

Leveraging the Thermodynamics of Protein Conformations in Drug Discovery

Bin W. Zhang,^{*,‡} Mikolai Fajer,^{*,‡} Wei Chen, Francesca Moraca, and Lingle Wang^{*}



Cite This: *J. Chem. Inf. Model.* 2025, 65, 252–264



Read Online

ACCESS |



Metrics & More

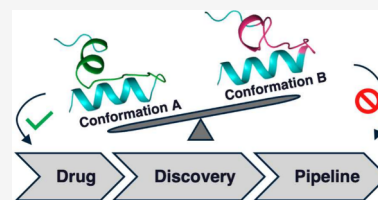


Article Recommendations



Supporting Information

ABSTRACT: As the name implies, structure-based drug design requires confidence in the holo complex structure. The ability to clarify which protein conformation to use when ambiguity arises would be incredibly useful. We present a large scale validation of the computational method Protein Reorganization Free Energy Perturbation (PReorg-FEP) and demonstrate its quantitative accuracy in selecting the correct protein conformation among candidate models in apo or ligand induced states for 14 different systems. These candidate conformations are pulled from various drug discovery related campaigns: cryptic conformations induced by novel hits in lead identification, binding site rearrangement during lead optimization, and conflicting structural biology models. We also show an example of a pH-dependent conformational change, relevant to protein design.



1. INTRODUCTION

Conformational changes are essential for many proteins to fulfill their biological functions. For example, the activation of protein kinases involves the conformational change of its activation loop, the DFG motif, and the α C helix.^{1,2} G protein-coupled receptors (GPCRs) rely on conformational changes upon ligand binding to pass messages through cell membrane and activate cellular responses.^{3,4} More generally, loop dynamics have been shown to be an important component of the catalytic mechanism of enzyme.² Therefore, accurate modeling of the thermodynamics of protein conformational changes is essential to understand the molecular mechanism of protein functions.^{5–7}

The concept of a conformational energy landscape dates back to the statistical mechanics of glasses, and was first applied to proteins by Wolynes et al. in 1991.⁸ The terminology can vary, especially between fields, so first we outline what we will use in this paper. A set of positions, for example from a X-ray crystal structure, will be referred to as a configuration. A configuration is static and has a potential energy associated with it. A conformational state, or conformation for short, is an ensemble of configurations. At a given temperature this conformational state will have a free energy.

Exploring the conformational energy landscapes of disease-related biological targets is also useful for rational drug design.^{9,10} The major objective of small molecule drug discovery program is to identify molecules that can potentially bind to the target protein while balancing other drug related properties. The most obvious contributions to small molecule binding potency are the interactions between the small molecule and the protein in the holo complex state as compared to the interactions with water in the apo state. If the protein structure rearranges upon binding, then the difference

between the conformational free energies for the apo and holo states, referred to as the reorganization free energy, will also contribute to the overall binding potency. Since the apo state is by definition the lowest free energy state in the absence of the ligand, this contribution will always be penalty. The “druggability” of a specific binding site protein conformation is thus a combination of the penalty to adopt that protein conformational state and the availability of favorable binding site interactions for ligands.

Most computational drug discovery implicitly handles the reorganization free energy by assuming that all investigated ligands bind to the same conformational state. In that case the reorganization free energies are the same for each ligand and either cancel out or can be estimated from a reference compound. This strategy has allowed relative and absolute binding free energy perturbation (RB-FEP and AB-FEP) to quantitatively predict binding affinities over a wide variety of systems and stages of drug discovery pipelines.^{11–22} The implicit assumption here is that the druggability of the protein conformational state has already been established, for example through crystallography of a known potent compound.

A novel target or a novel protein conformation for a known target will not have that luxury. If a high-throughput screen identifies a weak binder in a novel protein conformation, how does one assess if it is worth pursuing? Is a cryptic binding pocket found through computational methods like mixed

Received: September 26, 2024

Revised: November 20, 2024

Accepted: November 26, 2024

Published: December 16, 2024



solvent MD viable.^{23,24} Accurate modeling of the reorganization free energy could derisk pursuing these novel protein conformations, although the protein conformations would still need to expose favorable interactions to result in high potency compounds.

Other approaches have been tried to estimate the free energy differences between protein conformational states. The most straightforward method is to run a molecular dynamics (MD) simulation long enough to observe multiple transitions between these conformational states. The time scales of these transitions are frequently longer than practical simulation lengths, though.^{25,26} A more computationally efficient approach is to utilize advanced sampling algorithms.^{5,27,28} One class of algorithms divides the conformational landscape between the two states into multiple small regions. Umbrella Sampling forcibly explores each region by applying a biasing restraint, and then combines the regions together through reweighting the data.^{29–31} The Weighted Ensemble algorithm “splits” trajectories as they cross between regions, and by only progressing trajectories in the less populated regions one can achieve more frequent sampling of those regions than from brute force MD.^{32,33} Another class of algorithms like Metadynamics and Wang–Landau adaptively learns the conformational landscape, and proceeds to flatten out that landscape in order to make transitions more frequent.^{34,35} Most of these advanced sampling algorithms depend on identifying a set of collective variables, for example the distance between key hydrogen bond partners or protein backbone torsions. These are highly system-specific, and finding all of the important features for a conformational transition can be very challenging.

Protein Reorganization FEP (PReorg-FEP) is an algorithm that estimates the free energy difference between conformational states through an alchemical pathway.^{36–38} Unlike other enhanced sampling algorithms, PReorg-FEP does not require collective variables or transition pathways to be known beforehand. The only inputs for PReorg-FEP are the two configurations that represent different conformational states of the protein. We have tested PReorg-FEP on Abl kinase and HSP90 in a previous paper and obtained promising results.³⁸ In this paper, we validate PReorg-FEP by applying it to a large set of biological systems and comparing the computational results to experiments. We also demonstrate how these calculations can help inform rational drug design.

This paper is organized as follows. After a brief recap of the PReorg-FEP methodology detailed in a previous paper, we first report protein reorganization free energy calculations on 6 protein systems where the holo complexes require significant conformational changes as compared to the apo states. We demonstrate the quantitative accuracy of these reorganization free energies as they accurately reproduce the experimentally observed binding free energies when combined with the interaction free energies obtained from AB-FEP. Then we collect a number of systems where multiple protein conformations were reported either from different PDB entries corresponding to different crystallization conditions or from the same PDB entry corresponding to alternative conformations reflected in the diffraction data. We show that protein conformational free energy calculations among these different conformational states can help to differentiate the true native state from the other non-native states due to crystal contacts or other artifacts in the crystallization. Finally we demonstrate that the experimentally observed pH dependent conforma-

tional changes of the bovine β -lactoglobulin protein was accurately modeled by protein conformational free energy calculations combined with protein pK_a calculations. We then conclude the paper with a summary of key findings and future directions.

2. METHOD

Protein Reorganization Free Energy Perturbation (PReorg-FEP) is a method to calculate the free energy difference between two protein conformational states via an alchemical pathway. The protein region that undergoes the conformational change is defined as the alchemical part while the rest of the protein is the “shared” part. The free energy difference between the conformational state A and B is the sum of the free energy change of removing the alchemical region unique to conformational A from the system and the free energy change of adding the alchemical region unique to conformational B back to the system.^{36–38} The method only requires the two structures corresponding to different conformational states as inputs, and the free energy difference between the corresponding conformational ensembles is obtained from rigorous alchemical simulations.

The thermodynamic cycle of PReorg-FEP implemented in Schrodinger Software Suite is as follows.³⁸ In the first stage, we start from a single structure with one loop conformation (LoopA) and add a noninteracting copy of the alternative loop conformation (LoopB). The bond, angle, and improper torsion force field terms within loopB are retained, but a harmonic restraint replaces each torsion term with an equilibrium value determined by the input coordinates. Initially, the LoopB is broken into two halves by removing the bond, angle, and torsion terms involving the two peptide atoms in the middle of the loop. Reference [37] proved that using harmonic restraints gives the same free energy change of adding the noninteracting broken loop to a system regardless of equilibrium values. Therefore, the free energy change of adding LoopB will cancel when LoopA is removed from the system. The second stage is to replace each torsion term within loopA with a harmonic restraint, and the third stage is to reform the bond between the half-loops of LoopB with a soft-bond stretch potential.³⁹ Note that the system has two complete loops at the end of the third stage. LoopA interacts with the shared atoms, and LoopB does not. The torsion terms within both loops have been replaced by harmonic restraints. The fourth stage is to turn off the interactions between LoopA and the shared atoms and turn on the interactions between LoopB and the shared atoms. From this point on, stages one through three are reversed to complete the thermodynamic cycle. The noninteracting LoopA is broken into two half-loops. The harmonic restraints within the interacting loopB are replaced by the corresponding torsion force field terms, and finally, the noninteracting LoopA is removed from the system. See Figure S32 for the diagram of this thermodynamic cycle. Interested readers are referred to ref [37] and ref [38] for more details on the algorithm of PReorg-FEP.

In this study, many of the PReorg-FEP calculations were performed with restraints applied to the alchemical parts because we found that large movements of the alchemical part in PReorg-FEP calculations usually lead to slow convergence. Additionally, the conformational changes of some proteins are coupled with titratable states of several key residues. It is critical to apply restraints to the alchemical part so that the conformational free energy change and the state free energy

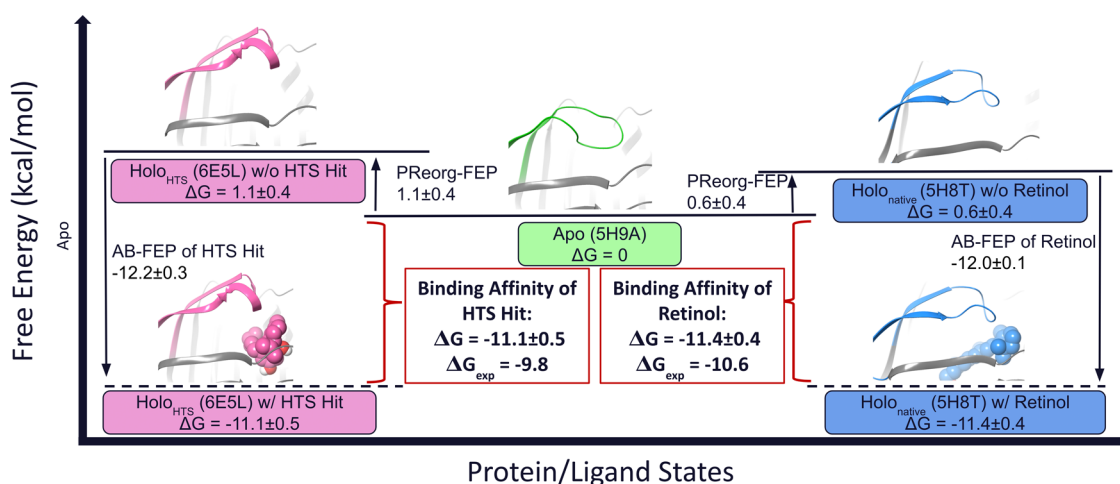


Figure 1. Free energy diagram for CRBP1. The apo (green, PDB 5H9A), native retinol-bound holo_{native} (blue, PDB 5H8T) and novel holo_{HTS} (pink, PDB 6E5L) states are each represented by a colored box and representative 3D structure. The colored portion of the 3D structure denotes the alchemical region for the PReorg-FEP calculation. The horizontal line associated with each state denotes the free energy for that state. The vertical arrows represent the PReorg-FEP or AB-FEP calculations. The AB-FEP calculations represent the ligand binding to a specific protein conformational ensemble, which can be combined with the apo-to-holo protein reorganization free energy to determine the *in vitro* absolute binding free energy.

change of titratable residues are separately estimated by PReorg-FEP and protein FEP, respectively. To obtain the free energy difference between the “free” initial and final conformational states, the restraints on the alchemical part were added gradually before PReorg-FEP and were removed gradually afterward. The free energy changes for adding and removing those restraints were reported with the result of PReorg-FEP together in the [Supporting Information](#).

The pK_a values of titratable residues were calculated by protein residue mutation FEP.^{22,40} The free energy difference ΔG between the protonated and deprotonated states of each titratable residue was estimated in both protein and solvent. Then the difference between ΔG in protein and solvent was used to calculate the pK_a in protein according to $\Delta\Delta G = RT \ln 10(pK_a - pK_a^0)$ where pK_a^0 is the intrinsic pK_a of the isolated residue in solvent. The free energy of deprotonating the residue in protein is $\Delta G = -RT \ln 10(pH - pK_a)$, where pH is the acidity of water.

3. RESULTS AND DISCUSSION

3.1. Protein Conformational Changes upon Ligand Binding. There are two prevailing models of protein–ligand binding that involve a rearrangement of the protein conformation: induced fit and conformational selection. In the conformational selection model there exists some population of proteins that adopt the holo-like conformation in the absence of any ligand, and when a ligand finds one of these proteins it “selects” it and binds to form the holo complex. This is a useful analogy for our computational approach, where we use PReorg-FEP to determine the reorganization free energy to adopt the holo-like conformational state of the protein in the absence of the ligand and then AB-FEP to determine the interaction free energy of the holo protein–ligand complex. AB-FEP has been extensively validated to accurately calculate the interaction free energies between a small molecule and a target protein conformational state and has delivered promising results in enhancing hit discovery in early stage of drug discovery programs.^{13,18–20,22} Previously, the reorganization free energy to adopt the holo-

like state was approximated from the offset between AB-FEP and experimental affinities for a large number of reference compounds. We will use this indirect estimation as a benchmark to quantitatively validate the PReorg-FEP method on a number of different protein systems. These cases range from high throughput screening (HTS) for hit selection to R-group modifications for lead optimization,^{41,42} and we show that we can quantitatively reproduce the experimental affinities. A comparison of the reorganization free energy to the strongest affinity ligand for each studied protein conformational state is also used to propose general guidelines for what constitutes too high of a reorganization free energy to pursue.

3.1.1. Conformational Changes Induced via HTS Hits. In the early stages of a drug discovery project it is typical for many diverse compounds to be tested against the target of interest. A high throughput screen (HTS) emphasizes a broad chemical space and the best hits may only reach the high μM potency range before optimization.⁴¹ For a novel protein conformation it can be difficult to know if the low potency is due to poor interactions that can be optimized or if it is due to high protein reorganization cost that is likely to persist during lead optimization. The ability to distinguish between these two cases, especially at an early stage, has the potential to save a lot of time and effort.

3.1.1.1. Cellular Retinol Binding Protein 1 (CRBP1). Cellular Retinol Binding Protein 1 (CRBP1) is the protein that carries and transports retinol (vitamin A₁) through the cytoplasm. CRBP1 is a potential therapeutic target to regulate the flux of retinoids,⁴³ and a possible carrier to deliver therapeutic agents due to its ability to bind and transport hydrophobic molecules.⁴⁴ The core structure of CRBP1 is a flattened β -barrel formed by two antiparallel β -sheets. The open end of this β -barrel is protected by its βC - βD , βE - βF hairpins and two α -helices, which is collectively referred to as the “portal region”. Conformational changes of the hairpin βE - βF region on CRBP1 are induced upon retinol binding (holo_{native}, PDB 5H8T) that opens access to the binding site relative to the apo state (PDB 5H9A).⁴⁴ HTS aiming for a nonretinoid inhibitor identified an ~ 67 nM compound that

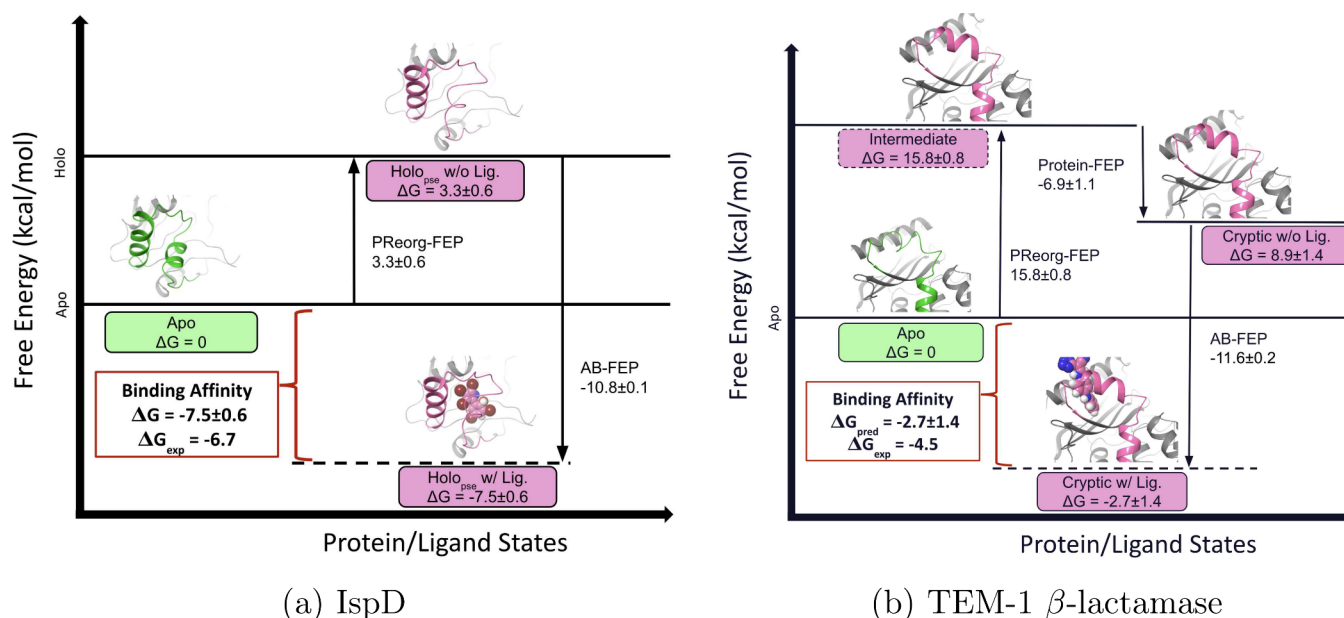


Figure 2. Free-energy diagram for the IspD (apo 4NAI and holo_{pse} 4NAK)) and TEM-1 β -lactamase (apo 1FQG and cryptic 1PZP) systems. Each conformation/ligand state is denoted by a colored box and representative 3D structure, with the vertical position denoting its free energy. The colored portion of the 3D structure denotes the alchemical region for the PREorg-FEP calculation. The vertical arrows represent PREorg-FEP or AB-FEP calculations. The AB-FEP calculations represent the ligand binding to a specific protein conformation, which can be combined with the apo-to-holo protein reorganization free energy to determine the *in vitro* absolute binding free energy. In the diagram of TEM-1 β -lactamase, the FEP calculations are separated into the PREorg-FEP part and the Protein FEP part.

shows a different holo conformation (holo_{HTS}, PDB 6ESL).⁴⁵ With the hit being so potent before optimization, one would expect the reorganization free energy of this novel conformational state to be minimal.

The calculations are summarized in Figure 1. PREorg-FEP predicts a small apo-to-holo_{HTS} reorganization free energy of $\Delta G_{a \rightarrow h_{HTS}} = 1.1 \pm 0.4$ kcal/mol. Combining the reorganization free energy with the AB-FEP calculated interaction free energy between the ligand and the protein, the calculated binding free energy of -11.1 ± 0.5 kcal/mol for the HTS compound is a close match to the experimental result of -9.8 kcal/mol. The apo-to-holo reorganization free energy for holo_{native} is predicted to be even smaller at $\Delta G_{a \rightarrow h_{native}} = 0.6 \pm 0.4$ kcal/mol. We also compared conformational reorganization free energy corrected AB-FEP results with experimental data for a nine other ligands binding to CRBP1, and the good agreement between calculated versus experimental results shown in Figure 3 further confirmed the small reorganization free energy of CRBP1.

As stated previously, the “druggability” of a specific protein conformation is a combination of a low reorganization free energy and favorable interactions available to the ligand. CRBP1 is an ideal case, where a very low reorganization free energy allows a binding site to open up with sufficient interactions for the greasy retinol-like small molecules. Therefore, it is not surprising that many potent inhibitors have been found for CRBP1.

3.1.1.2. IspD. IspD is the catalyst for the third step of the nonmevalonate pathway for isoprenoid biosynthesis in plants and pathogenic organisms. It is an attractive target for developing herbicides and anti-infective drugs, and no side-effects are expected for IspD inhibitors since the nonmevalonate pathway does not exist in mammals.^{46,47} In a previous HTS study by Kunfermann et al., a pseudilin-type

inhibitor was found as a μ M hit, and cocrystallized in a different conformation (holo_{pse}, PDB 4NAK) than the apo (PDB 4NAI). In the holo_{pse} conformation the flexible loop Glu255-Tyr268 loop moves away to open a cryptic binding pocket in close proximity to the active site (Figure S3).

The reorganization free energy for apo-to-holo_{pse} estimated by PREorg-FEP was $\Delta G_{a \rightarrow h_{pse}} = 3.3 \pm 0.3$ kcal/mol. In the left panel of Figure 2, we plotted the free energy levels of the apo, holo_{pse} with ligand, and holo_{pse} without ligand states according to our AB-FEP and PREorg-FEP calculations. Combining the apo-to-holo_{pse} reorganization free energy and the AB-FEP estimated interaction free energy between the ligand and the protein, our predicted binding free energy of -7.5 ± 0.6 kcal/mol is in very good agreement with the experimental result of -6.7 kcal/mol.

The apo-to-holo_{pse} reorganization free energy of 3.3 kcal/mol for IspD suggests a moderate state penalty to open the cryptic pocket, indicating potent binding inhibitors may still exist. Upon searching the literature we found the same group of researchers identified azolopyrimidine-type inhibitors with binding affinities in the nM range for IspD.⁴⁶ The X-ray structure of one of the azolopyrimidine-type inhibitors identified a slightly different conformation of the cryptic pocket, holo_{azo} (PDB 2YC5, Figure S3). The apo-to-holo_{azo} reorganization free energy estimated by PREorg-FEP was $\Delta G_{a \rightarrow h_{azo}} = 4.4 \pm 0.8$ kcal/mol. The calculated binding free energy of -9.6 ± 1.0 kcal/mol for the 2YC5 ligand after taking into account the protein reorganization free energy is a very close match to the experimental result of -10.2 kcal/mol (Figure S5).

To further validate the reorganization free energy difference for the above two holo protein conformations we performed AB-FEP calculations for the entire congeneric series of 5 pseudilin-type inhibitors⁴⁷ and the 14 azolopyrimidine-type

inhibitors⁴⁶ to their respective holo conformations. We found the indirect estimate of the protein reorganization free energy from AB-FEP (average offset between raw AB-FEP results and experimental binding free energy for each chemical series) to be 2.9 and 5.4 kcal/mol for the pseudilin-type and azolopyrimidine-type inhibitors, respectively. Both the trend of $\text{holo}_{\text{azo}} > \text{holo}_{\text{pse}}$ and the rough magnitudes match the direct PReorg-FEP predictions.

3.1.1.3. TEM-1 β -Lactamase (TEM). β -lactamase, a well-known therapeutic target, is an enzyme responsible for antibiotic resistance via breaking the β -lactam ring in antibiotics such as penicillins.⁴⁸ A previous study by Horn and Shoichet identified an inhibitor with affinity in high μM range which opens a cryptic site of *Escherichia coli* TEM β -lactamase (PDB 1PZP) which is about 16 Å away from the active site (Figure S6).⁴⁹ The cryptic site opens by helix 11 (residue Gly218-Ala224) shifting 3–7 Å away from helix 12 (residue Asp271-Lys286). Here we consider the apo state to be the acyl-enzyme intermediate of β -lactamase (PDB 1FQG) without the covalent ligand.

In addition to the conformational difference, we predicted the preferred protonation states of two titratable residues (Asp214 and Asp233) also changed upon the opening of the cryptic pocket. In the apo state, Asp214 is protonated and Asp233 is unprotonated, while in the holo state, Asp233 is protonated and Asp214 is unprotonated. The titratable states of these two residues are strongly coupled with the conformation of helix 11: in a MD simulation starting from the apo conformation but with deprotonated Asp214 and protonated Asp233, the protein will automatically transition to the holo conformational state during a 75 ns trajectory, while the protein will automatically transition from the holo conformation to the apo conformational state when the protonation states for these two residues are reversed. The average waiting times for these transitions taken over 4 different simulations are about 15 ns in either direction.

We applied PReorg-FEP and to estimate the free energy difference between the apo and holo states of TEM. Since the conformational change was coupled with the protonation state change of Asp214 and Asp233, we also perform pK_a calculation to estimate the free energy difference between the different protonation states for each protein conformational state. The detailed protocol for pK_a calculations and how it was combined with protein-Reorg FEP to estimate the free energy difference between the apo and holo states is provided in the [Supporting Information](#) along with the raw data. The estimated apo-to-holo free energy difference of 8.9 ± 1.4 kcal/mol (Figure 2b) indicates a large state penalty to open the cryptic pocket, suggesting finding inhibitors with strong binding affinity for this cryptic binding site might be challenging. Up to now, there is no potent inhibitors binding to this cryptic pocket reported in the literature. The calculated binding free energy of -2.7 ± 1.4 kcal/mol for the allosteric ligand is also in reasonable agreement with experimental value of -4.5 kcal/mol considering the large uncertainty in both calculations and experiment at such weak binding.⁴⁹

3.1.2. Conformational Changes Induced by Chemical Modifications on Known Ligands. During lead optimization the initial core compound undergoes chemical modifications from nitrogen walks to adding R-groups.⁴² In some cases these modifications cause a rearrangement of the binding site, which can confound structure based drug design since there is ambiguity in exactly what conformation each design should be

targeting for any specific modification. To accurately predict the affinity of such compounds it becomes necessary to accurately predict what conformational state a proposed compound will bind to.

3.1.2.1. Heat Shock Protein 90 (HSP90). Heat Shock Protein 90 (HSP90) is a molecular chaperone that has been implicated in cancer by stabilizing the underlying mutant of overexpressed signaling proteins.⁵⁰ This protection requires HSP90 to be able to convert ATP to ADP, and thus competitive inhibitors to the nucleotide binding site have shown preclinical success as cancer therapeutics.⁵¹ The nucleotide binding site is surrounded by one stable helix (Phe44-Thr65) and the ATP binding lid (Lys100-Gln123), the latter of which can undergo large conformational changes. Inhibitors have targeted the ADP-bound “open” state, but even there the binding lid shows multiple conformations ranging from “loop-in” to “helical” and “loop-out”. The most stable conformation appears to be the “loop-in”, resolved in both apo (PDB 3T0H) and small ligand crystal structures (PDB 2VCJ, 2VCI, 3B24, 5J64, 4LWG, 4LWI, 3BM9, and others). However, adopting the “helical” and “loop-out” states opens up a lipophilic pocket which has been shown to increase binding affinity in congeneric series.^{52–54} The reorganization free energy from “loop-in” (PDB 5J64) to “helical” (PDB 5J20) in the absence of any ligand was predicted to be 1.9 ± 0.7 kcal/mol (Figure S10). The binding affinity for seven compounds with “helical” crystal structures from four different chemical series was determined by combining raw AB-FEP and PReorg-FEP free energy resulting in a MUE of 1.3 kcal/mol, with five of the compounds within 0.8 kcal/mol (Figure 3). One caveat to this relative to the other systems is that an additional 4.7 kcal/mol correction was applied to compensate for AB-FEP overpredicting the reference “loop-in” compound (PDB 5J64). This specific example highlights how PReorg-FEP can still assist when there are doubts of the true apo state of a protein but there are known binders to a second protein conformation that can be used as a reference.

3.1.3. Checking Assumptions of Known Holo Complexes. Drug design frequently starts from a known holo complex, be it an inhibitor or the natural substrate. This can be attributed to the assumption that the previously observed holo complex has a relatively low protein reorganization free energy since it has already been shown to be “successful”. Assumptions like this can be directly tested using PReorg-FEP, and can either give confidence to move forward with the known holo complex or prompt a novel direction to pursue.

3.1.3.1. Cyclin-Dependent Kinase 2 (CDK2). Cyclin-dependent kinase 2 (CDK2) is involved in the cell proliferation, and is an attractive target for inhibiting tumor growth in cancer.⁵⁵ A substituted guanine core was identified as a potential ATP-competitive inhibitor with low μM affinity. However, crystallization of this core (PDB 1H1P) shows a P-loop conformation distinct from either the apo (PDB 1H27) or ATP-bound (PDB 1B38) proteins, as seen in Figure S8. We performed PReorg-FEP from the apo to the inhibitor-bound conformation in the absence of any ligand for the P-loop (Val7-Ala21). The protein reorganization free energy was only 1.5 ± 0.4 kcal/mol, which suggests that targeting a different P-loop conformation is not required to achieve good potency. Lead optimization of the substituted guanine core was able to achieve a single digit nM binder, confirming that conclusion.⁵⁶ Our PReorg-FEP also quantitatively agrees with the 1.2 ± 0.2 kcal/mol reorganization energy estimated from performing

AB-FEP on a set of 16 members of the congeneric series (Figure 3).²²

3.1.3.2. Abl Kinase (Abl). Abl is a tyrosine kinase which can become constitutively active in the BCR-ABL fusion protein and is a major contributor to several types of leukemias. The first generation inhibitor imatinib was crystallized in a “collapsed” P-loop (PDB 1IEP) conformation, and second generation inhibitors continued to target this conformation. There was no known apo structure, although the apo structure for the homologous c-Src kinase showed an “extended” P-loop conformation (PDB 1FMK).⁵⁷ We previously evaluated the reorganization free energy from the “extended” to “collapsed” P-loop conformational states for Abl kinase without a ligand in the DFG-in state using PReorg-FEP and determined that the “collapsed” P-loop conformation state was less stable by 4.5 ± 1.2 kcal/mol.³⁸ In conjunction with AB-FEP we showed that the affinity boost to sub-nM of dasatinib relative to other second generation DFG-in inhibitors was due to targeting the more stable “extended” P-loop conformation.

3.1.3.3. Antimethotrexate VHH Antibody (Anti-MTX VHH). Antibodies have strong affinity and selectivity for their specific targets, be it protein or small molecule. This has been leveraged beyond the typical immune response, from the detection of specific small molecules in a sample to the removal of toxic small molecules from the body.⁵⁸ A subset of IgG antibodies in the *camelidae* family completely lack light chains, resulting in a minimal antibody with a single variable domain (VHH) and are an attractive framework for diagnostic and therapeutic uses. One such example is the antimethotrexate VHH (anti-MTX), which was able to reach low nM affinity for the chemotherapy drug methotrexate despite the reduced size of a VHH.⁵⁹ The three hypervariable loops of the anti-MTX VHH were grafted onto a different VHH more amenable to crystallization (CDR1–3 graft) and solved in the apo (PDB 3QXU) and methotrexate-bound states (PDB 3QXT).⁶⁰ This construct was only able to reach μ M affinity, though. Comparison of the apo and holo structures shows two distinct conformational changes upon binding methotrexate. First, one of the hypervariable loops changes conformation to form the majority of the binding site (loop1, Ser23-Met36, Figure S12). Second, a nonhypervariable loop (loop2, Ile72-Gln84) unexpectedly moves, opening a deeper pocket.

We performed PReorg-FEP on the two loops in sequential calculations and determined an overall protein reorganization free energy of 1.8 ± 0.6 kcal/mol to open up the methotrexate binding site. Such a low reorganization free energy suggests this expanded binding site is not the reason for the μ M affinity of the construct. In order to reproduce the nM affinity of the original VHH, five residues in the nonhypervariable loop2 were mutated to reproduce the original VHH sequence (CDR1–4 graft). The holo conformation for the CDR1–4 complex (PDB 3QXV) has the same conformation as the CDR1–3 complex, suggesting the boost in affinity is coming from better side chain interactions with methotrexate rather than a new conformation. We confirmed this with AB-FEP, showing a gain of 4.3 ± 0.8 kcal/mol after mutating loop2. Taken together these calculations suggest that the expanded binding site is relevant for any small molecule VHH binding, and that mutations of the nonhypervariable loop2 are a novel source for both affinity and selectivity.

3.1.4. Summary of Validation. The PReorg-FEP and AB-FEP calculations for the systems studied in this section are summarized in Table 1 and Figure 3. The reorganization free

Table 1. Summary of Binding Affinity Predictions in the Study of Druggability of Different Protein Conformations (kcal/mol)^a

System	Reorg ΔG	No. Compounds	MUE w/o Reorg	MUE w/Reorg
CRBP1 (HTS)	1.1 ± 0.4	10	1.1 ± 0.8	0.9 ± 0.7
CRBP1 (Retinol)	0.6 ± 0.4	1	1.4	0.8
IspD (pse)	3.3 ± 0.3	5	2.9 ± 1.0	1.0 ± 0.4
IspD (azo)	4.4 ± 0.8	14	5.4 ± 1.5	1.5 ± 1.0
TEM	8.9 ± 1.4	1	7.1	1.8
HSP90	1.9 ± 0.7	7	2.3 ± 1.6	1.3 ± 1.6
CDK2	1.5 ± 0.4	16	1.4 ± 1.0	0.9 ± 0.7
anti-MTX VHH	1.8 ± 0.6	1	2.0	0.2

^aMUEs are the mean unsigned errors between our binding affinity predictions and the experimentally observed results.

energies report the mean and standard deviation from six PReorg-FEP calculations, three going from apo to holo and three going from holo to apo. All of the reorganization free energies are statistically significant from zero. The AB-FEP calculations for each ligand are a mean and standard deviation of three simulations. The mean unsigned error from experiment (MUE) for the compounds known to bind a specific protein conformational state is shown for the raw AB-FEP results, with the standard deviation of the unsigned error reported as an error. The raw AB-FEP results have poor performance for most of the systems. The reorganization free energy for each conformational state is then added to the raw AB-FEP results, and the new MUE calculated. The estimated affinity taken from combining AB-FEP and PReorg-FEP consistently lowers the MUE relative to the raw AB-FEP calculations. In the majority of cases the reorganization corrected MUEs drop to ≤ 1.0 kcal/mol.

3.1.5. Role of Apo-to-Holo Reorganization Free Energy in Druggability Assessment. Of the 9 distinct holo conformational states studied in this section, only the cryptic binding site for TEM-1 β -lactamase had a reorganization free energy exceeding 5 kcal/mol. It is also the protein conformational state where the strongest affinity for a known binder could not break at least low μ M. A conformation with a high reorganization free energy would have to provide excellent ligand interactions in order to support a high potency compound. A low reorganization free energy is not sufficient for a strong affinity compound either, as can be seen in the cases of anti-MTX and IspD_{pse}. If a compound cannot make strong interactions with a protein conformational state then it does not matter how easy it is to adopt that conformation. For a rigorous assessment of a protein conformational state's druggability both the reorganization and potential ligand interactions must be evaluated.

However, we anticipate situations where the reorganization free energy will be predicted before a project progresses into lead optimization and the binding site interactions are fully evaluated. In that situation it would be useful to have some general guidelines on what is an “acceptable” versus a “too high” reorganization free energy, even if it is an incomplete picture for any specific protein conformational state. To this end we plot the strongest affinity compound for each of the studied holo conformational states against the corresponding reorganization free energy in Figure 4. The criteria for an acceptable affinity will be program dependent, so we highlight a ≤ 1 nM and ≤ 100 nM desired K_d . A 1 nM K_d corresponds to

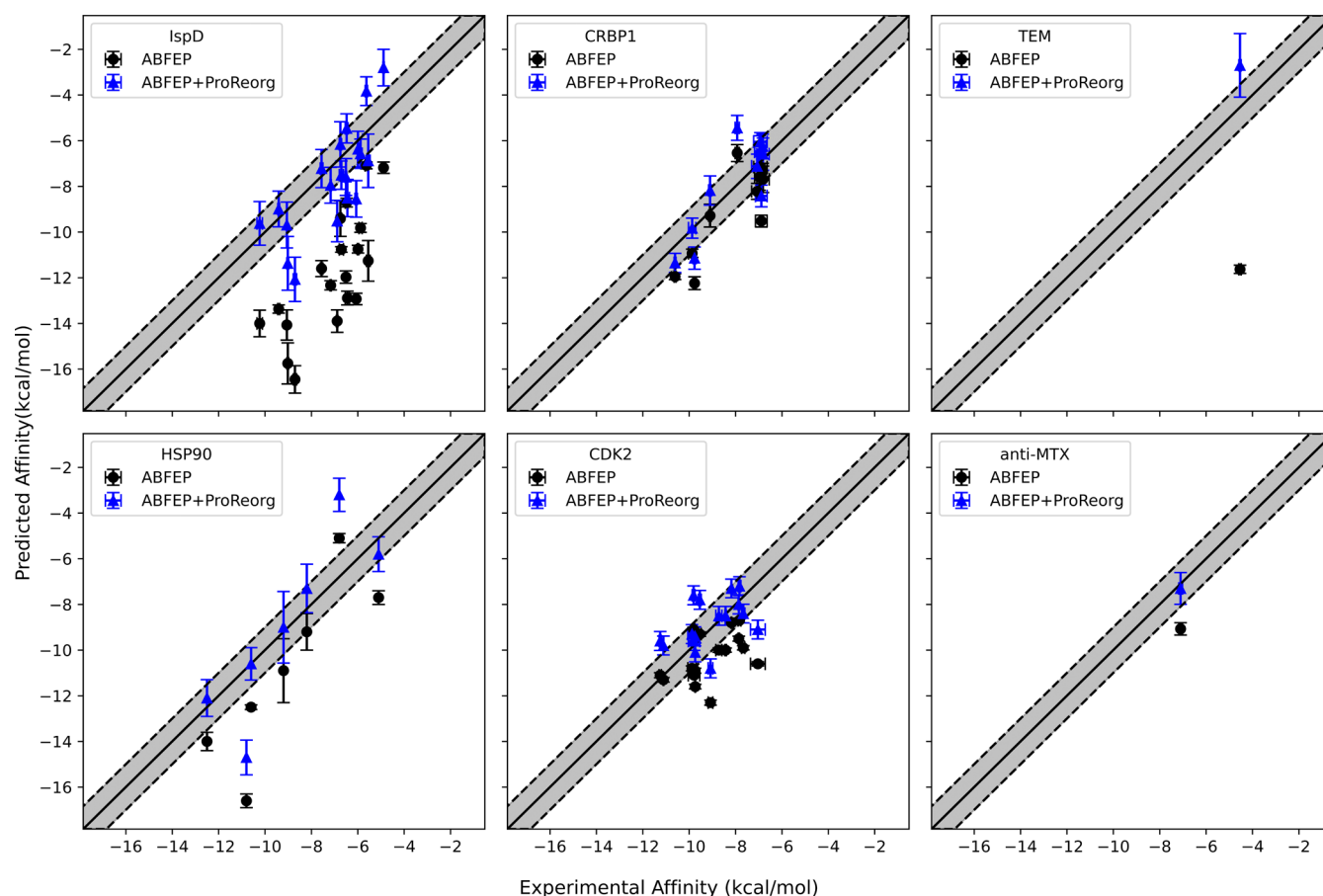


Figure 3. Comparison between ABFEP and experimental results. The raw ABFEP results and the ABFEP results with reorganization free energy corrections are compared with the experimental binding affinities for each system. As it shows, ABFEP usually overestimates the binding free energy. However, for systems with a known apo and holo conformation, PReorg-FEP successfully determines the reorganization free energies and significantly improves the agreements between our binding affinity predictions and the experimental results.

a ΔG of -12.3 kcal/mol, and the highest reorganization free energy to pass that threshold is HSP90 at 1.9 ± 0.7 kcal/mol. This establishes the upper limit of 2.6 kcal/mol reorganization free energy as a guideline for progressing a specific protein conformational state further in the drug discovery pipeline (dotted line in Figure 4). The same approach for 100 nM K_d results in a 5.7 kcal/mol upper limit (dashed line in Figure 4). These cutoffs represent the best case scenarios we have observed so far, and should be updated as additional systems are studied. It would also make sense to restrict this comparison to a family of proteins, for example different protein kinases, to increase the transferability within the family.

3.2. Selecting from Competing Structural Models. So far the majority of protein structures we have used were resolved via X-ray crystallography. In general, these structures are believed to be reliable representations of the native state within the cell. It is not hard to find conflicting cases, though, where experimental conditions such as high salt concentrations, low temperature, acidic or basic pH environment, or packing among crystal mates result in different structures. There are also PDB entries with alternative conformations that either can both fit or both are *required* to fit the available experimental data. The ability to differentiate between multiple candidate structures and identify the native structure would help focus drug design on the most relevant conformational states.

In this section we collect a number of systems where multiple conformations of the same protein were reported in the PDB, and perform protein reorganization free energy calculations between these conformational states. The results are summarized in Table 2. We find that alternative conformations reflecting the flexibility of the protein indicated by the diffraction data have relative small free energy differences, as should be expected. For conformational states coming from different crystallization conditions the reorganization free energy can be quite large, and the conformational states with the lowest free energies usually are closest to the true native environment with the least artifact from crystal packing or interactions with other components in the crystal structures. The proteins presented here are not all relevant to drug discovery, but do extend the quantitative validation of the PReorg-FEP method as well as show how these general issues in structural biology can be resolved.

3.2.1. Alternate Conformers in the Same PDB. Crystallography solves the structures of proteins in a highly ordered crystal lattice with a repeating asymmetric unit. The majority of diffraction experiments are run at 100 K, and the low temperature in conjunction with the crystal environment typically results in well-defined side chain rotamers in addition to backbone conformations.⁶¹ In some cases, though, multiple side chain or even loop backbone conformations can appear in the asymmetric unit and are labeled as “alternate conformers”. One possibility is that the experimental data is degenerate and

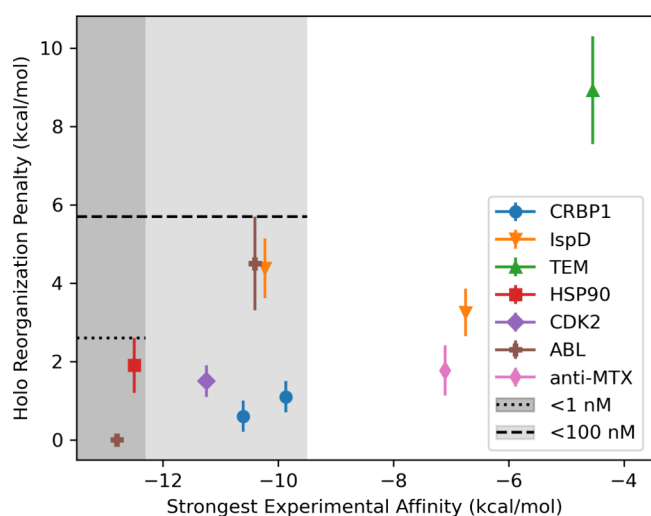


Figure 4. Strongest binding affinity compound for each specific protein conformational state studied is compared with the corresponding apo-to-holo reorganization free energy. Error bars for the reorganization free energy are based on the standard deviation of six independent calculations, three for each direction of the perturbation. The dark gray region denotes affinities with a $K_d \leq 1$ nM. The dotted horizontal line marks the highest observed reorganization free energy (including error bars) with a $K_d \leq 1$ nM compound, which is 2.6 kcal/mol. The same analysis is repeated for $K_d \leq 100$ nM, shown by the light gray region and the dashed line, and corresponds to a 5.7 kcal/mol reorganization free energy. The druggability of a protein conformational state is dependent on both the reorganization free energy and the potential ligand interactions available. If only the reorganization free energy is known, though, and it exceeds these threshold values then it seems less likely that a strong affinity compound will be found.

Table 2. Summary of Systems with Competing Structural Models

System	Reorg ΔG (kcal/mol)	Explanation
PTP1B	1.0 ± 0.6	Alternate conformer
eIF4E	0.7 ± 0.6	Alternate conformer
SpGH92	0.3 ± 1.0	Alternate conformer
Tim21	5.2 ± 0.4	Crystal packing artifact
GLTP	4.4 ± 0.7	Hydrocarbon binding
CIB1	6.9 ± 1.1	Contaminant from protein purification
AmpC	5.7 ± 0.5	Crystal packing artifact

can be fit by either conformation, so they are both provided for completeness. An alternative is that the different protein conformations actually coexist in the crystal lattice, in which case it is likely these conformations are both thermally accessible and have free energies within a few $k_B T$ of each other. The latter case offers a nice positive control for PReorg-FEP, where the reorganization free energy between the alternate conformers should be low. Here, we applied PReorg-FEP to estimate the free energy differences between the alternate conformers found in three systems: Protein-tyrosine phosphatase 1B (PTP1B), eukaryotic translation initiation factor 4E protein (eIF4E), and SpGH92 protein.^{62–64}

3.2.1.1. Protein-Tyrosine Phosphatase 1B (PTP1B). Protein tyrosine phosphatases (PTPs) and protein tyrosine kinases (PTKs) together control protein tyrosine phosphorylation, which is a key step of the signaling pathways in cell.⁶⁵ Although

PTPs are studied less often compared to PTKs, PTP1B is a well studied therapeutic target for diabetes, obesity, and cancer.⁶⁶ In PTP1B, three loops — the WPD loop (that contains Trp179, Pro180, and Asp181), the pTyr loop (that contains Tyr46), and the Q loop (that contains Gln262) — form a cleft containing the active site. It is known that substrate binding is accompanied by a closing movement of the WPD loop during the catalysis.^{65,67} The WPD loop also exhibits an open and closed conformations in the room-temperature crystal structure of apo PTP1B (PDB 6B8X),⁶² as shown in Figure S15. The free energy difference between the open and closed states of PTP1B estimated by PReorg-FEP is 1.0 ± 0.6 kcal/mol, well within the expected few $k_B T$.

3.2.1.2. Eukaryotic Translation Initiation Factor 4E (eIF4E). Eukaryotic translation initiation factor 4E (eIF4E) is one of the proteins that help ribosomes to recognize the capped 5'-end of mRNAs in eukaryotic cells. It has become a therapeutic target since research found that eIF4E is overexpressed in many human cancers.⁶³ In Figure S17, we compared the alternate conformations of the α_1 helix and adjacent loop observed in PDB 4TQB (chain A, Trp73-Gly88).⁶³ The free energy difference between these two conformational states estimated by PReorg-FEP is 0.7 ± 0.6 kcal/mol, also within a few $k_B T$.

3.2.1.3. SpGH92. SpGH92 protein is an α -(1,2)-mannosidase that plays an important role for the human pathogen *Streptococcus pneumoniae* to process N-linked glycans.⁶⁴ The loop Gly352-Gly356 in the C-terminal (α/α)₆ barrel domain of SpGH92 can be modeled in either a retracted (from the active site) conformation or an engaged conformation in PDB 5SWI (Figure S19). Our PReorg-FEP calculations estimates the free energy difference between the retracted and engaged conformational states is 0.3 ± 1.0 kcal/mol, which is also in the order of a few $k_B T$.

3.2.2. Different Conformations Due to Crystallization Conditions. Crystallization conditions are frequently varied for a given protein in an effort to improve experimental resolution or capture a distinct conformational state. For example, the pH of the solution can change the location and number of charged protein residues, stabilizing or destabilizing a specific conformation. These effects can compound and result in a different crystal lattice geometry, which alters the packing of each asymmetric unit with the others. Given two conformations that are putatively the same state, for example apo, and only differ in crystallization conditions it would be helpful to compare them and determine which has a lower free energy and thus more likely to represent the native conformation.

3.2.2.1. Tim21. Tim21 is a subunit of the TIM23 protein complex, a translocase of the mitochondrial inner membrane to import precursor proteins. The first X-ray structure in 2006 showed loop 2 of the Tim21 protein packed between neighboring crystal mates with very low B-factors (PDB 2CIU, Figure 5 “Crowded”).⁶⁸ Eight years later a solution NMR structure showed loop 2 as a highly flexible loop with multiple possible conformations fitting the available data (PDB 2MF7), none of which matched the original X-ray structure.⁶⁹ Bala et al. then demonstrated that by fusing Tim21 to the larger maltose binding protein a *crystal contact-free space* (CCFS) could be created for loop 2, resulting in a conformation similar to those in the NMR ensemble (PDB 6K7E, Figure 5 “CCFS”).⁷⁰

PReorg-FEP between the “crowded” and CCFS conformational states estimated that the CCFS conformation was more

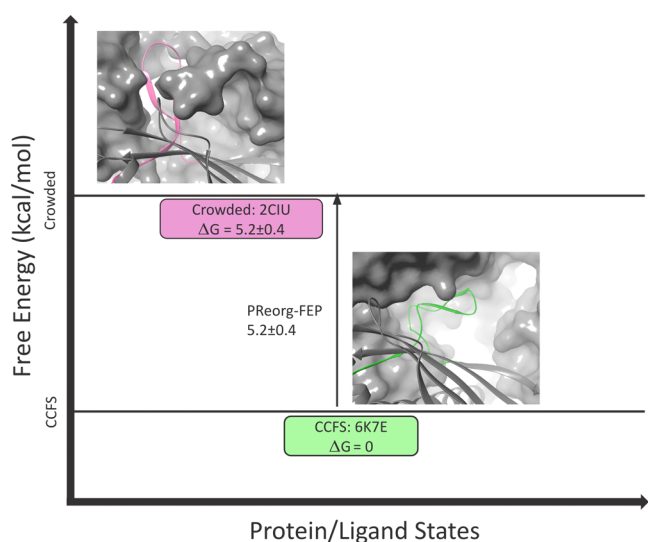


Figure 5. Crystal packing of Tim21. The tight crystal packing around the Tim21 loop in 2CIU (pink) forces it to adopt a high free energy conformation. A second conformation (green) was determined through a fusion protein. PReorg-FEP identifies the second conformation as a lower free energy state.

stable by 5.2 ± 0.4 kcal/mol. The prediction correctly identified the native-like CCFS conformation as the favorable state. The magnitude is suggestive of how large an effect crystal packing can play in structural biology, and the importance of having a tool to validate packed loops.

3.2.2.2. Glycolipid Transfer Protein (GLTP). GLTP is a protein that accelerates the transfer of glycolipids between membranes.⁷¹ The earliest X-ray structures in 2004 compared an “apo” (apo_{2004}) and a lactosylceramide holo complex (PDB 1SWX and 1SX6, respectively).⁷² These showed a binding site rearrangement of the loop connecting the $\alpha 1$ and $\alpha 2$ helices upon lactosylceramide binding (Figure S23, left panel). However, the apo_{2004} structure has a hexane molecule within the binding site. The effect of this hexane molecule was not appreciated until 2011 when a true apo structure (apo_{2011}) showed the loop in a similar conformation as the lactosylceramide holo complex, suggesting that there is no reorganization of this loop upon glycolipid binding after all (Figure S23, right panel).⁷³ We will refer to the lactosylceramide protein conformation as holo-like when the ligand is removed.

First, we used PReorg-FEP to compare the apo_{2011} and holo-like conformational states that visually look quite similar. The free energy difference was predicted to be -0.3 ± 0.6 kcal/mol, suggesting these are indeed the same conformational state and that no reorganization is required to bind lactosylceramide.

Second, we determined if PReorg-FEP would have helped identify the issue with only the 2004 X-ray crystal structures available. The free energy difference between the apo_{2004} and holo-like conformational states was -4.4 ± 0.7 kcal/mol, clearly showing that the hexane-bound apo_{2004} conformation was unfavorable. This illustrates how PReorg-FEP can be used to gain confidence in the structural biology before interpreting differences as meaningful, potentially leading a project astray.

3.2.2.3. Calcium and Integrin Binding Protein (CIB1). CIB1 is a small protein found throughout the body that has been associated with a broad set of functions, from regulating

platelet aggregation to calcium signaling and more recently cancer and cardiovascular disease.⁷⁴ One of the earliest crystal structures of CIB1 (PDB 1Y1A) had two chains in the asymmetric unit, one of which had a reduced glutathione (GSH) bound. Comparison of the apo to GSH-bound chains shows a short α -helix becoming a random coil (Val45-Arg50) (Figure S25).⁷⁵ The authors took this as evidence that CIB1 may be redox regulated, only briefly mentioning the possibility of contamination from the GSH-affinity column used during protein purification. PReorg-FEP between the apo and GSH-bound conformations shows the GSH-bound conformational state has a high free energy of 6.9 ± 1.1 kcal/mol. This points to purification artifact, more or less confirmed by the lack of any additional mention of CIB1 redox activity in the literature.

3.2.2.4. AmpC β -Lactamase. During lead optimization of non- β -lactam inhibitors for the class C β -lactamase AmpC three holo crystal structures were determined.⁷⁶ All three had two chains in the asymmetric unit, and at least one of the chains retained a binding site conformation similar to the apo structure (PDB 2BLS).⁷⁷ However, the second chain for two of the inhibitors adopted a distinct conformation for residues Asn289-Ala292 (Figure S27). Closer inspection of the crystal packing for these residues shows the formation of a short pair of parallel β -strands between the chain and the crystal mate that could be perturbing the structure (Figure S29).

In order to determine if this distinct conformation was relevant in a noncrystal environment we performed PReorg-FEP with the higher resolution holo structure (PDB 4JXV). Chain B has the apo-like conformation while chain A has the distinct conformation, and we will refer to them as 4JXV_B and 4JXV_A, respectively. The free energy difference between the apo and distinct 4JXV_A conformational states was 5.7 ± 0.5 kcal/mol. This high magnitude suggests that the 4JXV_A conformation is only relevant in that specific crystal environment and should not be used for drug discovery. By contrast the free energy difference between the apo and 4JXV_B conformational states was -0.7 ± 0.3 kcal/mol, predicting that they are roughly equivalent.

3.3. pH-Dependent Conformational Free Energy.

Proteins contain many titratable residues, and the pH of the environment is known to play an important role in regulating the protein function. In this section, we report an interesting system, bovine β -lactoglobulin, where the change of the pH in the environment was reported to trigger the protein conformational changes. Via protein conformational free energy and pK_a calculations, our simulations accurately recapitulated the experimentally observed conformational change triggered by the change of pH, providing an atomic level mechanism for pH dependent conformational change.

3.3.1. Bovine β -Lactoglobulin (BLG): Tanford Transition. β -lactoglobulin (BLG), the major whey protein in cow and sheep's milk, belongs to the lipocalin protein family that are able to bind to small hydrophobic molecules. The core structure of BLG is a β -barrel formed by eight antiparallel β -strands. The EF loop (Ile84-Asn90) serves as a lid of the β -barrel. Previous studies have showed that when the pH value changes around 7, this lid goes through a conformational change referred to as the Tanford transition. At low pH, the lid is in the closed state and blocks the access of the binding pocket in the β -barrel; while at high pH, the lid is in the open state and the binding pocket is accessible for ligands.^{78,79} It was suggested that this conformational change might relate to protecting ligands in acidic environment. The titratable residue

Glu89 in the EF loop has been identified as the key residue of the Tanford transition.⁷⁸

There are multiple variants of BLG. For the variant A of BLG, the Tanford transition from closed to open state happens when the pH increases from 6.2 to 7.1.⁷⁸ For a different variant of BLG, variant B, it was crystallized in the closed state at pH = 7.1, suggesting that the Tanford conformational transition for the variant B of BLG happens at a higher pH compared with the variant A.⁷⁹

We studied the Tanford transition of BLG variant B (BLGB) via PReorg-FEP and pK_a calculation calculations. Figure S30 shows the open and closed forms of BLGB. The pK_a calculations indicate a pK_a of 4.44 for Glu89 in the open conformation, but increases to 14.26 in the closed conformation since Glu89 is deeply buried in the close conformation. Then we applied PReorg-FEP to obtain the free energy changes from the closed state to the open state of BLGB when the residue Glu89 is protonated. The details of calculations are explained in the Supporting Information. Based on the PReorg-FEP and protein FEP results, the dependence of the free energy difference between the open and closed states of BLGB on pH is plotted in Figure 6. The calculations predict that the Tanford transition of BLGB happens at pH = 7.4 which is in good agreement with experimental findings.

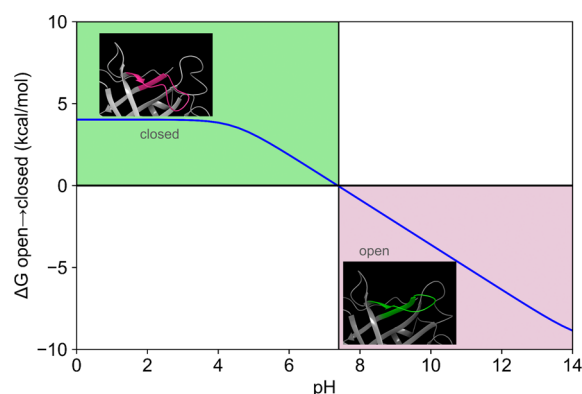


Figure 6. Tanford transition of β -lactoglobulin variant B (BLGB). The curve shows the dependence of the free energy difference between the open to the closed states on pH. The Tanford transition happens at pH = 7.4 when the free energy difference is zero.

4. CONCLUSIONS

Protein conformational changes are fundamental to the protein function and play a pivotal role in rational drug design. In this paper, we demonstrated that reorganization free energies predicted from Protein Reorganization FEP (PReorg-FEP) can provide novel assistance to computer-aided drug design. First, we applied PReorg-FEP to estimate the apo-to-holo reorganization free energies of several protein–ligand systems, and validated the accuracy of the method by comparing the reorganization corrected binding free energy with experimental affinity. The estimated protein reorganization free energy was able to discriminate between novel protein conformations discovered in the hit discovery phase that were able to achieve strong affinity compounds and one that was not. Applications to lead optimization and antibody design were also shown. The strongest affinity compound for each of these protein conformational states was used to establish a general cutoff

for reorganization free energies that are unlikely to produce strong affinity compounds. This is only an educated guess until a rigorous study of potential ligand interactions can be performed, but may still be helpful in focusing the limited resources available to a drug discovery project.

Second, we applied PReorg-FEP to discriminate between competing structural models. These competing models could come from different experimental structures due to different crystallization conditions, or candidate structure models that could fit well with the low resolution experimental data. We found small free energy differences for the alternate loop conformations reported in the PDB for three proteins due to their intrinsic local flexibility, on the order of a few $k_B T$ s. However, the “unnatural” structures from crystal packing or induced by other components in the crystal construct we studied were several kcal/mol higher than the correct structures. We also showed that PReorg-FEP and protein pK_a calculations can accurately model the Tanford conformational transition of Bovine β -lactoglobulin variant B observed in experiment.

The high level of accuracy in a broad class of protein systems for these different use cases demonstrates that PReorg-FEP is a practical and robust approach for calculating the free energy difference between conformational states of proteins. This provides drug discovery teams a tool that can test structural hypotheses in a relatively fast and cheap manner before following up with experimental confirmation. The impact on any specific project will obviously depend on the conformational heterogeneity of the specific protein, so while some projects may not benefit other projects may only become feasible with such a tool.

There are other potential applications of PReorg-FEP to drug discovery that we have not explored. One of particular note is the selectivity of ligands between different proteins.^{80,81} A ligand with similar interaction energies to two related proteins may still have selectivity if the on-target has a lower protein reorganization free energy to adopt the holo-like conformation as compared to the off-target. Comparing the conformational free energy landscapes among related proteins could provide insights to help narrow down the conformational ensembles that ligands should be designed against when achieving selectivity is challenging otherwise. It will be interesting to see how estimates of conformational free energies can be applied more broadly within the drug discovery pipeline.

■ ASSOCIATED CONTENT

Data Availability Statement

All calculations and analysis were performed within the Schrödinger Suite, which is available through commercial, academic or evaluation licenses. Prepared protein and ligand structures for the PReorg-FEP and AB-FEP calculations are available in a Github repository (https://github.com/schrodinger/PReorg-FEP_Validation_JCIM). The repository also includes the additional parameter files for the ligands beyond the standard OPLS4 force field.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.4c01612>.

Systems: CRBP1; IspD; TEM-1 β -lactamase; CDK2; HSP90; Abl Kinase; anti-MTX VHH; PTP1B; eIF4E;

SpGH92; Tim21; GLTP; CIB1; Ampc β -lactamase; Bovine β -lactoglobulin. Simulation Details (PDF)
AB-FEP and PReorg-FEP results for Figure 3 (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Bin W. Zhang – Schrödinger Inc., New York, New York 10036-4041, United States; orcid.org/0000-0003-3007-4900; Email: bin.zhang@schrodinger.com

Mikolai Fajer – Schrödinger Inc., New York, New York 10036-4041, United States; orcid.org/0000-0003-2116-3514; Email: mikolai.fajer@schrodinger.com

Lingle Wang – Schrödinger Inc., New York, New York 10036-4041, United States; orcid.org/0000-0002-8170-6798; Email: lingle.wang@schrodinger.com

Authors

Wei Chen – Schrödinger Inc., New York, New York 10036-4041, United States; orcid.org/0000-0003-2647-3888

Francesca Moraca – Schrödinger Inc., New York, New York 10036-4041, United States; orcid.org/0000-0003-0248-9834

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.4c01612>

Author Contributions

[‡]B.W.Z. and M.F. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Robert Abel, Greg Ross, Fiona McRobb, Martin Vögele, Di Cui, and Mee Shelley for helpful discussions.

REFERENCES

- (1) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. Large Conformational Changes in Proteins: Signaling and Other Functions. *Curr. Opin. Struc. Biol.* **2010**, *20*, 142–147.
- (2) Papaleo, E.; Saladino, G.; Lambrugh, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116*, 6391–6423.
- (3) Congreve, M.; de Graaf, C.; Swain, N. A.; Tate, C. G. Impact of GPCR Structures on Drug Discovery. *Cell* **2020**, *181*, 81–91.
- (4) Vögele, M.; Zhang, B. W.; Kaindl, J.; Wang, L. Is the Functional Response of a Receptor Determined by the Thermodynamics of Ligand Binding? *J. Chem. Theory Comput.* **2023**, *19*, 8414–8422.
- (5) Orellana, L. Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier. *Front. Mol. Biosci.* **2019**, *6*, DOI: 10.3389/fmolb.2019.00117.
- (6) Pietrucci, F. Strategies for the Exploration of Free Energy Landscapes: Unity in Diversity and Challenges Ahead. *Rev. Phys.* **2017**, *2*, 32–45.
- (7) Röder, K.; Joseph, J. A.; Husic, B. E.; Wales, D. J. Energy Landscapes for Proteins: From Single Funnels to Multifunctional Systems. *Adv. Theory Simul.* **2019**, *2*, 1800175.
- (8) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598–1603.
- (9) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632.
- (10) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-Aided Drug Design in Modern Drug Discovery. *Arch. Pharm. Res.* **2015**, *38*, 1686–1701.
- (11) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struc. Biol.* **2011**, *21*, 150–160.
- (12) Wang, L.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- (13) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153–4169.
- (14) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595–5623.
- (15) Song, L. F.; Merz, K. M. Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5308–5318.
- (16) Schindler, C. E. M.; et al. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.
- (17) Decherchi, S.; Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020**, *120*, 12788–12833.
- (18) Feng, M.; Heinzelmann, G.; Gilson, M. K. Absolute Binding Free Energy Calculations Improve Enrichment of Actives in Virtual Compound Screening. *Sci. Rep.* **2022**, *12*, 13640.
- (19) Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. Alchemical Absolute Protein–Ligand Binding Free Energies for Drug Design. *Chem. Sci.* **2021**, *12*, 13958–13971.
- (20) Alibay, I.; Magarkar, A.; Seeliger, D.; Biggin, P. C. Evaluating the Use of Absolute Binding Free Energy in the Fragment Optimization Process. *Commun. Chem.* **2022**, *5*, 105.
- (21) Ross, G. A.; Lu, C.; Scarabelli, G.; Albanese, S. K.; Houang, E.; Abel, R.; Harder, E. D.; Wang, L. The Maximal and Current Accuracy of Rigorous Protein–Ligand Binding Free Energy Calculations. *Commun. Chem.* **2023**, *6*, 222.
- (22) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beutrait, A.; Vendome, J.; Abel, R.; Friesner, R. A.; Wang, L. Enhancing Hit Discovery in Virtual Screening through Absolute Protein–Ligand Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2023**, *63*, 3171–3185.
- (23) 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.
- (24) Meller, A.; Ward, M.; Borowsky, J.; Kshirsagar, M.; Lotthammer, J. M.; Oviedo, F.; Ferres, J. L.; Bowman, G. R. Predicting Locations of Cryptic Pockets from Single Protein Structures Using the PocketMiner Graph Neural Network. *Nat. Commun.* **2023**, *14*, 1177.
- (25) Zwier, M. C.; Chong, L. T. Reaching Biological Timescales with All-Atom Molecular Dynamics Simulations. *Curr. Opin. Pharmacol.* **2010**, *10*, 745–752.
- (26) Makarov, D. E. *Single Molecule Science: Physical Principles and Models*; CRC Press, 2015; Chapter 5, p 59.
- (27) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40*, 41–62.
- (28) Maximova, T.; Moffatt, R.; Ma, B.; Nussinov, R.; Shehu, A. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLoS Comput. Biol.* **2016**, *12*, No. e1004619.
- (29) Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using non-Boltzmann Sampling: Application to the sub-critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.

- (30) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (31) Dickson, A.; Warmflash, A.; Dinner, A. R. Nonequilibrium Umbrella Sampling in Spaces of Many Order Parameters. *J. Chem. Phys.* **2009**, *130*, 074104.
- (32) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. Efficient and Verified Simulation of a Path Ensemble for Conformational Change in a United-residue Model of Calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18043–18048.
- (33) Russo, J. D.; et al. WESTPA 2.0: High-Performance Upgrades for Weighted Ensemble Simulations and Analysis of Longer-Timescale Applications. *J. Chem. Theory Comput.* **2022**, *18*, 638–649.
- (34) Wang, F.; Landau, D. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- (35) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.
- (36) He, P.; Zhang, B. W.; Arasteh, S.; Wang, L.; Abel, R.; Levy, R. M. Conformational Free Energy Changes via an Alchemical Path without Reaction Coordinates. *J. Phys. Chem. Lett.* **2018**, *9*, 4428–4435.
- (37) Arasteh, S.; Zhang, B. W.; Levy, R. M. Protein Loop Conformational Free Energy Changes via an Alchemical Path without Reaction Coordinates. *J. Phys. Chem. Lett.* **2021**, *12*, 4368–4377.
- (38) Fajer, M.; Borrelli, K.; Abel, R.; Wang, L. Quantitatively Accounting for Protein Reorganization in Computer-Aided Drug Design. *J. Chem. Theory Comput.* **2023**, *19*, 3080–3090.
- (39) Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *J. Chem. Theory Comput.* **2017**, *13*, 42–54.
- (40) Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *J. Mol. Biol.* **2017**, *429*, 948–963.
- (41) Wildey, M. J.; Haunso, A.; Tudor, M.; Webb, M.; Connick, J. H. In *Platform Technologies in Drug Discovery and Validation*; Goodnow, R. A., Ed.; Annual Reports in Medicinal Chemistry; Academic Press, 2017; Vol. 50; pp 149–195.
- (42) Klebe, G. Applying Thermodynamic Profiling in Lead Finding and Optimization. *Nat. Rev. Drug Discovery* **2015**, *14*, 95–110.
- (43) Plau, J.; Morgan, C. E.; Fedorov, Y.; Banerjee, S.; Adams, D. J.; Blaner, W. S.; Yu, E. W.; Golczak, M. Discovery of Nonretinoid Inhibitors of CRBP1: Structural and Dynamic Insights for Ligand-Binding Mechanisms. *ACS Chem. Biol.* **2023**, *18*, 2309–2323.
- (44) Silvaroli, J. A.; Arne, J. M.; Chelstowska, S.; Kiser, P. D.; Banerjee, S.; Golczak, M. Ligand Binding Induces Conformational Changes in Human Cellular Retinol-binding Protein 1 (CRBP1) Revealed by Atomic Resolution Crystal Structures. *J. Biol. Chem.* **2016**, *291*, 8528–8540.
- (45) Silvaroli, J. A.; Widjaja-Adhi, M. A. K.; Trischman, T.; Chelstowska, S.; Horwitz, S.; Banerjee, S.; Kiser, P. D.; Blaner, W. S.; Golczak, M. Abnormal Cannabidiol Modulates Vitamin A Metabolism by Acting as a Competitive Inhibitor of CRBP1. *ACS Chem. Biol.* **2019**, *14*, 434–448.
- (46) Witschel, M. C.; Höffken, H. W.; Seet, M.; Parra, L.; Mietzner, T.; Thater, F.; Niggeweg, R.; Röhl, F.; Illarionov, B.; Rohdich, F.; Kaiser, J.; Fischer, M.; Bacher, A.; Diederich, F. Inhibitors of the Herbicidal Target IspD: Allosteric Site Binding. *Angew. Chem. - Int. Ed.* **2011**, *50*, 7931–7935.
- (47) Kunfermann, A.; Witschel, M.; Illarionov, B.; Martin, R.; Rottmann, M.; Höffken, H. W.; Seet, M.; Eisenreich, W.; Knölker, H.-J.; Fischer, M.; Bacher, A.; Groll, M.; Diederich, F. Pseudilins: Halogenated, Allosteric Inhibitors of the Non-Mevalonate Pathway Enzyme IspD. *Angew. Chem. - Int. Ed.* **2014**, *53*, 2235–2239.
- (48) Drawz, S. M.; Bonomo, R. A. Three Decades of β -Lactamase Inhibitors. *Clin. Microbiol. Rev.* **2010**, *23*, 160–201.
- (49) Horn, J. R.; Shoichet, B. K. Allosteric Inhibition Through Core Disruption. *J. Mol. Biol.* **2004**, *336*, 1283–1291.
- (50) Neckers, L. Hsp90 Inhibitors as Novel Cancer Chemotherapeutic Agents. *Trends Mol. Med.* **2002**, *8*, S55–S61.
- (51) Butler, L. M.; Ferraldeschi, R.; Armstrong, H. K.; Centenera, M. M.; Workman, P. Maximizing the Therapeutic Potential of HSP90 Inhibitors. *Mol. Cancer Res.* **2015**, *13*, 1445–1451.
- (52) Chen, D.; et al. Discovery of Potent N-(isoxazol-5-yl)amides as HSP90 Inhibitors. *Eur. J. Med. Chem.* **2014**, *87*, 765–781.
- (53) Miura, T.; Fukami, T. A.; Hasegawa, K.; Ono, N.; Suda, A.; Shindo, H.; Yoon, D.-O.; Kim, S.-J.; Na, Y.-J.; Aoki, Y.; Shimma, N.; Tsukuda, T.; Shiratori, Y. Lead Generation of Heat Shock Protein 90 Inhibitors by a Combination of Fragment-based Approach, Virtual Screening, and Structure-based Drug Design. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 5778–5783.
- (54) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat. Commun.* **2017**, *8*, 2276.
- (55) Davies, T. G.; et al. Structure-based Design of a Potent Purine-based Cyclin-dependent Kinase Inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745–749.
- (56) Hardcastle, I. R.; et al. N²-Substituted O⁶-Cyclohexylmethyl-guanine Derivatives: Potent Inhibitors of Cyclin-Dependent Kinases 1 and 2. *J. Med. Chem.* **2004**, *47*, 3710–3722.
- (57) Xu, W.; Harrison, S. C.; Eck, M. J. Three-dimensional Structure of the Tyrosine Kinase c-Src. *Nature* **1997**, *385*, 595–602.
- (58) Fodey, T.; Leonard, P.; O'Mahony, J.; O'Kennedy, R.; Danaher, M. Developments in the Production of Biological and Synthetic Binders for Immunoassay and Sensor-based Detection of Small Molecules. *TrAC, Trends Anal. Chem.* **2011**, *30*, 254–269.
- (59) Alvarez-Rueda, N.; Behar, G.; Ferré, V.; Pugnère, M.; Roquet, F.; Gastinel, L.; Jacquot, C.; Aubry, J.; Baty, D.; Barbet, J.; Birklé, S. Generation of Llama Single-domain Antibodies Against Methotrexate, a Prototypical Hapten. *Mol. Immunol.* **2007**, *44*, 1680–1690.
- (60) Fanning, S. W.; Horn, J. R. An anti-hapten Camelid Antibody Reveals a Cryptic Binding Site with Significant Energetic Contributions from a nonhypervariable Loop. *Protein Sci.* **2011**, *20*, 1196–1207.
- (61) Garman, E. F.; Owen, R. L. Cryocooling and Radiation Damage in Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 32–47.
- (62) Keedy, D. A.; Hill, Z. B.; Biel, J. T.; Kang, E.; Rettenmaier, T. J.; Brandão-Neto, J.; Pearce, N. M.; von Delft, F.; Wells, J. A.; Fraser, J. S. An Expanded Allosteric Network in PTP1B by Multitemperature Crystallography, Fragment Screening, and Covalent Tethering. *eLife* **2018**, *7*, e36307.
- (63) Papadopoulos, E. Structure of the Eukaryotic Translation Initiation Factor eIF4E in Complex with 4EGI-1 Reveals an Allosteric Mechanism for Dissociating eIF4G. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E3187–E3195.
- (64) Robb, M.; Hobbs, J. K.; Woodiga, S. A.; Shapiro-Ward, S.; Suits, M. D. L.; McGregor, N.; Brumer, H.; Yesilkaya, H.; King, S. J.; Boraston, A. B. Molecular Characterization of N-glycan Degradation and Transport in Streptococcus pneumoniae and Its Contribution to Virulence. *PLOS Pathog.* **2017**, *13*, No. e1006090.
- (65) Tonks, N. K. PTP1B: From the Sidelines to the Front Lines! *FEBS Lett.* **2003**, *546*, 140–148.
- (66) Combs, A. P. Recent Advances in the Discovery of Competitive Protein Tyrosine Phosphatase 1B Inhibitors for the Treatment of Diabetes, Obesity, and Cancer. *J. Med. Chem.* **2010**, *53*, 2333–2344.
- (67) Jia, Z.; Barford, D.; Flint, A. J.; Tonks, N. K. Structural Basis for Phosphotyrosine Peptide Recognition by Protein Tyrosine Phosphatase 1B. *Science* **1995**, *268*, 1754–1758.
- (68) Albrecht, R.; Rehling, P.; Chacinska, A.; Brix, J.; Cadamuro, S. A.; Volkmer, R.; Guiard, B.; Pfanner, N.; Zeth, K. The Tim21 Binding Domain Connects the Preprotein Translocases of Both Mitochondrial Membranes. *EMBO Reports* **2006**, *7*, 1233–1238.

- (69) Bajaj, R.; Jaremko, L.; Jaremko, M.; Becker, S.; Zweckstetter, M. Molecular Basis of the Dynamic Structure of the TIM23 Complex in the Mitochondrial Intermembrane Space. *Structure* **2014**, *22*, 1501–1511.
- (70) Bala, S.; Shinya, S.; Srivastava, A.; Ishikawa, M.; Shimada, A.; Kobayashi, N.; Kojima, C.; Tama, F.; Miyashita, O.; Kohda, D. Crystal Contact-free Conformation of an Intrinsically Flexible Loop in Protein Crystal: Tim21 as the Case Study. *Biochim. Biophys. Acta - Gen. Subj.* **2020**, *1864*, 129418.
- (71) Brown, R. E.; Mattjus, P. Glycolipid Transfer Proteins. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2007**, *1771*, 746–760.
- (72) Malinina, L.; Malakhova, M. L.; Teplov, A.; Brown, R. E.; Patel, D. J. Structural Basis for Glycosphingolipid Transfer Specificity. *Nature* **2004**, *430*, 1048–1053.
- (73) Samygina, V. R.; Popov, A. N.; Cabo-Bilbao, A.; Ochoa-Lizarralde, B.; de Cerio, F. G.; Zhai, X.; Molotkovsky, J. G.; Patel, D. J.; Brown, R. E.; Malinina, L. Enhanced Selectivity for Sulfatide by Engineered Human Glycolipid Transfer Protein. *Structure* **2011**, *19*, 1644–1654.
- (74) Leisner, T. M.; Freeman, T. C.; Black, J. L.; Parise, L. V. CIB1: a Small Protein with Big Ambitions. *FASEB J.* **2016**, *30*, 2640–2650.
- (75) Blamey, C. J.; Ceccarelli, C.; Naik, U. P.; Bahnson, B. J. The Crystal Structure of Calcium- and Integrin-binding Protein 1: Insights into Redox Regulated Functions. *Protein Sci.* **2005**, *14*, 1214–1221.
- (76) Hendershot, J. M.; Mishra, U. J.; Smart, R. P.; Schroeder, W.; Powers, R. A. Structure-based Efforts to Optimize a non- β -lactam Inhibitor of AmpC β -lactamase. *Bioorgan. Med. Chem.* **2014**, *22*, 3351–3359.
- (77) Usher, K. C.; Blaszcak, L. C.; Weston, G. S.; Shoichet, B. K.; Remington, S. J. Three-Dimensional Structure of AmpC β -Lactamase from *Escherichia coli* Bound to a Transition-State Analogue: Possible Implications for the Oxyanion Hypothesis and for Inhibitor Design. *Biochemistry* **1998**, *37*, 16082–16092.
- (78) Qin, B. Y.; Bewley, M. C.; Creamer, L. K.; Baker, H. M.; Baker, E. N.; Jameson, G. B. Structural Basis of the Tanford Transition of Bovine β -Lactoglobulin. *Biochemistry* **1998**, *37*, 14014–14023.
- (79) Qin, B. Y.; Jameson, G. B.; Bewley, M. C.; Baker, E. N.; Creamer, L. K. Functional Implications of Structural Differences between Variants A and B of Bovine β -lactoglobulin. *Protein Sci.* **1999**, *8*, 75–83.
- (80) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J. Am. Chem. Soc.* **2017**, *139*, 946–957.
- (81) Albanese, S. K.; Chodera, J. D.; Volkamer, A.; Keng, S.; Abel, R.; Wang, L. Is Structure-Based Drug Design Ready for Selectivity Optimization? *J. Chem. Inf. Model.* **2020**, *60*, 6211–6227.



CAS INSIGHTS™

EXPLORE THE INNOVATIONS SHAPING TOMORROW

Discover the latest scientific research and trends with CAS Insights. Subscribe for email updates on new articles, reports, and webinars at the intersection of science and innovation.

Subscribe today

CAS
A division of the American Chemical Society