JCTC Journal of Chemical Theory and Computation

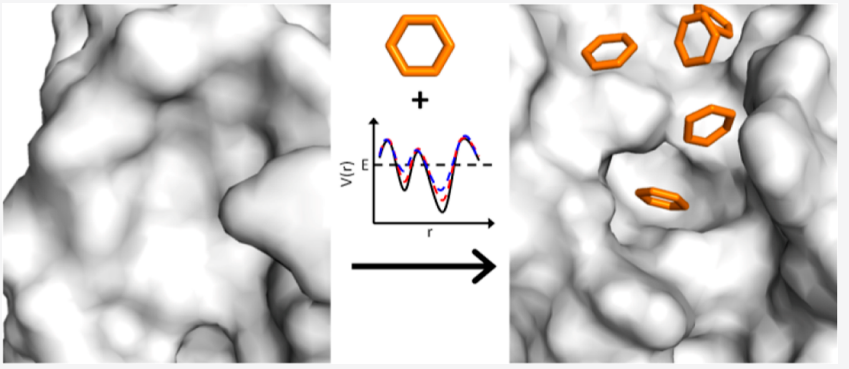
pubs.acs.org/JCTC Article

Accelerated Ligand-Mapping Molecular Dynamics Simulations for the Detection of Recalcitrant Cryptic Pockets and Occluded Binding Sites

Justin Tze-Yang Ng and Yaw Sing Tan*

Cite This: J. Chem. Theory Comput. 2022, 18, 1969–1981 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: The identification and characterization of binding sites is a critical component of structure-based drug design (SBDD). Probe-based/cosolvent molecular dynamics (MD) methods that allow for protein flexibility have been developed to predict ligand binding sites. However, cryptic pockets that appear only upon ligand binding and occluded binding sites with no access to the solvent pose significant challenges to these methods. Here, we report the development of accelerated ligand-mapping MD (aLMMD), which combines accelerated MD with LMMD, for the detection of these challenging binding sites. The method was validated on five proteins with what we term "recalcitrant" cryptic pockets, which are deeply buried pockets that require extensive movement of the protein backbone to expose, and three proteins with occluded binding sites. In all the cases, aLMMD was able to detect and sample the binding sites. Our results suggest that aLMMD could be used as a general approach for the detection of such elusive binding sites in protein targets, thus providing valuable information for SBDD.

1. INTRODUCTION

Structure-based drug design (SBDD) involves the exploitation of information from three-dimensional biomolecular structures to aid drug discovery efforts.[1] A key component of SBDD involves the identification and characterization of ligand binding sites in biomolecular targets. Such information allows the rational design and modification of ligands to enhance binding with their targets.[2] Additionally, modulation of a target protein's activity can be achieved via allostery when a ligand binds away from the orthosteric site.[3,4] One of the major challenges in SBDD is accounting for receptor flexibility.[5,6] Proteins frequently undergo a wide variety of conformational changes in their unbound states and upon ligand binding. As a consequence of receptor flexibility, binding sites that are not apparent in the unbound protein structure may form upon ligand binding. These so-called cryptic binding sites may enable the targeting of proteins previously thought to be undruggable, bind allosteric ligands, or be exploited to enhance the potency of existing inhibitors.[7]

Various computational methods for the identification and characterization of binding sites have been developed as cheaper and faster alternatives to experimental approaches. However, computational approaches that do not consider receptor flexibility[8] may not be able to detect binding sites that are absent in the input structures. This limitation can be addressed by the use of molecular dynamics (MD) simulations, which provide an atomistic description of the dynamics of a system.[9] The relaxed complex scheme[10] involves the use of MD simulations to generate a conformational ensemble of the receptor for small-molecule docking.[11] This method was most notably used to identify a novel pocket in HIV-1 integrase,[12] which was eventually utilized for inhibitor development.[13]

Received: November 22, 2021
Published: February 17, 2022

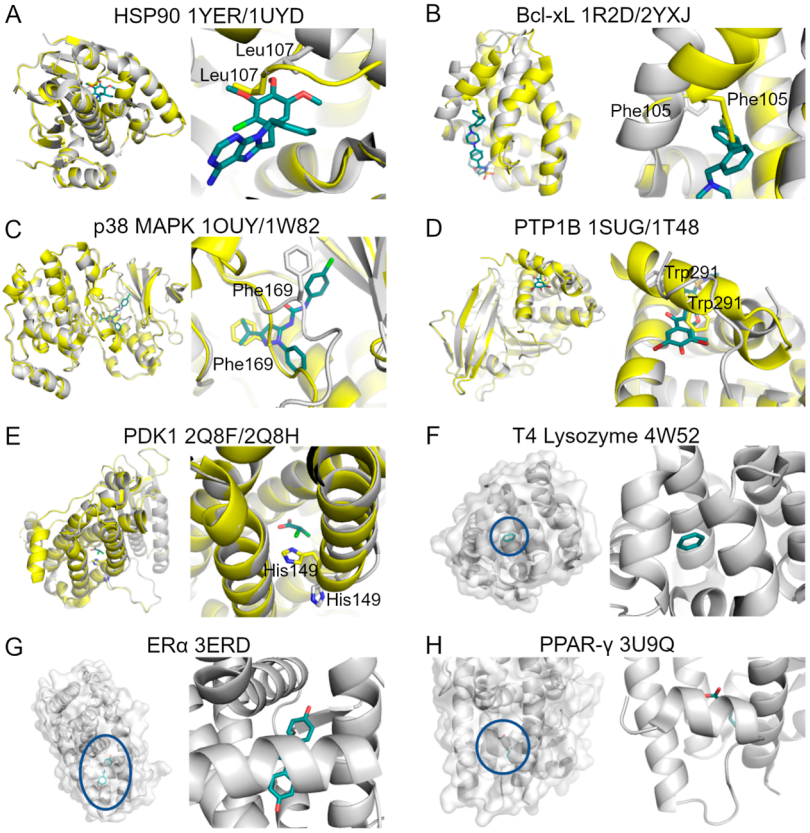


Figure 1. Structures of test proteins. Ligands are shown as teal sticks. For each subfigure, the overall protein structure is shown on the left, while a close-up of the cryptic or occluded binding site is shown on the right. (A–E) Superimposition of test protein structures with their cryptic pockets unoccupied (yellow, PDB code on left) and occupied (white, PDB code on right) by a ligand. Residues relevant to cryptic pocket formation are shown as sticks. (F–H) Crystal structures of test proteins with occluded binding sites. PDB codes of the holo protein structures are indicated.

However, cryptic binding sites are less amenable to detection by conventional MD (cMD) simulations, even when extended to microsecond timescales,[14] while enhanced sampling MD techniques such as metadynamics and umbrella sampling require a priori knowledge of the binding site for selection of suitable collective variables.[15] Enhanced sampling methods that do not require the use of collective variables, such as parallel tempering, have also been shown to be ineffective at opening cryptic binding sites.[14] Adaptive sampling algorithms have shown promise in revealing cryptic binding sites but do not indicate how ligands can interact with the pocket.[16,17] To overcome the limitations of these MD-based methods, probe-based/cosolvent MD methods for binding site detection have been developed.[18−23] In this approach, small-molecule probes are added into a water box with the protein of interest and the system is subjected to classical MD simulations to allow the probes to bind to the protein surface to generate ligand affinity maps. The use of small-molecule probes in MD simulations allows for receptor conformational changes induced by ligand binding. These probe-based MD methods include ligand-mapping MD (LMMD), which was developed by our group. It uniquely uses low concentrations of hydrophobic probes[22,24,25] to detect binding sites without the use of additional artificial forces to prevent probe aggregation and protein unfolding. It has been applied to the detection of cryptic pockets[22,26,27] and the design of small-molecule[28] and peptide-based[22,29,30] protein−protein interaction inhibitors. Most recently, a multiple-probe implementation of LMMD has been developed, which allows for simultaneous mapping of different types of binding sites by multiple types of probes.[31]

Although probe-based MD methods have been demonstrated to be successful at detecting most cryptic pockets, deeply buried cryptic pockets that require large movements of secondary structure elements and domains to expose still pose a formidable challenge. We term such pockets "recalcitrant cryptic pockets". Substantial sampling may be required to allow time for the probes to fully explore the protein surface and overcome high energy barriers of pocket opening.[32] Cimermancic et al.[33] curated a data set of proteins with 84 cryptic binding sites based on ligand-free and ligand-bound structures in the Protein Data Bank (PDB).[34] Eight of these proteins were chosen for systematic validation of a probe-based MD protocol.[35] Although all of the pockets were opened by the probes, half of them only opened partially. More recently, another study was carried out to apply probe-based MD simulations to the detection of cryptic pockets.[36] Partial pocket opening was observed in three out of the seven test cases. These are challenging cases for which pocket opening likely involves timescales beyond that of cMD simulations. Another class of binding sites known as occluded binding sites also poses a significant challenge for probe-based MD simulations. Occluded binding sites, in contrast to cryptic binding sites, pre-exist in the unbound structures of the protein, have largely the same conformation in the unliganded and liganded states, and are not accessible to the solvent.[37] Conformational changes such as movement of multiple side chains and helical segments[38,39] have to occur to allow ligands to enter these

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

binding cavities. Ligand binding at occluded binding sites has been shown to occur at microsecond timescales in cMD simulations.[39] These binding sites are found in human proteins such as nuclear receptors and G-protein-coupled receptors and in bacterial proteins such as quorum-sensing transcription factors.

The combination of enhanced sampling methods with probe-based MD simulations can help to reduce the timescale needed to map these challenging binding sites. The probe-based MD method SILCS has been coupled with a grand canonical Monte Carlo (GCMC) scheme[40] to facilitate the detection of occluded ligand binding sites,[41] while Gervasio and co-workers have used their Hamiltonian replica exchange method known as SWISH in combination with 1 M of benzene probes to detect cryptic pockets.[14,42] In both methods, repulsive inter-probe potentials and restraints on protein atoms are introduced to prevent probe aggregation and protein unfolding, respectively. However, these artificial forces could prevent the proper mapping of large pockets that can accommodate more than one benzene molecule and hinder the opening of recalcitrant cryptic pockets. Separately, isopropanol probes have been used in accelerated MD (aMD)[43] simulations, which enhance conformational sampling by reducing the energy barriers between low-energy states and do not require a priori information of the cryptic pocket, to generate druggable conformations of B-cell lymphoma-extra-large (Bcl-xL) for docking.[44] However, this work was only demonstrated on a single protein, and it was not clear on the extent to which the cryptic site was detected.[44] While our work was in progress, Smith and Carlson most recently reported the use of their probe-based method, MixMD, in combination with aMD[45] for the detection of cryptic binding sites. Multiple simulation runs were performed with different probe types, including isopropanol, acetonitrile, pyrimidine, and a methyl-ammonium + acetate combination, on 12 systems. In two cases, accelerated MixMD was necessary to detect the cryptic binding site, which was missed by standard MixMD. For five of the cases, the cryptic binding sites were not detected by either standard MixMD or accelerated MixMD.[45]

Here, in this work, we report the development of accelerated LMMD (aLMMD), which combines LMMD with aMD to enable the detection of recalcitrant cryptic pockets and occluded binding sites. This accelerated version of LMMD incorporates a low concentration of benzene molecules into aMD simulations to detect the aforementioned binding sites. Benzene was selected as the ligand of choice as these binding sites are predominantly hydrophobic in nature[32,46−48] and it has been shown to be very effective at exploring cryptic pockets.[14] Benzenes encourage the exposure of these hydrophobic binding sites by reducing the polarity of the aqueous solvent environment. We validated aLMMD on a test set comprising a diverse range of challenging protein systems, including five cases of recalcitrant cryptic pockets and three cases of occluded binding sites. By systematically applying aLMMD and standard LMMD to these targets and comparing their results, we demonstrate the value of aLMMD as a general method for the detection of cryptic and occluded binding sites.

2. METHODS

2.1. Choice of Systems. We selected a test set comprising five proteins (heat shock protein 90 [HSP90], Bcl-xL, p38 mitogen activated protein kinase [p38 MAPK], protein tyrosine phosphatase 1B [PTP1B], and pyruvate dehydrogenase kinase 1 [PDK1]) with deeply buried cryptic pockets that require major movements of the backbone to expose and three proteins (T4 lysozyme L99A mutant, estrogen receptor alpha [ERα], and peroxisome proliferator-activated receptor gamma [PPAR-γ]) with occluded binding sites (Figure 1). The cryptic pockets in HSP90, Bcl-xL, and PTP1B, and occluded binding sites in PPAR-γ and the T4 lysozyme L99A mutant were used as test cases in previous probe-based MD studies.[21,36,45,49,50] In particular, the cryptic pockets of HSP90, Bcl-xL, and PTP1B have been identified as especially challenging to detect using conventional probe-based MD simulations.[36,45] PDK1 was selected from the Cryptosite data set[33] and represents a special case of a buried cryptic pocket.

The following PDB structures were used as initial structures for MD simulations: 1YER[51] (HSP90), 1R2D[52] (Bcl-xL), 1OUY[53] (p38 MAPK), 1SUG[54] (PTP1B), 2Q8F[55] (PDK1), 4W51[56] (T4 lysozyme L99A mutant), 3ERD[57] (active ERα), and 3U9Q[58] (PPAR-γ). The cryptic pocket is in the unbound and closed state in the first five structures, while the ligand binding site is completely occluded by helical segments in the last three structures.

2.2. System Setup. All crystallographic ligands were removed. Missing loop residues were modeled using MODELLER.[59] The N- and C- termini of protein chains were capped with acetyl and *N*-methyl groups, respectively, if the first or last residue in the native sequence was missing. PDB2PQR[60] was used to determine residue protonation states and add hydrogen atoms. Using the LEaP module of AMBER18,[61] each system was solvated with TIP3P[62] water molecules in a periodic truncated octahedron box such that the distance between the protein and box edge was at least 10 Å. Sodium or chloride counterions were then added to neutralize the system. A 50 ns cMD simulation of the apo protein was performed to determine the appropriate number of benzene molecules to be added from the average box volume. Packmol[63] was used to generate 20 different random distributions of benzene probes around the protein for LMMD simulations. Each system was then solvated with TIP3P[62] water molecules followed by charge neutralization with sodium or chloride ions to give a final benzene concentration of approximately 0.2 M.

2.3. MD Simulations. Energy minimization and MD simulations were performed using the particle mesh Ewald MD (PMEMD) module of AMBER18.[61] The ff14SB[64] force field was used for proteins, while the generalized AMBER force field (GAFF)[65] was used for benzenes. Atomic charges for benzene were taken from previous work.[29] Bonds involving hydrogen atoms were constrained using the SHAKE[66] algorithm to enable a time step of 2 fs. A nonbonded interaction cutoff distance of 9 Å was used. Long-range electrostatic interactions were treated using the particle mesh Ewald[67] method 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 in the *NVT* ensemble before equilibration in the *NPT* ensemble at a constant pressure of 1 atm for another 50 ps. Weak harmonic positional restraints with a force constant of 2.0 kcal mol^{-1} Å^{-2} were imposed on heavy atoms during energy minimization and equilibration steps. Further, equilibration without positional restraints was performed for 2 ns, followed by the production run (50 ns for protein-only cMD and 200 ns for standard LMMD) at 300 K and 1 atm. The

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

temperature of the system was maintained using a Langevin[68] thermostat with a collision frequency of 2 ps^{-1}, while the pressure was maintained using the Berendsen barostat[69] with a pressure relaxation time of 2 ps.

2.4. aMD Simulations. The aMD enhanced sampling method enables a system to overcome energy barriers more easily. In aMD, a boost potential $\Delta V(r)$ is added to a system's potential energy when the system's potential energy is lower than the threshold energy value E.

$$V(r)^* = V(r) + \Delta V(r), \qquad V(r) < E$$
$$V(r)^* = V(r), \qquad V(r) \geq E \tag{1}$$

In this equation shown above, $V(r)^*$ is the modified potential energy, $V(r)$ is the unmodified original potential energy, and $\Delta V(r)$ is the boost potential. In the dual-boost[70] implementation of aMD, a boost potential is applied to the system's potential energy and dihedral energy as follows:

$$\Delta V(r) = \frac{(E_p - V(r))^2}{(\alpha_P + E_p - V(r))} + \frac{(E_d - V_d(r))^2}{(\alpha_D + E_d - V_d(r))} \tag{2}$$

where E_p is the threshold potential energy, E_d is the threshold dihedral energy, $V(r)$ is the system's potential energy, and $V_d(r)$ is the system's dihedral energy. α_P is the acceleration factor for the system's potential energy and α_D is the acceleration factor for the system's dihedral energy. The boost parameters were calculated as follows:

$$\alpha_P = 0.2 \times N_{atoms} \tag{3}$$

$$E_p = V_{PE\_avg} + 0.2N_{atoms} \tag{4}$$

$$\alpha_D = \frac{4N_{res}}{5} \tag{5}$$

$$E_d = V_{dihed\_avg} + 4N_{res} \tag{6}$$

where N_{atoms} is the total number of atoms, N_{res} is the number of protein residues, V_{PE_avg} is the average potential energy, and V_{dihed_avg} is the average dihedral energy. V_{PE_avg} and V_{dihed_avg} for aLMMD and standard aMD were obtained from the first 50 ns of the corresponding standard LMMD and cMD simulations, respectively.

The aLMMD workflow is outlined in Figure 2. Dual-boost aMD simulations, in which both the dihedral energy and total potential energy are boosted, were initiated from the final equilibrated structures of the corresponding MD simulation runs. All aMD simulations were performed at 300 K and 1 atm. In total, 20 independent 200 ns aLMMD and standard aMD simulations were performed for each protein.

2.5. Trajectory Analysis. For both standard LMMD and aLMMD simulations, the 20 individual runs for each protein were combined into a single trajectory for analysis. Using the CPPTRAJ[71] module in AMBER18,[61] the trajectory snapshots saved at 10 ps intervals from the last 70 ns of each simulation were aligned with respect to their Cα atoms before binning the benzene carbon atoms into 1 Å × 1 Å × 1 Å grid cells to generate benzene occupancy maps. Based on the inverse Boltzmann relationship, the binding free energy associated with a probe atom at voxel i is given by the following equation

$$\Delta G_i = -RT \ln \frac{N_i}{N_0} \tag{7}$$

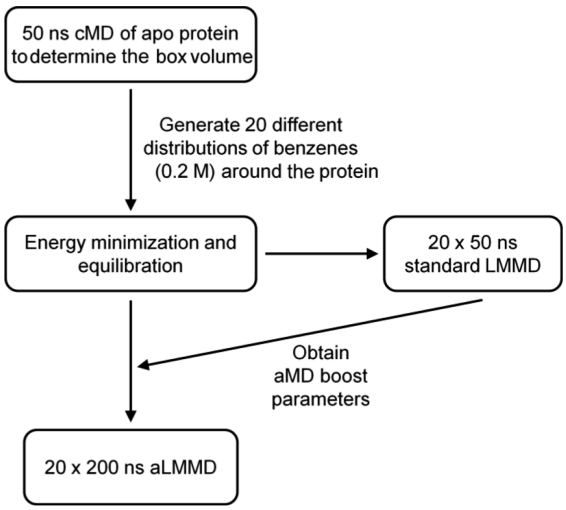


Figure 2. Flowchart of the aLMMD protocol.

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. The expected occupancy of a voxel by a probe atom can be obtained using the equation

$$N_0 = \frac{n_{atoms} \times n_{frames} \times C_p \times N_A}{10^{27}} \tag{8}$$

where n_{atoms} is the number of defined probe atoms per molecule (n_{atoms} = 6 for benzene), n_{frames} is the total number of trajectory frames being analyzed, C_p is the molar concentration of the probe (0.2 M), 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. Benzene occupancy maps were visualized at −1.5 kcal/mol as it has been found to be the best compromise between filtering out spurious binding sites and keeping the true positives.[31] Based on this cutoff, benzenes are approximately at least 12.4 times more likely to be found in the regions indicated by the benzene occupancy maps than the bulk solvent. To compare the overlap of the benzene occupancy maps with known ligands, ligand-bound structures were aligned using PyMOL[72] to the respective average protein structure sampled during the simulations.

The radius of gyration (R_g) of the Cα atoms of each protein was calculated using the CPPTRAJ[71] module in AMBER18.[61] Cavity analysis was performed using MDpocket[73] for the last 70 ns of the aLMMD and standard aMD trajectories. All water molecules, counterions, and benzene probes were removed from the trajectories, followed by structural alignment of the protein backbone atoms prior to MDpocket analysis.

2.6. Pocket Clustering and Ensemble Docking. Snapshots from the last 70 ns of the aLMMD runs of the five proteins with recalcitrant cryptic pockets were clustered using the k-means clustering algorithm in the CPPTRAJ[71] module of AMBER18. Clustering was based on the all-atom root-mean-square deviation (RMSD) of residues that are within 4 Å of the ligand in the respective holo protein structures shown in Figure 1. The number of clusters was set to 15, and the centroid of each cluster was selected as a representative for the docking.

Docking of crystallographic ligands to the representative cluster structures was carried out using AutoDock Vina 1.1.2.[74] The search space was defined by aligning all the 15 representative cluster structures to the holo protein and

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

including all residues that are within 4 Å of the ligand. Exhaustiveness was set to 64. The RMSD between the docked pose and the crystallographic pose was calculated using the rms_cur tool in PyMOL[72] after alignment of the binding site residues that are within 4 Å of the ligand in the crystal structure.

3. RESULTS AND DISCUSSION

In this study, we report the development of aLMMD for the detection of cryptic pockets and occluded binding sites. We first describe the modifications made to standard LMMD. We then apply aLMMD to our test set of five cryptic pockets and three occluded binding sites. To demonstrate the challenge these binding sites pose to conventional probe-based MD and enhanced sampling methods, we also performed the same number of runs and simulation length (20 × 200 ns) of standard LMMD and aMD as aLMMD on all the targets and compared the results.

3.1. Optimization of the aLMMD Protocol. Benzene concentration was retained at 0.2 M. One cryptic pocket test case and one occluded binding site test case were selected to optimize the number of independent simulation runs, simulation length, and analysis window for aLMMD. HSP90 was chosen for its highly challenging cryptic pocket, while the T4 lysozyme L99A mutant was chosen for its small size, thus ensuring rapid testing of multiple simulation protocols.

aLMMD was initially performed on the T4 lysozyme L99A mutant using 10 simulation runs of 50 ns each, similar to the multiple-probe implementation of LMMD.[31] However, no densities corresponding to the occluded binding site of the T4 lysozyme L99A mutant could be observed in the benzene occupancy map (Figure S1). To enhance the sampling of conformational space,[75] the number of simulation runs was increased to 20, but the T4 lysozyme L99A mutant binding site remained elusive (Figure S2). As previous studies have suggested that longer timescales may be needed in probe-based MD methods to detect binding sites occluded by the protein backbone,[25,36,45] we decided to increase the length of each simulation run to 200 ns. The occluded binding site of the T4 lysozyme L99A mutant was successfully detected when 20 aLMMD simulations of 200 ns each were performed (Figure S3A). By contrast, standard LMMD failed to detect the binding site (Figure S3B).

We applied this protocol to the detection of the cryptic binding site in HSP90. However, no densities corresponding to the cryptic pocket could be observed in the benzene occupancy maps generated from both aLMMD and standard LMMD simulations (Figure S4). The entire simulation length was used to generate the benzene occupancy maps for both types of simulations. We suspected that for such challenging cryptic pockets, pocket opening could occur at later timescales, as it takes time for the protein backbone to move away and expose the pocket for benzene binding. If this is the case, the sampling of data from earlier timescales could mask the presence of benzene densities within the pocket at later timescales. We examined data obtained from combining the 0−60, 60−130, and 130−200 ns segments of each aLMMD simulation run, each representing approximately one-third of the total simulation length. No benzene densities were observed within the cryptic pocket in benzene occupancy maps generated from the 0−60 and 60−130 ns (Figure S5) segments. Only the benzene occupancy maps generated from the 130−200 ns segment showed benzene densities within the cryptic pocket (Figure 3A). Not only does this show that pocket opening indeed occurs at later timescales for HSP90 but also shows that the simulation length of 200 ns is necessary for detection of the HSP90 cryptic pocket.

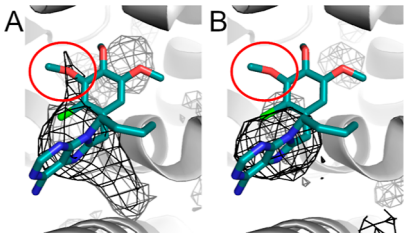


Figure 3. Benzene occupancy maps of HSP90 generated from (A) aLMMD and (B) standard LMMD simulations superimposed on the ligand-bound PDB structure 1UYD. Benzene occupancy maps are shown as black meshes. The red circles indicate the position of the cryptic pocket.

To assess the convergence of the aLMMD simulations, we divided the 20 200-ns trajectories of HSP90 and the T4 lysozyme L99A mutant into two sets of 10 trajectories each and generated benzene occupancy maps for each set from the 130−200 ns segment. The cryptic pocket in HSP90 and the occluded binding site in the T4 lysozyme L99A mutant were mapped in both sets of benzene occupancy maps (Figures S6 and S7), which indicates convergence of benzene densities in these challenging binding sites. Between both sets of occupancy maps, there are slight differences in the size and shape of the benzene densities within the pockets, which are refined by the combination of data from all 20 simulation runs. Therefore, we performed 20 aLMMD simulations of 200 ns each for the rest of the proteins in the test set and used the combined 130−200 ns segments of the trajectories for analysis.

As simulation length has significantly increased along with the implementation of aMD, there is concern over whether benzene-induced protein unfolding could occur during these long aLMMD simulations. The R_g is a measure of protein compactness and has been used as a metric to detect protein unfolding due to probe infiltration of the protein's hydrophobic core.[76] Here, we used R_g to monitor the stability of the protein fold over simulation time. In all the test cases, no sudden or large increases in R_g were observed (Figures S8−S15), which suggests that ligand-induced protein unfolding did not occur during the aLMMD simulations. This was confirmed by manual inspection of the trajectories. Small fluctuations in R_g in some of the runs can be attributed to movements of flexible loop regions and do not indicate unfolding. We did not use the RMSD of the protein atoms to evaluate protein unfolding as large fluctuations in RMSD values may reflect conformational changes in flexible regions associated with pocket opening. This is an important consideration as significant conformational changes are required to expose the binding sites in our test cases.

3.2. HSP90. HSP90 is a molecular chaperone that helps in protein folding and stabilization. It contains a cryptic site that requires the conformational transition of a loop segment into a helix for exposure. In the reference ligand-bound structure of HSP90 (PDB ID: 1UYD),[77] the 3-methoxy group of the inhibitor displaces Leu107 to occupy the cryptic site (Figure 1A). aLMMD was performed using an initial HSP90 structure with Leu107 blocking the cryptic site (PDB ID: 1YER).[51] Densities corresponding to the position of the 3-methoxy

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

group were observed in the aLMMD benzene occupancy maps (Figure 3A), while standard LMMD failed to detect this binding site (Figure 3B). This cryptic site was previously interrogated by a few probe-based MD studies, but either partial mapping was reported[36] or it was unclear whether the binding site was fully mapped based on the reference ligand-bound structures provided.[45,49]

3.3. Bcl-xL. Bcl-xL is an antiapoptotic protein that suppresses apoptosis by binding to the helical BH3 domains of proapoptotic proteins. It contains a cryptic binding site that is revealed by the movement of the helix α3 and changes in secondary structure of the residues N-terminal of α3. In one of the ligand-bound structures of Bcl-xL (PDB ID: 2YXJ),[78] the deepest part of the binding site is occupied by the 4-chlorophenyl moiety of the inhibitor (Figure 1B). aLMMD was performed using an initial structure in which this pocket is occupied by Phe105 (PDB ID: 1R2D).[52] Benzene densities that coincide with the location of the 4-chlorophenyl moiety were observed in the aLMMD occupancy maps (Figure 4A),

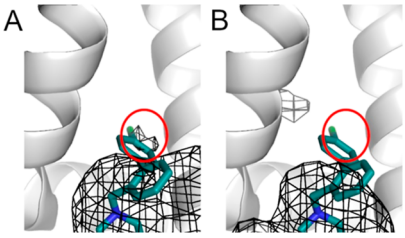


Figure 4. Benzene occupancy maps of Bcl-xL generated from (A) aLMMD and (B) standard LMMD simulations superimposed on the ligand-bound PDB structure 2YXJ. Benzene occupancy maps are shown as black meshes. The red circles indicate the position of the cryptic pocket.

while they were absent in the LMMD benzene occupancy maps (Figure 4B). Bcl-xL has been used as a test case in previous probe-based MD studies. LMMD simulations that used chlorobenzene probes were able to detect this cryptic site because the initial structure of Bcl-xL had the pocket in an open conformation.[24] Cosolvent MD simulations that used isopropanol probes in combination with aMD have been used to sample the conformations of apo Bcl-xL, but the extent of pocket opening was unclear.[44] A more recent study reported the use of 10% phenol as the cosolvent to detect this cryptic site starting from an apo structure and observed only partial pocket opening with incomplete displacement of Phe105.[36]

3.4. p38 MAPK. p38 MAPK is a protein kinase that controls important cellular processes such as proliferation, differentiation, and apoptosis. The activation loop containing the conserved Asp−Phe−Gly (DFG) motif[79] can adopt different conformations to bind different ligands. In the "DFG-in" conformation,[53] the DFG-Phe169 packs against the C-helix of the N-terminal lobe, while in the "DFG-out" conformation, Phe169 is flipped out, thus blocking the ATP binding site and exposing a cryptic pocket beneath (Figure 1C).[80] To evaluate the ability of the aLMMD simulations to map this cryptic pocket, we used a p38 MAPK structure in the DFG-in conformation for our simulations (PDB ID: 1OUY).[53] Densities corresponding to the position of a p38 MAPK inhibitor's *tert*-butyl group bound at the cryptic pocket could be observed in the benzene occupancy map generated from the aLMMD simulations (Figure 5A), while standard LMMD simulations failed to map this cryptic pocket (Figure 5B).

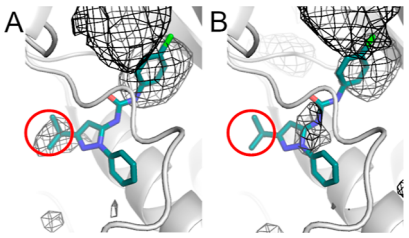


Figure 5. Benzene occupancy maps of p38 MAPK generated from (A) aLMMD and (B) standard LMMD simulations superimposed on the ligand-bound PDB structure 1W82. Benzene occupancy maps are shown as black meshes. The red circles indicate the position of the cryptic pocket.

Previous probe-based MD studies focused on mapping other binding sites of p38 MAPK, such as the ATP binding site[31,81] and the cryptic allosteric site formed by the surface loop backbone and side chain movements of residues 195−198.[21,31,36,42,49] These sites were also successfully mapped by the aLMMD simulations (Figure S16).

3.5. PTP1B. PTP1B is a negative regulator of the insulin receptor and has been shown to be a promising target for the treatment of type II diabetes and cancer. It contains a cryptic allosteric pocket that is occluded by the C-terminal helix α7 and occupied by the buried residue Trp291 in the ligand-free state (Figure 1D).[54] When a ligand binds at this allosteric site, α7 is displaced and becomes partially or fully disordered.[82] To evaluate whether aLMMD could identify the allosteric site when it is occluded by α7, we used a structure of apo PTP1B with α7 resolved for our simulations (PDB ID: 1SUG).[54] The allosteric pocket was successfully mapped by the aLMMD simulations but not by the standard LMMD simulations (Figure 6). Although previous probe-based MD studies have

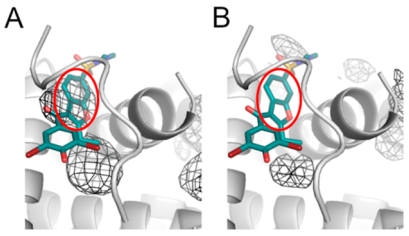


Figure 6. Benzene occupancy maps of PTP1B generated from (A) aLMMD and (B) standard LMMD simulations superimposed on the ligand-bound PDB structure 1T48. Benzene occupancy maps are shown as black meshes. The red circles indicate the position of the cryptic pocket.

also reported successful mapping of the PTP1B allosteric site,[21,50] we note that they used structures that did not contain a resolved helix α7 for simulation, resulting in an already exposed binding site that is accessible to the probe molecules. The challenge in detecting this binding site was highlighted by a recent study in which MixMD combined with aMD failed to map it.[45] In that study, 10 independent 100 ns simulations were carried out for each probe, none of which was fully hydrophobic in nature. The relative success of aLMMD in mapping this cryptic site could be attributed to the use of hydrophobic benzene probes, which create a favorable environment for hydrophobic pockets to open, and the implementation of a large number of simulation runs at moderately long timescales.

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

3.6. PDK1. PDK1 is a kinase that is found in the mitochondria and inhibits the activity of pyruvate dehydrogenase via phosphorylation of specific serine residues. It contains a buried cryptic pocket within the N-terminal domain.[55] The binding of the inhibitor dichloroacetate (DCA) to PDK1 involves the unwinding of a short segment of a helix bundle and the displacement and rotation of His149 toward the exterior of the helix bundle (Figure 1E). Simulations were carried out using an apo structure of PDK1 with His149 buried within the helix bundle and occupying the cryptic binding site (PDB ID: 2Q8F).[55] Densities corresponding to the location of the cryptic pocket were observed in the aLMMD benzene occupancy map (Figure 7A). The benzene densities overlapped with only one of the chlorine atoms of the bound DCA. The second chlorine atom could not be mapped, probably because it occupies a hydrophilic region of the pocket. Standard LMMD simulations were unable to map the pocket (Figure 7B).

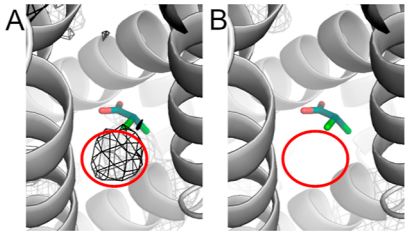


Figure 7. Benzene occupancy maps of PDK1 generated from (A) aLMMD and (B) standard LMMD simulations superimposed on the ligand-bound PDB structure 2Q8H. Benzene occupancy maps are shown as black meshes. The red circles indicate the position of the cryptic pocket.

3.7. T4 Lysozyme L99A Mutant. The L99A mutant of T4 lysozyme is a popular model system for studying the effect of buried cavities on protein stability, protein structure, and thermodynamics of ligand binding. It contains an occluded hydrophobic cavity[83] in the C-terminal domain that is created by the L99A mutation and can bind small hydrophobic ligands such as benzene (Figure 1F).[56] For benzene to enter and bind at this solvent-inaccessible pocket, subtle displacements of the surrounding helices have to occur.[39] Prominent benzene densities were observed within the cavity in the aLMMD benzene occupancy map (Figure 8A). They overlap well with the single benzene ligand. No benzene densities were found within the cavity in the standard LMMD benzene occupancy maps (Figure 8B). This is not surprising as it has been shown that benzene binding to the cavity is observed only in microsecond-long unbiased MD simulations.[39]

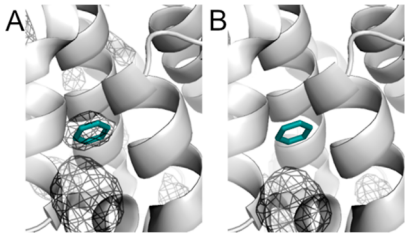


Figure 8. Benzene occupancy maps of the T4 lysozyme L99A mutant generated from (A) aLMMD and (B) standard LMMD simulations superimposed on the ligand-bound PDB structure 4W52. Benzene occupancy maps are shown as black meshes.

Although aLMMD simulations were able to detect the cavity, they were unable to map a narrow hydrophobic channel that extends from the main cavity and is occupied by the aliphatic moieties of alkyl benzenes.[56] This is presumably because the benzene probes are too bulky to enter this channel. This agrees with a previous study in which the hybrid probe-based MD method, SILCS-GCMC/MD, was able to map the channel with propane probes only.[40] No benzene densities were detected within the channel as well. This highlights a limitation of our current implementation of aLMMD in mapping channels that are too narrow for benzenes to access.

3.8. ERα. ERα is a nuclear receptor protein that adopts different conformations depending on whether it is bound to an agonist or antagonist.[38] In the antagonist-bound form, the ligand binding site is exposed and ERα is inactive. LMMD had been used to map this inactive conformation of ERα in a previous study.[31] Here, we are interested in the agonist-bound and active form of ERα, in which the C-terminal helix, helix 12, adopts a conformation that occludes the ligand binding site (Figure 1G).[58] This conformational change promotes homodimerization and the recruitment of coactivator proteins. aLMMD simulations were performed on the active form of ERα to determine if the occluded ligand binding site could be detected.

Although benzene densities were observed in the ERα ligand binding site in both aLMMD and standard LMMD benzene occupancy maps, there was more complete mapping of the ligand binding site by aLMMD (Figure 9). Benzene densities corresponding to both phenol moieties of the agonist diethylstilbestrol could be observed in the aLMMD benzene occupancy map (Figure 9A), while the standard LMMD benzene occupancy map covered only one of the two phenol moieties of the ligand (Figure 9B). Another agonist, *ortho*-trifluoromethylphenylvinyl estradiol,[84] was completely mapped by the aLMMD simulations (Figure 9C), while its trifluoromethylphenyl group was not mapped by standard LMMD (Figure 9D). Similarly, aLMMD was able to completely map a dimesitylborane ligand[85] (Figure 9E), while standard LMMD could not map the trifluoroethyl and mesityl groups of the ligand (Figure 9F). Our results here show that aLMMD offers an obvious advantage over standard LMMD in that it allows more complete and extensive mapping of the ERα ligand binding site.

3.9. PPAR-γ. PPAR-γ is a nuclear receptor involved in the regulation of lipid metabolism. Similar to other members of the nuclear receptor family,[38] the C-terminal helix of PPAR-γ is able to expose or occlude the ligand binding site depending on the binding of agonist or antagonist ligands. The simulations were performed using an initial structure of PPAR-γ with the ligand binding site occluded by the C-terminal helix and decanoic acid ligand removed (PDB ID: 3U9Q) (Figure 1H).[58] Both aLMMD and standard LMMD simulations were able to map the occluded binding site occupied by decanoic acid (Figure 10A,B). In addition, they also mapped a nearby second pocket that is occupied by the larger ligand GW409544[86] (Figure 10C,D).

The success of standard LMMD in mapping the ligand binding site of PPAR-γ suggests that it may not be as occluded as we thought it was. Indeed, the benzene occupancy maps indicate a pathway from the ligand binding site to the protein

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

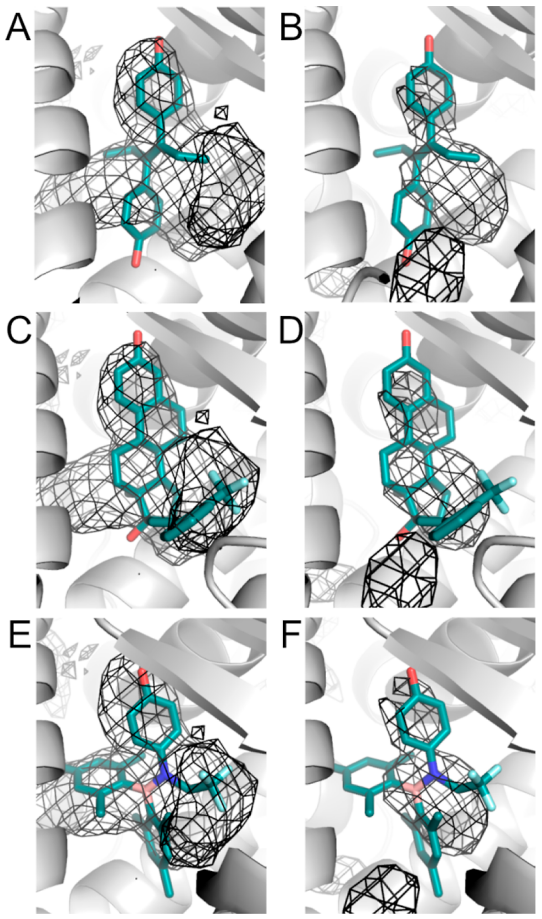


Figure 9. Benzene occupancy maps of ERα superimposed on ligand-bound structures. (A) aLMMD and (B) standard LMMD benzene occupancy maps superimposed on the PDB structure 3ERD. (C) aLMMD and (D) standard LMMD benzene occupancy maps superimposed on the PDB structure 2P15. (E) aLMMD and (F) standard LMMD benzene occupancy maps superimposed on the PDB structure 2Q6J. Benzene occupancy maps are shown as black meshes.

surface via the relatively solvent-exposed second pocket. Visualization of the simulations corroborates the pathway of benzene entry into the ligand binding site via the second pocket. This is also in line with observations reported in a previous study, in which probe densities traced a path from the ligand binding site to the second pocket and the protein surface in GCMC/MD simulations.[41]

To compare the timescale required for aLMMD and standard LMMD to detect the occluded binding site of PPAR-γ, we examined the benzene occupancy maps obtained from combining the 0−60, 60−130, and 130−200 ns segments of the respective simulation runs. The GW409544 ligand was completely mapped when the 0−60 and 60−130 ns analysis windows were used for the aLMMD simulations (Figure S17A,B). It was not mapped when the 0−60 ns analysis window was used and completely mapped when the 60−130 ns analysis window was used for the standard LMMD simulations (Figure S17C,D). This shows that aLMMD is able to completely map the PPAR-γ ligand binding site within a shorter timescale than standard LMMD.

3.10. Timescale of Binding Site Detection by aLMMD. Besides PPAR-γ, we were curious to find out if the binding sites in the rest of the test cases could be identified at an earlier time point, which would allow for shorter aLMMD simulations

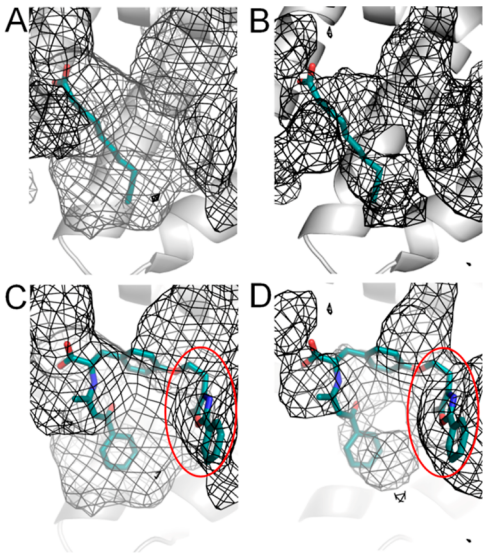


Figure 10. Benzene occupancy maps of PPAR-γ superimposed on ligand-bound structures. (A) aLMMD and (B) standard LMMD benzene occupancy maps superimposed on the PDB structure 3U9Q. (C) aLMMD and (D) standard LMMD benzene occupancy maps superimposed on the PDB structure 1K74. Benzene occupancy maps are shown as black meshes. The red circles indicate the location of the second pocket.

to be performed. All aLMMD trajectories were divided into analysis windows of 0−60, 60−130, and 130−200 ns, each representing approximately one-third of the total simulation length.

As shown earlier, the cryptic binding site for HSP90 could only be detected when the last 70 ns of the aLMMD simulations were analyzed (Figures 3A and S5). Interestingly, Bcl-xL represents another test case for which benzene densities corresponding to the cryptic pocket could be detected only when the last 70 ns of the aLMMD simulation was used for analysis. No benzene densities were observed within the Bcl-xL cryptic binding pocket for the two earlier analysis windows (Figure S18). By contrast, benzene densities at the cryptic pockets of p38 MAPK, PTP1B, PDK1, and the occluded binding site of the T4 lysozyme L99 mutant could be observed in benzene occupancy maps generated from the 60−130 and 130−200 ns analysis windows, but not the 0−60 ns analysis windows (Figures S19−S22). Similar to PPAR-γ, the occluded binding site in ERα could be detected in all three analysis windows. However, complete mapping of the diethylstilbestrol ligand was observed only for the 60−130 and 130−200 ns analysis windows (Figure S23).

Although our analysis shows that some of the binding sites were detected at earlier timescales, longer aLMMD simulation lengths of 200 ns were necessary to detect highly challenging binding pockets in two test cases (HSP90 and Bcl-xL). We recommend an aLMMD simulation length of 200 ns per run for all targets, especially in the absence of a priori knowledge about the presence of cryptic pockets and occluded binding sites in the protein.

3.11. aLMMD Generates Ligand-Accessible Conformations. It is clear from the benzene occupancy maps that benzene probes are able to access and occupy the cryptic pockets. To determine if aLMMD could generate cryptic pocket conformations that are accessible to known ligands, clustering of the aLMMD trajectory structures followed by

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

docking of crystallographic ligands to the representative cluster structures was carried out.

The docking pose with the lowest RMSD from the crystallographic pose was considered the best pose to assess whether the trajectories have sampled pocket conformations that can accommodate a known ligand. In all the five cases, the docked ligand has similar binding pose to the crystallographic ligand, with an RMSD of less than 3 Å (Figure 11). Visual

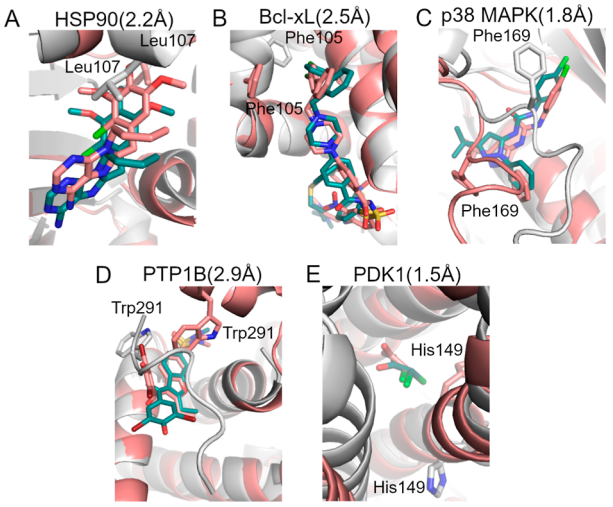


Figure 11. Docking poses of ligands at the cryptic pockets of (A) HSP90, (B) Bcl-xL, (C) p38 MAPK, (D) PTP1B, and (E) PDK1. The docked ligands (pink) and representative cluster structures (pink) they are docked to are shown superimposed on the respective ligand-bound crystal structure (ligand in teal and protein in white), with the ligand RMSD indicated. Residues relevant to cryptic pocket formation are shown as sticks.

inspection of the docked poses confirms that the docked ligand is able to access the cryptic pocket in all the test proteins. In the case of HSP90, aLMMD sampled cryptic pocket conformations that are larger than those in the crystal structures, which results in the ligand docking deeper into the pocket than its crystallographic pose.

Curiously, we note that all the p38 MAPK representative cluster structures are not in the DFG-out conformation, and yet the reference DFG-out ligand was able to access the cryptic pocket during docking. On visualization of the aLMMD trajectories, we found that although the activation loop remained mostly in the DFG-in conformation throughout all the simulation runs, there were a significant number of snapshots, including the one shown in Figure 11C, in which a substantial displacement of the DFG motif has occurred to allow access to the cryptic pocket. These conformations are reminiscent of the DFG-intermediate conformations that are observed in some kinase structures, in which the DFG-Phe is out of the C-helix pocket but not completely in a DFG-out conformation,[87,88] and could represent intermediate conformations of the DFG-in to DFG-out transition. This explains why the DFG-out ligand was able to dock into the cryptic pocket of the aLMMD snapshot even though the DFG-out conformation was not sampled. Our result agrees with a previous study in which a DFG-out ligand was able to dock into p38 MAPK DFG-in conformations generated by MD simulations,[89] and suggests that the transition from DFG-in to DFG-out conformation could be triggered by initial ligand binding.

3.12. Standard aMD Compared with aLMMD. It has been previously shown that aMD was unable to fully sample the ligand-bound conformations of Bcl-xL while cosolvent aMD using isopropanol allowed greater sampling of the initially buried binding pockets.[44] We were curious to see if aMD alone could detect any of the binding sites in our test set. Here, we used MDpocket[73] to determine if standard aMD could allow the exposure of cryptic pockets and the detection of occluded binding sites. MDpocket uses Voronoi tessellation and α-spheres to detect cavities in MD simulation trajectories.[73] The density map generated by MDpocket relates to the number of α-spheres around grid points normalized over simulation frames. Adjusting the isovalue for visualization would allow the detection of conserved cavities or transient binding channels, with a higher value allowing the visualization of more conserved cavities. MDpocket calculations were performed for both the aLMMD and aMD trajectories. To allow for a fair comparison, 20 independent 200 ns aMD runs were performed. For the recalcitrant cryptic pocket test cases, the highest isovalue at which densities corresponding to the cryptic pocket region are still visible in the aLMMD pocket density map was used as a reference value for visualization. No α-sphere densities were observed in all the cryptic pockets for the pocket density maps generated from the aMD simulations (Figures S24−S28), except for PDK1, which shows a tiny speck of density that becomes prominent at lower isovalues. This indicates that standard aMD is able to sample the open pocket conformation, but not as efficiently as aLMMD, and also that the PDK1 cryptic pocket is not as recalcitrant as the other cryptic pockets in the test set.

For the T4 lysozyme L99A mutant, ERα, and PPAR-γ, the highest isovalue at which densities corresponding to a transient channel from the protein surface to the occluded binding site are visible in the pocket density map generated from aLMMD simulations was used for visualization. No α-sphere densities corresponding to a transient channel could be observed in the pocket density maps generated from the aMD simulations of these proteins (Figures S29−S31). This means that someone who has no prior knowledge of the occluded cavity's ability to bind small ligands will not be able to detect it as a ligand binding site based on standard aMD simulations, as there is no visible pathway from the solvent to the cavity.

In summary, we found that aMD coupled with MDpocket analysis was unable to identify the binding site in seven out of eight test cases. Taken together, our analysis indicates that the combination of aMD with LMMD is essential for the detection of recalcitrant cryptic pockets and occluded binding sites.

4. CONCLUSIONS

We have presented an enhanced version of LMMD, called aLMMD, in which benzene probes are incorporated into aMD simulations, as a general and reliable method for the detection of "challenging" binding sites in proteins. These binding sites can be either recalcitrant cryptic pockets, which are deeply buried pockets that are closed in the absence of a bound ligand and open via large protein backbone movements on ligand binding, or occluded cavities that are inaccessible to the solvent in both the ligand-free and ligand-bound states. Although standard LMMD and other probe-based MD methods have been shown to be capable of identifying cryptic binding sites, there remain such elusive pockets that are not amenable to detection by these methods. Here, we describe eight cases of recalcitrant cryptic pockets and occluded binding

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

sites on which we validate the aLMMD method. The elusiveness of these binding sites was demonstrated by their inability to be fully mapped by standard LMMD, with the exception of PPAR-γ. Our results show that aLMMD is able to unequivocally map the target pockets. This is in contrast to previous probe-based MD studies, which either partially mapped or failed to detect cryptic pockets in HSP90, PTP1B, and Bcl-xL.[36,45] aLMMD also proved to be superior to standard LMMD in all the test cases. For six cases, standard LMMD was unsuccessful in detecting the binding site. In the case of ERα, aLMMD provides more complete mapping of the binding site than standard LMMD. For PPAR-γ, aLMMD was able to detect the binding site at an earlier time point than standard LMMD.

One limitation of aLMMD in its present form is that it cannot fully characterize the functional group affinity patterns of the detected binding sites. We have seen how the benzene probes were unable to map a narrow hydrophobic channel of the T4 lysozyme L99A mutant where aliphatic moieties bind. It is possible that multiple probe types can be introduced into aLMMD, like we did for mLMMD,[31] which incorporates multiple types of probes for simultaneous detection and characterization of binding sites. However, we foresee significant challenges in the characterization of occluded binding sites, for which ligand entry and egress is a rare event[39] that may result in either inadequate sampling by polar and charged probes or the appearance of spurious functional group densities due to probes that are "stuck" within the cavity.

We expect aLMMD to be a valuable tool for the structure-based drug discovery community as it can be applied to the comprehensive identification of ligand binding sites in proteins of pharmaceutical interest. New binding sites may be identified in established targets and binding site maps can be created for novel targets. The method may also find applications in ensemble docking and rational drug design. The ability of aMD to enhance conformational sampling has been harnessed to generate diverse structural ensembles for molecular docking.[90] In aLMMD, the incorporation of benzene probes into aMD simulations arguably generates a greater proportion of relevant ligand binding conformations that could be used for ensemble docking. The value of aLMMD to rational drug design also cannot be understated. As we have shown, aLMMD allows more extensive mapping of the conformational ensemble of binding sites compared to standard LMMD. After a binding site has been identified, the cutoff isovalue for visualization of the benzene occupancy maps can be lowered to better show the full extent of pocket space that can be filled by ligands. Hence, the shape and size of the benzene densities in the benzene occupancy maps may be used to guide SBDD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c01177.

Benzene occupancy maps of the T4 lysozyme L99A mutant generated by different aLMMD protocols, R_g plots, benzene occupancy maps generated from different analysis windows, and MDpocket analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yaw Sing Tan − *Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore;* orcid.org/0000-0002-2522-9421; Email: tanys@bii.a-star.edu.sg

Author

Justin Tze-Yang Ng − *Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore;* orcid.org/0000-0003-3862-0750

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c01177

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by A*STAR's Career Development Award (202D800022). We thank the National Supercomputing Centre of Singapore for computing facilities.

■ REFERENCES

(1) Anderson, A. C. The process of structure-based drug design. *Chem. Biol.* **2003**, *10*, 787−797.

(2) Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Comput. Struct. Biotechnol. J.* **2013**, *5*, No. e201302011.

(3) Nussinov, R.; Tsai, C.-J. Allostery in disease and in drug discovery. *Cell* **2013**, *153*, 293−305.

(4) Lu, S.; Li, S.; Zhang, J. Harnessing allostery: a novel approach to drug discovery. *Med. Res. Rev.* **2014**, *34*, 1242−1285.

(5) Feixas, F.; Lindert, S.; Sinko, W.; McCammon, J. A. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys. Chem.* **2014**, *186*, 31−45.

(6) Spyrakis, F.; Cavasotto, C. N. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Arch. Biochem. Biophys.* **2015**, *583*, 105−119.

(7) 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.

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

(9) Śledź, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93−102.

(10) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632−5633.

(11) Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693−705.

(12) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879−1881.

(13) Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglia, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.; Rowley, M. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* **2008**, *51*, 5843−5855.

(14) 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.
(15) Spiwok, V.; Sucur, Z.; Hosek, P. Enhanced sampling techniques in biomolecular simulations. *Biotechnol. Adv.* **2015**, *33*, 1130−1140.
(16) Bowman, G. R.; Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 11681−11686.
(17) Zimmerman, M. I.; Bowman, G. R. FAST conformational searches by balancing exploration/exploitation trade-offs. *J. Chem. Theory Comput.* **2015**, *11*, 5747−5757.
(18) Seco, J.; Luque, F. J.; Barril, X. Binding site detection and druggability index from first principles. *J. Med. Chem.* **2009**, *52*, 2363−2371.
(19) Guvench, O.; MacKerell, A. D., Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput. Biol.* **2009**, *5*, No. e1000435.
(20) Lexa, K. W.; Carlson, H. A. Full protein flexibility is essential for proper hot-spot mapping. *J. Am. Chem. Soc.* **2011**, *133*, 200−202.
(21) 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.
(22) Tan, Y. S.; Śledź, 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. Engl.* **2012**, *51*, 10078−10081.
(23) Ghanakota, P.; Carlson, H. A. Driving structure-based drug discovery through cosolvent molecular dynamics. *J. Med. Chem.* **2016**, *59*, 10383−10399.
(24) 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.
(25) 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.
(26) 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.
(27) Pradhan, M. R.; Siau, J. W.; Kannan, S.; Nguyen, M. N.; Ouaray, Z.; Kwoh, C. K.; Lane, D. P.; Ghadessy, F.; Verma, C. S. Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket. *Nucleic Acids Res.* **2019**, *47*, 1637−1652.
(28) Kaan, H. Y. K.; Sim, A. Y. L.; 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*, No. e0178381.
(29) 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.
(30) 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.
(31) Tan, Y. S.; Verma, C. S. Straightforward incorporation of multiple ligand types into molecular dynamics simulations for efficient binding site detection and characterization. *J. Chem. Theory Comput.* **2020**, *16*, 6633−6644.
(32) Kuzmanic, A.; Bowman, G. R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F. L. Investigating cryptic binding sites by molecular dynamics simulations. *Acc. Chem. Res.* **2020**, *53*, 654−661.
(33) Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Woldeyes, R. A.; Schneidman-Duhovny, D.; Demerdash, O. N.; Mitchell, J. C.; Wells, J. A.; Fraser, J. S.; Sali, A. CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. *J. Mol. Biol.* **2016**, *428*, 709−719.
(34) 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.
(35) 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.
(36) 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.
(37) Liang, J.; Woodward, C.; Edelsbrunner, H. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*, 1884−1897.
(38) Huang, P.; Chandra, V.; Rastinejad, F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu. Rev. Physiol.* **2010**, *72*, 247−272.
(39) Mondal, J.; Ahalawat, N.; Pandit, S.; Kay, L. E.; Vallurupalli, P. Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLoS Comput. Biol.* **2018**, *14*, No. e1006180.
(40) Lakkaraju, S. K.; Raman, E. P.; Yu, W.; MacKerell, A. D., Jr. Sampling of organic solutes in aqueous and heterogeneous environments using oscillating excess chemical potentials in grand canonical-like Monte Carlo-molecular dynamics simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2281−2290.
(41) Lakkaraju, S. K.; Yu, W.; Raman, E. P.; Hershfeld, A. V.; Fang, L.; Deshpande, D. A.; MacKerell, A. D., Jr. Mapping functional group free energy patterns at protein occluded sites: nuclear receptors and G-protein coupled receptors. *J. Chem. Inf. Model.* **2015**, *55*, 700−708.
(42) Comitani, F.; Gervasio, F. L. Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH. *J. Chem. Theory Comput.* **2018**, *14*, 3321−3331.
(43) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(44) Kalenkiewicz, A.; Grant, B.; Yang, C.-Y. Enrichment of druggable conformations from apo protein structures using cosolvent-accelerated molecular dynamics. *Biology* **2015**, *4*, 344−366.
(45) Smith, R. D.; Carlson, H. A. Identification of cryptic binding sites using MixMD with standard and accelerated molecular dynamics. *J. Chem. Inf. Model.* **2021**, *61*, 1287−1299.
(46) Wurtz, J.-M.; Bourguet, W.; Renaud, J.-P.; Vivat, V.; Chambon, P.; Moras, D.; Gronemeyer, H. A canonical structure for the ligand-binding domain of nuclear receptors. *Nat. Struct. Biol.* **1996**, *3*, 87−94.
(47) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular signatures of G-protein-coupled receptors. *Nature* **2013**, *494*, 185−194.
(48) Li, Z.; Nair, S. K. Quorum sensing: how bacteria can coordinate activity and synchronize their response to external signals? *Protein Sci.* **2012**, *21*, 1403−1417.
(49) Martinez-Rosell, G.; Lovera, S.; Sands, Z. A.; De Fabritiis, G. PlayMolecule CrypticScout: predicting protein cryptic sites using mixed-solvent molecular simulations. *J. Chem. Inf. Model.* **2020**, *60*, 2314−2324.
(50) Ghanakota, P.; Carlson, H. A. Moving beyond active-site detection: MixMD applied to allosteric systems. *J. Phys. Chem. B* **2016**, *120*, 8685−8695.
(51) Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell* **1997**, *89*, 239−250.

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

(52) Manion, M. K.; O'Neill, J. W.; Giedt, C. D.; Kim, K. M.; Zhang, K. Y. Z.; Hockenbery, D. M. Bcl-XL mutations suppress cellular sensitivity to antimycin A. *J. Biol. Chem.* **2004**, *279*, 2159−2165.
(53) 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 p38alpha MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. *Nat. Struct. Biol.* **2003**, *10*, 764−769.
(54) Pedersen, A. K.; Peters, G. H.; Møller, K. B.; Iversen, L. F.; Kastrup, J. S. Water-molecule network and active-site flexibility of apo protein tyrosine phosphatase 1B. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1527−1534.
(55) Kato, M.; Li, J.; Chuang, J. L.; Chuang, D. T. Distinct structural mechanisms for inhibition of pyruvate dehydrogenase kinase isoforms by AZD7545, dichloroacetate, and radicicol. *Structure* **2007**, *15*, 992−1004.
(56) Merski, M.; Fischer, M.; Balius, T. E.; Eidam, O.; Shoichet, B. K. Homologous ligands accommodated by discrete conformations of a buried cavity. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 5039−5044.
(57) 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.
(58) Malapaka, R. R. V.; Khoo, S.; Zhang, J.; Choi, J. H.; Zhou, X. E.; Xu, Y.; Gong, Y.; Li, J.; Yong, E.-L.; Chalmers, M. J.; Chang, L.; Resau, J. H.; Griffin, P. R.; Chen, Y. E.; Xu, H. E. Identification and mechanism of 10-carbon fatty acid as modulating ligand of peroxisome proliferator-activated receptors. *J. Biol. Chem.* **2012**, *287*, 183−195.
(59) Webb, B.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinf.* **2016**, *54*, 5.6.1.
(60) 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.
(61) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; SalomonFerrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2018*; University of California: San Francisco, 2018.
(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(63) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(64) 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.
(65) Wang, J.; 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.
(66) 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.
(67) 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.
(68) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090−2098.
(69) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(70) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J. Chem. Phys.* **2007**, *127*, 155102.
(71) Roe, D. R.; Cheatham, T. E., 3rd PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(72) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.
(73) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, *27*, 3276−3285.
(74) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455.
(75) Caves, L. S. D.; Evanseck, J. D.; Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **1998**, *7*, 649−666.
(76) 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*, 1880−1891.
(77) Wright, L.; Barril, X.; Dymock, B.; Sheridan, L.; Surgenor, A.; Beswick, M.; Drysdale, M.; Collier, A.; Massey, A.; Davies, N.; Fink, A.; Fromont, C.; Aherne, W.; Boxall, K.; Sharp, S.; Workman, P.; Hubbard, R. E. Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. *Chem. Biol.* **2004**, *11*, 775−785.
(78) Lee, E. F.; Czabotar, P. E.; Smith, B. J.; Deshayes, K.; Zobel, K.; Colman, P. M.; Fairlie, W. D. Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ.* **2007**, *14*, 1711−1713.
(79) Taylor, S. S.; Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **2011**, *36*, 65−77.
(80) Gill, A. L.; Frederickson, M.; Cleasby, A.; Woodhead, S. J.; Carr, M. G.; Woodhead, A. J.; Walker, M. T.; Congreve, M. S.; Devine, L. A.; Tisi, D.; O'Reilly, M.; Seavers, L. C. A.; Davis, D. J.; Curry, J.; Anthony, R.; Padova, A.; Murray, C. W.; Carr, R. A. E.; Jhoti, H. Identification of novel p38alpha MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* **2005**, *48*, 414−426.
(81) Ustach, V. D.; Lakkaraju, S. K.; Jo, S.; Yu, W.; Jiang, W.; MacKerell, A. D., Jr. Optimization and evaluation of Site-Identification by Ligand Competitive Saturation (SILCS) as a tool for target-based ligand optimization. *J. Chem. Inf. Model.* **2019**, *59*, 3018−3035.
(82) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; McDowell, R. S.; Hansen, S. K. Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat. Struct. Mol. Biol.* **2004**, *11*, 730−737.
(83) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **1992**, *355*, 371−373.
(84) Nettles, K. W.; Bruning, J. B.; Gil, G.; O'Neill, E. E.; Nowak, J.; Hughs, A.; Kim, Y.; DeSombre, E. R.; Dilis, R.; Hanson, R. N.; Joachimiak, A.; Greene, G. L. Structural plasticity in the oestrogen receptor ligand-binding domain. *EMBO Rep.* **2007**, *8*, 563−568.
(85) Zhou, H.-B.; Nettles, K. W.; Bruning, J. B.; Kim, Y.; Joachimiak, A.; Sharma, S.; Carlson, K. E.; Stossi, F.; Katzenellenbogen, B. S.; Greene, G. L.; Katzenellenbogen, J. A. Elemental isomerism: a boron-nitrogen surrogate for a carbon-carbon double bond increases the chemical diversity of estrogen receptor ligands. *Chem. Biol.* **2007**, *14*, 659−669.
(86) Xu, H. E.; Lambert, M. H.; Montana, V. G.; Plunket, K. D.; Moore, L. B.; Collins, J. L.; Oplinger, J. A.; Kliewer, S. A.; Gampe, R. T., Jr.; McKee, D. D.; Moore, J. T.; Willson, T. M. Structural determinants of ligand binding selectivity between the peroxisome

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981

proliferator-activated receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13919−13924.
(87) Ung, P. M.-U.; Rahman, R.; Schlessinger, A. Redefining the protein kinase conformational space with machine learning. *Cell Chem. Biol.* **2018**, *25*, 916−924.
(88) Modi, V.; Dunbrack, R. L. Defining a new nomenclature for the structures of active and inactive kinases. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 6818−6827.
(89) Frembgen-Kesner, T.; Elcock, A. H. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. *J. Mol. Biol.* **2006**, *359*, 202−214.
(90) Miao, Y.; Goldfeld, D. A.; Moo, E. V.; Sexton, P. M.; Christopoulos, A.; McCammon, J. A.; Valant, C. Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E5675−E5684.

https://doi.org/10.1021/acs.jctc.1c01177
J. Chem. Theory Comput. 2022, 18, 1969−1981