

Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics

S. Roy Kimura,^{*,#,V} Hai Peng Hu,[±] Anatoly M. Ruvinsky,[§] Woody Sherman,^{§,¶} and Angelo D. Favia^{*,±,⊥,ID}

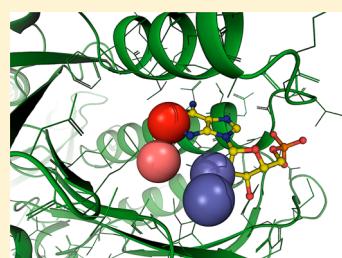
[#]Schrödinger KK, 17th Fl, Marunouchi Trust Tower North, 1-8-1 Marunouchi, Chiyoda-ku, Tokyo, Japan

[§]Schrödinger LLC, 222 Third Street, Suite 2230, Cambridge, Massachusetts 02142, United States

[±]Lilly China Research and Development Center (LCRDC), Eli Lilly and Company, Building 8, 338 Jia Li Lue Road, Shanghai 201203, PR China

Supporting Information

ABSTRACT: In recent years, molecular dynamics simulations of proteins in explicit mixed solvents have been applied to various problems in protein biophysics and drug discovery, including protein folding, protein surface characterization, fragment screening, allosteric, and druggability assessment. In this study, we perform a systematic study on how mixtures of organic solvent probes in water can reveal cryptic ligand binding pockets that are not evident in crystal structures of apo proteins. We examine a diverse set of eight PDB proteins that show pocket opening induced by ligand binding and investigate whether solvent MD simulations on the apo structures can induce the binding site observed in the holo structures. The cosolvent simulations were found to induce conformational changes on the protein surface, which were characterized and compared with the holo structures. Analyses of the biological systems, choice of probes and concentrations, druggability of the resulting induced pockets, and application to drug discovery are discussed here.



■ INTRODUCTION

In structure-based drug design, knowledge of the ligand binding sites (orthosteric and allosteric) can be central to the design and discovery of molecules that bind in a potent and selective fashion. In many cases, the binding sites are well-defined regions, naturally engineered to host substrates, and are partially enclosed within the 3D fold of the macromolecule.¹ Such pockets are commonly present in both apo and holo X-ray structures of the protein of interest. However, other times a binding site reveals itself only upon ligand binding, thus making more difficult the structure-assisted design of modulators, especially for those proteins with limited structural information available. Those so-called “cryptic sites” sometimes have great biological relevance and are widely exploited by nature both in protein–protein recognition processes and orthosteric and allosteric modulation.^{2–4} Targeting those sites could potentially offer a number of advantages over traditional pockets, such as enhanced selectivity (in the case of allosteric sites) and the opportunity to modulate targets that were previously deemed undruggable.⁵

When experimental determination of cryptic sites is challenging or not feasible, computational methods represent a less expensive and prompter alternative means of characterization. To this end, long time scale molecular dynamics (MD) simulations have demonstrated the ability to unveil cryptic sites for certain molecular systems.^{2,3,6,7} On the down side, those calculations are computationally intensive and their general applicability to a wide array of biological systems has yet to be established. On the other hand, bioinformatics methods that rely on sequence conservation and coarse-grained potentials,

although computationally appealing, might overlook the fine atomistic details of the induced pocket, thus limiting their applicability in prospective drug design scenarios.⁸

Here, we explore molecular dynamics simulations in various mixed solvents to induce openings of cryptic sites in a timely fashion (ie, 100 ns production time per probe). The use of mixed-solvent simulations has been widely exploited to assess druggability of proteins with a high degree of accuracy.^{9–20} Within this MD framework, increased local concentrations of organic solvent molecules are detected and used to compute estimates of affinity that could be attained through binding of exogenous molecules. One such recent study suggests that use of a mixture of a consensus set of organic probes enables prediction of a theoretically maximal achievable binding affinity that is consistent with existing compound affinity data.⁹ Given the grid-based detection method employed, the underlying assumption (that usually holds true) enabling accurate mapping of the surface of the protein is that the system does not undergo major conformational changes within the studied time frame (i.e., usually 20–40 ns). However, in our preliminary studies on a data set of diverse proteins, with mixed-solvent simulations extended to 100 ns, we have observed recurrent conformational changes. Hence, building on the work of others,^{13,17} we asked ourselves whether those occurrences, rather than being systematic artifacts, had some biological relevance instead and could be traced back to the presence of cryptic sites. To that end, we have thoroughly examined a

Received: October 14, 2016

Published: May 24, 2017

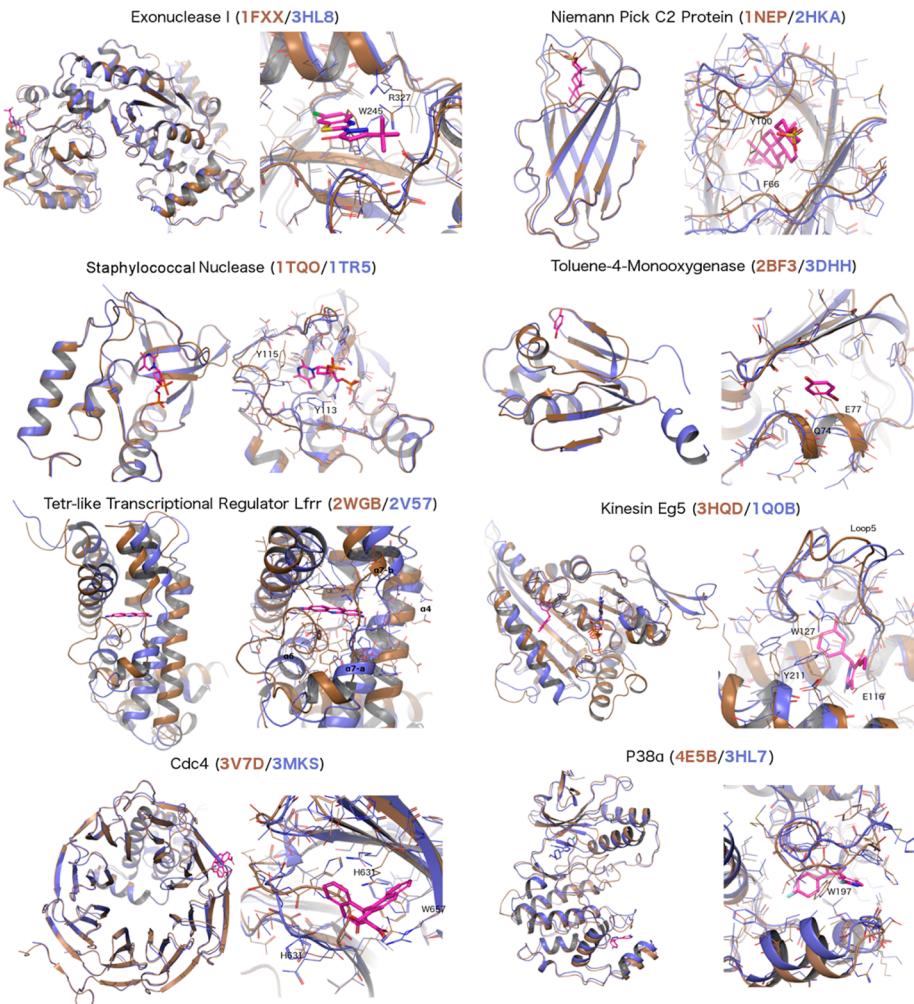


Figure 1. Structures of the apo/holo protein pairs. The superimposed apo and holo structures are shown in brown and blue, respectively. The zoom-in views show conformational changes induced by each ligand (C-colored magenta). PDB IDs are also shown.

diverse set of proteins extracted from the work of Cimermancic et al.,⁸ taken from the PDB²¹ (www.wwpdb.org), solved in apo and holo states that show pocket opening induced by ligand binding and applied the protocol to the apo structures. The driving hypothesis, initially based on circumstantial observations, was that the presence of organic solvent could help reveal the presence of a cryptic site by promoting conformational changes within an affordable time frame.

In our exploratory survey, we asked two questions regarding induced cryptic site opening: (a) whether mixed-solvent MD simulations could reliably identify the correct locations of cryptic binding sites and (b) whether simulations with organic probes could induce opening of the cryptic sites.

We used a mixed-solvent simulation protocol involving small organic probes in separate water–probe binary mixture simulations to reduce the potential complications of interprobe interactions. We opted to assess whether commonly available state of the art tools and simple protocols on a realistic data set would aid the identification and characterization of cryptic sites and investigate the utility of such methods in drug discovery research.

In our analysis, we used an energetic analysis of probe hotspots based on the work of Bakan et al.⁹ However, because we used binary mixtures involving a single probe type and water, our energies do not predict the theoretical maximal

binding affinity but rather approximately describe the affinity of each employed probe toward their binding sites, given that certain underlying assumptions are met.

To the best of our knowledge, this is among the first reports of a systematic exploration of mixed-solvent simulations to investigate cryptic sites and complements nicely the recently reported works on allosteric and cryptic site induction via MixMD¹⁷ and via enhanced sampling methods.²² Here, we present a detailed case-by-case analysis on use of the protocol on experimentally known cryptic sites, taking into account, the detailed biology and biochemistry of the induced protein structures as evidenced by the apo/holo structure pairs. Other topics associated with mixed-solvent simulations including convergence, choice of probes and concentrations, druggability of the resulting induced pockets, and application to drug discovery are discussed.

RESULTS AND DISCUSSION

Data Set of Proteins and Solvent Probes. For our data set, we selected a wide array of protein systems for which apo/holo forms pairs were publicly available. A common feature of those systems was the presence of a ligand in the holo form that revealed a cryptic site. The final selection of diverse proteins, extracted from a recently published more comprehensive data

Table 1. Results of the Hotspot Analysis^a

system	apo structure	holo structure	cosolvent probe	hotspot rank	hotspot energy	pocket opening
exonuclease I	1FXX	3HL8	isopropanol	6	-4.2	yes
Niemann-Pick C2 protein	1NEP	2HKA	resorcinol	3	-11.9	yes
Staphylococcal nuclease	1TQO	1TRS	acetic acid	1	-7.3	partial
toluene-4-monooxygenase	2BF3	3DHH	resorcinol	3	-6.0	yes
TETR-like transcriptional regulator LFRR	2WGB	2VS7	isopropanol	3	-4.7	partial
kinesin Eg5	3HQD	1Q0B	resorcinol	4	-9.3	partial
Cdc4	3V7D	3MKS	acetid acid	2	-7.5	partial
P38 α	4ESB	3HL7	isopropanol	1	-6.0	yes

^aPDB IDs of apo and holo structures used, results of cosolvent probe occupancy analysis, and degree of induced fit pocket opening as judged by visual inspection of probe occupied trajectory frames. Energies are given in kilocalories per mole.

set,⁸ encompassing eight diverse systems is highlighted in Figure 1.

Separate simulations were performed using 5% molar ratios of (a) acetic acid, (b) isopropanol, and (c) resorcinol in water. At the same time, as a means of comparison, standard simulations in water were also performed. Each system was simulated for 100 ns. The cumulative simulation time for the eight systems was 3.2 μ s.

Overall Results. Table 1 summarizes the overall results of probe occupancy hotspot analysis to assess whether mixed-solvent simulations could correctly identify the locations of the cryptic sites. The table includes results from solvent probe simulations that exhibited high occupancy hotspots near the cryptic site as observed in the holo structure. As stated earlier, energetic ranking was performed only with respect to the affinity for the particular probes used. Despite the employment of a lean protocol, we note that all but one of our cryptic sites were found among the top four clusters identified through the probe occupancy analysis.

Figure 2 shows the locations of high occupancy probe hotspots that best coincided with the cryptic sites on the proteins from the simulations. The open-like state was selected, as representative, from the trajectories. Again, we only show the results from those probes that resulted in the highest density hotspots near the cryptic site.

In Figure 3 we show selected results of PCA on the cryptic site side chain relative positions visited during the MD simulation. The rest of the PCA plots are shown in the Supporting Information (Figure S2). Here, we plot the density distribution of relative side chain positions visited during each MD simulation in the first two dominant principal components.

From those plots, it is quite evident that different mixed solvents induce different motions. It is also interesting to note that frames from similar solvents (e.g., hydrophilic solvents water and acetic acid versus those with hydrophobic moieties, isopropanol and resorcinol) often appear to cluster together. We found that each probe caused different induced fit effects on the proteins. Some probe types caused greater pocket opening while others opened pockets to a lesser extent or not at all, and this was highly target dependent. We do remark here that, while visually useful in a retrospective analysis like this one, the PCA analysis here employed can find limited applicability in a real-case scenarios, due to the lack of initial information on the area to focus on. Nevertheless, the analysis served the purpose to show how different organic mixtures trigger distinct conformational changes in a consistent way. To assess the reproducibility of those findings, we have run independent simulations for a chosen system (i.e., Kinesin Eg5) and then compared the final results in terms of exploration of

the PC space. While not perfect and dependent upon the explained variance of the first two principal components, results from this analysis looked encouraging as the covered PC space looks nicely dependent upon the employed probes and fairly consistent across runs. Figure S3 in the Supporting Information summarizes the outcome of the PCA analysis.

For the sake of completeness, RMSD plots for the entire data set can also be found in the Supporting Information (Figure S1). In the following, we present a detailed case by case analysis of each of the target proteins examined, taking into account the potential biological and atomistic details of induced pocket formation at the cryptic sites.

Exonuclease I (ExoI) is a protein that regulates ssDNA-binding proteins (SSBs) in *E. coli*.^{23,24} The interaction between those proteins is well-understood and is mediated through the SSBs C-terminal segment. The X-ray structure of ExoI bound to a C-terminal fragment has shed light at the atomistic level on the features needed for interaction. To better assess the biological relevance of this partnership, a number of tool molecules have been identified. As one might anticipate, most of those chemicals bind at the orthosteric C-terminal binding site on ExoI. However, interestingly enough, Keck and co-workers have identified one additional allosteric site that is not evident in the apo structure of ExoI.²⁴

Binding to this site triggers conformational changes that indirectly hamper the binding of SSBs. The main differences between the holo (PDB ID: 3HL8) and apo (PDB ID: 1FXX) structures are due to side-chain movements of W245 and R327. The 70° shift in dihedral angle along the CB-CG bond of W245 opens up a hydrophobic pocket defined by P228, L264, C330, L331, and L334. This pocket has been shown to be capable of hosting a variety of nonpolar groups. Furthermore, the ligand found in the holo structure (i.e., BCBP) is stabilized by a water mediated interaction with R327 (whose hydrophobic side chain contributes to the pocket definition) and by close favorable van der Waals contacts with N313 and P317.

While at first glance this might not look as a major structural rearrangement, our pure water simulations failed to recapture it consistently within the simulation time (Figure 3). The four investigated systems were well behaved as the backbone RMSD remained stable throughout the simulations. The trajectories in the 5% acetic acid solution and pure water closely resembled each other. Given the hydrophobic nature of the cryptic site, this is unsurprising. On the other hand, the resorcinol probe trajectory, while briefly exploring around the closed state, pushed the sampling mainly toward a region equidistant from the holo and the apo structures. Eventually, the isopropyl alcohol simulation explored a region of the free energy landscape closer to the holo structure. Deeper analysis showed

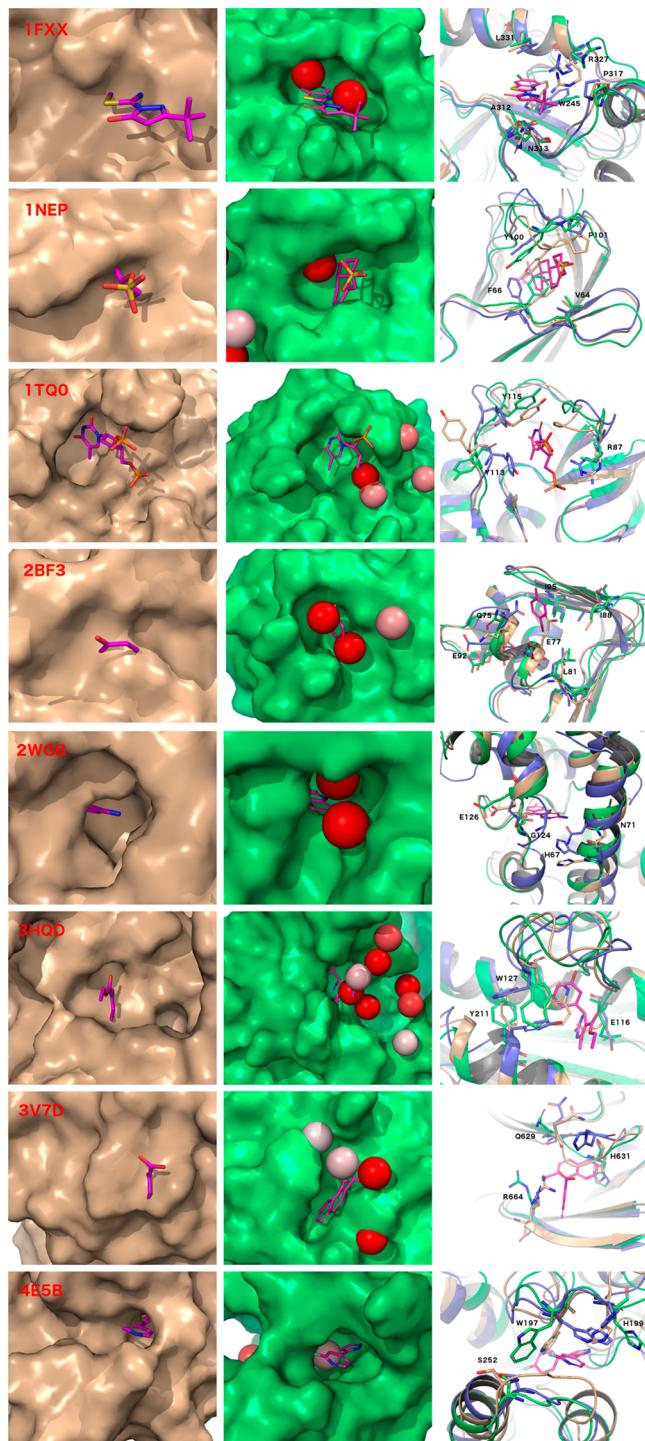


Figure 2. Apo structures (brown, left column) and structures taken from the MD trajectories (green, middle column) showing pocket opening. The middle column shows the locations of high density probe occupancy hotspots colored by density relative to bulk solvent density (blue). The density increases from blue to red (white corresponds to 25% above bulk density and red is greater than 50% above density). The right column shows a superposition of the apo (brown), holo (blue), and MD-generated, open-like state (green). The ligand, as present in the holo structure, is overlaid for reference.

that this favorable behavior was due to the aliphatic part of the probe that pushed the W245 away, thus fully exposing the BCBP enclosing pocket. In this case, there was a full pocket opening with the isopropyl probe and the cryptic site was

ranked number 6 with an overall ΔG of -4.2 kcal/mol. The orthosteric site ranked number 1 with a predicted binding free energy of -8.4 kcal/mol instead. It is here worth noting the cryptic site opening induced the appearance of an additional pocket through the movements of R327 and I325 (data not shown). The appearance of this putative site, whose biological relevance is not known, could be due to some concerted movements that propagate from the allosteric site itself.²⁵

A moderate rearrangement of the helix supporting L331 and R327 can be observed from a representative snapshot of the mixed-solvent simulation (Figure 2). At the same time, it is worth noting the position of W245 (behind the ligand) that very closely mimics the holo reference. The position of R327 does not differ markedly from the apo configuration and likely was responsible for a suboptimal scoring of the cryptic site (this system proved to be the more challenging for our protocol). Some minor rearrangements can be observed for the loop supporting A312, N313, and P317.

NPC2 is a small protein whose deficiency causes accumulation of cholesterol in lysosomes (i.e., Niemann-Pick type C2 disease).^{26,27} NPC2 moves cholesterol (as well as other sterols) from lysosomes to other locations to meet specific cellular needs. Unlike similar structures belonging to the same MD-2-related lipid recognition domain family, NPC2 lacks a preformed hydrophobic cavity for ligand binding. Despite solid evidence of the formation of a stable 1:1 complex with cholesterol, early attempts of crystallization of a complex have proven unfruitful. In 2007, Stock and collaborators succeeded where others had failed by using cholesterol-3-O-sulfate, a tight binder of the protein, as the interacting ligand.²⁷ In this case, the studied site is orthosteric, however it is not detectable by simply looking at the apo structure (PDB ID: 1NEP) since its presence becomes evident only upon ligand binding. As a matter of fact, early experimental attempts²⁸ as well as the here reported simulation in pure water failed to highlight the binding pocket.

The X-ray model of the holo (PDB ID: 2HKA) form shows a deep tunnel that is perfectly molded to accommodate the partner ligand. The opening of the cavity is facilitated by an extensive rearrangement of residues F66, Y100, P101, and I103 that lie on the β D, β E– β F loops. The cryptic site, outlined by hydrophobic residues, can fully envelope the cholesterol sulfate leaving only the hydrophilic moiety exposed to the solvent. A notable difference between NPC2 and other proteins that act as lipid carriers is the absence of a molecular “lid” that facilitates the ligand entrance through membrane inclusion. This might partially explain the absence of a preformed pocket that otherwise would be exposed to the solvent leading to an energetically unfavorable state. Under physiological conditions, the ratio between open and closed states likely changes with respect to a simpler *in vitro* environment, due to the presence of assisting partners such as proteins, lipid aggregates or membranes.

Because of the striking differences between the holo and apo form, this represents a very challenging case. Nonetheless, the employed protocol successfully ranked the cholesterol binding pocket as number 3 with a theoretical ΔG value of -11.9 kcal/mol using the resorcinol probe. The overall structure of the protein behaved well during the four simulations (Figure S1) with the main shifts localized within the cryptic pocket. As reflected in the PCA plots in Figure 3, the pure water and acetic acid probe simulations produced similar results. On the other hand, probes with a more marked hydrophobic character clearly

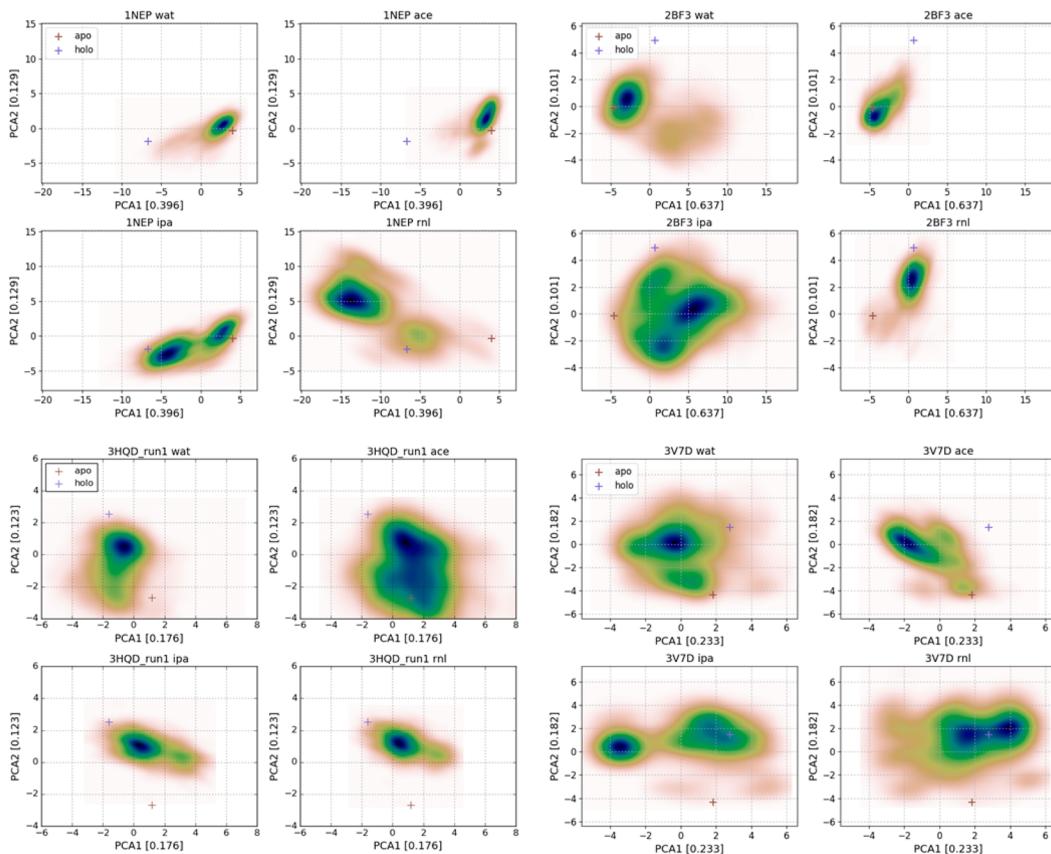


Figure 3. PCA analysis of the relative side chain positions at the cryptic sites. Shown are density distributions of the MD frames using different mixed-solvent probes, plotted against the first 2 PCA vectors. See [Results and Discussion](#) and [Methods](#) sections for details. The locations of the apo and holo structures are highlighted by blue and brown “+” symbols, respectively.

sampled different regions of the free energy landscape. The isopropyl alcohol simulation explored regions in between the closed and open forms. The water/resorcinol mixture after an initial approach to the target structure pushed the sampling substantially away from the holo reference. Nonetheless, the latter probe was the top performing one. In-depth analysis of the trajectories showed that the bulkier character of the ring played a key role in triggering the conformational changes needed to unveil the pocket. Moreover, displacement of F66 and Y100 (at the edge of the site) was greatly facilitated by the aromatic nature of the probe. While the pocket of interest was not the top ranked one, given the promiscuous nature of this protein (in terms of assisting partners), we also cannot exclude the biological significance of some of the other hotspots found on the surface.

A representative frame taken from the mixed-solvent trajectory ([Figure 2](#)) nicely shows the inward movement of the side chain of F66 that opens up the pocket. A major rearrangement pointing toward the holo reference can be noted for the loop supporting Y100 and P101 as well.

Staphylococcal nuclease is a calcium dependent enzyme produced by certain strains of *Staphylococcus aureus*. The role of the enzyme is to hydrolyze DNA and RNA at the 5' position of the phosphodiester bond yielding a free 5'-hydroxyl group and a 3'-phosphate monoester.²⁹ The structure pair analyzed in our study consisted of variants (I92E) purposely designed to study the thermodynamics effects of burying a charged amino acid within a lipophilic binding site.³⁰

The here analyzed site is the orthosteric pocket that hosts the substrate. We see that in fact this particular structure appears to be more open in the apo form (PDB ID: 1TQO), with a pair of tyrosines, Y113 and Y115, rearranging in the holo (PDB ID: 1TRS) structure to form hydrophobic interactions with the ligand. We noted that a short run of molecular dynamics of the apo structure in pure explicit water caused the pocket to close up, mainly due to large movements in the loop region defined by residues K84 to R87 (data not shown).

One of the phosphate groups of the ligand coordinates the catalytically relevant calcium ion whose allocation is sustained through further coordination with the side chains of D21, D40, and E43 and the carbonyl's backbone of T41. The other phosphate group engages the protein through interaction with K84 and Y85. Notably, these residues belong to a loop that faces the movable 112–119 loop, thus contributing to the boundaries' definition of the orthosteric pocket. The thymine part of the ligand fits deeper into the pocket and is stabilized through a complex network of water-mediated hydrogen bonds between the enzyme and the polar functionalities (i.e., the hydroxyl group and the N of the pyrimidine ring). Interestingly enough, the polar part of the thymine ring points to a markedly hydrophobic region of the enzyme defined by three closely stacked leucine residues (L37, L38, and L89). The ligand's ring does not seem to be directly involved in $\pi-\pi$ interactions.

While the apo/holo pair lies fairly close in the PCA space ([Figure 3](#)), this system has proven to be very challenging for the proposed methodology. The simple water simulation highlighted two main basins. The secondary basin appears in

proximity of the apo structure while the primary basin explored a wider region of the landscape located far from both studied X-ray structures. The four simulated systems appeared quite stable within the simulation time with minimal RMSD fluctuations throughout. The isopropyl alcohol trajectory, centered on the apo structure, was unable to unveil the cryptic site. The acetic acid trajectory sampled closer to the holo structure and yielded relevant hot-spots in proximity of the cryptic site. On the other hand, the resorcinol based system sampled a region at the edge of the apo structure. The predicted hotspot energy at the cryptic site was -7.3 kcal/mol and ranked overall as the top solution using the acetic acid probe. One of the reasons behind this result could be the hydrophilic/charged nature of the cryptic site. In fact, the organic solvent with enhanced lipophilic features (i.e., resorcinol) was capable to highlight only part of the pocket. Namely, the one characterized by the presence of more hydrophobic amino-acids. Nonetheless, even at low affinity, for resorcinol, a partial pocket opening was observed and characterized. A representative structure from the acetic acid probe simulation (Figure 2) shows a slight opening of the thymine pocket due to rearrangement of Y115, and the backbone of Y113 closing in toward the ligand binding site. The two side chains appear to be well positioned to adopt the induced holo conformation through a rotation of their α - β bonds.

Toluene-4-monooxygenase catalytic effector protein (T4moD) is part of a bigger macromolecular complex that catalyzes the interconversion between toluene and *p*-cresol.^{31,32}

T4moD has been shown to alter the kinetics and the regiospecificity (among other effects) of the reaction by molecular association with T4moH. This effector protein is made of 102 amino acids and does not contain cofactors or ions. The atomistic basis of its modulatory effect on the NADH-dependent hydroxylation activity were poorly understood until 2008, when Fox and co-workers solved the structure of the effector protein–hydroxylase complex of toluene-4-monooxygenase in two redox states.³¹ The analyzed holo (PDB ID: 3DHH) structure represents T4moD in complex with 4-bromophenol. The core of the enzyme is almost entirely hydrophobic with the sole exceptions of Q74 and E77, right at the edge of the cryptic site. The first layer of hydrophobic residues in contact with the ligand is composed of L57, L77, L81, I88, I95, and F97. A second layer of residues defining the inner pocket comprises, among other residues, P16, I18, V27, L59, and M74.

The main difference between the holo and apo (PDB ID: 2BF3) structures can be observed in the 3D arrangement of the helix 72–80 that opens up upon interaction with the ligand through the outward shift of Q74 and E77. Stabilization of the complex seems to occur mainly through unspecific hydrophobic contacts. Taking into account the nature of the cryptic site and the chemical features of the cocrystallized ligand (as well as the ones of the natural substrate), this case study offers an ideal scenario for the proposed methodology. Unsurprisingly, as it can be observed by analyzing the trajectory in the PCs space (Figure 3), both the water and the acetic acid simulations allowed for an exploration of the free energy landscape that substantially differed from the apo structure. The water trajectory occasionally drifted toward a space further away from both studied X-rays. Interestingly, although the conformational space near the crystal structures was somehow explored by the isopropyl alcohol trajectory, this probe did not trigger a

complete conformational transition to the holo state. The resorcinol trajectory differed substantially from the others, as the system appeared to quickly move toward the holo structure, where it sampled extensively a well-defined region of the landscape at the edge of the target configuration. As a result of this effective sampling, the resorcinol hotspot analysis correctly highlighted the cryptic site ranking it as number 3 with a ΔG value of -5.96 . The number 1 ranked cluster, positioned in proximity of cluster number 2, defined a volume that encompassed the interaction area for the macromolecular complex formation. The remaining clusters were by far less energetically favored with ΔG values estimated to be around -3 kcal/mol or above. As it can be inferred from the RMSD plots, the four simulated systems were well behaved with minimal fluctuations around the average value. Part of the success of this case is due to the fact that the resorcinol probe successfully managed to engage the only two polar residues of the site (i.e., Q75 and E77) inducing an outward shift of the underlying helix that led to a positive opening. Once exposed, the cryptic site, due to its hydrophobic features, could easily attract and bind the probe molecules.

The most important conformational shift that was instrumental to the pocket detection is due to the reallocation of the helix supporting M74, Q75, and E77 (Figure 2). The position of the side chain of Q75 appears distinct with respect to the apo reference and seems drifted away from the holo configuration. On the other hand, the opening of side chain of E77 (as in the holo X-ray) was instrumental to the entrance of probe molecules.

TetR-like transcription factor LfrR is a transcriptional repressor that controls the expression of the efflux pump LfrA in *Mycobacterium smegmatis* much like the TetR (tetracycline repressor) protein that controls expression of the bacterial efflux pump, TetA.³³ The LfrA pump removes various antibiotics and foreign compounds such as fluoroquinolones, cationic dyes, and anthracyclines from the cellular interior. Overexpression of LfrA in mycobacteria is known to confer resistance to various antibiotics and deletion of the lfrR gene significantly increases expression of the lfrA gene which in turn causes increased antibiotic resistance. Moreover, it has been shown that various LfrA substrates directly promote the dissociation of the LfrR repressor–DNA operator complex.³³

Here we examined the structure of LfrR apo protein (PDB ID: 2WGB) and its holo structure (PDB ID: 2V57) in complex with proflavin.³⁴ The LfrR protein consists of nine α -helices, the first three defining the DNA binding domain that includes a helix–turn–helix motif, and the last two comprising the dimerization domain. The protein functions as a homodimer and is unique among TetR-like repressors in that it forms an asymmetric dimer with each monomer adopting a different conformation (RMSD 3.4 Å). In the apo structure, the monomer chains differ the most in the α 6 (P100–P09) and α 7-a (P115–D125) regions; in chain A, these regions adopt mostly nonhelical loop conformations, whereas in chain B they are helical despite the identical sequence. In addition, the end of α 7-a and most of α 7-b helices are disordered in chain B with no electron density observed.

The asymmetric unit of the holo structure is comprised of two homodimers, but with only one proflavin molecule bound per dimer. The binding site of proflavin is completely defined by residues from the same monomer, i.e., there are no intermonomer residue interactions with the bound ligand. Therefore, the dimers in the holo structure each contain

monomers in both ligand-bound and unbound states. Interestingly, the $\alpha 6$ and $\alpha 7\text{-}a$ regions for both forms of the monomer retain their α -helical nature, unlike the case for the A chain in the native apo structure.

The proflavine binding site in the fully apo structure is mostly covered up by the loop region of $\alpha 6$ (P100–P109). In addition the unraveling of $\alpha 7\text{-}a$ (P115–E125) in the apo structure appears to cause a larger gap between helices $\alpha 7\text{-}a$ (P115–D125) and $\alpha 4$ (E54–D78) compared to the holo structure. This results in a water filled “side channel” near G124 of $\alpha 7\text{-}a$ and S70 of $\alpha 4$ that does not exist in the holo structure but leads to the deepest part of the proflavine pocket.

For our investigations, we had the option of examining the unbound monomer of the holo crystal, as this structure is more similar to the bound form. However, we thought that a more typical situation encountered in drug discovery research would be that only the fully apo crystal structure (PDB ID: 2WGB in this case) is available for analysis. Since chain B of this structure showed no electron density for the $\alpha 7\text{-}b$ helix region which comprises part of the proflavine binding site, we used chain A from 2WGB as our native apo structure in our mixed-solvent simulations, even though this presented a highly challenging case with significant helix unraveling in the $\alpha 6$ and $\alpha 7\text{-}a$ regions compared to the holo structure. Furthermore, we also decided to simulate just the monomer in our calculations instead of the functional dimer to save on computational expense.

Despite these challenges, our mixed-solvent simulations with isopropanol probe showed a high density hotspot cluster with a value of -4.7 kcal/mol, ranked third by energy criteria at the proflavin binding pocket. Interestingly, this cluster was divided into two subclusters, one showing high isopropanol probe occupancy closer to the opening of the holo binding site, and the other showing significant overlap with one of the rings in the tricyclic ligand in the deepest part of the binding pocket. This latter pocket is connected to the aforementioned side channel that only exists in the apo form between the $\alpha 7\text{-}a$ and $\alpha 4$ helices. Inspection of the MD trajectories revealed that indeed the isopropanol probe entered the pocket from this side channel. However, the opening of the proflavine pocket observed in the holo crystal structure remained occluded by the unraveled $\alpha 6$ and $\alpha 7\text{-}a$ regions in all our apo MD simulations.

Interestingly, our side chain PCA analysis indicates that our two hydrophobic mixed probe simulations (resorcinol, isopropanol) sampled pocket side chain pair distances closer to the apo state. The polar simulations (pure water, acetic acid) sampled closer to the holo state. Upon inspection of all MD trajectories, the $\alpha 6$ and $\alpha 7\text{-}a$ helices remained unwound as in the apo crystal structure throughout all the simulations. Thus, it appears that although the binding pocket in the apo form is cryptic and adopts a very different structure compared to the bound complex, the site exhibits affinity toward hydrophobic organic probes. However, within our 100 ns long simulations, the protein was not able to undergo a conformational change to adopt the final holo form even with the binding of our probe molecules. This may be due to insufficient sampling and to the small size and low complexity of the probe molecules.

A representative snapshot from the isopropanol mixed-solvent simulation (see Figure 2) indeed does not show significant changes in the vicinity of the proflavin pocket, with possibly a slight opening of the side channel between the $\alpha 7\text{-}a$ and $\alpha 4$ helices owing to the outward rotation of H67. It is

tempting to speculate that opening up of this side channel is a state along the pathway toward ligand binding.

Eg5 kinesin, also known as kinesin-5 or kinesin spindle protein (KSP), is an ATP powered molecular motor that moves along microtubules within the cell.³⁵ It is localized on mitotic spindles and spindle poles and is essential for mitosis. It is a member of a larger ATPase family of microtubule associated kinesin motor proteins that enables critical cellular functions including mitosis, meiosis, and cellular cargo transport. Inhibition of KSP has received much clinical and pharmaceutical interest in recent years as a strategy for controlling tumor growth.³⁵

Eg5 kinesin functions as a homotetramer with each monomer consisting of four functional domains: the N-terminal globular motor head, the neck linker, a coiled-coil stalk region, and the C-terminal tail. The structure of the motor head domain has been studied extensively and contains elements implicated in mechanical motion such as switch 1 (approximately T222–S235), switch 2 (approximately D265–D279), the $\alpha 4$ relay helix (D279–E304), all involved in ATP hydrolysis-induced conformational change, loop 5 (E116–G134), implicated in ADP release and inhibitor binding, and the neck linker (N358 and up to stalk domain) involved in actuating head domain movements relative to the stalk. In addition, the motor domain directly interacts with tubulin at the $\alpha 4\text{--}\alpha 5\text{--}\alpha 6$ binding face.

As our holo reference, we used the motor domain structure of Eg5 bound to the small molecule allosteric inhibitor, monastrol (PDB ID: 1Q0B).³⁶ This structure contains ADP and a Mg²⁺ ion in the nucleotide binding site. The binding site of monastrol, is distinct from the nucleotide binding site and is located approximately 12 Å away, sandwiched between Loop 5, and the $\alpha 2$ and $\alpha 3$ helices near switch 1. Monastrol inhibits the ATPase activity of Eg5 with an IC₅₀ of 6.1 μM and an estimated k_d of 1.2 μM measured through a FRET assay.³⁷ Monastrol achieves its binding affinity through several key interactions within the allosteric binding site. The phenolic hydroxyl group of the inhibitor forms a hydrogen bond with the backbone carbonyl group of E118, and one of the NH groups of the tetrahydropyrimidine-2-thione ring forms a second hydrogen bond with the backbone carbonyl group of E116. The phenol ring potentially forms an offset cation– π interaction with the guanidino group of R119 and an edge-to- π interaction with the side chain of Y211. In addition, the thione sulfur fits into a hydrophobic pocket formed by I136, L214, and the alkyl stem of E116.

The apo structure used in our MD simulations contains a nonhydrolyzable ATP analog, AMPPNP, in the nucleotide binding site (PDB ID: 3HQD).³⁸ The allosteric site in this structure is nearly completely buried due to a nearly 3 Å shift of the loop 5 main chain and also side chain rearrangements including the approximately 120 degree rotations around the $\text{C}\alpha\text{--C}\beta$ bonds of E116, W127, and Y211, and also the roughly 1.5 Å inward shifts of the side chains of R119 and L211. Moreover, the side chains of W127 and Y211 adopt the opposite relative stacking order: in the apo structure, Y211 is positioned toward the outer surface of the protein and W127 is relatively buried under loop 5, whereas in the holo structure, W127 is flipped outward, and Y211 buried (Figure 2).

This “apo” structure is quite unique in that it is trapped in the pre-ATP-hydrolyzed state through binding of AMPPNP. Therefore, this structure exhibits some differences compared to other kinesin crystal structures. For example, switch 1 adopts a

closed conformation that involves the unraveling and outward extension of the switch 1 loop that forms a short 2-stranded β sheet instead of the α 3-a helix seen in other kinesin structures. Second, likely owing to interactions with the γ -phosphate in AMPPNP, the switch 2 loop is stabilized and is completely resolved and visible, unlike other kinesin structures where this loop is unresolved. Switch 2 is presumed also to be in a closed state, and interacts with the switch 1 loop through a backbone hydrogen bond between L227 and T349 and a salt-bridge between R234 and E270. These interactions appear to alter the orientation of the α 4 relay helix and the microtubule interaction face. And finally, the α 0 helix (N27–A35) near the nucleotide binding site is shifted away from the protein core by approximately 6 Å.

Mixed-solvent simulations using a 5% resorcinol probe concentration on the apo structure resulted in partial opening of this allosteric pocket (Figure 2). Our probe density cluster analysis revealed this inhibitor pocket as the fourth top ranked with one of the high-density hotspots occupying a region overlapping with the tetrahydropyrimidine-2-dione ring of the ligand in the holo structure. The cluster of hotspots extended further toward the switch 1 end of the α 2 and α 3 helices. Side chain PCA (Figure 3) indicated that the isopropyl and resorcinol based simulations sampled inhibitor pocket side chain positions closest to the holo state while the more polar probes did not.

A representative snapshot from our mixed-solvent simulations using resorcinol probe (Figure 2) shows substantial backbone movement of Loop5. The pocket is still partially covered by Y211, while an outward flipping of W127 exposes it to the solvent probes.

Bakan and colleagues⁹ also studied this system using mixed-solvent simulations. They used a different apo protein structure (PDB ID: 1II6) with ADP and Mg²⁺ bound in the nucleotide binding site and also used multiple organic probe types in a single simulation. Their results indicate that different probes consistently found the correct allosteric pocket, but the combination was required to correctly assess the theoretical maximally achievable binding affinity of the site. In line with our findings, they also pointed out the large induced effects observed in their simulations involving up to 10 Å displacements of W127.

Cdc4 is an F-box protein that forms part of the SCF (Skp1-Cdc53/Cullin-F-box) family of protein complexes.³⁹ The SCFs constitute the largest class of E3 enzymes involved in the ubiquitin-proteasome intracellular protein degradation pathway. Together with E1 and E2, these enzymes activate and link ubiquitin proteins to their substrates, marking them for subsequent proteasomal degradation. Within this pathway, the E3 enzymes, also known as ubiquitin ligases, are involved in the recognition of phosphorylation-dependent sequence motifs called degrons in their substrates. Cdc4 consists of 8 WD40 repeat units forming a classic beta-propeller structure. The WD40 repeat is a conserved structural motif found in many proteins involved in a variety of cellular processes.⁴⁰

The holo crystal structure (PDB ID: 3MKS) contains a small molecule inhibitor, SKF-I2, bound in a well-defined allosteric pocket formed between blades 5 and 6 of the Cdc4 beta-propeller. SKF-I2 or 1-(2-carboxynaphth-1-yl)-2-napthoic acid is a biplanar atropoisomer containing two linked naphthyl bicycles and two carboxylic acids. It exhibits axial chirality and exists as a mixture of enantiomers even at high temperatures.

The apo protein structure (PDB ID: 3V7D) shows no obvious pocket in the area (Figures 1 and 2). Binding of SKF-I2 causes the main chains of blades 5 and 6 to separate, and a large shift in the loop connecting beta strands 21 and 22. In addition, H631 is buried in the apo structure, but is completely flipped into a solvent-exposed orientation in the holo structure. One of the naphthyl rings of SKF-I2 makes hydrophobic contacts with I594, L628, L634, A649, and W657 and one of the carboxylic groups forms a hydrogen bond with the side chain of W657 and salt-bridges with R664. The other carboxyl group interacts with H631 and R655. In addition, van der Waals and stacking interactions are observed between the other naphthyl group and R664 and S667.

In our mixed-solvent simulations, we found that only the use of the acetic acid probe showed high-density hotspot clusters near the allosteric binding site. This is somewhat expected given the importance of the carboxylic acids in the binding interactions of SKF-I2. The hotspot cluster energy was -7.5 kcal/mol and was ranked second among the clusters found on the surface. However, our acetic acid probes were able to only open up the induced pocket partially (Figure 2) and, for example, did not cause H631 to flip out into its holo crystal structure position. This may be in part due to the lack of an aromatic ring in our probe which may be required for full induced pocket opening. In addition, the large energy barrier expected between the two histidine conformations suggests that observing such a flip likely would require much longer simulation time scales.

Interestingly, the highest ranked occupancy cluster from our acetic acid simulations was located on a protein surface patch within the interior of the beta-propeller structure, directly opposite the cryptic allosteric site located on the external side. Although the significance of this cluster is unknown, one might speculate that induced fit binding at the allosteric site may require complex concerted ligand interactions from the interior side of the beta-propeller as well.

A representative snapshot from the acetic acid probe (Figure 2) shows significant opening of the pocket due to backbone movements of Q629, G630, and H631, although H631 did not flip outward as in the holo structure. We also observed that the amide bond between Q629 and G630 is in a cis conformation in the holo structure, whereas it remained in its trans conformation throughout the mixed-solvent simulation as in the apo structure. Since the energetic barrier to this transition is quite large, we do not expect to observe it within the 100 ns simulation. Nevertheless, it appears that our mixed-solvent probe induced the pocket to an intermediate state along the way to the holo conformation.

PCA analysis of the four simulations with different probes shows that the isopropanol and resorcinol probes enabled sampling closest to the holo structure and the acetic acid probe appeared to sample wells that were the most distant from it. This, together with the hotspot cluster distribution result above suggests that the carboxylic probes likely function to initially steer the molecule to the active site, while the naphthyl group in the ligand likely is required for the subsequent induced fit. It is also interesting to note that among all the targets examined in this study, Cdc4 showed among the smallest amount of variance explained (44.9%) using the top two PCs, suggesting higher complexity in the induced pocket dynamics relative to the other systems.

P38 α is a mitogen activated protein kinase (MAPK), a class of serine-threonine kinases involved in extracellular signal

transduction and a variety of downstream cellular processes including cell proliferation, differentiation, inflammation, and apoptosis.⁴¹ P38 α is primarily activated by dual phosphorylation of the activation loop by the MAPK kinases (MKKs). In recent years, three other activation modes for p38 have been identified, and one of them involves a potential lipid binding site, known as the MAPK insert (MKI) pocket on the C-terminal lobe, distal to the N-terminal catalytic nucleotide binding site.^{42,43}

Figure 1 shows the MKI site in the apo and holo forms. Note that the holo structure (PDB ID: 3HL7) contains the inhibitor SD-0006 bound to the nucleotide binding site and a diaryl pyrazole compound (2-fluoro-4-[4-(4-fluorophenyl)-1*H*-pyrazol-3-yl]pyridine) in the MKI site, while the apo structure (PDB ID: 4ESB) contains no ligands in either the nucleotide binding or MKI sites. The apo form has the MKI pocket mostly buried by the backbone shift of the loop region between M194 and Y200 and the inward rotation of W197.

The cryptic MKI site was revealed as the top ranked hotspot cluster with the isopropanol probe, with an energy of -6.0 kcal/mol. Interestingly, the orthosteric nucleotide binding site was ranked second with a comparable energy value of -5.9 kcal/mol. This appears to reveal an equal preference for the isopropanol probe to bind to either the allosteric or nucleotide pocket. A representative snapshot from the mixed-solvent simulation using isopropanol probe (Figure 2) shows substantial pocket opening at the MKI site due to significant backbone and side chain movements of W197, H199, and S252. Although the side chains of W197 and H199 did not completely open up, their respective holo conformations appear to be achievable with simple rotation around their $C\alpha-C\beta$ bonds.

The PCA analysis of the MKI pocket side chains reveal that the isopropanol/water mixture simulation brought the sampled side chain distribution closer to the holo structure with respect to the other mixtures. The pure water and the acetic acid mixture sampled the free energy landscape in a similar fashion. The resorcinol trajectory, while resembling to some extent the isopropanol one, never really crossed paths with the holo reference. Because the putative endogenous binding ligand for this site is a phosphatidyl inositol ether with a long lipid tail and the holo structure we used for reference contained a diaryl pyrazole compound, pocket opening likely requires ligands with higher complexity and size compared to our cosolvent probes.

Detection of Cryptic Sites via SiteMap. We asked whether less computationally intensive methods could be used to find cryptic sites with a comparable degree of reliability. To that end, we ran a comprehensive SiteMap analysis^{44,45} on the studied protein data set. Results of the analysis are summarized in Table 2. As in our mixed-solvent simulations, we ran the calculations on the apo structures. By using standard settings, the protocol was able to correctly identify and rank within the top five solutions, the investigated cryptic site two times out of eight. We note that the two structures for which SiteMap successfully found the correct binding sites had apo binding pocket that were partially pre-existing and thus were not fully “cryptic”. In particular, for Staphylococcal nuclease, the cryptic site was ranked number 1 in the SiteMap analysis, although with a fairly low druggability score. On the other hand, for the remaining six systems, the protocol was unable to pinpoint the cryptic sites. To fairly assess the mixed-solvent outcomes and also to assess the potential of using the two methods in conjunction, we ran SiteMap analysis on the representative

Table 2. Results of the SiteMap Analysis Performed on the Apo Structure and on the Corresponding Mixed-Solvent-Generated Representative Structure^a

system	apo (rank, Dscore) ^b	mixed-solvent (rank, Dscore)
exonuclease I ^b	NF	9, 0.81
Niemann-Pick C2 protein	NF	1, 0.93
Staphylococcal nuclease	1, 0.81	1, 0.80
toluene 4-monoxygenase	NF	2, 0.74
TETR-like transcriptional regulator LFRR	NF	1, 1.02
kinesin Eg5	NF	1, 0.75
Cdc4 ^c	NF	13, 0.32
P38 α	5, 0.78	1, 1.78

^aUnless otherwise noted, default SiteMap settings were used to run the calculations. ^bSearch extended to the top 20 sites. ^cSearch extended to the top 20 sites and at least 10 points required to define a site. ^dNF = not found.

snapshots shown in Figure 2 extracted from the mixed-solvent simulation trajectories. In all the cases the protocol improved or, like in the above-mentioned case did not alter substantially the SiteMap results. In six out of eight cases, the cryptic sites were correctly identified and ranked as either top or second best. Two systems proved to be particularly challenging, namely Exonuclease I and Cdc4; for those systems, more relaxed criteria (see Methods for details) were needed to identify the low ranking cryptic sites. Although the purpose of this analysis was not to compare the performance between different methods, we note how representative snapshots derived from our mixed-solvent simulations scored exceptionally well using SiteMap. SiteMap is only one particular method within the large class of static methods for binding site detection, nonetheless these results suggest that coupling MD based methods with such less demanding pocket detection protocols may help streamline the systematic identification of druggable sites. It is expected that further analysis of the protein dynamics and of the interactions between probes and structures could offer more insights on the subtleties involved in ligand recognition and conformational transitions.

CONCLUSIONS

Mixed-solvent molecular dynamics simulations have emerged in the past years as a powerful tool to assess the druggability of proteins.^{10,46–48} The atomic-level insights and speed of this approach make its use amenable at the very early stage of drug discovery projects, when limited investments are desirable and more time demanding experimental options⁴⁹ are not viable. Building on the work of Barril, Bakan, Carlson, and others,^{9,10,13,17,50} we have examined in this paper the potential of mixed-solvent simulations as a tool to systematically induce biologically relevant conformational transitions in a practical time frame. To test this hypothesis, we employed a data set extracted from a recently published work,⁸ consisting of systems that are reported to undergo structural adaptations only in the presence of a binding counterpart. The studied data set was structurally diversified and posed ascending degrees of challenges that spanned from simple side-chain movements to secondary structure rearrangements and from orthosteric to allosteric sites. By using only three molecular probes, we demonstrated that encouraging results could be observed in the majority of cases. In particular, for all of our systems but two, the location of the cryptic site was nicely recapitulated and

ranked among the top three solutions, although, at times, some degree of manual intervention was necessary to combine neighboring clusters that together defined the site. It is worth emphasizing that cryptic sites are commonly present on proteins with other distinct functional areas. Hence, a cryptic site identified through this approach should not necessarily be the top ranked site as scored by approximate probe binding affinity or other theoretical measurements. As a general guideline, hotspot analyses should be limited to the top 5–7 clusters, unless calculated energy values or any other biology background knowledge call for a broader evaluation. In many cases, additional trajectory analysis clearly showed that the presence of small organic solvent molecules was instrumental to explore relevant regions of the conformational landscape that would otherwise have been overlooked. In real-world applications, this kind of outcome would at least prompt deeper investigation.

Discerning true cryptic sites within the identified set of patches is a great hurdle for this class of methods. To that end, we have explored the possibility of addressing the issue using mixed-solvent simulation in conjunction with less computationally intensive methods that rely on static structures, such as SiteMap. The outcome appears very promising for prospective applications, as it offers a practical and systematic way to highlight hot-spots of interest that were initially hidden. As such, rather than looking at absolute druggability values, we were more interested in the difference between initial and probe-induced structures' results. In our hands, apart from a couple of very challenging cases (i.e., Exonuclease I and Cdc4), SiteMap was able to rank as top (or in one case, second best) druggable site the cryptic one. In all the cases but one (i.e., Staphylococcal nuclease) the analysis run on the probe-induced structure considerably improved the druggability scores. Such analysis could be automated to map pocket openings throughout the trajectory. Furthermore, one can always extend the investigation to literature and alternative methods that rely on sequence and folding conservation to mark, among the top solutions, sites that are already known or have enhanced probability to play a biological role (i.e., naturally engineered to interact with partnering ligands). While experimental assessment of the sites' druggability would be the ultimate way to assess druggability methods' predictive power, in Table S1 we highlight the biological relevance (if available) of the top six ranked hot-spots for each investigated system with the consideration that those sites with no known relevance might be true positive as well.

The use of complementary methods and exploratory parameters (i.e., increased simulation time, probe diversification and concentration, accelerated sampling, etc.) in order to get better agreement with experimental data was beyond the scope of this paper. In this context, the analysis of the simulations offers semiquantitative insights on the studied spots. Taking into account the transient nature of the cryptic sites and the manageable selection of probes and simulation time employed (arbitrarily set to 100 ns), we did not anticipate to achieve full convergence. We've included in this manuscript results obtained using a value of -1.5 kcal/mol for the clustering step. Indeed, using the -1.0 kcal/mol cutoff recommended in the Bakan et al. work in our case resulted in scattered and fuzzy clustering. This discrepancy could be ascribed partly to the use of a different force field. The overall importance of the energy cutoff for clustering along with the intrinsic limitations of a grid-based approach has still to be investigated thoroughly and

might be subject of a future work. It has to be highlighted that the selection of probes here proposed is not meant to be representative of a bigger set and is not expected to outperform other choices of small organic solvent in the majority of cases.^{9,10,50} In principle, one could use any organic solvent to probe the system's plasticity, as long as it is miscible with water.⁵¹ In a few cases, the pure water simulations alone already offered glimpses of the cryptic sites (although within a limited time frame). Nonetheless, the results highlight how only the employment of organic probes (acting as "catalysts") effectively enabled the exploration of regions of the free energy landscape distinct from the apo and closer to the holo structures for a substantial amount of time in several cases.

In contrast to the work of others,⁹ we've decided to rely on binary systems only (i.e., water + one probe) in order to simplify the final analysis and to avoid the risk of solvent–solvent interactions that could have masked some functional groups and altered the outcome. Along this reasoning, we also chose to analyze the systems separately without merging the hotspots originating from different probes.

To the best of our knowledge, accounting for a recent work on allosteric modulation,¹⁷ this is among the first attempts to utilize mixed-solvent simulations for the identification of cryptic sites. As such, it paves the way to explore the protocol systematically in combination with complementary/synergistic options such as general 3D methods,^{44,45,52–57} bioinformatics tools,^{8,58–61} and enhanced sampling techniques,^{62–64} just to name a few. It is worth noting that, very recently, Oleinikovas et al. have reported about the use of enhanced sampling simulations to understand cryptic pocket formation.²² The authors tested a methodology that relies on the combination of a Hamiltonian replica exchange method (termed SWISH) and mixed-solvent simulations on four biological systems. The work highlights all the intrinsic limitations of standard methods (such as brute force MD simulations) in inducing cryptic pockets' opening for a significant amount of time. Surprisingly enough, even parallel tempering⁶⁵ was not a viable way to trigger the naturally occurring conformational changes. The authors nicely offered an energy-based explanation for this and proposed a protocol that promotes the initial exposure of the hidden site (SWISH) followed by its ligand-induced stabilization (via mixed-solvent simulations). In our simulations, in concordance with some of their conclusions, while clear signals were registered at the cryptic sites, not always the opening of the pocket was complete. Moreover, while the PCA analysis offered some insights on the proximity to the holo reference, a more detailed RMSD analysis on the pocket of interest (Table S2) with respect to the holo structures, also revealed some of the limitation of the employed sampling time. While in most cases, the mixed-solvent induced structures were sensibly closer to the holo references than the starting apo structure there is clearly room for improvement. Our findings are generally in agreement with the ones illustrated in the work of Oleinikovas et al. and converge on a key role of induced-fit effects which might be especially relevant in the case of allosteric modulations where saturation phenomena might take place.⁶⁶

Finally, from a pragmatic perspective, the protocol offers the possibility to easily extract the probes' configuration in the framework of the studied systems, thus promptly enabling the rational structure-assisted design of suitable modulators in the absence of holo-like experimental structures.

METHODS

Preparation of Crystal Structures. We selected structure pairs from a comprehensive cryptic binding site database curated by Cimermancic et al.⁸ The database consists of apo/ holo pairs of protein structures exhibiting conformational differences at ligand binding sites. These sites are typically not the endogenous ligand binding site but are often protein–protein interaction interfaces or allosteric sites. We visually inspected the entries in the database, analyzed B-factors and electron densities to avoid ambiguous atomic positions, and chose eight apo/ holo pairs (Table 1) exhibiting varying levels of conformational change upon ligand binding.

Coordinates for all protein structures were downloaded from the PDB²¹ (www.wwpdb.org) and prepared using the Schrödinger Maestro software.⁶⁷ Hydrogen positions, protonation states of histidines, and conformational flips of asparagine and glutamine side chains were optimized using the default protocols available in the Maestro Protein Preparation Wizard.^{68,69} Missing side chains were built using the Schrödinger Prime package using default settings.⁷⁰

Probe Molecules. Three organic molecules fully miscible with water (namely, acetic acid, resorcinol, and isopropyl alcohol) were employed in this study. Values for the pair correlation function $g(r)$ at a distance of 10 Å were 0.97, 1.01, and 1.10 for acetic acid, isopropyl alcohol, and resorcinol, respectively.

Initial Coordinates for Mixed-Solvent Systems. We performed molecular dynamics simulations with the apo form of the protein immersed in an explicit binary water–probe mixture. Only one probe type was used per simulation. As a control we also ran simulations using pure water solvent.

Coordinates for the apo protein solvated in pure explicit water were first prepared using Schrödinger Maestro. The protein was placed in an explicit water box allowing for a 10 Å buffer around the protein. Overlapping water molecules were removed and neutralizing Na^+ or Cl^- ions were placed in the system using default Maestro protocols. Starting with the protein–water system, initial coordinates for the binary solvent–probe system were generated by randomly replacing water molecules within the pure-water solvated system, while ensuring a certain buffer distance from the protein and other probe molecules. This was done to avoid the occurrence of deeply buried organic molecules that could have affected the protein folds and/or lead to a decrease in the cosolvent effective concentration. The probe molecules were randomly oriented, and any overlapping water molecules were deleted. This process was repeated until the desired concentration was reached. In all simulations, we used a 5% probe:water molar ratio. A Python GUI panel in the Maestro interface was created to prepare systems solvated in binary probe–water mixtures for this purpose and is available free of charge from the authors of this paper.

Molecular Dynamics Simulation Details. All simulations were conducted through the GPU-enabled Desmond molecular dynamics software from D.E. Shaw Research using the OPLS3 force field.^{71,72} Simulation parameters and protocols are summarized in Table 3.

Analysis of MD Trajectories. Root mean square deviations (RMSD) from the initial crystal structure coordinates were computed using the Simulation Event Analysis tool available in Schrödinger Maestro package.

Table 3. Molecular Dynamics Simulation Parameters and Protocols

simulation ensemble	NPT
production simulation time	100 ns
timestep: bonded, near, far	2 fs, 2 fs, 6 fs
thermostat	Nose–Hoover chain
barostat	Martyna–Tobias–Klein
temperature	300 K
pressure	1 atm
heating and equilibration protocol	100 ps, $T = 10$ K, Brownian dynamics, heavy atoms restrained 12 ps, NVT MD, $T = 10$ K, heavy atoms restrained 12 ps, NPT MD, $T = 10$ K, heavy atoms restrained 12 ps, NPT MD, $T = 300$ K, heavy atoms restrained 24 ps, NPT MD, $T = 300$ K, no restraints
hardware	NVIDIA GTX780

In addition, a new Desmond trajectory analysis program was written to perform a probe grid-density analysis based on an algorithm described by Bakan et al.⁹ Briefly, probe centroid occupancies within 0.5 Å cubic grid cells were calculated from the MD trajectories, and smoothing was applied by averaging each occupancy value with its 26 neighbor cells. The smoothed grid occupancies were then converted into free energies using the formula $\Delta G = -kT \ln(N_i/N_o)$ where N_i is the occupancy of the grid cell and N_o is the expected grid occupancy in the bulk solvent. N_o values were computed by running MD simulations of mixed-solvents at 5% molar ratio without any protein involved. Next, the grid was filtered by removing from consideration cells with free energies greater than a certain free energy cutoff (-1.5 kcal/mol) and further filtered by removing all hotspots within the radius of the probe compound used around each local energy minimum hotspot. Finally, the filtered hotspots were clustered by starting from the lowest energy hotspot, merging the next lowest energy hotspot to the cluster if that hotspot was within 6.2 Å (as defined in Bakan et al.)⁹ and repeating this procedure until there were no more hotspots to merge. For the probe occupancy hotspot analysis, we used a cutoff energy of -1.5 kcal/mol rather than -1.0 kcal/mol as stipulated by Bakan et al. This was because we observed that a -1.0 kcal/mol cutoff resulted in excessively large clusters, which we speculate is due to use of a different force-field (OPLS3 in this study rather than CHARMM in Bakan et al.). Occasionally, we found that a few hotspot clusters were found near the induced-fit pocket and that adjusting the free energy cutoff used resulted in the merging of such clusters. In such cases, we did not attempt to arbitrarily fine-tune the free energy cutoff, but rather manually merged these neighboring clusters, and added their corresponding energies to obtain the total energy for the merged cluster. A Python GUI panel in the Maestro interface was created to analyze the trajectories and is available free of charge from the authors of this paper.

To visualize the induced-fit effects, we extracted frames for which probes were present at high occupancy hotspots near the induced-fit site. We manually chose frames for which the pocket opening was most apparent as representatives.

The principal component analysis (PCA) was performed on each simulation trajectory as follows. Residues with at least one atom within 4 Å of the bound ligand in each holo structure

were identified, and the closest distances between all N pairs of these residues were measured for each frame in the apo mixed-solvent simulations. These distances were used to define N -dimensional feature vectors for each saved frame. PCA analysis was performed on these N -dimensional feature vectors pooling the four simulations with different solvent types used for each target. For visualization, we plotted the first two PC values for each frame from the MD trajectories, and next used a Gaussian kernel-based density estimator to obtain a smooth density distribution for each solvent type used per target. These were computed using the Python Scipy package,⁷³ and Scott's Rule was used to determine the Gaussian bandwidth.⁷⁴ In addition, feature vectors corresponding to the apo and holo crystal structures were also computed and their locations relative to the first two PCs are shown on the same plot.

SiteMap Druggability Assessment. SiteMap analyses were performed using standard settings as available in the 2016-3 version of the Schrödinger suite.^{44,45,75} In particular, (a) the analyses were not restricted to a specific binding site region, (b) at least 15 points were required to define a site, and (c) the top five sites were reported. Moreover, a more restrictive definition of hydrophobicity and standard grids were used. The site maps were cropped at 4 Å from the nearest site point and the “detect shallow binding sites” options was unchecked. In 2 cases (see notes in Table 2), the analyses were modified to include the top 20 sites and to require at least 10 points to define a site.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jcim.6b00623](https://doi.org/10.1021/acs.jcim.6b00623).

Root mean square deviation analyses for all the investigated systems as well as the remaining PCA plots ([PDF](#))

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: +1-317-9670830. E-mail: favia_angelo@lilly.com, ad.favia@gmail.com (A.D.F.).

*Phone: +81-3-6860-8316. Fax: +81-3-6273-4722. E-mail: leadingelement@gmail.com (S.R.K.).

ORCID

Angelo D. Favia: [0000-0003-3013-2912](https://orcid.org/0000-0003-3013-2912)

Present Addresses

[†]S.R.K: Modulus Discovery, Inc., Resona Kudan Building 5F KS Floor, 1-5-6 Kudan-Minami, Chiyoda-ku, Tokyo, Japan.

[‡]W.S.: Silicon Therapeutics, 300 A St., Boston, MA 02210.

[§]A.D.F.: Elanco, 2500 Innovation way, Greenfield, IN 46140.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors wish to thank Phani Ghanakota and Jim Durbin for valuable recommendations and Davide Branduardi for useful discussion and revision of the manuscript.

■ ABBREVIATIONS

MSS, mixed-solvent simulations; MD, molecular dynamics; PCA, principal component analysis; RMSD, root mean square deviation

■ REFERENCES

- Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. Protein Clefts in Molecular Recognition and Function. *Protein Sci.* **1996**, *5*, 2438–2452.
- 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.
- Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71.
- Diskin, R.; Engelberg, D.; Livnah, O. A Novel Lipid Binding Site Formed by the Map Kinase Insert in P38 Alpha. *J. Mol. Biol.* **2008**, *375*, 70–79.
- Johnson, D. K.; Karanicolas, J. Selectivity by Small-Molecule Inhibitors of Protein Interactions Can Be Driven by Protein Surface Fluctuations. *PLoS Comput. Biol.* **2015**, *11*, e1004081.
- Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S. Discovery of Multiple Hidden Allosteric Sites by Combining Markov State Models and Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2734–2739.
- Johnson, D. K.; Karanicolas, J. Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation Than the Rest of the Protein Surface. *PLoS Comput. Biol.* **2013**, *9*, e1002951.
- 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.
- 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.
- Seco, J.; Luque, F. J.; Barril, X. Binding Site Detection and Druggability Index from First Principles. *J. Med. Chem.* **2009**, *52*, 2363–2371.
- Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53*, 5858–5867.
- Kozakov, D.; Hall, D. R.; Napoleon, R. L.; Yueh, C.; Whitty, A.; Vajda, S. New Frontiers in Druggability. *J. Med. Chem.* **2015**, *58*, 9063–9088.
- Basse, N.; Kaar, J. L.; Settanni, G.; Joerger, A. C.; Rutherford, T. J.; Fersht, A. R. Toward the Rational Design of P53-Stabilizing Drugs: Probing the Surface of the Oncogenic Y220c Mutant. *Chem. Biol.* **2010**, *17*, 46–56.
- Yang, C. Y.; Wang, S. Analysis of Flexibility and Hotspots in Bcl-XL and Mcl-1 Proteins for the Design of Selective Small-Molecule Inhibitors. *ACS Med. Chem. Lett.* **2012**, *3*, 308–312.
- Yang, C. Y.; Wang, S. Computational Analysis of Protein Hotspots. *ACS Med. Chem. Lett.* **2010**, *1*, 125–129.
- Ghanakota, P.; Carlson, H. A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. *J. Med. Chem.* **2016**, *59*, 10383–10399.
- Ghanakota, P.; Carlson, H. A. Moving Beyond Active-Site Detection: Mixmd Applied to Allosteric Systems. *J. Phys. Chem. B* **2016**, *120*, 8685–8695.
- Ung, P. M.; Ghanakota, P.; Graham, S. E.; Lexa, K. W.; Carlson, H. A. Identifying Binding Hot Spots on Protein Surfaces by Mixed-Solvent Molecular Dynamics: Hiv-1 Protease as a Test Case. *Biopolymers* **2016**, *105*, 21–34.
- Alvarez-Garcia, D.; Barril, X. Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites. *J. Med. Chem.* **2014**, *57*, 8530–8539.
- Lexa, K. W.; Carlson, H. A. Improving Protocols for Protein Mapping through Proper Comparison to Crystallography Data. *J. Chem. Inf. Model.* **2013**, *53*, 391–402.
- Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980.

- (22) 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.
- (23) Matthews, B. W.; Breyer, W. A. Structure of Escherichia Coli Exonuclease I Suggests How Processivity Is Achieved. *Nat. Struct. Biol.* **2000**, *7*, 1125–1128.
- (24) Lu, D.; Bernstein, D. A.; Satyshur, K. A.; Keck, J. L. Small-Molecule Tools for Dissecting the Roles of Ssb/Protein Interactions in Genome Maintenance. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 633–638.
- (25) Ma, X.; Meng, H.; Lai, L. Motions of Allosteric and Orthosteric Ligand-Binding Sites in Proteins Are Highly Correlated. *J. Chem. Inf. Model.* **2016**, *56*, 1725–1733.
- (26) Dixit, S. S.; Sleat, D. E.; Stock, A. M.; Lobel, P. Do Mammalian Npc1 and Npc2 Play a Role in Intestinal Cholesterol Absorption? *Biochem. J.* **2007**, *408*, 1–5.
- (27) Xu, S.; Benoff, B.; Liou, H. L.; Lobel, P.; Stock, A. M. Structural Basis of Sterol Binding by Npc2, a Lysosomal Protein Deficient in Niemann-Pick Type C2 Disease. *J. Biol. Chem.* **2007**, *282*, 23525–23531.
- (28) Friedland, N.; Liou, H. L.; Lobel, P.; Stock, A. M. Structure of a Cholesterol-Binding Protein Deficient in Niemann-Pick Type C2 Disease. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2512–2517.
- (29) Cotton, F. A.; Hazen, E. E., Jr.; Legg, M. J. Staphylococcal Nuclease: Proposed Mechanism of Action Based on Structure of Enzyme-Thymidine 3',5'-Bisphosphate-Calcium Ion Complex at 1.5- α Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 2551–2555.
- (30) Nguyen, D. M.; Leila Reynald, R.; Gittis, A. G.; Lattman, E. E. X-Ray and Thermodynamic Studies of Staphylococcal Nuclease Variants I92e and I92k: Insights into Polarity of the Protein Interior. *J. Mol. Biol.* **2004**, *341*, 565–574.
- (31) Bailey, L. J.; McCoy, J. G.; Phillips, G. N., Jr.; Fox, B. G. Structural Consequences of Effector Protein Complex Formation in a Diiron Hydroxylase. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19194–19198.
- (32) Lountos, G. T.; Mitchell, K. H.; Studts, J. M.; Fox, B. G.; Orville, A. M. Crystal Structures and Functional Studies of T4mod, the Toluene 4-Monoxygenase Catalytic Effector Protein. *Biochemistry* **2005**, *44*, 7131–7142.
- (33) Buroni, S.; Manina, G.; Guglierame, P.; Pasca, M. R.; Riccardi, G.; De Rossi, E. Lfrr Is a Repressor That Regulates Expression of the Efflux Pump Lfra in Mycobacterium Smegmatis. *Antimicrob. Agents Chemother.* **2006**, *50*, 4044–4052.
- (34) Bellinzoni, M.; Buroni, S.; Schaeffer, F.; Riccardi, G.; De Rossi, E.; Alzari, P. M. Structural Plasticity and Distinct Drug-Binding Modes of Lfrr, a Mycobacterial Efflux Pump Regulator. *J. Bacteriol.* **2009**, *191*, 7531–7537.
- (35) Wojcik, E. J.; Buckley, R. S.; Richard, J.; Liu, L.; Huckaba, T. M.; Kim, S. Kinesin-5: Cross-Bridging Mechanism to Targeted Clinical Therapy. *Gene* **2013**, *531*, 133–149.
- (36) Yan, Y.; Sardana, V.; Xu, B.; Homnick, C.; Halchenko, W.; Buser, C. A.; Schaber, M.; Hartman, G. D.; Huber, H. E.; Kuo, L. C. Inhibition of a Mitotic Motor Protein: Where, How, and Conformational Consequences. *J. Mol. Biol.* **2004**, *335*, 547–554.
- (37) Maliga, Z.; Kapoor, T. M.; Mitchison, T. J. Evidence That Monastrol Is an Allosteric Inhibitor of the Mitotic Kinesin Eg5. *Chem. Biol.* **2002**, *9*, 989–996.
- (38) Parke, C. L.; Wojcik, E. J.; Kim, S.; Worthy lake, D. K. Atp Hydrolysis in Eg5 Kinesin Involves a Catalytic Two-Water Mechanism. *J. Biol. Chem.* **2010**, *285*, 5859–5867.
- (39) Tang, X.; Orlicky, S.; Mittag, T.; Csizmok, V.; Pawson, T.; Forman-Kay, J. D.; Sicheri, F.; Tyers, M. Composite Low Affinity Interactions Dictate Recognition of the Cyclin-Dependent Kinase Inhibitor Sic1 by the Scfcdc4 Ubiquitin Ligase. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3287–3292.
- (40) Smith, T. F.; Gaitatzes, C.; Saxena, K.; Neer, E. J. The Wd Repeat: A Common Architecture for Diverse Functions. *Trends Biochem. Sci.* **1999**, *24*, 181–185.
- (41) Eckert, R. L.; Efimova, T.; Balasubramanian, S.; Crish, J. F.; Bone, F.; Dashti, S. P38 Mitogen-Activated Protein Kinases on the Body Surface—a Function for P38 Delta. *J. Invest. Dermatol.* **2003**, *120*, 823–828.
- (42) Tzarum, N.; Eisenberg-Domovich, Y.; Gills, J.; Dennis, P.; Livnah, O. Lipid Molecules Induce P38 α Activation Via a Novel Molecular Switch. *J. Mol. Biol.* **2012**, *424*, 339–353.
- (43) Beenstock, J.; Ben-Yehuda, S.; Melamed, D.; Admon, A.; Livnah, O.; Ahn, N. G.; Engelberg, D. The P38beta Mitogen-Activated Protein Kinase Possesses an Intrinsic Autophosphorylation Activity, Generated by a Short Region Composed of the Alpha-G Helix and Mapk Insert. *J. Biol. Chem.* **2014**, *289*, 23546–23556.
- (44) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146–148.
- (45) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- (46) Guvench, O.; MacKerell, A. D., Jr. Computational Fragment-Based Binding Site Identification by Ligand Competitive Saturation. *PLoS Comput. Biol.* **2009**, *5*, e1000435.
- (47) 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.
- (48) Arcon, J. P.; Defelipe, L. A.; Modenutti, C. P.; Lopez, E. D.; Alvarez-Garcia, D.; Barril, X.; Turjanski, A. G.; Marti, M. A. Molecular Dynamics in Mixed Solvents Reveals Protein-Ligand Interactions, Improves Docking, and Allows Accurate Binding Free Energy Predictions. *J. Chem. Inf. Model.* **2017**, *57*, 846.
- (49) 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.
- (50) 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.
- (51) Lexa, K. W.; Goh, G. B.; Carlson, H. A. Parameter Choice Matters: Validating Probe Parameters for Use in Mixed-Solvent Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 2190–2199.
- (52) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (53) van Westen, G. J.; Wegner, J. K.; Bender, A.; Ijzerman, A. P.; van Vlijmen, H. W. Mining Protein Dynamics from Sets of Crystal Structures Using "Consensus Structures". *Protein Sci.* **2010**, *19*, 742–752.
- (54) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic Analysis of Water Molecules at the Surface of Proteins and Applications to Binding Site Prediction and Characterization. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 871–883.
- (55) 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.
- (56) Benson, M. L.; Smith, R. D.; Khazanov, N. A.; Dimcheff, B.; Beaver, J.; Dresslar, P.; Neroth, J.; Carlson, H. A. Binding Moad, a High-Quality Protein-Ligand Database. *Nucleic Acids Res.* **2007**, *36*, D674–678.
- (57) Schuyler, A. D.; Carlson, H. A.; Feldman, E. L. Computational Methods for Predicting Sites of Functionally Important Dynamics. *J. Phys. Chem. B* **2009**, *113*, 6613–6622.
- (58) Chartier, M.; Adriansen, E.; Najmanovich, R. Isomif Finder: Online Detection of Binding Site Molecular Interaction Field Similarities. *Bioinformatics* **2016**, *32*, 621–623.
- (59) Chartier, M.; Najmanovich, R. Detection of Binding Site Molecular Interaction Field Similarities. *J. Chem. Inf. Model.* **2015**, *55*, 1600–1615.
- (60) Glaser, F.; Morris, R. J.; Najmanovich, R. J.; Laskowski, R. A.; Thornton, J. M. A Method for Localizing Ligand Binding Pockets in Protein Structures. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 479–488.

- (61) Khazanov, N. A.; Carlson, H. A. Exploring the Composition of Protein-Ligand Binding Sites on a Large Scale. *PLoS Comput. Biol.* **2013**, *9*, e1003321.
- (62) Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the Density of States by History-Dependent Metadynamics. *Phys. Rev. Lett.* **2004**, *92*, 170601.
- (63) Swendsen, R. H.; Wang, J. S. Replica Monte Carlo Simulation of Spin Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
- (64) 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.
- (65) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (66) van Westen, G. J.; Gaulton, A.; Overington, J. P. Chemical, Target, and Bioactive Properties of Allosteric Modulation. *PLoS Comput. Biol.* **2014**, *10*, e1003559.
- (67) Maestro, version 10.6; Schrödinger, LLC: New York, NY, 2016.
- (68) Protein Preparation Wizard, suite 2016-2; Schrödinger, LLC: New York, NY, 2016.
- (69) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234.
- (70) Prime, version 4.4; Schrödinger, LLC: New York, NY, 2016.
- (71) Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kollossvary, I.; Moraes, M.; Sacerdoti, F.; Salmon, J.; Shan, Y.; Shaw, D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing 2006*, 43.
- (72) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. Opls3: A Force Field Providing Broad Coverage of Drug-Like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (73) Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. The Numpy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.
- (74) Scott, D. W. On Optimal and Data-Based Histograms. *Biometrika* **1979**, *66*, 605–610.
- (75) Sitemap, version 4.0.014; Schrödinger, LLC: New York, NY, 2016.