JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

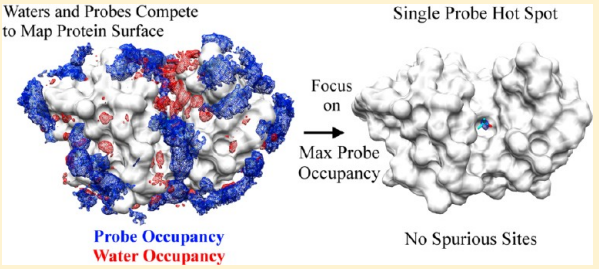
Improving Protocols for Protein Mapping through Proper Comparison to Crystallography Data

Katrina W. Lexa and Heather A. Carlson*

Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church Street, Ann Arbor, Michigan 48109-1065, United States

Supporting Information

ABSTRACT: Computational approaches to fragment-based drug design (FBDD) can complement experiments and facilitate the identification of potential hot spots along the protein surface. However, the evaluation of computational methods for mapping binding sites frequently focuses upon the ability to reproduce crystallographic coordinates to within a low RMSD threshold. This dependency on the deposited coordinate data overlooks the original electron density from the experiment, thus techniques may be developed based upon subjective—or even erroneous—atomic coordinates. This can become a significant drawback in applications to systems where the location of hot spots is unknown. On the basis of comparison to crystallographic density, we previously showed that mixed-solvent molecular dynamics (MixMD) accurately identifies the active site for HEWL, with acetonitrile as an organic solvent. Here, we concentrated on the influence of protic solvent on simulation and refined the optimal MixMD approach for extrapolation of the method to systems without established sites. Our results establish an accurate approach for comparing simulations to experiment. We have outlined the most efficient strategy for MixMD, based on simulation length and number of runs. The development outlined here makes MixMD a robust method which should prove useful across a broad range of target structures. Lastly, our results with MixMD match experimental data so well that consistency between simulations and density may be a useful way to aid the identification of probes vs waters during the refinement of future multiple solvent crystallographic structures.



INTRODUCTION

The development of fragment-based drug design (FBDD) as a complement to traditional structure-based drug design (SBDD) has allowed discovery efforts to probe binding across a broad range of protein structures, including many traditionally difficult cases. FBDD was first introduced as a viable tool for drug discovery in 1996 with the structure−activity relationship (SAR) by nuclear magnetic resonance (NMR) method and multiple solvent crystal structure (MSCS) method.[1−3] In the first SAR-by-NMR study of protein−ligand binding, small organic molecules with low molecular weight were screened for binding affinity to FKBP, and their bound positions were determined from ^{15}N heteronuclear single quantum correlation (HSQC) spectra.[3] The authors identified low-affinity binding sites along the protein surface, optimized the resultant hits, and then linked the fragments together, to create ligands with nanomolar affinity for FKBP. In the original MSCS study, cross-linked crystals of elastase were soaked in a solution of acetonitrile (ACN), allowing ACN to displace water molecules and form favorable interactions along the protein surface. Then, the crystals were washed to remove unbound ACN, enabling the crystallographers to identify optimal binding sites for amphiphilic nitrogen-containing fragments. The authors concluded that the MSCS results from several organic probe types could be superimposed and translated into a template, or pharmacophore, for drug design. Both experimental methods aim to identify hotspots, the sites where a small number of protein residues contribute significantly to the free energy of binding.[4,5] These hot spots signify regions where important protein−ligand interactions may be formed and thus represent a tool for the design of novel therapeutics. FBDD has been effectively implemented in the development of clinical candidates, including ABT-263, AT7519, and AT9283.[6,7]

Computational techniques can be used to complement the existing experimental methods for FBDD by offering a less costly approach to the examination of potential binding interactions in the target system. Several groups have recently developed computational techniques that incorporate concepts from MSCS into molecular dynamics (MD). The first of these computational approaches to probe mapping used a single 16-ns simulation of protein solvated by a box of 20% volume/volume (v/v) isopropanol (IPA) and water to identify hot spots based on grid occupancies and calculated binding free energies.[8] The second MD-based approach, SILCS, relied upon a ternary solvent box of 1 M benzene, 1 M propane, and water to calculate probe occupancies and map the potential energy surface.[9,10] The third MD-based approach utilized MD

Received: September 11, 2012

   dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

simulations with IPA or multiple probes (acetamide, acetic acid, isopropylamine, and IPA) at 20% v/v probe:water to determine grid occupancies and binding free energies for several difficult cases. The authors showed that calculation of the binding free energy required careful development in order to avoid overestimation.[11] However, each of these methods has been limited by the identification of irrelevant local minima with equal weight as the true binding site(s). With this problem, extrapolation to cases where the answer is unknown becomes very difficult. An improved sampling protocol that reduces spurious minima is essential for robust application. Spurious local minima are also a common problem among the traditional computational probe-mapping techniques like GRID[12] and MCSS.[13]

Though it does not use MD to sample full dynamics, the FTMAP algorithm[14] deserves mention because it was also based on experimental methods for fragment mapping. It applies a fast Fourier Transform approach to the docking of billions of probe conformations to a rigid protein. Like MSCS, it places highest significance on hot spots that are mapped by many different probe molecules. It does not identify as many spurious minima as other probe-mapping methods.[14] This technique has recently been expanded to include a minimization step for selected side chains after performing FTMAP on protein−protein interfaces.[15] Two nondynamic approaches recently demonstrated success in mapping the hot spot of HEWL using multiple rigid structures with several probe types; however, accounting for flexibility is critical to the proper exploration of most druggable targets, including those under study here.[19,20] The results from Bahar et al. demonstrated the necessity of employing MD to adequately describe the conformational states of bound proteins like LFA-1, Eg5, and p38.[11]

In our development of mixed-solvent MD (MixMD), we focused on providing a computational tool that would preferentially locate the most relevant hot spots along a protein surface.[16] MSCS studies can be difficult to perform with fragile crystals or at a high concentration of organic solvent, and the results can be influenced by the crystallization conditions. Computational studies avoid these limitations and enable the detailed study of protein−probe interactions for a wider range of proteins. Our original MixMD study was the first computational technique to definitively show the need to include full protein flexibility to significantly reduce extraneous minima and correctly map a protein surface. Our initial studies of HEWL in ACN and water demonstrated the utility of MixMD for hot-spot mapping, and we were interested in incorporating additional functional groups. Here, we have performed MixMD for a range of protein cases with the most common MSCS probes: IPA and ACN. We particularly focused on the impact of length and number of simulations on solvent behavior and protein structure. Our motivation was establishing the appropriate protocol that can be accurately employed in new systems where binding sites are unknown.

Hot-spot data were available for IPA and ACN with the following proteins: elastase, HEWL, p53 core, RNase A, subtilisin, and thermolysin. These probes were originally selected for MSCS based on their miscibility with water, their interaction type, and the ease of distinguishing their crystallographic density from that of water.[1,17] Due to the subjective nature of assigning experimental density to specific atomic coordinates, the most valid comparison between simulation and MSCS is between occupancy grids and experimental density data. Therefore, we focused primarily on cases with electron density data available: elastase+IPA, HEWL+ACN, HEWL +IPA, p53 core+IPA, RNase A+IPA, and thermolysin+IPA. Structure factor files were not available for elastase+ACN (coordinates obtained from authors or ref 1), subtilisin+ACN (1SCB[18]), or thermolysin+ACN (1FJU[19]), but our simulations of these systems showed that our results agreed with the known binding sites of these proteins (see Supporting Information (SI) Figure S1−S3). These proteins vary in size and active-site composition, thus providing a variety of interaction types to explore through MixMD to develop the most robust protocol with the greatest potential for application to new proteins. Through this analysis, we have formulated an optimized approach for comparing the simulation trajectory from MD with crystallographic data.

METHODS

HEWL+ACN was the focus of our previous study,[16] and here, we added IPA as a probe. We concentrated on the systems with experimental density data available for this study. The starting structures for elastase+IPA (2FOF[20]), HEWL+IPA (1LY0[21]), p53 core+IPA (2IOM[22]), RNase A+IPA (3EV2[23]), and thermolysin+IPA (7TLI[17]) were obtained from the PDB.[24] Crystallographic waters and structural ions were retained to maintain their stabilization of the protein conformation, but all probe molecules and crystallographic ions were removed. We emphasize that no simulations were initiated with probes occupying the locations seen experimentally. Molprobity[25] was used to check the side-chain orientations of ASN, GLN, and HIS, and these results were confirmed by a visual examination.

The crystallographic structure of RNase A contained two copies of the protein within the asymmetric unit. Chain B was selected for use in our simulations because chain A was missing several residues. Furthermore, chain B contained two IPA molecules located in the active site, while chain A contained two IPA molecules that were located along a packing interface. Parameters for IPA were based on the OPLS-AA parameters for pure alcohols from the work of Jorgensen et al.[26] Parameters for ACN were implemented as described in our established MixMD protocol.[16]

The atomic coordinates from the MSCS studies were used for each system. Using the AMBER10/AMBERTOOLS 1.2 package[27] and FF99SB parameter set,[28] hydrogens were added to the protein using *tLeAP* and minimized in *sander*. The protein was solvated in an 18-Å pre-equilibrated box of 50% w/w probe and TIP3P water,[29] and then, Na^+ or Cl^- ions were added to neutralize the system charge. Simulations used a 2 fs time step and SHAKE[30] to restrain bonds to hydrogen. Nonbonded, van der Waals interactions were subjected to a 10-Å cutoff, and the Particle Mesh Ewald[31] approximation was applied. Each protein−probe system was simulated with AMBER10 for several, independent MD, each initiated with a different random number seed. Temperature was regulated through an Anderson thermostat.[32]

Each system underwent 250 cycles of steepest-descent minimization followed by 4750 cycles of conjugate-gradient minimization with the protein fixed. Then, each system was gradually heated under constant volume from 10 to 300 K over 80 ps while the protein was gently restrained by a harmonic force constant of 10 kcal/mol·Å^2. These restraints were gradually removed over 500 ps of equilibration under constant pressure until the protein was fully flexible. Five independent, 50-ns simulations in the NPT ensemble were performed for

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

each MixMD system; for elastase and thermolysin, a total of 10 independent, 50-ns simulations were generated.

Our binary-solvent simulations commenced from the crystal coordinates of the MSCS, so we could use a common frame of reference to accurately compare each independent simulation with the crystallographic density data. CCP4i[33] was used to derive crystallographic density for each MSCS probe and to rotate maps of the simulation density into the crystal structure orientation. In CCP4i, the pdb and structure factor files were used to derive the mtz file, which contains the reflection data, including HKL index, intensity, and amplitude. Each small molecule probe was individually removed from the structure file, and then, the map was refined through one complete round of Buster[34] (five cycles) to produce a new refined map.

The results of our MD simulations were analyzed using *ptraj*, a module within the AMBERTOOLS package. The final 5 ns of each individual run were combined into one set of data. The simulation data was first imaged and fit to the crystal conformation by Cα-RMS using CCP4i; then, the solvent occupancies of water and probe were calculated using the grid command with a 0.5 Å × 0.5 Å × 0.5 Å spacing over the entire box. The probe "density" for each point was determined by summing the occupancy over the last 5 ns of all five runs combined. Simulation RMSD and RMSF demonstrate that solvation by 50% w/w probe does not result in unrealistic sampling over the simulation trajectory (SI Figures S4−S5). In our comparison with experimental data, occupancies were examined for each solvent heavy atom, each solvent center, both solvent centers together, and all atoms. The occupancy grid for all atoms is the equivalent of the electron density in the crystal structure, and like density data, it removes the identity of which atoms are populating the grid points.

The primary focus of our analysis is upon the grid points with the highest occupancies, literally the positions where probe molecules spend the overwhelming majority of their time during the simulation. The shape and volume of the set of high-occupancy grid points must also be appropriate for the probe chosen, ACN or IPA. Systems with a clear single site have only the maximum-occupancy region shown in the figures; these are the most occupied grid points of the entire map (1YL0 and 2IOM). Here, we define a probe site or binding region as a single, contiguous cluster of high-occupancy points. The occupancies are noted in the figure captions. The top-8 most-occupied sites were explored as hot spots for all other protein systems because they are known to have several binding sites. Furthermore, a detailed examination of the "extra" sites was essential to understand how to optimize our routine to down-weight spurious sites. Elimination of false positives is crucial for extrapolation to new protein targets where the answers are unknown. Furthermore, all mapping techniques are based on a limited number of probe types, and an intrinsic goal is that they can still map interactions appropriate for other functional groups, albeit more weakly than they map their own interactions. In fact, our analysis of weaker sites mapped on elastase and RNase A are compared to alternate MSCS (i.e., structures produced using probe solvents other than ACN and IPA). The top-8 were used for all probe types and all systems under study. *The eight most-occupied hot spots delineate regions that have a significant probe population of at least five standard deviations above the mean density.*

RESULTS AND DISCUSSION

Understanding Electron Density and the Influence of Crystal Packing. The most accurate comparison between experiment and computation involves a joint examination of the MSCS electron density and the occupancy grids from the MixMD simulation. Although the MSCS probes were specifically chosen to aid in identification of crystallographic density, placement of probes is necessarily subjective. Electron density cannot be unambiguously assigned in many cases, *so occupancy grids for solvent should not be solely compared to atomic coordinates*. It is often unclear whether density represents the position of a probe or a water molecule, and the placement of one of these versus the other can affect the final density map.

Electron density maps are typically presented as $2F_O-F_C$ and F_O-F_C maps, where F_O refers to the observed phasing information and F_C refers to the calculated phasing information. The $2F_O-F_C$ map illustrates well occupied, stable structural features when viewed with a 1.5σ contour. The F_O-F_C map is a difference map highlighting the dissimilarity between the observed and expected features. When examined at 3.0σ, noise in the F_O-F_C map is minimized, and it shows positive peaks that indicate structure features which may be missing in the model as well as negative peaks that represent features in the model which may be in error. The atomic coordinates for all probes must be compared to the refined $2F_O-F_C$ and F_O-F_C maps to ensure that each is supported by appropriate electron density based on the 1.5σ/3.0σ criteria listed above (Table 1).

Table 1. Electron Density from the Refined Cycles of Buster Used to Evaluate Each Organic Molecule in the Noted MSCS Data[a]

Protein-Probe System	Unsupported Density	Supported Density	Supported at an Interface
HEWL (1YL0)		2836 (0.934)	2837 (0.741)
p53 core (2IOM)		3001 (0.889)	
RNase A (3EV2)	905A (0.934)*, 917A (0.881), 903B (0.485)*	902B (0.723)	
Elastase (2FOF)		1001 (0.823), 1003 (0.871)	1002 (0.898), 1002x
Thermolysin (7TLI)	2004 (0.838), 2005 (0.742), 2007 (0.703), 2008 (0.693)*	2001 (0.837), 2002 (0.872)	2003 (0.494), 2006 (0.921)

*Density does not support a probe site, but is instead a potential water site. [a]The real-space correlation coefficient from the refinement is shown in parentheses following the residue number for each organic solvent molecule in the MSCS. A value of 1.0 means that the experimental map exactly matches the map calculated from the model (perfect correlation), while a value of −1.0 means that the experimental map is perfectly anti-correlated to a map calculated directly from the model. A value of 0.0 means that there is no correlation between the experimental map and model map.

There are several caveats inherent to this comparison. Crystallographic density may include artificial features influenced by experimental conditions and crystal contacts. As Allen et al. acknowledged in their MSCS study of elastase +ACN,[1] several of the ACN probes were involved in crystal-packing interactions and were not believed to indicate true hot spots for binding. Therefore, probes near crystal-packing interfaces required additional scrutiny. Depending upon the

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

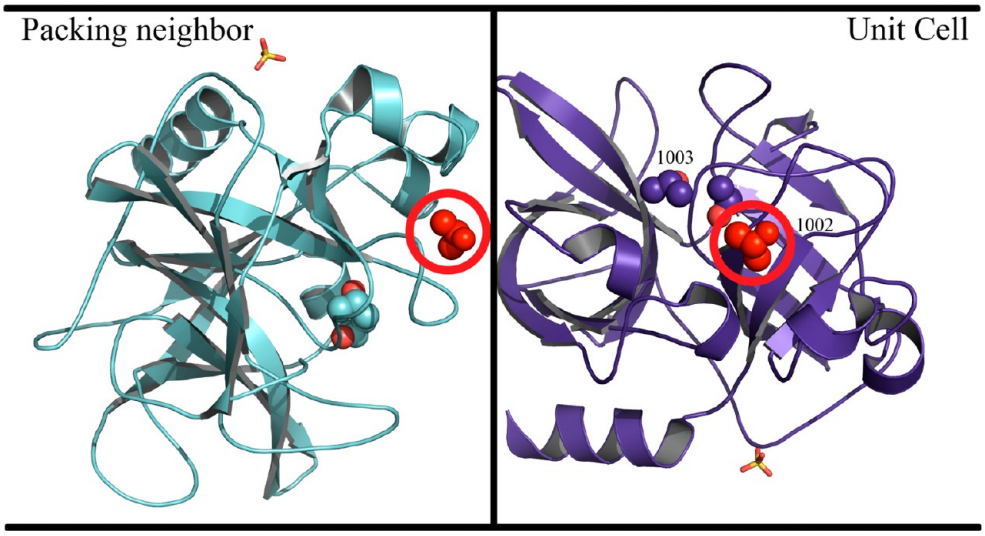


Figure 1. MSCS of elastase+IPA is shown in violet (right) with a neighboring symmetry partner in cyan (left), according to their crystallographic orientations. The two symmetry-related, equivalent IPA positions are circled in red. Both IPA are within 4 Å of the protein in the initial unit cell. Probes at the contact interfaces may be irreproducible, but when they are observed, the probe may map to either or both of the circled sites on the protein.

protein system, it may not be possible to reproduce the probe location at an interface in MixMD because it is not favorable in the true solution phase.

Identifying Probes with Appropriate Density. It must be stressed that crystallographic structures can be subject to phase bias and experimental uncertainty, and accurate comparisons between theory and experiment can only be made when the limitations of both are understood. To properly confirm that a probe site is supported with appropriate density, a refined map must be calculated where the F_O-F_C map is recalculated with that particular probe excluded from the set of atomic coordinates. This removes the bias of the probe itself on the determination of the density. This procedure was important for confirming hot spots in the crystal structures, and it was just as important for our analysis of any "spurious" sites from MixMD observed in the occupancy grids. In some cases, "spurious" sites from MixMD actually reflected the position of a valid probe contact from the neighboring unit cell of the crystal (Figure 1). Thus, it was important to examine the surrounding symmetry partners present in the crystalline environment for probe sites in contact with the central protein. When refined density maps were generated, several outcomes were possible for a probe molecule. A probe could be well-placed in positive F_O-F_C density (within the crystal or along a crystalline interface), the F_O-F_C density at the probe site could support placement of a water molecule, the probe could be located within negative F_O-F_C density, or there could be insufficient density to justify the placement of any molecule. Figure 2 illustrates probes that are well justified by density and others that are more likely to be water molecules based on the F_O-F_C density in the refined map (SI Figures 6–8 detail the refined density for all other probes).

To verify the probe locations in MSCS, each probe was removed from the structure individually along with all of the low density waters, as is appropriate for the creation of a refined map. The density in these figures has been generated through a refinement cycle, and therefore, it is not equivalent to the ready-made mtz file that is electronically available through the Uppsala Electron Density Server.[35] Chimera,[36] COOT,[37] and

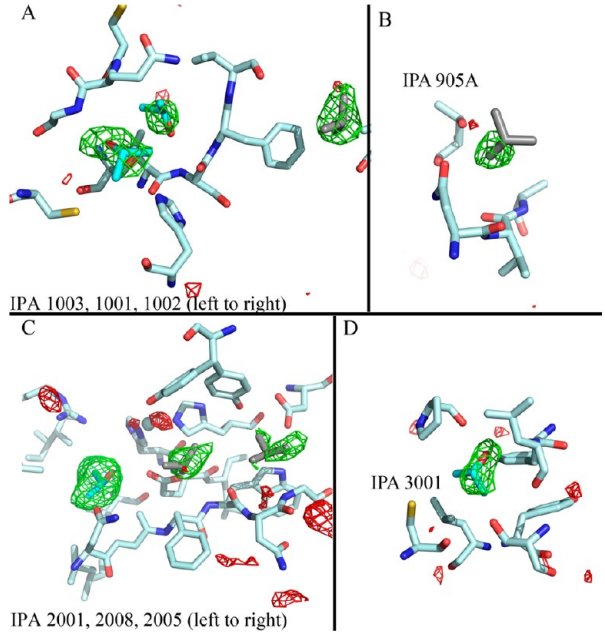


Figure 2. Density in the Buster refinement maps for several of the MSCS structures. The refined maps were based on the coordinate file for the MSCS structure, with probes and close-contact waters/ions removed. The positive F_O-F_C density is contoured in green at 3.0σ, the negative F_O-F_C density in red. This illustrates regions where crystallographic probes are justified or not, respectively. In A, the positive F_O-F_C density from the MSCS of elastase+IPA (2FOF) clearly agrees with the probe placement, where IPA.1001 and IPA.1003 are located in the active site with good agreement between coordinates and density (IPA.1002 is located near the crystal interface). The positive F_O-F_C density in B indicates that the site for IPA.905A in RNase A+IPA (3EV2) may be a water molecule. In C, the F_O-F_C density along the active site of thermolysin+IPA (7TLI) indicates that the site for IPA.2001 is justified while IPA.2008 is improperly placed in the density and IPA.2005 might best support a water molecule as opposed to a probe. In D, the F_O-F_C density in p53 core (2IOM) illustrates that this probe is justified in the crystal structure.

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

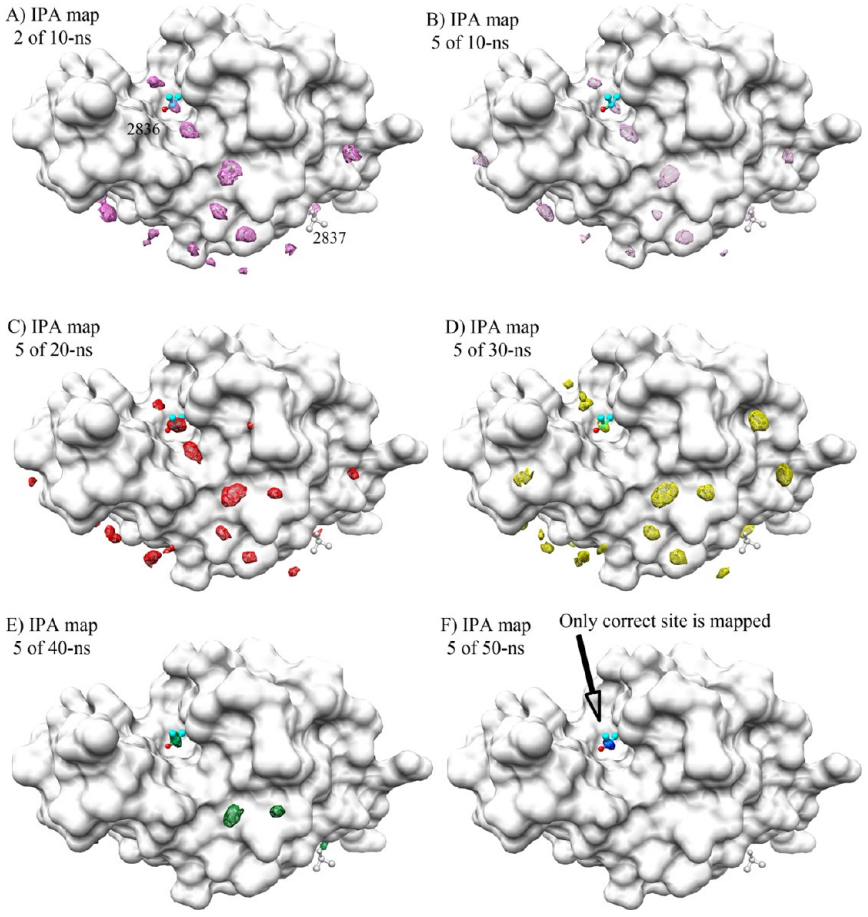


Figure 3. MixMD data combined for the last 5 ns of 5 individual 50-ns runs of fully flexible HEWL in 50% w/w IPA and water. All snapshots were superimposed on the starting crystal structure (PDB ID 1YL0, white surface). The MSCS coordinate positions for IPA.2836 are shown in cyan and IPA.2837 is shown in gray as ball-and-stick. The occupancy level in the frames is 0.164 ± 0.005 probe/grid point. (A) Simulation IPA density does map the IPA.2836 site using a 2-ns window after 10 ns of simulation; however, many other spurious minima are mapped as well. (B−F) Using a 5-ns window, we find that the spurious, local minima are eliminated over time as the simulation converges to the maximally occupied hot spot.

PyMol[38] were used for visualization of results. Displaying multiple density occupancy layers in the figures is confusing to the eye, so the figures are simplified by showing the MSCS probes whose positions were confirmed by the F_O-F_C density from Buster in cyan and probes at the interface or without adequate density in gray as ball-and-stick coordinates. However, we stress that our analysis was driven by actual electron density, and other methods should be evaluated with a similar approach.

Establishing the Appropriate Simulation Length: Adequate Sampling and Convergence. We were interested in establishing the appropriate simulation length for MixMD to allow widespread application of hot-spot mapping, including cases where the active site might be unknown. In order to determine the appropriate methodology to be used with a protic probe, which should diffuse through simulation more slowly than an aprotic probe, we first had to define appropriate system behavior. All of the mixed-solvent simulations were stable over the nanosecond time scales examined here. Longer time scales should reveal unfolding of the protein and are less desirable.[39] As shown previously, a probe density of 50% w/w achieved the best balance of efficiently mapping the protein surface with ACN.[16] To assess whether adequate mixing of the two solvents had occurred, we examined the number ratio of probe to water (N_p/N_w) at the edges of the box. In a properly mixed system, N_p/N_w at the edges of the box will correspond with N_p/N_w of the whole system, since the presence of the protein should not impact the distribution of solvent at the box edges. Within 5 Å of the box edges, we found ratios of 0.95−1.13, where 1.0 is perfect; therefore, our binary solvents mixed appropriately in the bulk with insignificant variation around 1.0 due to the time scale.

HEWL was chosen as our initial validation protein because it is a simple canonical system with a wealth of available structural and binding data. Since IPA is a protic solvent with a slower diffusion rate than ACN, we were interested in comparing our results from HEWL+ACN to the HEWL+IPA system. The MSCS of HEWL contained only a single ACN bound at the active site, but two bound probes were identified in the MSCS of HEWL+IPA: IPA.2836 was bound in the active site, and IPA.2837 was located on the opposite surface at the crystal interface.

Although our previous MixMD study used the final 2 ns from a set of 10-ns simulations to identify hot spots, we expected that a wider window and longer simulation time would be needed for obtaining converged data with a protic solvent. Indeed, we found that the simulation time used for HEWL+ACN was too short to account for complete sampling of IPA. To calculate optimal window size, we examined the data combined across the final 2, 5, 10, 20, 30, and 40-ns segments of each 50-ns simulation. For example, the last 5 ns from five independent 50-

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

ns trajectories of HEWL+IPA were combined by the grid function in *ptraj* to yield solvent occupancies for 25 ns of total simulation time. We found that the final 5 ns, combined across all the independent trajectories, best represented converged probe positions in the active site, while larger sampling windows displayed additional local minima for each run (SI Figure S9). Therefore, all of the data presented in this study has combined the final 5-ns timeframes for the calculation of simulation occupancy grids.

Length of the simulation was also examined. We analyzed the solvent occupancy from the first 10 ns (2-ns window), 20 ns (5-ns window), 30 ns (5-ns window), 40 ns (5-ns window), and 50 ns (5-ns window). We found that by 50 ns the spurious minima had been sufficiently reduced such that only the relevant hot spots were located (Figure 3). At the point that the spurious minima were lost, IPA.2837 was no longer mapped. This was entirely expected. That IPA had only a small spherical density which appeared more appropriate for a water, and it was located at a crystal packing interface. In fact, the occupancy maps for water identified IPA.2837 as well as most of the low B-factors crystallographic waters (SI Figure S10). Thus, in cases where the refined density map indicates that the organic probe might actually be a water molecule, simulation density can aid in the proper assignment. Mapping water density and examining lifetimes of water at subsites have been established techniques for some time; our ability to map low B-factor waters demonstrates that MixMD retains this functionality. This was also shown in our previous paper on HEWL+ACN. Therefore, we focus our discussion of water occupancy to only those probe sites where the experimental density appears to support a water molecule.

To determine the general applicability of using five 50-ns simulations as the optimal sampling with a protic solvent, we also applied MixMD to the MSCS of p53 core (PDB ID 2IOM). Although the crystal structure was missing residues 92−96 and 285−294, these residues are part of the disordered domain linkers and were not expected to impact mapping results. Experimentally, a single IPA molecule was bound to p53 core at the active site, and we validated this with the refined map from Buster. As we found for HEWL+IPA, examination of 5-ns windows over the course of the trajectory showed that 50 ns of simulation time was more than sufficient for mapping hot spots with IPA while minimizing spurious minima along the p53 core surface (Figure 4). The crystallographic site matched our maximally occupied location from the simulation IPA density, demonstrating the capacity of MixMD to capture important binding sites in DNA-binding proteins.

Through MixMD simulations of HEWL+IPA and p53 core +IPA, we determined an appropriate protocol for IPA to map the protein hot spots correctly without simultaneous identification of numerous irrelevant minima, a difficulty for other computational methods. We could show other minima by changing the contour value, but that does not change the fact that the correct location is clearly identified by the maximum occupancy site. These sites are at least five standard deviations above the mean solvent density, by the highest-occupancy grid points. Our studies based on these two proteins indicated an optimal simulation time of 5 independent simulations of 50 ns and an analysis window of 5 ns when running MixMD with a protic solvent.

Determining the Appropriate Number of Simulations. To study the consequence of the number of independent simulations, we compared the results of 5 runs versus 10 runs of

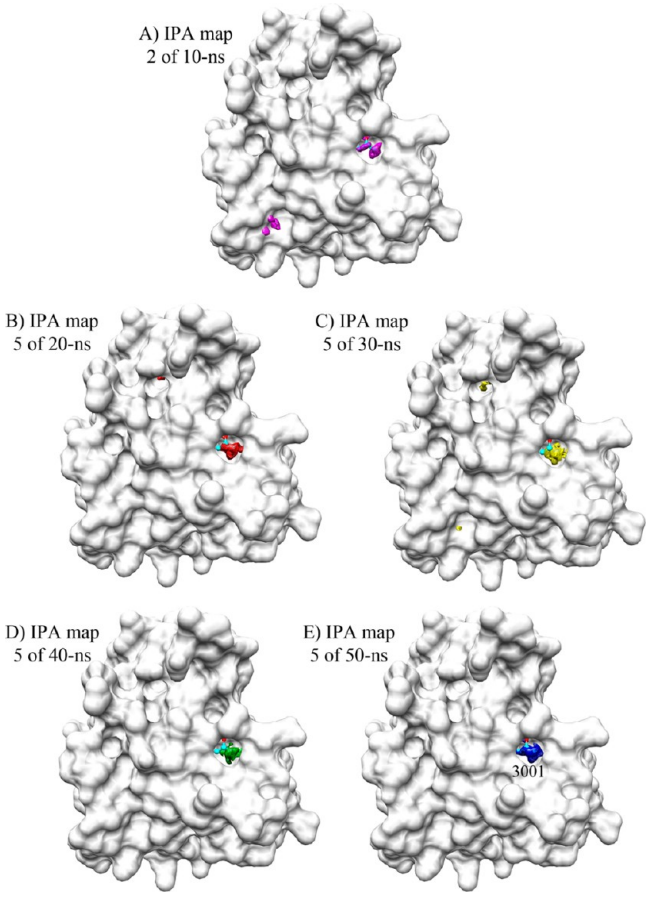


Figure 4. MixMD data combined for the last 5 ns for 5 individual 50-ns runs of fully flexible p53 core in 50% w/w IPA and water. All snapshots were superimposed on the original crystal structure (PDB ID 2IOM, white surface). The occupancy level in the frames is 0.434 ± 0.011 probe/grid point. The MSCS coordinate position for IPA.3001 is shown in cyan as ball-and-stick. The maximally occupied site for simulation IPA density clearly agrees with the crystallographic probe position. (A) Simulation IPA density does map the IPA.3001 site using a 2-ns window after 10 ns of simulation; however, a few other spurious minima are mapped as well. (B−E) Using a 5-ns window, we find that the local minima are eliminated over time as the simulation quickly converges to the maximally occupied hot spot.

50 ns each for the larger protein systems, elastase and thermolysin. The success of MD simulations is firmly based on the assumption of the ergodic hypothesis: given adequate sampling time, all relevant states will be reached. We prefer many shorter simulations run in parallel to increase sampling, instead of one long simulation, because they increase the potential space that may be explored. This is particularly true for computational studies of solvent mapping, where effective mapping must be balanced with an efficient use of computational cycles. Furthermore, very long simulations should eventually denature the protein in 50% w/w ACN or IPA. Many short simulations focus the maps on the more biologically relevant conformation of the protein. However, we needed to determine if it is best for probe mapping.

MSCS of elastase were solved in the presence of acetone, ACN, benzene with IPA, cyclohexane with IPA, dimethylformamide, ethanol, IPA, and trifluoroethanol. This allowed us to compare our MixMD to a large body of crystallographic data. The MSCS structure of elastase+IPA contained three probe molecules; two were bound in the active site (IPA.1001 and

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

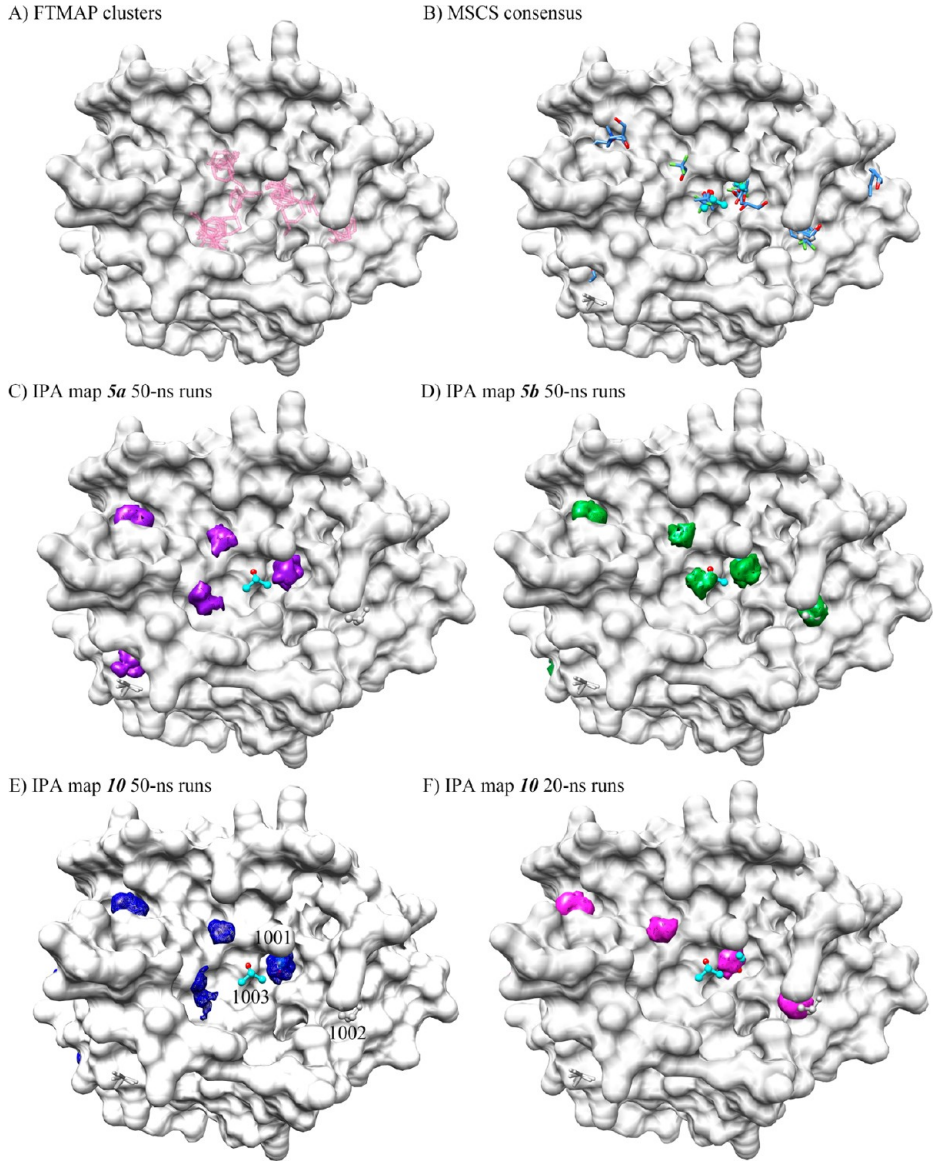


Figure 5. MixMD data combined for the last 5 ns of 10 individual 50-ns runs of fully flexible elastase in 50% w/w IPA and water. All snapshots were superimposed on the original crystal structure (PDB ID 2FOF, white surface). The MSCS coordinate positions for IPA.1001, IPA.1002, and IPA.1003 are shown in cyan as ball-and-stick. The crystal interface probe site IPA.1002x is shown in gray. Alternate probes from other MSCS by the same authors are shown in blue. Density contour levels were chosen such that the top eight hot spots were represented at a "hide dust" level of 3.5. (A) Results from performing FTMAP against PDB ID 2FOF, with functional group clusters shown in pink. (B) All probe sites from MSCS with elastase. (C) Simulation IPA density from the last 5 ns of 50 ns total from independent runs 1−5 (purple). (D) The last 5 ns of 50 ns total from independent runs 6−10 (green) do not perform as well as the results from all 10 runs. (E) The maximally occupied site for simulation IPA density from the last 5 ns of 50 ns total from 10 independent runs (blue) corresponds with all three of the crystallographic probe positions as well as two other MSCS sites. (F) Simulation IPA density from the last 5 ns of 20 ns total simulation time over 10 individual runs performs better than the IPA density after 50 ns over 5 runs. The occupancy level in frames C−F is 0.270 ± 0.056 probe/grid point.

IPA.1003), and the third was located near the crystal packing interface (IPA.1002). Inspection of the symmetry partners revealed that IPA.1002 could also be located on an opposite face of elastase (referred to as IPA.1002x). Results from density refinement in CCP4i showed that all four locations were supported by the F_O-F_C density (Table 1).

Our MixMD results for elastase+IPA were highly similar to the MSCS data of the corresponding crystal structure as well as the consensus MSCS data of many probe types across many crystal structures. Overlaying all of the MSCS structures and examining the electron densities of the probes indicated that IPA.1001 was the most-populated consensus site, which confirmed our identification of IPA.1001 as the maximally occupied hot spot. The other IPA positions in the binding site were identified as sites 9 and 10. Simulation density sites 3, 6, and 7 corresponded with known hot spots in the active region. IPA 1002x was mapped as site 8. We found that solvent occupancies calculated from the middle of 10 independent trajectories (15−20 ns) resulted in comparable mapping when measured against the occupancies from the final 5 ns (45−50 ns) of 5 runs (Figure 5). Although 5 simulations performed over 50 ns were sufficient to show convergence, doubling the number of simulations decreased the amount of simulation required: 250 ns were required for 5 runs compared to 200 ns

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

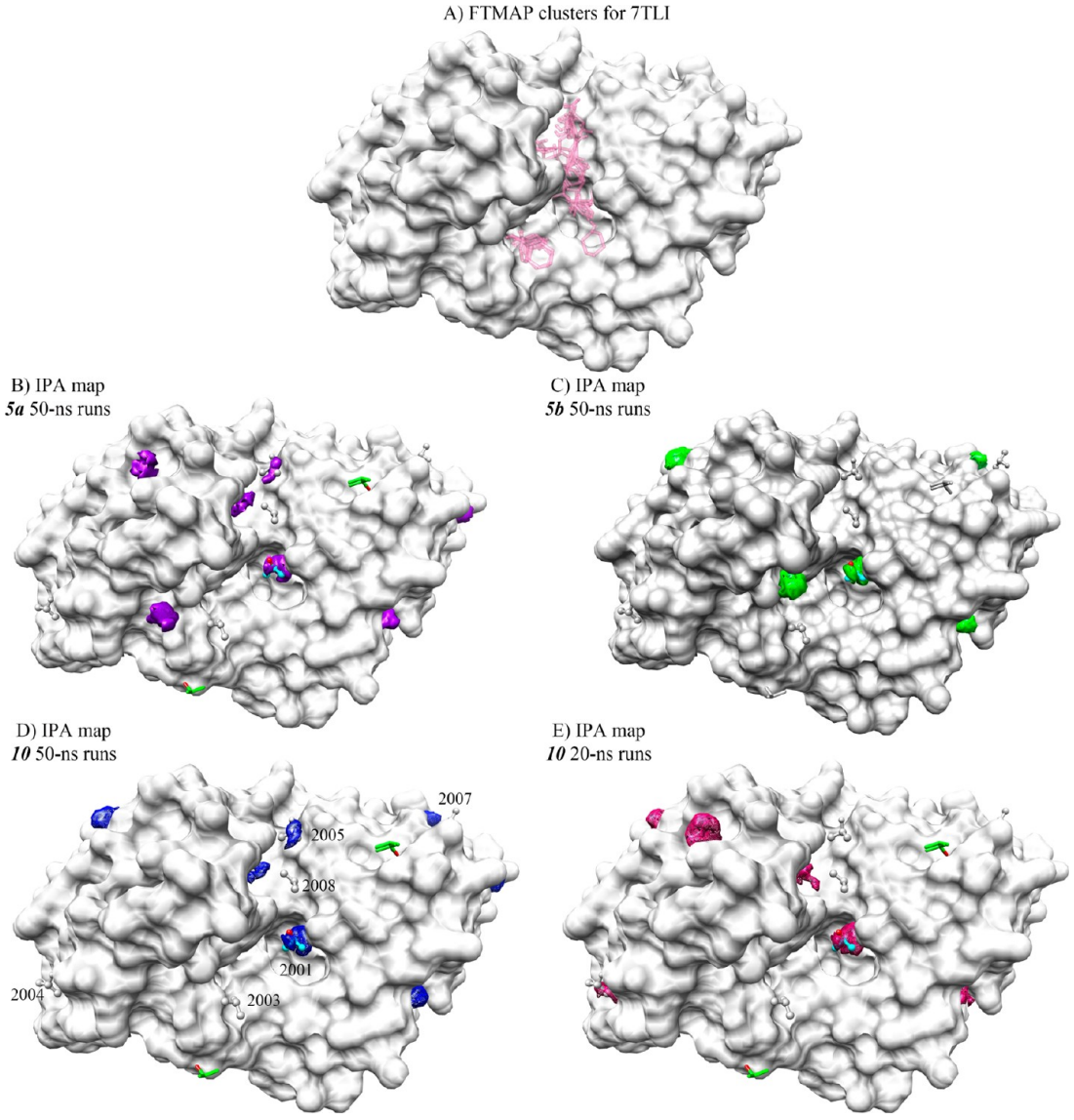


Figure 6. MixMD data combined for the last 5 ns of 10 individual 50-ns runs of fully flexible thermolysin in 50% w/w IPA and water. All snapshots were superimposed on the original crystal structure (PDB ID 7TLI, white surface). The MSCS coordinate positions for IPA.2001, IPA.2006, and IPA.2008 are shown in cyan as ball-and-stick while the unjustified probe sites are shown in gray. The crystal interface probe sites are shown in green. Density contour levels were chosen such that the top eight hot spots were represented, and a "hide dust" level of 3.5 was used. The occupancy level in frames B−E is 0.310 ± 0.056 probe/grid point. (A) Results from performing FTMAP against PDB ID 7TLI, with functional group clusters shown in pink. (B) Simulation IPA density from the last 5 ns of 50 total ns from independent runs 1−5 (purple). (C) The last 5 ns of 50 total ns from independent runs 6−10 (green) do not perform as well as the results from all 10 runs. (D) The maximally occupied site for simulation IPA density from the last 5 ns of 50 total ns from 10 independent runs (blue) corresponds with all three of the crystallographic probe positions as well as two other MSCS sites. (E) Simulation IPA density from the last 5 ns of 20 ns total simulation time over 10 individual runs performs better than the IPA density after 50 ns over 5 runs.

for 10 runs. After the first 20 ns for each of the 10 runs, the hot spot at IPA.1001 had been identified as the highest populated site and continued to be the primary hot spot for the duration of simulation time. Examination of the top-8 hot spots showed that IPA.1001, IPA.1002, and IPA.1003 were each mapped, in addition to a pocket close to IPA.1002x and several hot spots from other MSCS results, including sites with justified electron density for ethanol, dimethylformamide, and IPA (Figure 5). Although structure factor files were not available for elastase +ACN, our MixMD simulations of this system preferentially located the active site as well as several MSCS probe locations (see SI Figure S1).

Like elastase, thermolysin has also been used extensively for MSCS studies[17,19] and computational FBDD studies,[8,40,41] allowing us to compare our MixMD results to a wealth of experimental and computational data. The results from our simulations of thermolysin lent strong support to our emphasis on comparing simulation occupancy grids to the experimental density instead of to the crystallographic coordinates. Although several of the MSCS systems we have discussed had more than one solvent probe along the protein surface, the refined electron density has not always given a clear indication that the coordination position of the probe was justified. This was particularly true for thermolysin+IPA: although eight IPA molecules were specified in the PDB file, comprehensive

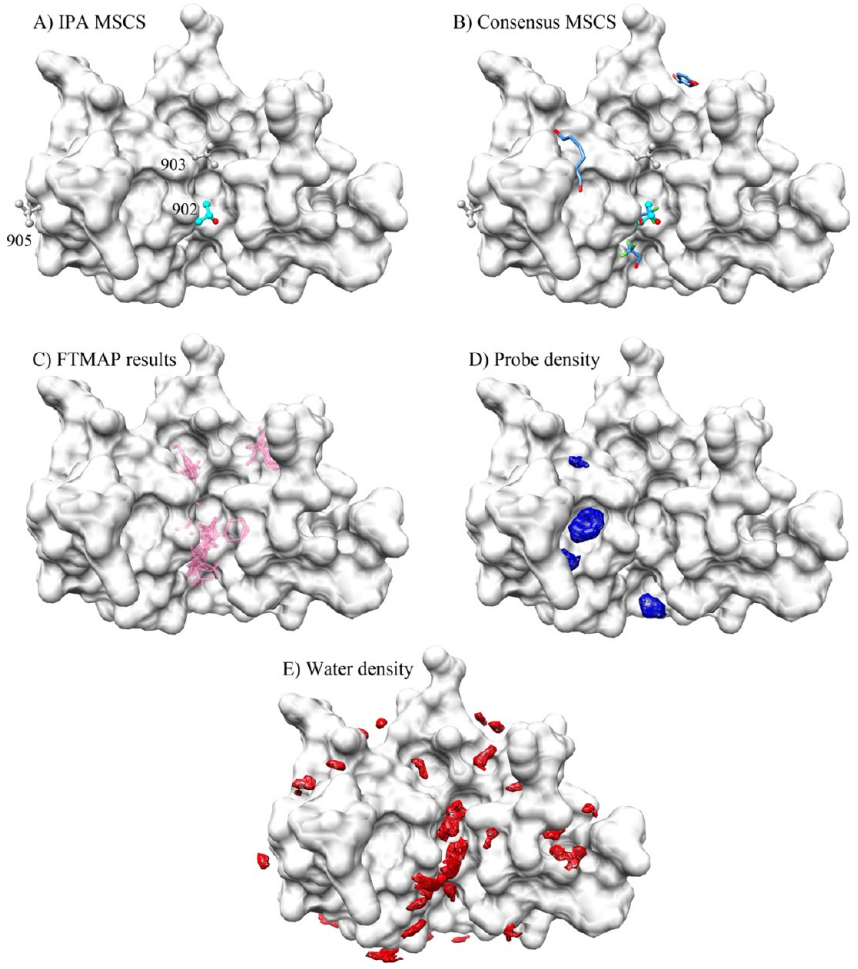


Figure 7. MixMD data combined for the last 5 ns of five individual 50 ns runs of fully flexible RNase A in 50% w/w IPA and water. All snapshots were superimposed on the crystal structure (PDB ID 3EV2, white surface). Crystallographic waters are colored black for B-factors <33 Å and yellow for B-factors >33 Å. Water molecules along the crystallographic boundary are not shown. The MSCS coordinate positions for IPA.905A, IPA.917A, IPA.902B, and IPA.903B are shown in cyan as ball-and-stick. MSCS with other functional groups obtained by the same authors are shown in blue, FTMap sites are shown in pink. (A) MSCS probe sites from 3EV2 chain B. (B) Consensus MSCS probe sites with a variety of different solvent probes. (C) Cluster sites from performing FTMAP against chain B of 3EV2. (D) Simulation density for the probe occupancy over the final 5 ns of the combined runs (0.243 probe/grid point). (E) Simulation density for the water occupancy of the final 5 ns of the combined runs (0.175 probe/grid point).

analysis of the refined structure factor file and the crystal contact region showed that only two of the eight IPA molecules were placed in justified density (Table 1). A number of the solvent-mapping procedures that have been published have used the probe positions in thermolysin to validate their technique, which further emphasizes the necessity of understanding the limitations of crystallographic data prior to performing validation analyses.

F_O-F_C density in the refined map revealed that the only justified probe sites were IPA.2001 and IPA.2002. Of these two sites, IPA.2001 corresponded to the Zn^{2+}-containing active site, while IPA.2002 corresponded to a small, buried pocket distant from the active site. The maximally occupied sites for IPA molecules correctly agreed with the position of the greatest probe density in the active site, the position of IPA.2001. We identified several other hot spots in the thermolysin active site, including the position of IPA.2005 and the alternate TYR157 position A, which occupies a shallow pocket near IPA.2008 and is close to the catalytic histidines. As we saw for elastase, the use of 10 runs of 20 ns each performed equally to 5 runs of 50 ns each (Figure 6).

Our results for the appropriate number of simulation runs to perform for MixMD were particularly relevant to a common assumption in studies of hot-spot mapping, which holds that individual simulation densities should converge to the correct hot spot. We found this assumption to be invalid. Individual simulations all showed some occupancy at the principal sites of probe interaction, but they diverged in their identification of the global minimum. They also diverged in their positions for spurious sites. Combining the simulation data best represented the true binding potential and eliminated irrelevant minima. We found that for both elastase and thermolysin, solvent occupancy grids from 10 runs out to 20 ns (200-ns total) were comparable or outperformed simulation occupancy from 5 runs out to 50 ns (250-ns total), where performance was defined as the number of crystallographic probe sites found within the top-ranked hot spots. This highlighted the efficiency of multiple, shorter simulations. We focused our examination of hot spots on the top-8 probe sites that were well-occupied by simulation density and required that they correspond with most of the crystallographic hot spots to demonstrate success.

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Allowing for Conformational Change. There are occasions when a MixMD map may not reproduce the MSCS data. In cases where conformational rearrangement of the protein alters the solvent accessible surface area in the binding site, the available interactions may change relative to the original crystal structure. The crystal structure of RNase A was solved such that a salt bridge was formed between Asp121 and a neighboring residue in a symmetry partner. When MD simulations were performed in the absence of the crystalline environment, this salt bridge no longer existed, and therefore, Asp121 shifted toward the active site. This caused the binding cavity to narrow, with a new distance of 10.2 Å from Val43-Asp121 compared to an original distance of 11.6 Å. Furthermore, the interactions available at the binding site were altered over the course of simulation due to a median RMSD shift of 4.54 Å from the crystal structure, based on the final 5 ns from five independent simulations of 50 ns each for RNase A+IPA (SI Figure S4).

RNase A binds and cleaves an RNA substrate in a deep cleft flanked by the two catalytic residues, His12 and His119, through a transition state complex that is stabilized by the presence of three basic residues: Lys9, Lys41, and Lys66. The MSCS of RNase A (chain B) had coordinates for two IPA probe molecules at the active site, each with B-factors over 60 Å. IPA.902 was principally associated with His12 and Thr45, while IPA.903 interacted with His12 through a bridging water (Wat921). The refined map of 3EV2 showed that the density for IPA.902 was more compelling as a probe site than the density for IPA.903 (Table 1). We found that the refined F_O-F_C density indicated a justified probe site at IPA.902B and the expected density for a water molecule at IPA.903B.

The maximally occupied site for simulation IPA was located in a deep pocket next to Asp121 (Figure 7), where the probe was able to form a stable hydrogen-bonding interaction due to the conformational shift that occurred in four of the five simulations. The second site identified by probe density overlapped with the position of a cytosolic RNase inhibitor that has been experimentally determined to bind across the active site (PDB ID 3MWQ[42]); we stress that this inhibitor was not present in our simulations nor did we start our simulations from a similar bound structure. Without experimental confirmation of the structural changes observed in our simulation, it is not possible to judge the validity of these findings. These observations highlight the difference in conformations that may be accessible to a biomolecule in a simulation versus a crystalline environment.

We found that our occupancy grid for water mapped the binding site, including the position of IPA.902 and a secondary binding site between IPA.902B and IPA.903B, which was well-positioned to take advantage of contacts to His12 and Lys41. Over the course of our MD simulation, the conformational shift triggered by movement of Asp121 caused the β-sheet that contained His119 to move downward, which oriented His119 into the position occupied by Wat921 in the crystal structure. This affected mapping of the binding site, and the secondary site is likely the equivalent of the density assigned to IPA.903B in the crystal structure. Our combined solvent occupancy (IPA +water) mapped the bound position of IPA.917A of 3EV2, which lies along the protein surface in a shallow binding pocket near Thr70. Only sixteen of the crystallographic water positions were within 5 Å of the protein and had B-factors of <33 Å; all but one of these positions was strongly mapped by our water occupancy grids.

These results for RNase A+IPA indicated that a lower concentration or different probe may be required for mapping in order to correctly identify the active site. Guvench and coauthors presented an excellent computational study on balancing the potential for probe-induced denaturation with the need for protein flexibility while mapping IL-2.[39] It is unclear based on the available experimental data whether or not the conformation observed in our simulation is valid for this protein system, or whether the alcohol concentration has influenced the conformational ensemble toward an unfolded state.[11,39] Future studies will examine the impact of a lower concentration of solvent on conformational sampling and probe mapping in our simulations.

■ CONCLUSIONS

Our results demonstrated the utility of MixMD for identifying the hot spots present in a broad range of pharmaceutically relevant receptor targets. We have shown that both protic and aprotic solvent types can be used in MixMD to specifically model the consensus sites for binding as a complementary technique to experimental FBDD. Evaluation of the F_O-F_C density generated based on the refined structure with probe(s) removed showed the importance of thoroughly examining the experimental data in a test set to ensure fair and accurate comparison. In all cases, true hot spots in the crystal structure were located by the simulation density. The additional hot spots that were identified by solvent occupancies from MixMD were judged to be reasonable based on other crystal structures, the surrounding binding surface, and analysis of the crystalline interface. Probe concentration should be explored in order to determine its influence on true conformational change or partial unfolding of the protein.

We have stressed the importance of performing careful validation studies when developing an approach to solvent mapping like MixMD. Detailed analysis suggests that MixMD may be used to aid in the assignment of crystallographic density. Additionally, the optimal simulation length and run number for mapping probe sites without overabundant local minima was determined so as to enable MixMD to be broadly applied to diverse protein cases. Our study found that an analysis of computational solvent mapping that is based on simulation density *can* result in clean identification of binding sites that are not scattered among numerous, spurious local minima.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Supplementary analyses include results from elastase+ACN, subtilisin+ACN, and thermolysin+ACN, core RMSD and RMSF, the refined density maps, and water occupancy maps. Figures S1–S3 display results from MixMD with ACN, Figures S4 and S5 demonstrate the stability of the proteins over the course of the MD trajectory, Figures S6–S8 illustrate the F_o-F_c density for each organic probe following its omission from the data and the subsequent refinement of the electron density, Figure S9 presents the results for simulation solvent occupancy over different time windows, and Figure S10 depicts the correspondence between MixMD simulation occupancy and crystallographic water. This material is available free of charge via the Internet at http://pubs.acs.org.

dx.doi.org/10.1021/ci300430v | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

AUTHOR INFORMATION

Corresponding Author

*E-mail: carlsonh@umich.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Charlie Brooks for generously sharing his computer resources. This work has also been supported by the TeraGrid project (TG-MCB110089); TeraGrid is part of XSEDE (NSF OCI-1053575). We thank Dagmar Ringe for sharing the coordinates for her MSCS of elastase+ACN. We also thank Jeanne Stuckey, Oleg Tsodikov, and Thomas Goddard for their advice and assistance in interpreting electron density data. This work has been supported by the National Institutes of Health (GM65372). K.W.L. thanks Rackham Graduate School, the Pharmacological Sciences Training Program (GM07767), and the American Foundation for Pharmaceutical Education for funding. Molecular graphic images were produced using COOT, PyMOL, and the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

REFERENCES

(1) Allen, K. N.; Bellamacina, C. R.; Ding, X.; Jeffery, C. J.; Mattos, C.; Petsko, G. A.; Ringe, D. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. *J. Phys. Chem. B* **1996**, *100*, 2605−2611.

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

(3) 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.

(4) Clackson, T.; Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267*, 383−6.

(5) Bogan, A. A.; Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1−9.

(6) Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **2009**, *1*, 187−192.

(7) Howard, S.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Curry, J.; Devine, L. A.; Early, T. R.; Fazal, L.; Gill, A. L.; Heathcote, M.; Maman, S.; Matthews, J. E.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Rees, D. C.; Reule, M.; Tisi, D.; Williams, G.; Vinkovic, M.; Wyatt, P. G. Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. *J. Med. Chem.* **2009**, *52*, 379−388.

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

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

(10) Raman, E. P.; Yu, W.; Guvench, O.; Mackerell, A. D. Reproducing Crystal Binding Modes of Ligand Functional Groups Using Site-Identification by Ligand Competitive Saturation (SILCS) Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 877−896.

(11) 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.

(12) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(13) Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29−34.

(14) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621−627.

(15) Kozakov, D.; Hall, D. R.; Chuang, G. Y.; Cencic, R.; Brenke, R.; Grove, L. E.; Beglov, D.; Pelletier, J.; Whitty, A.; Vajda, S. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13528−13533.

(16) Lexa, K. W.; Carlson, H. A. Full Protein Flexibility Is Essential for Proper Hot-Spot Mapping. *J. Am. Chem. Soc.* **2011**, *133*, 200−202.

(17) English, A. C.; Done, S. H.; Caves, L. S.; Groom, C. R.; Hubbard, R. E. Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 628−640.

(18) Fitzpatrick, P. A.; Steinmetz, A. C.; Ringe, D.; Klibanov, A. M. Enzyme crystal structure in a neat organic solvent. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8653−8657.

(19) English, A. C.; Groom, C. R.; Hubbard, R. E. Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.* **2001**, *14*, 47−59.

(20) Mattos, C.; Bellamacina, C. R.; Peisach, E.; Pereira, A.; Vitkup, D.; Petsko, G. A.; Ringe, D. Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J. Mol. Biol.* **2006**, *357*, 1471−1482.

(21) Deshpande, A.; Nimsadkar, S.; Mande, S. C. Effect of alcohols on protein hydration: crystallographic analysis of hen egg-white lysozyme in the presence of alcohols. *Acta Crystallogr. D Biol. Crystallogr.* **2005**, *61*, 1005−1008.

(22) Ho, W. C.; Luo, C.; Zhao, K.; Chai, X.; Fitzgerald, M. X.; Marmorstein, R. High-resolution structure of the p53 core domain: implications for binding small-molecule stabilizing compounds. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 1484−1493.

(23) Dechene, M.; Wink, G.; Smith, M.; Swartz, P.; Mattos, C. Multiple solvent crystal structures of ribonuclease A: an assessment of the method. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 861−881.

(24) 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.

(25) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375−W383.

(26) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(27) Case, D. A.; Darden, T. A.; Cheatham, I., T.E. ; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California, San Francisco: 2008.

(28) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.

(29) 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.

(30) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(31) 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.
(32) Andrea, T. A.; Swope, W. C.; Andersen, H. C. The role of long ranged forces in determining the structure and properties of liquid water. *J. Chem. Phys.* **1983**, *79*, 4576−4584.
(33) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235−242.
(34) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P; Sharff, A.; Smart, O. S.; Vonrhein, C.; T.O., W. *Buster 2.11.1*; Global Phasing Ltd: Cambridge, United Kingdom, 2011.
(35) Kleywegt, G. J.; Harris, M. R.; Zou, J.-y.; Taylor, T. C.; Wahlby, A.; Jones, T. A. The Uppsala Electron-Density Server. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2240−2249.
(36) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(37) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486−501.
(38) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, CA, 2008.
(39) 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.
(40) Englert, L.; Silber, K.; Steuber, H.; Brass, S.; Over, B.; Gerber, H.-D.; Heine, A.; Diederich, W. E.; Klebe, G. Fragment-Based Lead Discovery: Screening and Optimizing Fragments for Thermolysin Inhibition. *ChemMedChem* **2010**, *5*, 930−940.
(41) Wang, S. M.; Yang, C. Y. Computational Analysis of Protein Hotspots. *ACS Med. Chem. Lett.* **2010**, *1*, 125−129.
(42) Arnold, U.; Leich, F.; Neumann, P.; Lilie, H.; Ulbrich-Hofmann, R. Crystal structure of RNase A tandem enzymes and their interaction with the cytosolic ribonuclease inhibitor. *FEBS J.* **2011**, *278*, 331−340.