JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

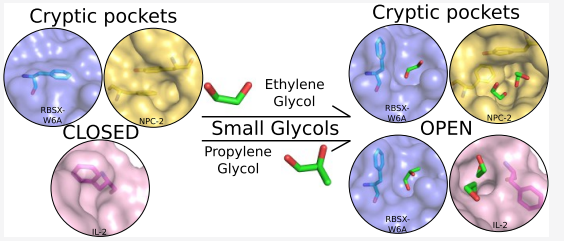
Small Glycols Discover Cryptic Pockets on Proteins for Fragment-Based Approaches

Harsh Bansia,* Pranjal Mahanta, Neela H. Yennawar, and Suryanarayanarao Ramakumar*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c01126 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cryptic pockets are visible in ligand-bound protein structures but are occluded in unbound structures. Utilizing these pockets in fragment-based drug-design provides an attractive option for proteins not tractable by classical binding sites. However, owing to their hidden nature, they are difficult to identify. Here, we show that small glycols find cryptic pockets on a diverse set of proteins. Initial crystallography experiments serendipitously revealed the ability of ethylene glycol, a small glycol, to identify a cryptic pocket on the W6A mutant of the RBSX protein (RBSX-W6A). Explicit-solvent molecular dynamics (MD) simulations of RBSX-W6A with the exposed state of the cryptic pocket (ethylene glycol removed) revealed closure of the pocket reiterating that the exposed state of cryptic pockets in general are unstable in the absence of ligands. Also, no change in the pocket was observed for simulations of RBSX-W6A with the occluded state of the cryptic pocket, suggesting that water molecules are not able to open the cryptic pocket. "Cryptic-pocket finding" potential of small glycols was then supported and generalized through additional crystallography experiments, explicit-cosolvent MD simulations, and protein data set construction and analysis. The cryptic pocket on RBSX-W6A was found again upon repeating the crystallography experiments with another small glycol, propylene glycol. Use of ethylene glycol as a probe molecule in cosolvent MD simulations led to the enhanced sampling of the exposed state of experimentally observed cryptic sites on a test set of two proteins (Niemann-Pick C2, Interleukin-2). Further, analyses of protein structures with validated cryptic sites showed that ethylene glycol molecules bind to sites on proteins (Bcl-xL, G-actin, myosin II, and glutamate receptor 2), which become apparent upon binding of biologically relevant ligands. Our study thus suggests potential application of the small glycols in experimental and computational fragment-based approaches to identify cryptic pockets in apparently undruggable and/or difficult targets, making these proteins amenable to drug-design strategies.



INTRODUCTION

Use of structural information concerning binding site(s) on validated protein targets[1] is often the starting point of any drug-design process. However, many of the validated targets are not easily druggable owing to the presence of undesirable traits in them such as featureless binding sites and the lack of complementary hydrogen bonding partners.[2,3] Adding to this notion of undruggability is the observation that many pharmaceutically important targets have cryptic binding sites,[4,5] which are occluded in unbound proteins but become apparent in ligand-bound structures[6,7] thereby making structural information available only on the unbound protein inadequate for drug-design purposes. Nonetheless, finding and targeting cryptic binding sites presents an attractive opportunity for many of the targets that are not tractable by traditional drug-design strategies, thereby expanding the druggable genome.[8] For years, efforts to develop inhibitors against K-RAS, an oncogene mutated in human cancers, were unsuccessful until site-directed tethering identified a new cryptic allosteric pocket that was exploited to successfully target K-RAS,[4,9] thus emphasizing the importance of finding cryptic sites in the course of drug discovery. Consequently, identifying and utilizing cryptic binding sites for therapy has gained momentum in the past few years.[10] However, identification of cryptic binding sites is a daunting task owing to their hidden nature and also because molecular mechanism(s) by which cryptic sites are formed are not properly understood.[4]

Serendipity has accounted for revelation of cryptic sites in several protein structures[9,11−13] enabling characterization of these sites which in turn has been helpful in developing methods for identification of cryptic sites. Current approaches to cryptic-site discovery include computational methods such as, CryptoSite,[6] long-timescale molecular dynamics (MD) simulations combined with Markov state models[14,15] or fragment probes,[4,5] mixed-solvent MD simulations,[16] tools for mapping small-molecule binding hot spots,[17] and experimental methods, such as extensive screening of small

Received: September 27, 2020

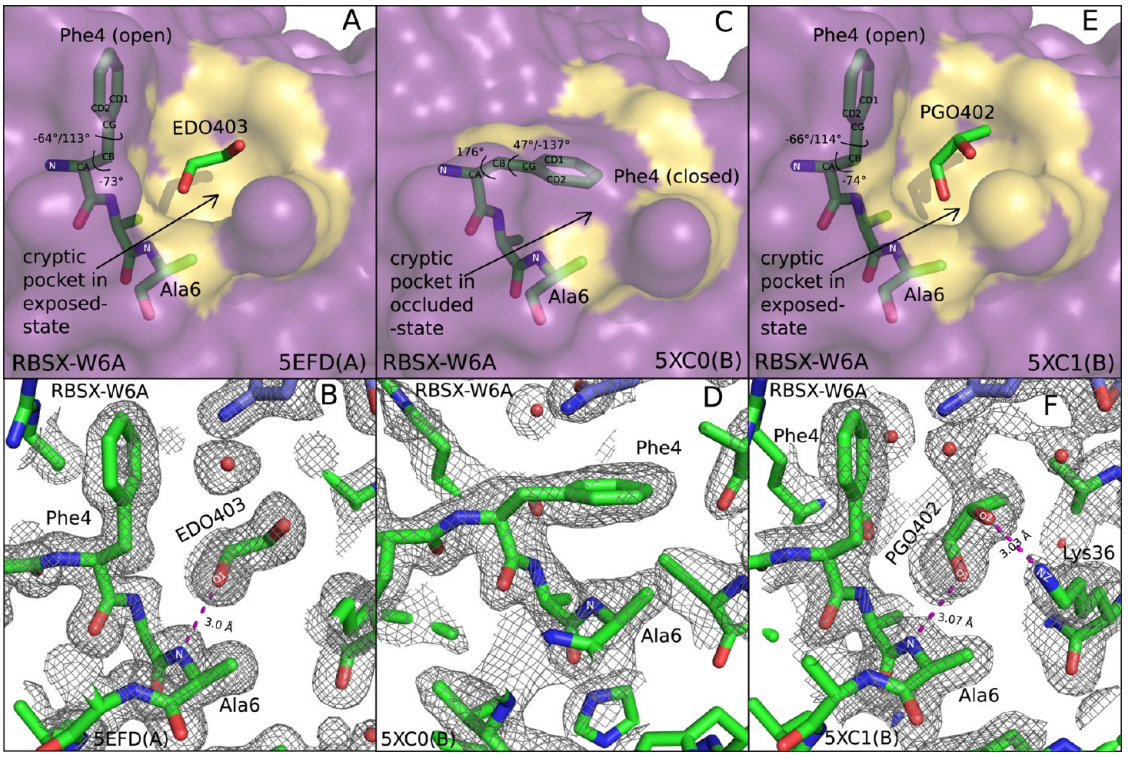


Figure 1. Small glycols, ethylene glycol and propylene glycol, identify a cryptic pocket on RBSX-W6A. Amino acid residues (green) and bound ligands (green) are shown in a stick model. Nitrogen and oxygen atoms are colored blue and red, respectively, while carbon atoms are colored according to the color mentioned for residues or ligands. Hydrogen bonds are represented as dashed lines (magenta) labeled with the donor−acceptor distance in Å. The surface of RBSX-W6A (violetpurple) is drawn and is superposed with its residue Phe4 to illustrate the open/closed conformation of Phe4 and the associated exposed state/occluded state of the cryptic surface pocket (lightorange) caused by the presence/absence of small glycols. The open and closed conformations of Phe4 are defined here in terms of its side-chain χ^1 and χ^2 (measured along N, CA, CB, CG and CA, CB, CG, CD1/CD2 atoms, respectively) dihedral angles. (A) Exposed state of the cryptic surface pocket containing a molecule of ethylene glycol (1,2-ethanediol, Ligand ID EDO) with Phe4 in the open state is shown for RBSX-W6A (PDB ID 5EFD, chain A) when it was crystallized in the presence of ethylene glycol. (C) Occluded state of the cryptic surface pocket with Phe4 in the closed state is shown for RBSX-W6A (PDB ID 5XC0, chain B) when it was crystallized in the absence of ethylene glycol. (E) Exposed state of the cryptic surface pocket containing a molecule of propylene glycol (*S*-1,2-propanediol,Ligand ID PGO) with Phe4 in the open state is shown for RBSX-W6A (PDB ID 5XC1, chain B) when it was crystallized in the presence of propylene glycol. (B), (D), and (F) Electron density maps (gray) corresponding to (A), (C), and (E) are shown. The *2mFo-DFc* electron density maps contoured at the 1.0 σ-level show a well-defined electron density for EDO and PGO in (B) and (F), respectively.

fragments[18] and site-directed tethering.[19,20] Although, CryptoSite, a cryptic-site prediction tool based on machine learning and trained on a representative data set of protein structures with validated cryptic sites, is a promising approach, it may be limited by the availability of experimentally determined cryptic sites, which constitute the training set. The success of mixed-solvent MD simulations and hot spot mapping tools depends upon the nature of probe molecules used for cryptic-site identification and are not always successful. Fragment screening experiments are expensive, time-consuming, and may not always have the desired outcomes. Use of probe molecules with the known "cryptic-site finding" potential may possibly reduce time, efforts, and expenditure in fragment screening experiments for identification of cryptic sites. If such probe molecules are also amenable to mixed-solvent MD simulations and hot spot mapping protocols, these computational methods incorporating the cryptic-site finding probe molecules can be used to identify cryptic sites on experimentally determined or modeled structures of protein targets.

In the present work, we show that small glycols find cryptic pockets on various proteins. Initial serendipitous observation made through crystallography experiments conducted on a protein system (RBSX-W6A, Trp6 to the Ala mutant of a recombinant xylanase from *Bacillus* sp. *NG-27*) showed the ability of ethylene glycol, a small glycol, to discover a cryptic pocket on RBSX-W6A. Explicit-solvent MD simulations of RBSX-W6A with the exposed state of the cryptic pocket revealed closure of the pocket in the absence of the glycol molecule, whereas no change in the cryptic pocket was observed in MD simulations of RBSX-W6A with the occluded state of the cryptic pocket, showing that cryptic pockets are unstable without ligands and prefer to stay in a closed state in their absence.[4,21] Upon repeating the crystallography experiments with propylene glycol, another small glycol, the cryptic pocket of RBSX-W6A was rediscovered further supporting the finding that small glycols can discover cryptic pockets. The cryptic pocket of RBSX-W6A showed properties similar to those of cryptic sites characterized in other protein systems,[6] and interacted with the glycols through hydrogen bonds and van der Waals (Vdw) contacts. The combined crystal structure and simulation results thus justify the role of glycols in identifying the cryptic pocket of RBSX-W6A. Further, using ethylene glycol as a probe molecule in cosolvent MD simulations, we demonstrate its ability to induce the opening of experimentally validated cryptic sites on a test set of two proteins (Niemann-Pick C2, Interleukin-2). Finally, through data set construction and analysis of protein structures with

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

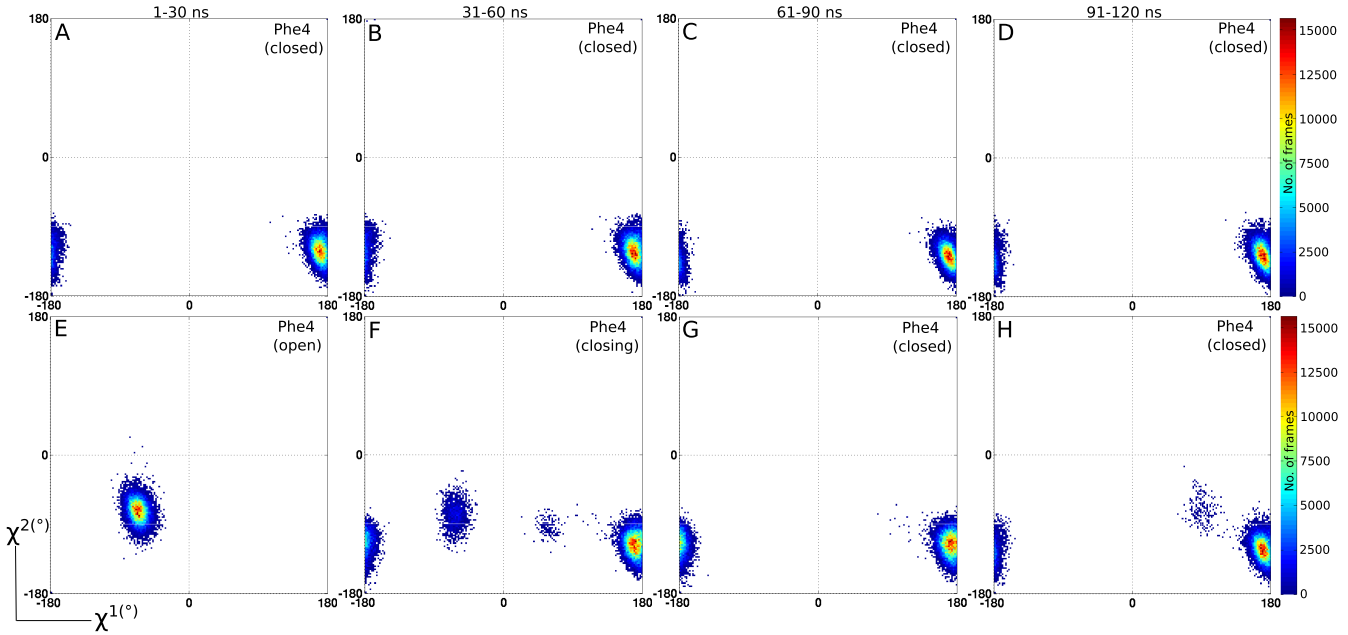


Figure 2. Two-dimensional (2D) histogram plots showing population distribution of residue Phe4 between the closed and open states during 120 ns long explicit-solvent MD simulations of RBSX-W6A. The distribution measured every 30 ns in terms of χ^1 and χ^2 side-chain dihedral angles of residue Phe4 shows that for the entire simulation time (A)−(D) Phe4 remains in the closed state and therefore the cryptic surface pocket remains in the occluded state when the RBSX-W6A structure (PDB ID 5XC0, chain B) with Phe4 in the closed state occluding the cryptic surface pocket was used to initialize the explicit-solvent MD simulations. For explicit-solvent MD simulations, started with the RBSX-W6A structure (PDB ID 5EFD, chain A, EDO removed) in which Phe4 is in the open state exposing the cryptic surface pocket, the same analysis shows that (E) for approximately the first 30 ns Phe4 remains in the open state and therefore the cryptic surface pocket remains in the exposed state, then (F) Phe4 transitions to the closed state with the cryptic surface pocket transitioning to the occluded state, and after that (G) and (H) Phe4 stays closed for the rest of the simulation and the cryptic surface pocket remains occluded.

validated cryptic sites, we show that ethylene glycol molecules bind to cryptic sites in other proteins including targets of pharmaceutical interest for which the same cryptic sites become apparent upon binding of biologically relevant ligands. Our study thus validates and generalizes the "cryptic-pocket finding" potential of small glycols in proteins.

RESULTS AND DISCUSSION

Ethylene Glycol, a Small Glycol, Identifies a Cryptic Pocket in the RBSX-W6A Crystal Structure. During the course of studies aimed at characterizing the importance of aromatic residues in the stability of a recombinant xylanase from *Bacillus* sp. *NG-27* (RBSX),[22] we observed the presence of an ethylene glycol (1,2-ethanediol, Ligand ID EDO) molecule occupying a surface pocket in the crystal structure of RBSX Trp6 to the Ala mutant (RBSX-W6A) (Protein Data Bank (PDB) ID 5EFD, Table S1) (Figure 1 A). EDO was used as a cryoprotectant in diffraction experiments (Methods, Supporting Information) and was clearly detected in the *2mFo-DFc* electron density map contoured at the 1.0 σ-level (Figure 1B). It must be noted that this surface pocket was not observed in the native RBSX structure (PDB ID 4QCE)[23] (Figure S1A). Analysis of atoms lining the surface pocket in 5EFD (Chain A) and their atomwise solvent accessibilities (AwSA (Å^2)) revealed that this surface pocket is composed of both polar and nonpolar atoms (Table S2). Further, the EDO molecule interacts with the pocket atoms by forming a hydrogen bond (3.0 Å) with a polar, solvent accessible (AwSA: 1.059 Å^2), backbone amide nitrogen atom of Ala6 along with a number of van der Waals (Vdw) contacts (18, $d \leq 4$ Å) (Figure S2A).

Studies have shown that the interaction of surface water also called hydration water, with the protein surface affects the protein structure, stability, and function.[24−27] The above observations led us to wonder if in the absence of the EDO molecule, a water molecule would interact with and occupy the amphiphilic surface pocket found in 5EFD, prompting us to redetermine the crystal structure of RBSX-W6A without the use of cryoprotectant ethylene glycol in the diffraction experiment (PDB ID 5XC0, Table S1). Much to our surprise, the particular surface pocket observed in 5EFD was not seen in 5XC0 and the backbone amide nitrogen atom of Ala6 was also found to be solvent inaccessible (AwSA: 0.000 Å^2).

Further analysis revealed that in the absence of the EDO molecule, the aromatic side-chain of the neighboring Phe4 residue moves and covers the surface pocket (Figure 1C,D), occluding it and thereby making the backbone amide nitrogen atom of the Ala6 solvent inaccessible. Based on the occluded state and exposed state of this surface pocket in 5XC0 and 5EFD, respectively, we term it as a cryptic surface pocket, which, otherwise masked by the side-chain of the Phe4 residue, becomes apparent only when an EDO molecule binds to it. Residue Phe4 is observed to have two distinct conformations in 5XC0 and 5EFD, characterized by its side-chain χ^1 (measured along N, CA, CB, and CG atoms) and χ^2 (measured along CA, CB, CG, and CD1/CD2 atoms) dihedral angles. We term the conformation of Phe4 in 5XC0 as the "closed state" ($\chi^1 = 176°$, $\chi^2 = 47/-137°$) (Figure 1C), where it occludes the cryptic surface pocket in the absence of the EDO molecule, and in 5EFD as the "open state" ($\chi^1 = -73°$, $\chi^2 = -64/113°$) (Figure 1A), where it gets displaced exposing the cryptic surface pocket upon binding of the EDO molecule.

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

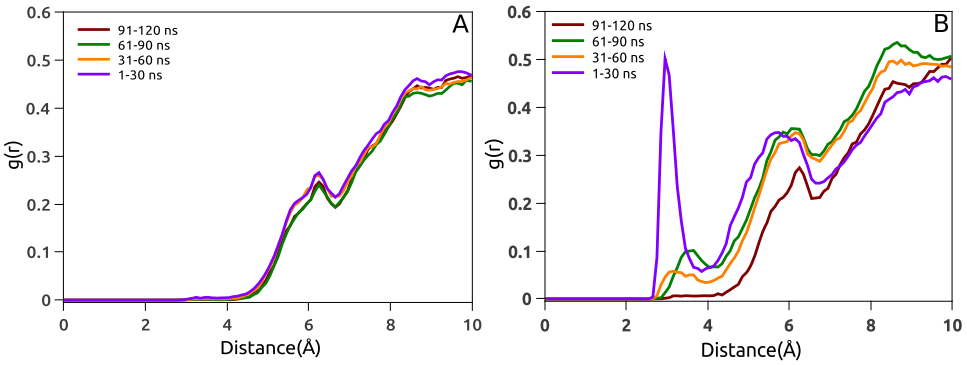


Figure 3. Radial distribution function of water around the backbone amide nitrogen atom of Ala6 calculated from the MD simulation trajectories of RBSX-W6A. The peaks in the radial distribution function of water (represented as $g(r)$, where r stands for the distance from the selected atoms) representing hydration shells around the selected atoms are used here as indicators of change in the solvent accessibility of the backbone amide nitrogen atom of Ala6 during the MD simulation of RBSX-W6A. (B) Hydration shell, which occurs at 2.95 Å from the backbone amide nitrogen atom of Ala6 when Phe4 is in the open state exposing the cryptic surface pocket (1−30 ns), is not observed in (A) 1−120 ns and (B) 31−120 ns, indicating that the backbone amide nitrogen atom of Ala6 is solvent inaccessible when Phe4 is in the closed state occluding the cryptic surface pocket.

Further, a measurement of the solution affinity constant between ethylene glycol molecules and RBSX/RBSX-W6A was performed using isothermal titration calorimetry (ITC) experiments (Methods, Supporting Information). No measurable binding was seen between ethylene glycol and the RBSX while weak binding was detected for the RBSX-W6A (Figure S3, Table S3) suggesting binding of ethylene glycol molecules to RBSX-W6A in solution under physiological conditions. It must be noted that small molecular probes binding with weak affinity such as ethylene glycol may be inaccessible to some of the current biophysical methods of detection.[28] Also, compared to other methods, X-ray crystallography can detect binding of MiniFrags, ultralow molecular weight compounds binding with very low affinities,[29] an observation reiterated here through crystallography for binding of ethylene glycol to RBSX-W6A.

Explicit-Solvent MD Simulations of RBSX-W6A Reveals Closure of the Cryptic Pocket in the Absence of Ethylene Glycol Molecules. We further computationally explored the possibility whether water molecules, in a way similar to an EDO molecule, can induce the transition of Phe4 from the closed to the open state and the associated opening of the cryptic surface pocket. For this, we conducted all-atom explicit-solvent molecular dynamics (MD) simulations starting with the RBSX-W6A structure PDB ID (5XC0, chain B) in which Phe4 was observed in the closed state occluding the cryptic surface pocket (Methods, Supporting Information). The side-chain χ^1 and χ^2 dihedral angles for the Phe4 residue and radial distribution function of water around the backbone amide nitrogen atom of Ala6, as indicators of transition of Phe4 from the closed to the open state and the corresponding transition of the cryptic surface pocket from the occluded state to the exposed state, were monitored from the simulation trajectories. The peaks in the radial distribution function of water (represented as $g(r)$, where r stands for the distance from the selected atoms) represent the hydration shells around the selected atoms. The population distribution of Phe4 measured every 30 ns in terms of its side-chain χ^1 and χ^2 dihedral angles from the simulation trajectories revealed that the Phe4 remained in the closed state and therefore the cryptic surface pocket remained in the occluded state throughout the length (120 ns) of the MD simulations with the most populated states of Phe4 being defined by $\chi^1 = -170$ to 160° (centered at 180°) and $\chi^2 = -140$ to −100° (centered at −120°) (Figure 2A−D).

Interestingly, upon conducting all-atom explicit-solvent MD simulations starting with the RBSX-W6A structure (PDB ID 5EFD, chain A, EDO removed) where Phe4 is present in the open state exposing the cryptic surface pocket (Methods, Supporting Information), similar analysis on the simulation trajectories showed that Phe4 stayed in the open state and therefore the cryptic surface pocket remained in the exposed state for approximately the first 30 ns, $\chi^1 = -70$ to −50° (centered at −60°) and $\chi^2 = -80$ to −50° (centered at −70°), (Figure 2E, 1−30 ns). Thereafter, Phe4 started transitioning to the closed state with the cryptic surface pocket transitioning to the occluded state (Figure 2F, 31−60 ns). After that, Phe4 stayed closed and therefore the cryptic surface pocket stayed occluded for the remaining simulation time, $\chi^1 = -170$ to 160° (centered at 180°) and $\chi^2 = -130$ to −90° (centered at −110°) (Figure 2G,H, 61−120 ns).

Further, the radial distribution function of water calculated from the simulation trajectories of 5EFD (chain A) reveals that the hydration shell that occurs at 2.95 Å from the backbone amide nitrogen atom of Ala6 when Phe4 is in the open state exposing the cryptic surface pocket (1−30 ns) (Figure 3B), disappears when Phe4 transitions from the open to the closed state occluding the cryptic surface pocket. Thereafter, (31−120 ns) (Figure 3B) the radial distribution function of water in MD simulations of 5EFD (chain A) resembles the radial distribution function of water calculated from the simulation trajectories of 5XC0 (chain B) (Figure 3A), reiterating observation made for the crystal structure 5XC0 that the backbone amide nitrogen atom of Ala6 is solvent inaccessible when Phe4 is in the closed state occluding the cryptic surface pocket. Thus, the simulation results support our experimental findings that in the hydrated environment without EDO molecules, the closed state of Phe4 occluding the cryptic surface pocket is preferred and the presence of EDO molecules stabilizes the open state of Phe4 exposing the cryptic surface pocket. It has also been shown through the use of various force fields in MD simulations of different protein systems with known cryptic sites that cryptic pockets are unstable without ligands and prefer to stay in the closed state in their absence.[4,21]

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Propylene Glycol, Another Small Glycol Identifies the Cryptic Pocket in the RBSX-W6A Crystal Structure. Based on the results obtained so far, we hypothesized that small glycols can discover the cryptic surface pocket of RBSX-W6A by displacing the aromatic side-chain of the Phe4 residue. To test the hypothesis, we redetermined the crystal structure of RBSX-W6A in the presence of another small glycol, propylene glycol (1,2-propanediol, racemic mixture) (PDB ID 5XC1, Table S1) and re-examined the status of the cryptic surface pocket. As anticipated, we reobserved the cryptic surface pocket in the exposed state, this time containing a molecule of *S*-1,2-propanediol (Ligand ID PGO) with the Phe4 in the open state ($\chi^1 = -74°$, $\chi^2 = -66/114°$) (Figure 1E), thus vindicating our hypothesis. The cryptic surface pocket in 5XC1 shares similar attributes with the cryptic surface pocket in 5EFD (Table S2). Similar to the scenario observed for the EDO molecule in 5EFD, in 5XC1 also, the identified cryptic pocket was found to be amphiphilic in nature (Table S2) and the PGO molecule was observed to participate in the hydrogen bonding interaction (3.07 Å) with the solvent accessible polar backbone amide nitrogen atom of Ala6 (AwSA: 1.393 Å^2), which has unsatisfied hydrogen bonding potential (Figure S2B). PGO in the pocket also forms a hydrogen bond with Lys36 (Figure S2B) and forms a number of van der Waals (Vdw) contacts (21, $d \leq 4$ Å) with the pocket atoms. A recent comprehensive study, characterizing cryptic sites in different protein systems, has also reported that cryptic sites tend to be less hydrophobic and more flexible.[6] The finding that these small glycols (EDO and PGO) identify and interact with the cryptic surface pocket in a similar manner suggests a certain level of selectivity to their mode of interaction with this pocket. Moreover, in 5XC1, of the several surface pockets, we were able to model PGO (*2mFo-DFc* electron density map (Figure 1F)) only in this particular surface pocket emphasizing the nonrandom nature of the interaction. It is interesting to note that the rotational freedom about the carbon−carbon single bond enables small glycols, EDO and PGO, to adopt different conformations out of which these glycols display amphiphilic character in gauche+ and gauche− conformations. Accordingly, the inherent flexibility, amphiphilicity of small glycols would enable an interplay between the dynamics of the flexible, less hydrophobic cryptic binding site and the dynamics of the glycol molecule interacting with the site. Although not observed in 5XC1, we cannot rule out the possibility of *R*-1,2-propanediol (Ligand ID PGR) acting in a way similar to that of PGO and EDO.

We summarize that in RBSX-W6A, the substitution of bulky Trp6 to Ala creates a surface pocket that remains occluded by the aromatic side-chain of the Phe4 residue as observed in 5XC0 (Figure S1B). When EDO/PGO molecules are allowed to interact with RBSX-W6A, they not only displace the aromatic side-chain of Phe4 exposing the cryptic surface pocket but also interact with the pocket in a similar manner as seen in 5EFD and 5XC1 (Figures S1A and C, S2). Thus, the combined crystal structure and simulation results justify the role of small glycols in identifying the cryptic pocket of RBSX-W6A. Importantly, cryptic cavity/cavity generated on the protein surface at the mutation site has been successfully used to target the oncogenic mutants, K-Ras(G12C)[9] and p53-(Y220C),[30] assuming that binding of small molecules, while inhibiting the oncogenic mutant, would not affect the functioning of the wild-type protein lacking the mutational cavity. This suggests that cryptic cavities/cavities arising due to mutations are equally important for therapy in addition to such sites inherently present in wild-type proteins. Based on our work, it may be surmised that determination of protein crystal structures in the presence and absence of small glycols (ethylene glycol and propylene glycol) and a systematic comparison of the protein structures could help in the identification of cryptic sites in the protein under study.

Cryptic-Pocket Finding Potential of Small Glycols in Other Proteins: Cosolvent Simulations and Data Set Construction and Analysis. Our so far discussed results concluded from the studies conducted on RBSX-W6A, used as a model protein, are thus suggestive of a role of the small glycols in finding cryptic pockets on proteins. To investigate the cryptic-pocket finding potential of the small glycols in other proteins with experimentally validated cryptic sites, we conducted ethylene glycol-based cosolvent simulations on a test set of two proteins (described below) and constructed and analyzed the data set of proteins (Methods, Supporting Information) from the Protein Data Bank (PDB). Our data set construction (Table S5, Supporting Information) and analysis revealed the ability of ethylene glycol molecules to bind and occupy the cryptic sites of proteins: Bcl-xL (Figure S6), actin (Figure S7), myosin II (Figure S8), and glutamate receptor 2 (Figure S9) in which the same cryptic sites become apparent upon binding of biologically relevant ligands. Further, in our analysis of the above four proteins, we observed overlapping features between the interaction profile of ethylene glycol and the interaction profile of the biologically relevant ligand bound at the same cryptic site (Figures S6D, S7D, S8D and S9D) thus, pointing toward a certain level of specificity and selectivity in the interaction of the ethylene glycol with the cryptic sites. Our analysis thus validates and generalizes the cryptic-pocket finding potential of small glycols in proteins.

Enhanced Sampling of the Exposed State of Experimentally Validated Cryptic Sites of Other Proteins by Ethylene Glycol Molecules in Explicit-Cosolvent MD Simulations. Motivated by the application of cosolvent MD simulations in identifying cryptic sites,[16,31] we wondered whether the use of ethylene glycol as a cosolvent in MD simulations could unveil cryptic sites on other proteins. A variety of small molecules containing aromatic and aliphatic moieties substituted with different chemical functional groups have been used as probes in cosolvent MD simulations.[32] However, it is important to note that the use of ethylene glycol as a cosolvent in MD simulations to identify cryptic sites and/or binding hot spots has not been reported earlier to the best of our knowledge. Over the past few years, the cosolvent-based MD simulations approach has been increasingly employed to assess druggability of targets of pharmaceutical interest by identifying binding sites and hot spots on proteins[33] and has also been shown to enhance the sampling of exposed states of cryptic sites on a range of protein targets.[16,31] We chose two proteins, Niemann-Pick C2 (NPC-2) and interleukin-2 (IL-2), for ethylene glycol-based cosolvent MD simulations. The choice was motivated by the observation that these proteins are pharmaceutically important, harbor experimentally validated cryptic sites, and have been previously subjected to cosolvent MD simulations using different probes (other than ethylene glycol) to identify druggable cryptic binding sites.[16,31] Most importantly, to the best of our knowledge small glycols studied in this report have not been used in the crystallization experiments of NPC-2 and IL-2. For cosolvent MD simulations of NPC-2 and IL-2, we used a 5% v/v ethylene

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

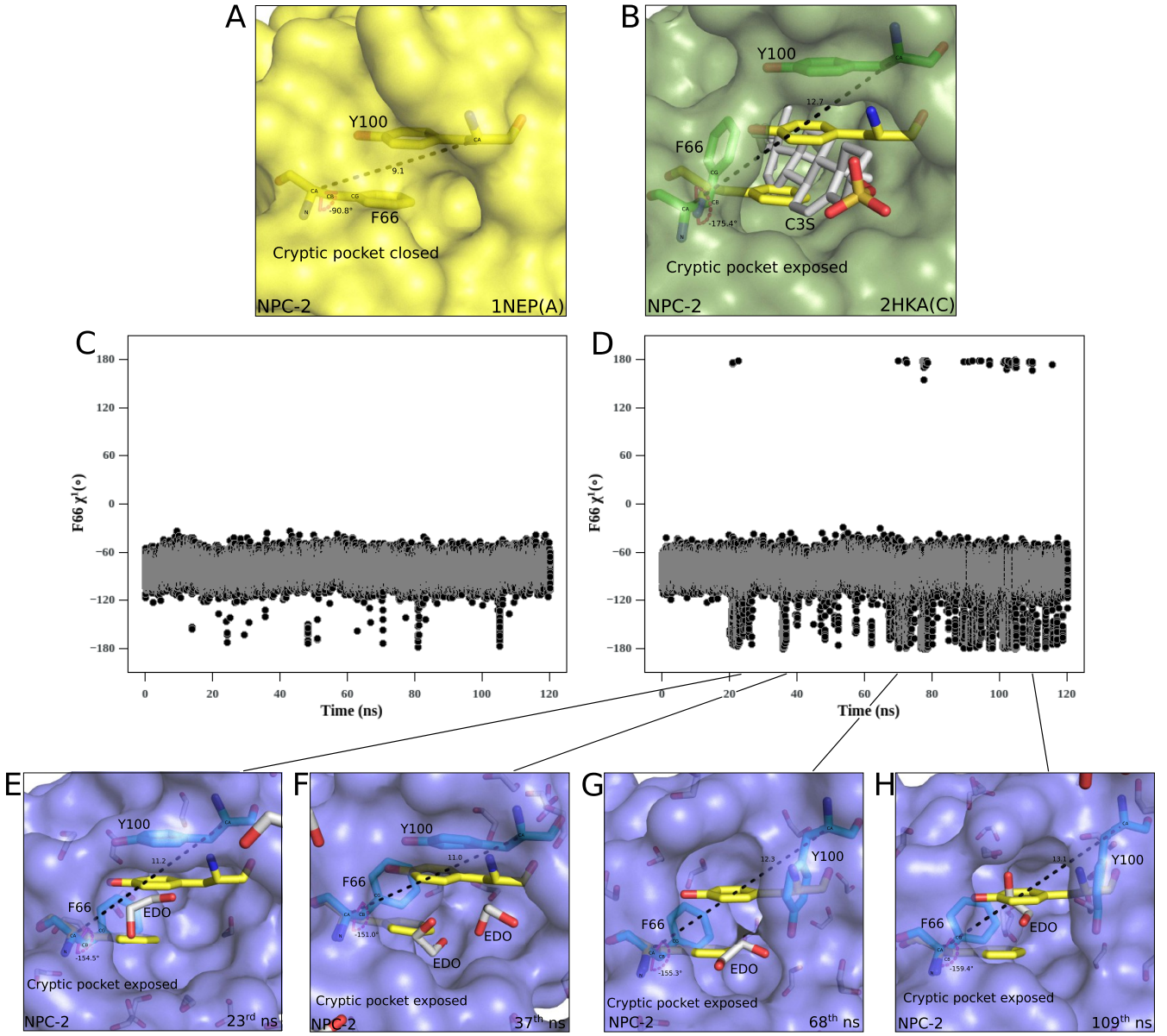


Figure 4. Enhanced sampling of the exposed state of the cryptic pocket of NPC-2 by ethylene glycol in cosolvent MD simulations. Amino acid residues and bound ligands are shown in a stick model. Nitrogen, oxygen, and sulfur atoms are colored blue, red, and orange, respectively, while carbon atoms are colored according to the color mentioned for residues or ligands. Residue side-chain χ^1 (measured along N, CA, CB, CG atoms) dihedral angles are shown as labeled magenta arcs. Residue displacements are represented as black dashed lines labeled with the distance in Å. The surface of NPC-2 is drawn and is superposed with its residues Phe66 and Tyr100 to illustrate the rearrangements of these residues and the associated opening of the cryptic pocket upon ligand binding. (A) Closed state of the cryptic pocket and associated arrangement of residues Phe66 and Tyr100 (yellow sticks) is shown in the unbound NPC-2 structure (yellow, PDB ID 1NEP, chain A). (B) Rearrangements of residues Phe66, Tyr100 (protruding, yellow sticks; displaced, green sticks) and opening up of the cryptic pocket upon binding of cholesterol-3-*O*-sulfate (Ligand ID C3S, gray) to the NPC-2 structure (smudge, PDB ID 2HKA, chain C) is shown. Time series plots showing distribution of the χ^1 side-chain dihedral angle of Phe66, as one of the parameters indicating the distribution of the closed/open states of the cryptic pocket, in (C) explicit-solvent and (D) 5% ethylene glycol-based explicit-cosolvent MD simulations of the apo NPC-2 structure (PDB ID 1NEP, chain A). Snapshots sampled from the cosolvent simulation trajectory at (E) 23rd ns (F) 37th ns (G) 68th ns, and (H) 109th ns show rearrangements of residues Phe66, Tyr100 (protruding, yellow sticks; displaced, blue sticks) and associated opening up of the cryptic pocket upon binding of ethylene glycol molecule(s) (Ligand ID EDO, gray) to the NPC-2 structure (slate).

glycol−water solution. Use of high cosolvent concentrations (15% or above) have been shown to induce protein unfolding and therefore are discouraged.[31] Also, a 5% concentration is close to concentrations used in experiments.[34] Details of system preparation and subsequent simulations of NPC-2 and IL-2 for ethylene glycol-based cosolvent MD simulations is given in Methods, Supporting Information. Here, we present the results of cosolvent simulations from the point of view of the potential of ethylene glycol molecule(s) to unmask the experimentally validated cryptic sites of NPC-2 and IL-2.

Niemann-Pick C2 (NPC-2) is a small protein involved in transport of cholesterol and other sterols from lysosomes to different cellular locations.[35] The deficiency of NPC-2 causes fatal Niemann-Pick type C2 disease characterized by accumulation of cholesterol in lysosomes.[36] In the holo-form of the NPC-2 structure (PDB ID 2HKA, chain C) (Figure 4B),

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

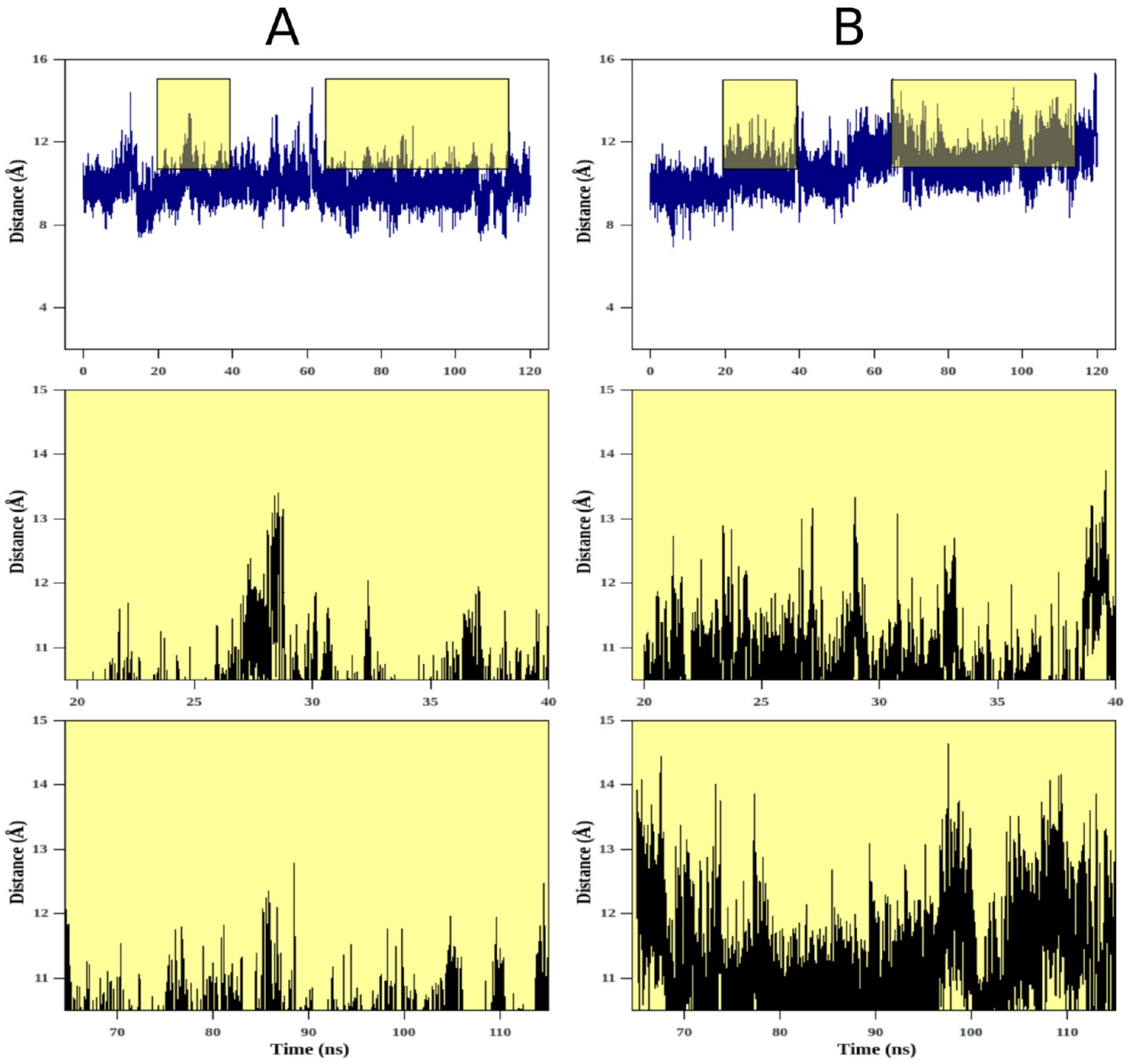


Figure 5. Time series plots showing distribution of the distance between CA atoms of residues Phe66 and Tyr100, as one of the parameters indicating the distribution of the closed/open states of the NPC-2 cryptic pocket, in (A) explicit-solvent and (B) 5% ethylene glycol-based explicit-cosolvent MD simulations of the apo NPC-2 structure (1NEP(A)). Plots are zoomed at specific time-intervals highlighting the frequent sampling of the open states in (B).

binding of cholesterol-3-*O*-sulfate (Ligand ID C3S) disrupts the π-stacking interaction and displaces the side-chains of residues, Phe66 on the βD strand and Tyr100 on the βE-βF loop, thereby revealing a deep cavity not observed in the apo-form of the NPC-2 structure (PDB ID 1NEP, chain A) (Figure 4A).[37] Thus, the cryptic site on the NPC-2 structure is an orthosteric site and represents a challenging test case for our ethylene glycol-based cosolvent simulations due to the buried hydrophobic nature of the cryptic pocket.[5] The apo-form of the NPC-2 structure was subjected to explicit-solvent and cosolvent MD simulations with ethylene glycol as probe molecules (Methods, Supporting Information) to see whether ethylene glycol molecules could enhance sampling of the open state of the NPC-2 cryptic pocket. Comparison of apo and holo structures of NPC-2 shows that binding of cholesterol-3-*O*-sulfate to NPC-2 and opening of the cryptic pocket is accompanied by a significant change in the χ^1 (measured along N, CA, CB, CG atoms) dihedral angle of Phe66 from −90.8° in the apo NPC-2 structure (Figure 4A) to −175.4° in the holo NPC-2 structure (Figure 4B) and displacement of residues Phe66, Tyr100 as measured here by an increase in the distance between CA atoms of residues Phe66 and Tyr100 from 9.1 Å in the apo NPC-2 structure (Figure 4A) to 12.7 Å in the holo NPC-2 structure (Figure 4B). Therefore, transition in the Phe66 χ^1 side-chain dihedral angle and distance between the CA atoms of residues Phe66 and Tyr100, as indicators of the dynamics of the cryptic orthosteric site, were monitored from the explicit-solvent and cosolvent simulation trajectories of NPC-2.

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

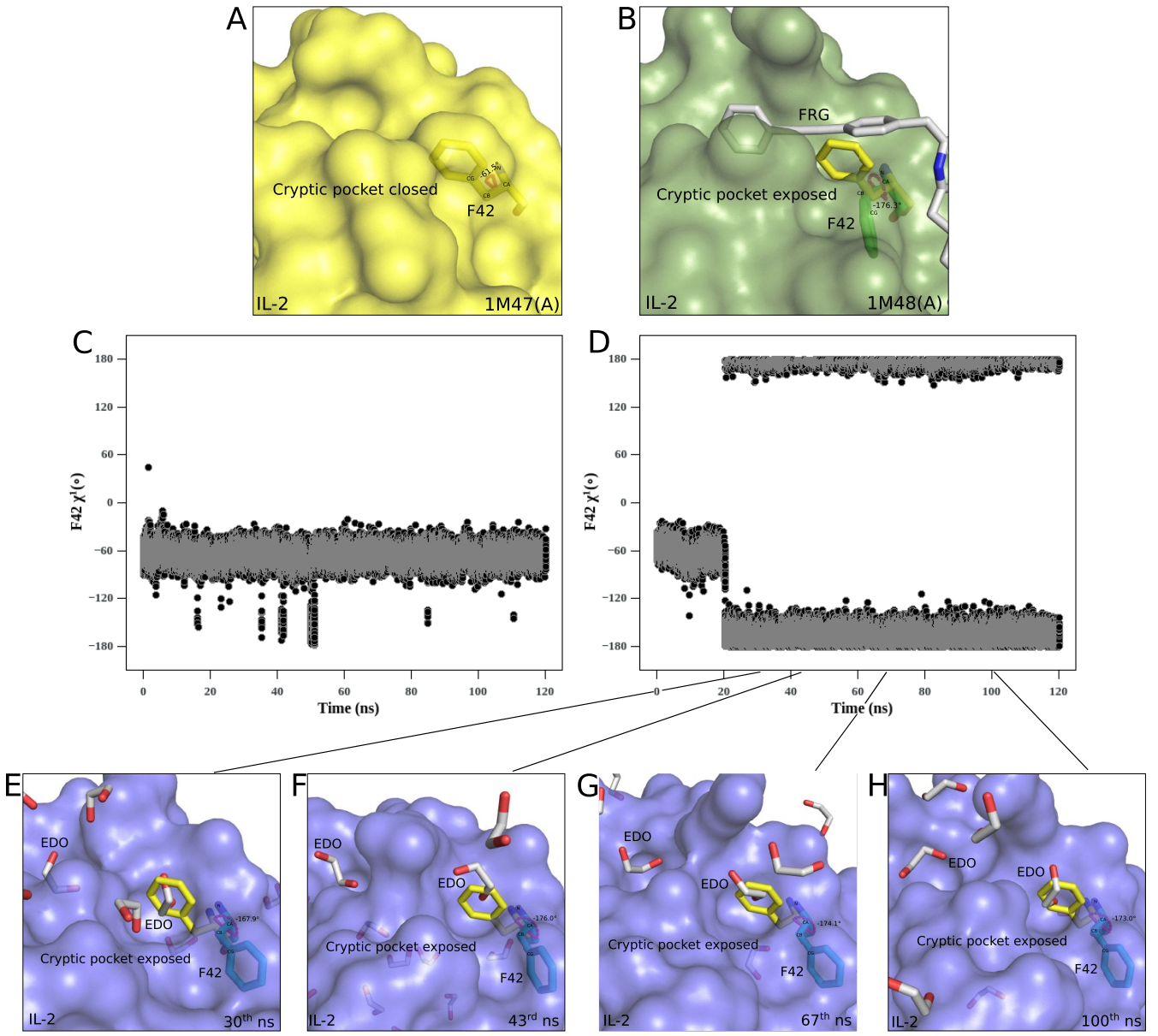


Figure 6. Enhanced sampling of the exposed state of the cryptic pocket of IL-2 by ethylene glycol in cosolvent MD simulations. Amino acid residues and bound ligands are shown in a stick model. Nitrogen and oxygen atoms are colored blue and red, respectively, while carbon atoms are colored according to the color mentioned for residues or ligands. Residue side-chain χ^1 (measured along N, CA, CB, CG atoms) dihedral angles are shown as labeled magenta arcs. The surface of IL-2 is drawn and is superposed with its residue Phe42 to illustrate the conformational change in this residue and the associated opening of the cryptic pocket upon ligand binding. (A) Closed state of the cryptic pocket and associated conformation of residue Phe42 (yellow stick) is shown in the unbound IL-2 structure (yellow, PDB ID 1M47), chain A. (B) Conformational change in residue Phe42 (protruding, yellow sticks; displaced, green sticks) and opening up of the cryptic pocket upon binding of the ligand (Ligand ID FRG, gray) to the IL-2 structure (smudge, PDB ID 1M48, chain A) is shown. Time series plots showing distribution of the χ^1 side-chain dihedral angle of Phe42, as one of the parameters indicating the distribution of the closed/open states of the cryptic pocket, in (C) explicit-solvent and (D) 5% ethylene glycol-based explicit-cosolvent MD simulations of the apo IL-2 structure (PDB ID 1M47, chain A). Snapshots sampled from the cosolvent simulation trajectory at the (E) 30th ns (F) 43rd ns (G) 67th ns, and (H) 100th ns shows the conformational change in the residue Phe42 (protruding, yellow sticks; displaced, blue sticks) and associated opening up of the cryptic pocket upon binding of ethylene glycol molecule(s) (Ligand ID EDO, gray) to the IL-2 structure (slate).

Time series plots show that, in comparison to the explicit-solvent simulation of NPC-2 (Figures 4C, 5A), the presence of ethylene glycol molecules in cosolvent simulation considerably shifts the distribution of the Phe66 χ^1 side-chain dihedral angle and distance between CA atoms of residues Phe66 and Tyr100 more toward the values defining the exposed state of the NPC-2 cryptic pocket (Figures 4D, 5B). Representative structural snapshots sampled from the cosolvent simulation trajectory at 23rd, 37th, 68th, and 109th ns (Figure 4E−H) show that ingression of the cosolvent, ethylene glycol molecules, at the cryptic site causes residues, Phe66 and Y100, to assume conformations similar (**ns: Phe66 χ^1, Phe66CA-Tyr100CA distance;** 23rd: −154.5°, 11.2 Å; 37th: −151.0°, 11.0 Å; 68th: −155.3°, 12.3 Å; 109th: −159.4°, 13.1 Å, Figure 4E−H) to the site-permissible conformations of these residues (Phe66 χ^1 = −175.4°, Phe66CA-Tyr100CA distance = 12.7 Å, Figure 4B)

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

observed in the holo-form of the NPC-2 structure, thereby identifying the cryptic orthosteric site (Figure 4E−H). A previous cosolvent MD simulation study of NPC-2 in the presence of 5% isopropyl alcohol only explored regions intermediate to the closed and open forms whereas 5% resorcinol-based cosolvent simulation initially sampled the exposed state of the cryptic pocket but later on drove the sampling away from the holo structure.[16] Our 5% ethylene glycol-based cosolvent simulation of NPC-2 echoes results similar to that of 5% resorcinol wherein the distribution of the Phe66 χ^1 side-chain dihedral angle and distance between CA atoms of residues Phe66 and Tyr100, though considerably shifted toward the open state in comparison to water only simulations of NPC-2 (Figures 4C, 5A), toggles back and forth between the open and closed states of the cryptic pocket (Figures 4D, 5B) and might point to the limitation of smaller probe molecules in sampling cryptic sites revealed by binding of huge ligands. However, in spite of the size and chemical nature of ethylene glycol similar to isopropyl alcohol, it fared better than isopropyl alcohol and similar to bulkier aromatic resorcinol in cosolvent studies to identify the cryptic pocket of NPC-2. The comparison made is only qualitative and not quantitative owing to the variations in the cosolvent simulation methodology. In conclusion, it may be seen that ethylene glycol molecules expose the cryptic orthosteric site on NPC-2, also sampled by the aromatic probe, resorcinol.[16]

Interleukin-2 (IL-2) is a cytokine that acts by binding to various IL-2 receptors (IL-2Rs).[38] It is of considerable pharmaceutical interest as it plays a significant role in the activation of T-cells and rejection of tissue grafts.[39] The cryptic site on IL-2 is located at the IL-2Rα binding site and is occluded by the protrusion of the side-chain of the residue Phe42 in the apo-structure (PDB ID 1M47), chain A (Figure 6A). Binding of a small molecule (Ligand ID FRG) displaces the Phe42 side-chain thereby revealing the cryptic site in the holo structure of IL-2 (PDB ID 1M48), chain A(Figure 6B).[40] Similar to NPC-2, the apo-form of the IL-2 structure was subjected to explicit-solvent and cosolvent MD simulations with ethylene glycol as probe molecules (Methods, Supporting Information) to see whether ethylene glycol molecules could enhance sampling of the open state of the IL-2 cryptic pocket. Comparison of apo and holo structures of IL-2 shows that binding of the ligand FRG to IL-2 and opening of the cryptic pocket is accompanied by a significant change in the χ^1 (measured along N, CA, CB, CG atoms) dihedral angle of Phe42 from −61.5° in the apo IL-2 structure (Figure 6A) to −176.3° in the holo IL-2 structure (Figure 6B). Therefore, transition in the Phe42 χ^1 side-chain dihedral angle, as an indicator of the dynamics of the cryptic site, was monitored from the explicit-solvent and cosolvent simulation trajectories of IL-2.

Time series plots show that, in comparison to the explicit-solvent simulation of IL-2 (Figure 6C), the presence of ethylene glycol molecules in cosolvent simulation substantially shifts the distribution of the Phe42 χ^1 side-chain dihedral angle more toward the values defining the exposed state of the IL-2 cryptic pocket (Figure 6D). Representative structural snapshots sampled from the cosolvent simulation trajectory at the 30th, 43rd, 67th, and 100th ns (Figure 6E−H) show that ingression of the cosolvent, ethylene glycol molecules, at the cryptic site causes the residue, Phe42, to assume conformations similar (**ns: Phe42 χ^1**; 30th: −167.9°; 43rd: −176.0°; 67th: −174.1°; 100th: −173.0°, Figure 6E−H) to the site-permissible conformation of this residue (Phe42 χ^1 = −176.3°, Figure 6B) observed in the holo-form of the IL-2 structure, thereby unveiling the cryptic pocket (Figure 6E−H). Though the cryptic site of IL-2 opens to a smaller extent in the absence of probe molecules, a previous cosolvent simulation study of IL-2 has shown that the 10% phenol-based cosolvent simulation enhanced the sampling of the open state of the IL-2 cryptic site.[31] Our 5% ethylene glycol-based cosolvent simulation of IL-2 shows that the distribution of the Phe42 χ^1 side-chain dihedral angle is significantly shifted toward the open state in comparison to water only simulations of IL-2 (Figure 6C,D). Here, the observation of interest is that a 5% concentration of ethylene glycol performs on par with a 10% phenol concentration in cosolvent studies to identify the cryptic pocket of IL-2. Thus, it is seen from our work that ethylene glycol molecules open the cryptic druggable site on the pharmaceutical target IL-2.

Though the focus of our cosolvent simulation analysis is to highlight the cryptic-pocket finding potential of ethylene glycol molecules, they were also observed to sample physiologically important sites on the protein and therefore serve as potential candidates for their inclusion as probe molecules in the FTMap probe library for computational mapping of ligand-binding hot spots on proteins.[17] We observed that although the nonterminal phenyl moiety of the ligand FRG displaces Phe42 and binds to the cryptic pocket (Figure 6B), the scenario also seen for ethylene glycol molecules (Figure 6E−H), we further observed in our cosolvent simulations of IL-2 that ethylene glycol molecules were found to bind to a pocket adjacent to the cryptic pocket (Figure 6E−H), which in the holo-form of IL-2 is bound by the terminal phenyl moiety of the ligand FRG (Figure 6B). This adjacent pocket might serve as an anchor point for the terminal phenyl moiety of the ligand FRG enabling the nonterminal phenyl moiety to exploit flexibility at the cryptic site thereby revealing it and pointing to the induced fit mechanism of cryptic pocket identification. Existence of such pockets/sites, termed as hot spots, in the vicinity of cryptic sites, has been suggested as one of the prerequisites for formation of druggable cryptic sites[7] and are exploited by ligands to identify cryptic sites. In fact, a hot spot mapping study on IL-2 has identified this pocket neighboring the cryptic site as one of the strongest hot spot.[7] Thus, it can be seen from our cosolvent simulations that ethylene glycol molecules not only sample the cryptic sites but also the neighboring hot spots crucial for the identification of druggable cryptic sites.

CONCLUSIONS

Among the effects of polyhydric alcohols on proteins, it has been reported that glycerol and poly(ethylene glycol) interact with features in proteins.[41] It is also known that ethylene glycol and propylene glycol bind to proteins and are used as cryoprotectants in protein crystallography experiments. However, that small glycols, such as ethylene glycol and propylene glycol, display a positive and desirable effect of uncovering cryptic sites in proteins has not been elaborated earlier. The present study has for the first time systematically explored and demonstrated the cryptic-site finding potential of these small glycols in proteins. Stand-alone small cryptic pockets formed solely by the movement of side-chains in general may not be useful for drug-design,[7] however, cryptic adjacent pockets that occur in the vicinity of functional sites of proteins can aid in the design of ligand molecules with an increased affinity and

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

specificity, given their identification by some means. In that respect, small probe molecules such as ethylene glycol and propylene glycol, which in the present work are shown to discover cryptic sites in proteins mainly by inducing side-chain movements, can be used to probe the vicinity of functional sites of pharmaceutically important protein targets for potential cryptic sites and/or binding hot spots. Such cryptic sites and/or hot spots, identified vicinal to functional sites, can then be used in a structure-aided chemical elaboration process to design ligands, which bind with enhanced affinity to those targets.

In fact, upon screening protein CK2α, a Ser/Thr kinase an important target in cancer therapy, with the fragment library, a fragment 3,4 dichlorophenethylamine identified a cryptic pocket, termed the αD pocket, adjacent to the ATP binding site by displacing the side-chain of Tyr125 (PDB ID 5CLP).[42] This cryptic site was then used to develop a new CK2α inhibitor with a high nanomolar affinity.[42] Interestingly, in one structure of CK2α (PDB ID 3WAR), it was observed that two ethylene glycol molecules were bound at the entrance of the partially opened αD pocket[42,43] further supporting our concept of using small glycols to identify cryptic sites in protein targets, which can subsequently be used for drug-design purposes. Another crystallography study of protein CK2α[43] has shown several ethylene glycol molecules occupying physiologically significant sites of CK2α suggesting their use in computational methods for binding site determination. Thus, it must be noted that, for a given protein (in this case CK2α), ethylene glycol molecules are able to identify a cryptic site in the protein as well as are able to map the binding hot spots of the protein. Similarly, through crystallography experiments, propylene glycol has been shown to map hot spots of binding in functional sites of various proteins and has been suggested as a relevant seed for further design.[44]

The present study that includes crystal structure analysis, cosolvent molecular dynamics simulations, and data set construction and analysis, reveals multiple instances of small glycols, ethylene glycol and propylene glycol, identifying cryptic pockets in a variety of proteins, some of which are pharmaceutically important. Since it is evident from our work that small glycols can unveil cryptic sites, they can be used as fragments in experimental fragment screening methods with the added advantage of identifying cryptic sites thereby enhancing the repertoire of minifragment libraries. Further, computationally these small glycols can be included in the probe set of hot spot mapping protocols and mixed-solvent MD simulations to identify cryptic sites and map binding hot spots on available structures of protein targets as well as on modeled structures of protein targets, which are difficult to crystallize. Importantly, our work argues for the use of small glycols in both experimental as well as computational protocols for the identification of cryptic sites that could be eventually exploited in drug-design endeavors.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c01126.

Methods; crystallization and data collection; X-ray data processing, structure determination and refinement; analysis of structures; isothermal titration calorimetry (ITC); system preparation for explicit-solvent and cosolvent molecular dynamics simulations; molecular dynamics simulations; analysis of simulation trajectories; cryptic-pocket finding potential of small glycols in other proteins: data set construction and analysis; Tables: statistics of X-ray diffraction data collection and structure refinement (Table S1); atoms lining the cryptic surface pocket and their atomwise solvent accessibilities (AwSA) in 5EFD (Chain A), 5XC0 (Chain B), and 5XC1 (Chain B) (Table S2); thermodynamic parameters describing the binding of RBSX and RBSX-W6A to ethylene glycol measured at 298 K in 25 mM Tris, pH 8.5 (Table S3); ratio of ethylene glycol and water molecule in 5% v/v ethylene glycol−water solutions for cosolvent simulations of NPC-2 and IL-2 proteins (Table S4); the PDB identifiers of protein systems in unbound, ligand-bound, and EDO-bound forms and the corresponding Ligand-IDs where the ligand and EDO molecules identify the same cryptic site not apparent in the unbound structure (Table S5); Figures: the cryptic surface pocket discovered by the small glycols on RBSX-W6A is not observed in the native RBSX structure (Figure S1); schematic representation of the interactions made by different small glycols at the exposed cryptic surface pocket of RBSX-W6A (Figure S2); isothermal titration calorimetry (ITC) monitors the thermodynamics of (A) RBSX and (B) RBSX-W6A binding interactions with ethylene glycol (Figure S3); system setup for ethylene glycol-based explicit-cosolvent MD simulations (Figure S4); backbone RMSD to the starting structure of (A) apo RBSX-W6A (B) holo RBSX-W6A (C) apo NPC-2 and (D) apo NPC-2 with EDO molecules (E) apo IL-2 (F) apo IL-2 with EDO molecules, over 120 ns explicit-solvent and cosolvent MD simulations (Figure S5); ethylene glycol identifies a cryptic site in Bcl-xL (Figure S6); ethylene glycol identifies a cryptic site in actin (Figure S7); ethylene glycol identifies a cryptic site in myosin II (Figure S8); and ethylene glycol identifies a cryptic site in glutamate receptor 2 (Figure S9) (PDF)

Accession Codes

The atomic coordinates and structure factors of all of the three structures have been deposited by us in the Protein Data Bank (PDB) and are available under the accession codes 5EFD, 5XC0, and 5XC1.

■ AUTHOR INFORMATION

Corresponding Authors

Harsh Bansia − *Department of Physics, Indian Institute of Science, Bengaluru 560012, India;* orcid.org/0000-0002-6738-8754; Email: hbansia@gc.cuny.edu, harsh@iisc.ac.in, harsh3b@gmail.com

Suryanarayanarao Ramakumar − *Department of Physics, Indian Institute of Science, Bengaluru 560012, India;* Phone: +91 (80) 2293 2312; Email: ramak@iisc.ac.in; Fax: +91 (80) 2360 2602

Authors

Pranjal Mahanta − *Department of Physics, Indian Institute of Science, Bengaluru 560012, India*

Neela H. Yennawar − *The Huck Institutes of the Life Sciences, The Pennsylvania State University, State College,*

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Pennsylvania 16802, United States; orcid.org/0000-0001-7278-659X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c01126

Author Contributions

H.B., P.M., and S.R. conceptualized the study. H.B. and P.M. crystallized and solved the crystal structures. H.B. carried out MD simulations and other computational studies. N.H.Y. performed ITC experiments. H.B., P.M., N.H.Y., and S.R. analyzed and interpreted the data. H.B. wrote the paper that was reviewed, edited, and approved by all of the co-authors. In the opinion of all authors, H.B. and P.M. should be considered as first authors.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank CSIR-UGC and the Indian Institute of Science (IISc) for providing financial support to H.B., UGC for Emeritus Fellowship to S.R., X-ray Facility at the Molecular Biophysics Unit (IISc), Bengaluru, India, and European Synchrotron Radiation Facility (ESRF), Grenoble, France for X-ray data collection, staff for providing access to and support on the beamline BM14 at ESRF, DBT for funding the trip to ESRF, DST sponsored computational facility in the Department of Physics (IISc), and Supercomputer Education and Research Center (IISc) for providing access to CRAY XC40-"SAHASRAT" supercomputer and other computational resources for MD simulations and Dr. V.S. Reddy and Dr. Amit Bharadwaj for providing protein samples of RBSX and RBSX-W6A. The authors thank Julia Fecko for the ITC binding study. The ITC work was supported by the NIH grant S10OD025145 to Dr. Yennawar for the TA Instruments Low Volume Auto Affinity ITC, housed in the Automated Biological Calorimetry Core Facility at the Penn State, Huck Institutes of the Life Sciences.

■ ABBREVIATIONS USED

RBSX, recombinant xylanase from *Bacillus* sp. *NG-27*; RBSX-W6A, RBSX tryptophan6 to the alanine mutant; MD, molecular dynamics; PDB, Protein Data Bank; EDO, 1,2-ethanediol (ethylene glycol); PGO, S-1,2-propanediol (propylene glycol); PGR, R-1,2-propanediol (propylene glycol); AwSA, atomwise solvent accessibilities; Bcl-xL, B-cell lymphoma-extra large; ITC, isothermal titration calorimetry; IL-2, interleukin-2; NPC-2, Niemann-Pick C2

■ REFERENCES

(1) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discovery* **2006**, *5*, 993−996.
(2) Surade, S.; Blundell, T. L. Structural Biology and Drug Discovery of Difficult Targets: the Limits of Ligandability. *Chem. Biol.* **2012**, *19*, 42−50.
(3) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. *J. Med. Chem.* **2005**, *48*, 2518−2525.
(4) 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.
(5) Comitani, F.; Gervasio, F. L. Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH. *J. Chem. Theory Comput.* **2018**, *14*, 3321−3331.
(6) 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.
(7) Beglov, D.; Hall, D. R.; Wakefield, A. E.; Luo, L.; Allen, K. N.; Kozakov, D.; Whitty, A.; Vajda, S. Exploring the Structural Origins of Cryptic Sites on Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E3416−E3425.
(8) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727−730.
(9) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K-Ras(G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* **2013**, *503*, 548−551.
(10) Vajda, S.; Beglov, D.; Wakefield, A. E.; Egbert, M.; Whitty, A. Cryptic Binding Sites on Proteins: Definition, Detection, and Druggability. *Curr. Opin. Chem. Biol.* **2018**, *44*, 1−8.
(11) Thanos, C. D.; Randal, M.; Wells, J. A. Potent Small-Molecule Binding to a Dynamic Hot Spot on IL-2. *J. Am. Chem. Soc.* **2003**, *125*, 15280−15281.
(12) Śledź, P.; Stubbs, C. J.; Lang, S.; Yang, Y. Q.; McKenzie, G. J.; Venkitaraman, A. R.; Hyvonen, M.; Abell, C. From Crystal Packing to Molecular Recognition: Prediction and Discovery of a Binding Site on the Surface of Polo-Like Kinase 1. *Angew. Chem., Int. Ed.* **2011**, *50*, 4003−4006.
(13) Horn, J. R.; Shoichet, B. K. Allosteric Inhibition Through Core Disruption. *J. Mol. Biol.* **2004**, *336*, 1283−1291.
(14) 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.
(15) 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.
(16) Kimura, S. R.; Hu, H. P.; Ruvinsky, A. M.; Sherman, W.; Favia, A. D. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 1388−1401.
(17) Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The FTMap Family of Web Servers For Determining and Characterizing Ligand-Binding Hot Spots of Proteins. *Nat. Protoc.* **2015**, *10*, 733−755.
(18) Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of Secondary Binding Sites in Proteins Using Fragment Screening. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 15910−15915.
(19) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. Site-Directed Ligand Discovery. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9367−9372.
(20) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. Tethering: Fragment-Based Drug Discovery. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199−223.
(21) Tan, Y. S.; Śledź, P.; Lang, S.; Stubbs, C. J.; Spring, D. R.; Abell, C.; Best, R. B. Using Ligand-Mapping Simulations to Design a Ligand Selectively Targeting a Cryptic Surface Pocket of Polo-Like Kinase 1. *Angew. Chem., Int. Ed.* **2012**, *51*, 10078−10081.
(22) Bhardwaj, A.; Leelavathi, S.; Mazumdar-Leighton, S.; Ghosh, A.; Ramakumar, S.; Reddy, V. S. The Critical Role of N- and C-Terminal Contact in Protein Stability and Folding of a Family 10 Xylanase Under Extreme Conditions. *PLoS One* **2010**, *5*, No. e11347.
(23) Mahanta, P.; Bhardwaj, A.; Kumar, K.; Reddy, V. S.; Ramakumar, S. Structural Insights into N-Terminal to C-Terminal Interactions and Implications for Thermostability of a (β/α)8-Triosephosphate Isomerase Barrel Enzyme. *FEBS J.* **2015**, *282*, 3543−3555.

(24) Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics of Water at the Protein-Solvent Interface. *J. Phys. Chem. B* **2002**, *106*, 6617−6633.
(25) Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Parak, F. G. Slaving: Solvent Fluctuations Dominate Protein Dynamics and Functions. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16047−16051.
(26) Fogarty, A. C.; Laage, D. Water Dynamics in Protein Hydration Shells: the Molecular Origins of the Dynamical Perturbation. *J. Phys. Chem. B* **2014**, *118*, 7715−7729.
(27) Levy, Y.; Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389−415.
(28) Fuglestad, B.; Kerstetter, N. E.; Bédard, S.; Wand, A. J. Extending the Detection Limit in Fragment Screening of Proteins Using Reverse Micelle Encapsulation. *ACS Chem. Biol.* **2019**, *14*, 2224−2232.
(29) Osborne, J.; Panova, S.; Rapti, M.; Urushima, T.; Jhoti, H. Fragments: Where Are We Now? *Biochem. Soc. Trans.* **2020**, *48*, 271−280.
(30) 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.
(31) Schmidt, D.; Boehm, M.; McClendon, C. L.; Torella, R.; Gohlke, H. Cosolvent- Enhanced Sampling and Unbiased Identification of Cryptic Pockets Suitable for Structure-Based Drug Design. *J. Chem. Theory Comput.* **2019**, *15*, 3331−3343.
(32) 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.
(33) Ghanakota, P.; Carlson, H. A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. *J. Med. Chem.* **2016**, *59*, 10383−10399.
(34) Ung, P. M. U.; 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.
(35) 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.
(36) 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.
(37) 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.
(38) Arenas-Ramirez, N.; Woytschak, J.; Boyman, O. Interleukin-2: Biology, Design and Application. *Trends Immunol.* **2015**, *36*, 763−777.
(39) Wells, J. A.; McClendon, C. L. Reaching for High-Hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, *450*, 1001−1009.
(40) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. Binding of Small Molecules to an Adaptive Protein-Protein Interface. *Proc. Natl. Acad. Sci. U.S A.* **2003**, *100*, 1603−1608.
(41) Dym, O.; Song, W.; Felder, C.; Roth, E.; Shnyrov, V.; Ashani, Y.; Xu, Y.; Joosten, R. P.; Weiner, L.; Sussman, J. L.; Silman, I. The Impact of Crystallization Conditions on Structure-Based Drug Design: A Case Study on the Methylene Blue/Acetylcholinesterase Complex. *Protein Sci.* **2016**, *25*, 1096−1114.
(42) Brear, P.; De Fusco, C.; Hadje Georgiou, K.; Francis-Newton, N. J.; Stubbs, C. J.; Sore, H. F.; Venkitaraman, A. R.; Abell, C.; Spring, D. R.; Hyvonen, M. Specific Inhibition of CK2α from an Anchor Outside the Active Site. *Chem. Sci.* **2016**, *7*, 6839−6845.
(43) Kinoshita, T.; Nakaniwa, T.; Sekiguchi, Y.; Sogabe, Y.; Sakurai, A.; Nakamura, S.; Nakanishi, I. Crystal Structure of Human CK2α at 1.06 Å Resolution. *J. Synchrotron Radiat.* **2013**, *20*, 974−979.
(44) Behnen, J.; Köster, H.; Neudert, G.; Craan, T.; Heine, A.; Klebe, G. Experimental and Computational Active Site Mapping as a Starting Point to Fragment-Based Lead Discovery. *ChemMedChem* **2012**, *7*, 248−261.

https://dx.doi.org/10.1021/acs.jcim.0c01126
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX