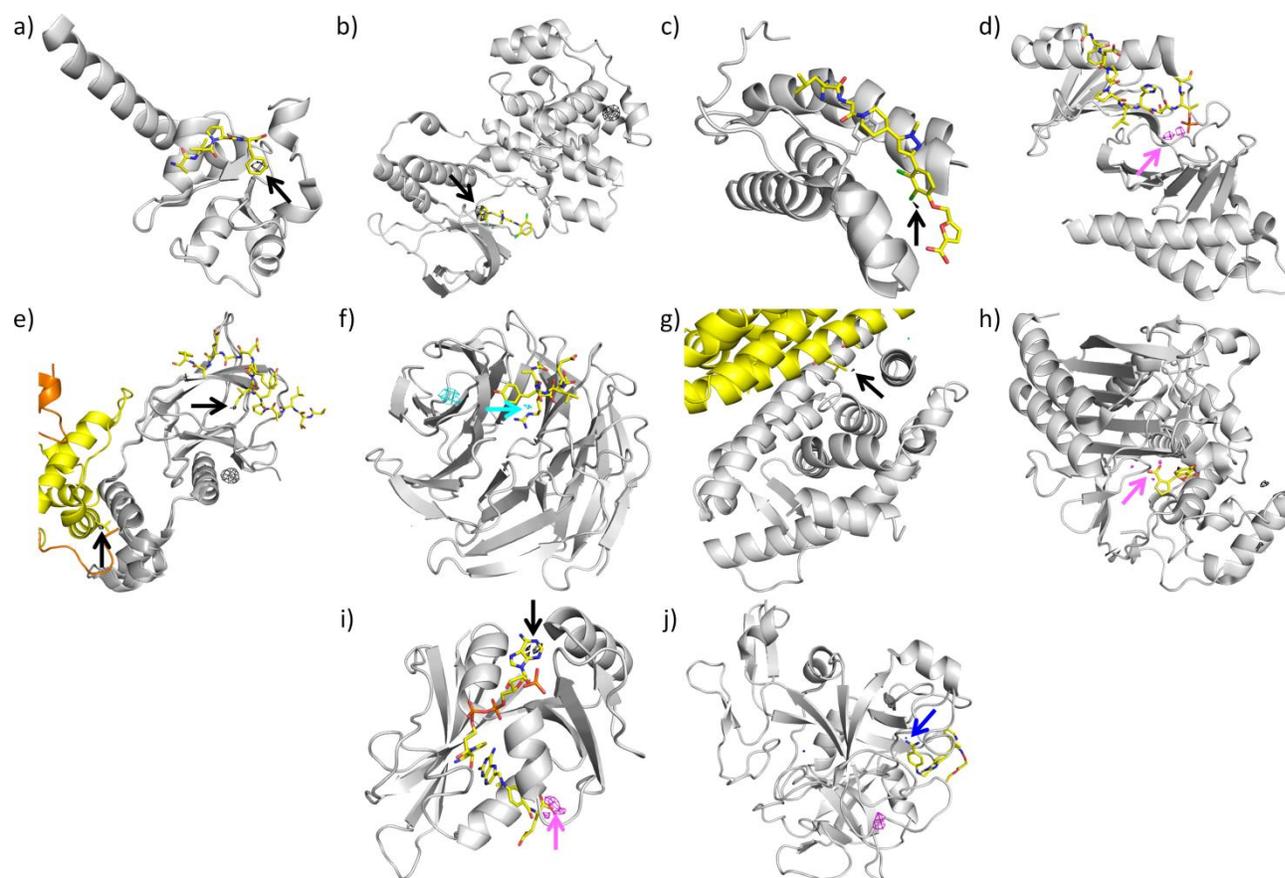
Figure 12. Ligand occupancy maps of (a) XIAP BIR3, (b) p38 MAPK, (c) IL-2, (d) PLK1 PBD, (e) VHL, (f) WDR5, (g) ERα, (h) AmpC, (i) DHFR, and (j) factor Xa superimposed on their ligand-bound structures (PDB codes 2OPZ, 1OUY, 1PY2, 3P37, 6BVB+3ZRF, 5SXM, 1ERR, 1XGJ, 3DFR, and 1G2L, respectively) and represented at high isovalues to show the top-ranked binding sites. All ligands are shown in yellow. The colored arrows indicate ligand densities that are found at known binding sites. The ratio for benzene, methanol, acetaldehyde, methylammonium and acetate isovalues is maintained at 6:3:3:3:2 for visualization.

4. CONCLUSION

We have presented a new multiple-ligand implementation of the LMMD method called mLMMD for efficient detection of diverse binding sites. Multiple types of small-molecule probes with different functional groups are incorporated into MD simulations to map the positions of ligand binding sites on proteins. The method was extensively validated on 10 proteins from different families. The results confirm that mLMMD is able to correctly identify the locations of hydrophobic, polar, and charged binding sites. The cryptic pocket detection ability of benzene is also not compromised by the presence of other probes, as cryptic pockets in p38 MAPK, IL-2, and PLK1 were successfully identified. This is an important development for this field as multiple types of binding sites can now be mapped with a single set of simulations, which is much more efficient than multiple sets of simulations that use only one type of ligand for mapping each time. mLMMD is also the only MD-based pocket detection method that allows for the simultaneous use of hydrophobic, hydrogen-bonding, and charged probes without the need for introduction of artificial interligand repulsive potentials. We have shown in this study that it is able to detect adjacent negative charge sites in DHFR and factor Xa, which may prove challenging to map when repulsive potentials between probes with the same charge are present.[21]

We note that some known hydrogen bond donor sites in proteins involved in PPIs (IL-2, PLK1 and VHL) are detected only at a higher energy cutoff of -1.5 kcal/mol, compared to -1.75 kcal/mol for other donor sites. This could be because hydrogen bond donor sites found at shallow protein–protein interfaces tend to be more exposed to the solvent compared to those found deep within the active sites of enzymes. Hydrogen bonds formed at protein–protein interfaces with hydrogen bond donor probes are relatively weaker as they are more easily replaced by those with water. We recommend that a less stringent cutoff of -1.5 kcal/mol be used for visualizing hydrogen bond donor sites on protein–protein interfaces, while -1.75 kcal/mol should be used for

enzymes and receptor proteins to minimize the appearance of spurious densities. It was observed in the case of ERα that the acetaldehyde probe could generate false positives for hydrogen bond acceptor sites because its methyl group may be drawn towards hydrophobic regions. It is easy to eliminate these false positives, however, simply by checking for chemical complementarity of residues in the immediate environment of the identified interaction sites.

There are several potential applications of mLMMD besides binding site detection. The ability of mLMMD to enhance sampling of ligand-binding conformations, especially those with open cryptic pockets, is particularly useful in ensemble docking, which makes use of multiple protein conformations for virtual screening of compounds. Indeed, it has been shown that protein structures generated by cosolvent MD are more suitable than conventional MD for ensemble docking.[73] The ligand occupancy maps generated by mLMMD could also be used to guide rational drug design. New interaction sites that have been identified close to where existing inhibitors are bound could be used to improve their potency. In this study, we have highlighted a few of such promising interaction sites in VHL, WDR5 and factor Xa. When multiple types of interactions are possible at a binding site, as indicated by the overlap of more than one ligand affinity map, the cutoff isovalue or free energy associated with each type of interaction can help suggest which functional group is the most preferred. For example, based on the affinity maps generated for the test proteins, we predict that a hydrogen bond donor is preferred to a positive charge in a region just outside the ERα agonist binding pocket, and that positive charges form stronger interactions than hydrogen bond donors at the bottom of the DHFR active site. Finally, mLMMD could be used to identify displaceable crystallographic water molecules in the binding cavity, as we have done for WDR5 and factor Xa. Ligands could then be designed to occupy the positions of these displaceable waters and mimic their hydrogen bonding contacts with the

protein. We are currently looking into exploiting this capability of mLMMD in structure-based drug design.

ASSOCIATED CONTENT

Supporting Information.

The Supporting Information is available free of charge.

RMSD of LMMD simulations, radial distribution function plots, comparison of XIAP BIR3 ligand occupancy maps, and probe parameters (PDF)

AUTHOR INFORMATION

Corresponding Author

* Email: tanys@bii.a-star.edu.sg

Notes

C.S.V is the founder and director of SiNOPSEE Therapeutics and Aplomex.

REFERENCES

(1)   Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** *274*, 1531-1534.

(2)   Mattos, C.; Ringe, D., Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* **1996,** *14*, 595-599.

(3)   Henrich, S.; Salo-Ahen, O. M. H.; Huang, B.; Rippmann, F.; Cruciani, G.; Wade, R. C., Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **2010,** *23*, 209-219.

(4)   Konc, J.; Janežič, D., Binding site comparison for function prediction and pharmaceutical discovery. *Curr. Opin. Struct. Biol.* **2014,** *25*, 34-39.

(5)   Vajda, S.; Beglov, D.; Wakefield, A. E.; Egbert, M.; Whitty, A., Cryptic binding sites on proteins: definition, detection, and druggability. *Curr. Opin. Chem. Biol.* **2018,** *44*, 1-8.

(6)   Kubinyi, H., Hydrogen bonding: the last mystery in drug design. *Pharmacokinetic optimization in drug research* **2001**, 513-24.

(7)   Eyrisch, S.; Helms, V., Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **2007,** *50*, 3457-3464.

(8)   Schmidtke, P.; Bidon-Chanal, A.; Javier Luque, F.; Barril, X., MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011,** *27*, 3276-3285.

(9)   Jukič, M.; Konc, J.; Janežič, D.; Bren, U., ProBiS H2O MD approach for identification of conserved water sites in protein structures for drug design. *ACS Med. Chem. Lett.* **2020,** *11*, 877-882.

(10) Seco, J.; Luque, F. J.; Barril, X., Binding site detection and druggability index from first principles. *J. Med. Chem.* **2009,** *52*, 2363-2371.

(11) Guvench, O.; MacKerell, A. D., Jr., Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comp. Biol.* **2009,** *5*, e1000435.

(12) Lexa, K. W.; Carlson, H. A., Full protein flexibility is essential for proper hot-spot mapping. *J. Am. Chem. Soc.* **2011,** *133*, 200-202.

(13) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I., Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J. Chem. Theory Comput.* **2012,** *8*, 2435-2447.

(14) Tan, Y. S.; Sledz, P.; Lang, S.; Stubbs, C. J.; Spring, D. R.; Abell, C.; Best, R. B., Using ligand-mapping simulations to design a ligand selectively targeting a cryptic surface pocket of polo-like kinase 1. *Angew. Chem. Int. Ed.* **2012,** *51*, 10078-10081.

(15) Tan, Y. S.; Spring, D. R.; Abell, C.; Verma, C., The use of chlorobenzene as a probe molecule in molecular dynamics simulations. *J. Chem. Inf. Model.* **2014,** *54*, 1821-1827.

(16) Kimura, S. R.; Hu, H. P.; Ruvinsky, A. M.; Sherman, W.; Favia, A. D., Deciphering cryptic binding sites on proteins by mixed-solvent molecular dynamics. *J. Chem. Inf. Model.* **2017,** *57*, 1388-1401.

(17) Schmidt, D.; Boehm, M.; McClendon, C. L.; Torella, R.; Gohlke, H., Cosolvent-enhanced sampling and unbiased identification of cryptic pockets suitable for structure-based drug design. *J. Chem. Theory Comput.* **2019,** *15*, 3331-3343.

(18) Yang, Y.; Mahmoud, A. H.; Lill, M. A., Modeling of halogen–protein interactions in co-solvent molecular dynamics simulations. *J. Chem. Inf. Model.* **2018,** *59*, 38-42.

(19) Prakash, P.; Sayyed-Ahmad, A.; Gorfe, A. A., pMD-membrane: a method for ligand binding site identification in membrane-bound proteins. *PLoS Comp. Biol.* **2015,** *11*.

(20) Graham, S. E.; Leja, N.; Carlson, H. A., MixMD Probeview: robust binding site prediction from cosolvent simulations. *J. Chem. Inf. Model.* **2018,** *58*, 1426-1433.

(21) Raman, E. P.; Yu, W.; Lakkaraju, S. K.; MacKerell, A. D., Jr., Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J. Chem. Inf. Model.* **2013,** *53*, 3384-3398.

(22) Sayyed-Ahmad, A.; Gorfe, A. A., Mixed-probe simulation and probe-derived surface topography map analysis for ligand binding site identification. *J. Chem. Theory Comput.* **2017,** *13*, 1851-1861.

(23) Graham, S. E.; Smith, R. D.; Carlson, H. A., Predicting displaceable water sites using mixed-solvent molecular dynamics. *J. Chem. Inf. Model.* **2018,** *58*, 305-314.

(24) Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L., Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *J. Am. Chem. Soc.* **2016,** *138*, 14257-14263.

(25) Horn, J. R.; Shoichet, B. K., Allosteric inhibition through core disruption. *J. Mol. Biol.* **2004,** *336*, 1283-1291.

(26) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M., K-Ras (G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013,** *503*, 548-551.

(27) Sledz, P.; Stubbs, C. J.; Lang, S.; Yang, Y. Q.; McKenzie, G. J.; Venkitaraman, A. R.; Hyvonen, M.; Abell, C., From crystal packing to molecular recognition: prediction and discovery of a binding site on the surface of polo-like kinase 1. *Angew. Chem. Int. Ed.* **2011,** *50*, 4003-4006.

(28) Hyde, J.; Braisted, A. C.; Randal, M.; Arkin, M. R., Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. *Biochemistry* **2003,** *42*, 6475-6483.

(29) Thanos, C. D.; Randal, M.; Wells, J. A., Potent small-molecule binding to a dynamic hot spot on IL-2. *J. Am. Chem. Soc.* **2003,** *125*, 15280-15281.

(30) Tan, Y. S.; Spring, D. R.; Abell, C.; Verma, C. S., The application of ligand-mapping molecular dynamics simulations to the rational design of peptidic modulators of protein–protein interactions. *J. Chem. Theory Comput.* **2015,** *11*, 3199-3210.

(31) Foster, T. J.; Mackerell, A. D., Jr.; Guvench, O., Balancing target flexibility and target denaturation in computational fragment-based inhibitor discovery. *J. Comput. Chem.* **2012,** *33*, 1880-91.

(32) Tan, Y. S.; Reeks, J.; Brown, C. J.; Thean, D.; Ferrer Gago, F. J.; Yuen, T. Y.; Goh, E. T. L.; Lee, X. E. C.; Jennings, C. E.; Joseph, T. L.; Lakshminarayanan, R.; Lane, D. P.; Noble, M.

E. M.; Verma, C. S., Benzene probes in molecular dynamics simulations reveal novel binding sites for ligand design. *J. Phys. Chem. Lett.* **2016,** *7*, 3452-3457.

(33) Lama, D.; Brown, C. J.; Lane, D. P.; Verma, C. S., Gating by tryptophan 73 exposes a cryptic pocket at the Protein-Binding interface of the oncogenic eIF4E protein. *Biochemistry* **2015,** *54*, 6535-6544.

(34) Iegre, J.; Brear, P.; Baker, D. J.; Tan, Y. S.; Atkinson, E. L.; Sore, H. F.; O' Donovan, D. H.; Verma, C. S.; Hyvönen, M.; Spring, D. R., Efficient development of stable and highly functionalised peptides targeting the CK2α/CK2β protein–protein interaction. *Chem. Sci.* **2019,** *10*, 5056-5063.

(35) Kaan, H. Y. K.; Sim, A. Y.; Tan, S. K. J.; Verma, C.; Song, H., Targeting YAP/TAZ-TEAD protein-protein interactions using fragment-based and computational modeling approaches. *PLoS One* **2017,** *12*, e0178381.

(36) Harish, M.; Kannan, S.; Puttagunta, S.; Pradhan, M. R.; Verma, C. S.; Venkatraman, P., A novel determinant of PSMD9 PDZ binding guides the evolution of the first generation of super binding peptides. *Biochemistry* **2019,** *58*, 3422-3433.

(37) Rajagopalan, M.; Balasubramanian, S.; Ramaswamy, A., Insights into the RNA binding mechanism of human L1-ORF1p: a molecular dynamics study. *Mol. Biosyst.* **2017,** *13*, 1728-1743.

(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28*, 235-242.

(39) Ndubaku, C.; Varfolomeev, E.; Wang, L.; Zobel, K.; Lau, K.; Elliott, L. O.; Maurer, B.; Fedorova, A. V.; Dynek, J. N.; Koehler, M., Antagonism of c-IAP and XIAP proteins is required for efficient induction of cell death by small-molecule IAP antagonists. *ACS Chem. Biol.* **2009,** *4*, 557-566.

(40) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C., Binding of small molecules to an adaptive protein-protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100*, 1603-1608.

(41) Yun, S. M.; Moulaei, T.; Lim, D.; Bang, J. K.; Park, J. E.; Shenoy, S. R.; Liu, F.; Kang, Y. H.; Liao, C. Z.; Soung, N. K.; Lee, S.; Yoon, D. Y.; Lim, Y.; Lee, D. H.; Otaka, A.; Appella, E.; McMahon, J. B.; Nicklaus, M. C.; Burke, T. R.; Yaffe, M. B.; Wlodawer, A.; Lee, K. S., Structural and functional analyses of minimal phosphopeptides targeting the polo-box domain of polo-like kinase 1. *Nat. Struct. Mol. Biol.* **2009,** *16*, 876-882.

(42) Min, J.-H.; Yang, H.; Ivan, M.; Gertler, F.; Kaelin, W. G.; Pavletich, N. P., Structure of an HIF-1α-pVHL complex: hydroxyproline recognition in signaling. *Science* **2002,** *296*, 1886-1889.

(43) Aho, E. R.; Wang, J.; Gogliotti, R. D.; Howard, G. C.; Phan, J.; Acharya, P.; Macdonald, J. D.; Cheng, K.; Lorey, S. L.; Lu, B., Displacement of WDR5 from chromatin by a WIN site inhibitor with picomolar affinity. *Cell Rep.* **2019,** *26*, 2916-2928. e13.

(44) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L., The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998,** *95*, 927-937.

(45) Fitzgerald, C. E.; Patel, S. B.; Becker, J. W.; Cameron, P. M.; Zaller, D.; Pikounis, V. B.; O'Keefe, S. J.; Scapin, G., Structural basis for p38α MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. *Nat. Struct. Mol. Biol.* **2003,** *10*, 764-769.

(46) Babaoglu, K.; Shoichet, B. K., Deconstructing fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2006,** *2*, 720-723.

(47) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J., Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate. *J. Biol. Chem.* **1982,** *257*, 13650-13662.

(48) Nar, H.; Bauer, M.; Schmid, A.; Stassen, J.-M.; Wienen, W.; Priepke, H. W.; Kauffmann, I. K.; Ries, U. J.; Hauel, N. H., Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure* **2001,** *9*, 29-37.

(49) Fiser, A.; Sali, A., ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003,** *19*, 2500-2501.

(50) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004,** *32*, W665-W667.

(51) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.;

Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*, University of California, San Francisco: 2014.

(52) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.

(53) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157-1174.

(54) Pang, Y.-P.; Xu, K.; El Yazal, J.; Prendergast, F. G., Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci.* **2000,** *9*, 1857-1865.

(55) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz Jr, K. M., Rational design of particle mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit solvent. *J. Chem. Theory Comput.* **2013,** *9*, 2733-2748.

(56) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y., R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011,** *39*, W511-W517.

(57) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A., Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993,** *115*, 9620-9631.

(58) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision B.1*; Gaussian, Inc.: Wallingford CT, 2009.

(59) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23*, 327-341.

(60) Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: an N•log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.

(61) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D., Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001,** *114*, 2090-2098.

(62) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.

(63) Alvarez-Garcia, D.; Barril, X., Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites. *J. Med. Chem.* **2014,** *57*, 8530-8539.

(64) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002.

(65) Lexa, K. W.; Goh, G. B.; Carlson, H. A., Parameter choice matters: validating probe parameters for use in mixed-solvent simulations. *J. Chem. Inf. Model.* **2014,** *54*, 2190-2199.

(66) Braisted, A. C.; Oslob, J. D.; Delano, W. L.; Hyde, J.; McDowell, R. S.; Waal, N.; Yu, C.; Arkin, M. R.; Raimundo, B. C., Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J. Am. Chem. Soc.* **2003,** *125*, 3714-3715.

(67) Liu, F.; Park, J.-E.; Qian, W.-J.; Lim, D.; Graeber, M.; Berg, T.; Yaffe, M. B.; Lee, K. S.; Burke, T. R., Jr., Serendipitous alkylation of a Plk1 ligand uncovers a new binding channel. *Nat. Chem. Biol.* **2011,** *7*, 595-601.

(68) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P., Halogen bonding: the sigma-hole. *J. Mol. Model.* **2007,** *13*, 291-296.

(69) Yu, W.; Lakkaraju, S. K.; Raman, E. P.; Fang, L.; MacKerell, A. D., Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules. *J. Chem. Inf. Model.* **2015,** *55*, 407-420.

(70) Bennett, B.; Langan, P.; Coates, L.; Mustyakimov, M.; Schoenborn, B.; Howell, E. E.; Dealwis, C., Neutron diffraction studies of Escherichia coli dihydrofolate reductase complexed with methotrexate. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 18493-18498.

(71) Cocco, L.; Roth, B.; Temple Jr, C.; Montgomery, J. A.; London, R. E.; Blakley, R. L., Protonated state of methotrexate, trimethoprim, and pyrimethamine bound to dihydrofolate reductase. *Arch. Biochem. Biophys.* **1983,** *226*, 567-577.

(72) Padmanabhan, K.; Padmanabhan, K.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D.; Cardin, A.; Kisiel, W., Structure of human des (1-45) factor Xa at 2· 2 Å resolution. *J. Mol. Biol.* **1993,** *232*, 947-966.

(73) Uehara, S.; Tanaka, S., Cosolvent-based molecular dynamics for ensemble docking: practical method for generating druggable protein conformations. *J. Chem. Inf. Model.* **2017,** *57*, 742-756.

Table of Contents graphic