

# Computational Analysis of Energy Landscapes Reveals Dynamic Features That Contribute to Binding of Inhibitors to CFTR-Associated Ligand

Published as part of *The Journal of Physical Chemistry virtual special issue "David N. Beratan Festschrift"*.

Graham T. Holt,<sup>†,‡,∇</sup> Jonathan D. Jou,<sup>†,∇</sup> Nicholas P. Gill,<sup>§</sup> Anna U. Lowegard,<sup>†,‡</sup> Jeffrey W. Martin,<sup>†</sup> Dean R. Madden,<sup>§</sup> and Bruce R. Donald<sup>\*,†,||,⊥</sup>

<sup>†</sup>Department of Computer Science, Duke University, Durham, North Carolina 27708, United States

<sup>‡</sup>Program in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina 27708, United States

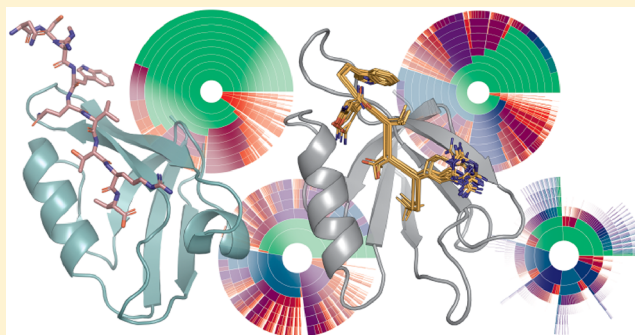
<sup>§</sup>Department of Biochemistry & Cell Biology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire 03755, United States

<sup>||</sup>Department of Biochemistry, Duke University, Durham, North Carolina 27710, United States

<sup>⊥</sup>Department of Chemistry, Duke University, Durham, North Carolina 27710, United States

## Supporting Information

**ABSTRACT:** The CFTR-associated ligand PDZ domain (CALP) binds to the cystic fibrosis transmembrane conductance regulator (CFTR) and mediates lysosomal degradation of mature CFTR. Inhibition of this interaction has been explored as a therapeutic avenue for cystic fibrosis. Previously, we reported the ensemble-based computational design of a novel peptide inhibitor of CALP, which resulted in the most binding-efficient inhibitor to date. This inhibitor, kCAL01, was designed using OSPREY and evinced significant biological activity in *in vitro* cell-based assays. Here, we report a crystal structure of kCAL01 bound to CALP and compare structural features against iCAL36, a previously developed inhibitor of CALP. We compute side-chain energy landscapes for each structure to not only enable approximation of binding thermodynamics but also reveal ensemble features that contribute to the comparatively efficient binding of kCAL01. Finally, we compare the previously reported design ensemble for kCAL01 vs the new crystal structure and show that, despite small differences between the design model and crystal structure, significant biophysical features that enhance inhibitor binding are captured in the design ensemble. This suggests not only that ensemble-based design captured thermodynamically significant features observed *in vitro*, but also that a design eschewing ensembles would miss the kCAL01 sequence entirely.



## 1. INTRODUCTION

Interactions between proteins and short linear motif peptides are important in many cellular contexts.<sup>1</sup> One such class of peptide-binding proteins is the PDZ (PSD-95, discs large, ZO-1) domain family, characterized by an 80–90 residue motif<sup>2</sup> that adopts a conserved fold composed of 2–3  $\alpha$ -helices and 5–6  $\beta$ -strands and binds C-terminal peptides through  $\beta$ -sheet interactions.<sup>3</sup> These domains commonly modulate protein localization and complex assembly<sup>4–6</sup> and regulate cell signaling,<sup>7,8</sup> thereby playing critical roles in auditory and visual systems;<sup>9,10</sup> epilepsy, pain, and addiction;<sup>6,11</sup> synapse formation;<sup>5</sup> cancer;<sup>7,8,12,13</sup> and cystic fibrosis.<sup>14–17</sup>

Protein–peptide interactions have been implicated as therapeutic targets in cystic fibrosis (CF), a genetic disease characterized by defects in the cystic fibrosis transmembrane

conductance regulator (CFTR) that result in impaired chloride ion transport.<sup>14</sup> Approximately 90% of CF patients are homozygous or heterozygous for the *F508del* (*c.1521\_1523delCTT*) mutation,<sup>14,18,19</sup> which encodes a protein variant F508del-CFTR (p.Phe508del) with severe loss of function. This variant exhibits impaired folding,<sup>20</sup> increased degradation by endoplasmic reticulum (ER) quality control machinery,<sup>21</sup> reduced capacity for Cl<sup>−</sup> transport,<sup>14</sup> and decreased half-life at the plasma membrane.<sup>22</sup> CFTR is recycled from the cell membrane and preferentially targeted for lysosomal degradation by interaction of the CFTR C-

**Received:** July 30, 2019

**Revised:** November 5, 2019

**Published:** November 7, 2019

terminus with the CFTR-associated ligand PDZ domain (CALP).<sup>15,16</sup> CALP has been implicated in both decreasing concentration of CFTR at the membrane<sup>16</sup> and arresting CFTR trafficking in the ER,<sup>17</sup> and knockdown of CALP has been shown to rescue transepithelial chloride transport in polarized CFBE41o- cells expressing F508del-CFTR by increasing the concentration of F508del-CFTR at the plasma membrane.<sup>23</sup> Hence, inhibition of the interaction between the CFTR C-terminal peptide and CALP is a potential therapeutic avenue for CF. Understanding of the CALP:CFTR binding interaction is critical for the development of therapeutic inhibitors.

Previous work toward inhibitor development<sup>24–27</sup> resulted in extensive characterization of the structural and stereochemical components of CALP binding. The structure of CALP bound to the CFTR C-terminal peptide was solved by solution NMR<sup>24</sup> with well-resolved interactions between the 4 C-terminal peptide residues (P<sup>-3</sup>–P<sup>0</sup>) and CALP. This structure revealed canonical class 1 PDZ interactions<sup>28</sup> including those between Leu P<sup>0</sup> and a hydrophobic pocket between secondary-structure elements  $\alpha 2$  and  $\beta 2$ , and the essential hydrogen bond between Thr P<sup>-2</sup> and a histidine residue in helix  $\alpha 2$ . Peptide screening and iterative optimization using substitutional analysis<sup>25</sup> revealed significant affinity effects of residues at other peptide positions up to P<sup>-9</sup> and resulted in a decapeptide inhibitor (iCAL36, ANSRWPT-SII) with an affinity of  $22.6 \pm 8.0 \mu\text{M}$ <sup>25,26</sup> that rescued functional CFTR activity as assessed by *in vitro* Ussing chamber assays.<sup>29</sup> Crystal structures of iCAL36 (and substituted peptide variants) in complex with CALP<sup>26,27</sup> revealed structural features that influence CALP binding and selectivity. In particular, shifts in peptide orientation and location, along with conformational shifts in the carboxylate-binding loop (characterized by a  $X\Phi_1G\Phi_2$  sequence motif, where  $\Phi_i$  represents a hydrophobic and X is any residue), affect the binding geometry and specificity of the peptide P<sup>0</sup> residue,<sup>26</sup> allowing CALP to accommodate both Leu and Ile at P<sup>0</sup>. Additionally, side-chain interactions at P<sup>-1</sup>, P<sup>-3</sup>, P<sup>-4</sup>, and P<sup>-5</sup> modulate affinity and specificity of CALP binding.<sup>27</sup> Finally, despite the fact that CALP:CFTR binding is thought to be primarily driven by enthalpic effects,<sup>30</sup> NMR data and molecular dynamics (MD) simulations suggest that entropy may play a role in modulating CALP binding,<sup>24</sup> a hypothesis which is reflected in studies of other PDZ domains.<sup>31–34</sup>

Previously,<sup>35</sup> we developed the most binding-efficient<sup>36</sup> inhibitor of CALP to date using the OSPREY<sup>37</sup> protein design software package, suggesting that components of CALP binding can be effectively captured using provable, ensemble-based computational protein design algorithms. Starting from the solution NMR structure of CALP:CFTR,<sup>24</sup> we used the  $K^*$  algorithm<sup>38</sup> to compute approximations to  $K_a$ —the  $K^*$  score—for CALP binding to  $\approx 8000$  hexameric C-terminal peptides (residue positions P<sup>-5</sup>–P<sup>0</sup>). Retrospective predictions on 6223 previously characterized sequences<sup>25,30</sup> showed that our algorithm was able to effectively classify sequences by binding affinity to CALP, with an area under the receiver operating characteristic (ROC) curve of 0.84. Additionally, OSPREY designs on 2166 sequences resulted in novel peptides that bind tightly to CALP. All of the top 11 prospective predictions were experimentally shown to bind with high affinity to CALP, and the tightest binding hexamer, kCAL01 (Ac-WQVTRV), bound with  $K_i = 2.3 \pm 0.2 \mu\text{M}$ .<sup>35</sup> kCAL01 bound more tightly than both the previous best hexamer

(iCAL35, WQTSII,  $K_i = 14.0 \pm 1 \mu\text{M}$ )<sup>35</sup> and decamer (iCAL36, ANSRWPTSII,  $K_i = 22.6 \pm 8.0 \mu\text{M}$ )<sup>26</sup> peptide inhibitors. Despite its small size (MW 829 Da), kCAL01 binds with an affinity comparable to a much larger (MW 1502 Da) fluorescein-modified version of iCAL36 (F\*-iCAL36,  $K_d = 1.3 \pm 0.1 \mu\text{M}$ ),<sup>25</sup> yielding a much better binding efficiency for kCAL01 (for molecular weights and inhibition constants of various inhibitors see Table S1). Furthermore, kCAL01 rescued chloride ion transport activity of F508del-CFTR in cell-based assays.<sup>35,39</sup> Ensemble-based design algorithms were shown to be critical for the success of this design: Ranking by energy of the global minimum energy conformations (GMECs) resulted in poor prediction accuracy and little overlap with the ensemble-based predictions.<sup>35</sup> These data not only suggest that computational structural protein design (CSPD) algorithms can capture features that contribute to CALP:peptide binding, but also that ensemble-based or entropic effects are critical for prediction accuracy.

Indeed, computational designs are more biophysically accurate when they model protein thermodynamic ensembles.<sup>35,38,40–47</sup> The objectives of CSPD algorithms are to (1) compute biophysical or thermodynamic properties of a protein or protein complex and (2) efficiently search for optimal sequences given an objective function. Without loss of generality, we choose binding affinity as our biophysical property of interest. CSPD algorithms search over a user-specified *input model* (*viz.*, a structural model, allowed side-chain and backbone flexibility, allowed mutations, energy function, etc.<sup>37</sup>). Because proteins exist as thermodynamic ensembles,<sup>41,48</sup> principled algorithms should exploit statistical thermodynamics of non-covalent binding, and therefore require approximation of the partition function.<sup>41,49</sup> However, because the conformation space available to proteins *in vivo* and *in vitro* is massive and grows exponentially with the number of flexible amino acid residues, protein design algorithms often make simplifying modeling assumptions to allow tractable computation. Such assumptions often include (1) modeling only rigid, discrete side-chain configurations, or *rotamers*,<sup>50</sup> and a small set of discrete backbone conformations,<sup>51–53</sup> (2) considering (or approximating) only a single global minimum energy conformation (GMEC),<sup>51,54–59</sup> and (3) approximating the partition function using stochastic, heuristic sampling methods.<sup>60–62</sup> However, these assumptions (1) fail to model small, commonly observed side-chain and backbone movements, (2) entirely omit conformational entropy, and (3) often fail to find even the GMEC.<sup>53</sup> Algorithms in the OSPREY software package efficiently solve protein design problems without these simplifications.<sup>37</sup>

Recently we developed the MARK\* algorithm, which, in addition to provably and efficiently approximating partition functions for input model states (*i.e.*, bound complex, unbound protein, and ligand), allows visualization of the entire energy landscape accessible to a protein input model.<sup>63</sup> MARK\* provably bounds the energy and statistical weight of every conformation in the input model conformation space, allowing designers to compute and visualize changes in *conformation distribution*, instead of merely analyzing changes to a small set of low-energy conformations, or to ensemble averaged values like free energy or  $K_a$ . By computing both a conformation distribution and provably approximating energies for every conformation, MARK\* enables visualization of the energy landscape. This novel capability complements traditional

structural analysis by providing insight into entropic and dynamic contributions to binding.

In this work, we report a 1.7 Å resolution crystal structure of a decapeptide variant of the peptide inhibitor kCAL01 bound to CALP (PDB ID: 6OV7). To evaluate the structural basis for the enhanced binding efficiency of kCAL01, we compare this structure to that of a previously developed decapeptide inhibitor of CALP, iCAL36 (PDB ID: 4E34).<sup>26</sup> In addition to performing traditional structural analysis, we compute energy landscapes for bound and unbound structural models for CALP:kCAL01 and CALP:iCAL36 using MARK\*. From these landscapes we compute approximations to the free energy, internal energy, and entropy for each model state (i.e., bound complex, unbound protein, unbound ligand) and use these quantities to model thermodynamics of binding for CALP:kCAL01 and CALP:iCAL36. Additionally, we analyze the energy landscapes for dynamic effects and show that these energy landscapes foreground important structural features of binding and reveal dynamic features that may contribute to the efficient binding of kCAL01. We conclude that investigation of energy landscapes complements traditional analysis of one or few low-energy structures represented in crystal structures and provides important information about the entire conformational ensemble that is available to a protein structure model. Finally, to assess the extent to which designs reported in ref 35 are a result of accurate modeling of structural and dynamic components of CALP binding, we compare the design output ensemble for kCAL01 to the newly solved crystal structure. We show that, despite notable differences between the NMR-based CALP structure used as design input vs the bound crystallographic conformation of CALP, many significant crystal structure features are captured in the design output ensemble. This suggests that the success of the ensemble-based computational design of kCAL01<sup>35</sup> was a result of effective modeling of structural and dynamic features of binding.

An overview of our system, model, and data is as follows. Table 1 shows data collection and refinement statistics for the crystal structure of CALP:kCAL01. Figure 1 depicts the crystal structure of CALP:kCAL01 and important structural features. Figure 2 shows a detailed structural comparison between the carboxylate-binding loops of CALP:kCAL01 and CALP:iCAL36. Figure 3 details structural views of predicted conformational heterogeneity at P<sup>0</sup>. Figure 4 depicts energy landscape diagrams for CALP:kCAL01. Figure 5 shows energy landscape diagrams for CALP:iCAL36. Figure 6 shows a schematic diagram of structural comparisons performed in this work. Figure 7 shows the original design model reported in ref 35 and highlights its similarity to the crystal structure of CALP:kCAL01.

## 2. METHODS

**2.1. Structure Determination of CALP:kCAL01.** Recombinant CAL PDZ (CALP; UniProt accession number Q9HD26-2; residues 278–362) was expressed and purified as described previously.<sup>64</sup> Briefly, an expression construct was engineered in pET16b containing an N-terminal decahistidine tag, a short linker, and a human rhinovirus (HRV) 3C protease cleavage site, followed by the CALP sequence. The construct was transformed into *E. coli* BL21 (DE3) RIL cells; expression was induced as previously described<sup>64</sup> (except that TB medium was used), and protein was purified by nickel-nitrilotriacetic acid (NiNTA) affinity chromatography and size-exclusion chromatography (SEC). Following removal of the affinity tag

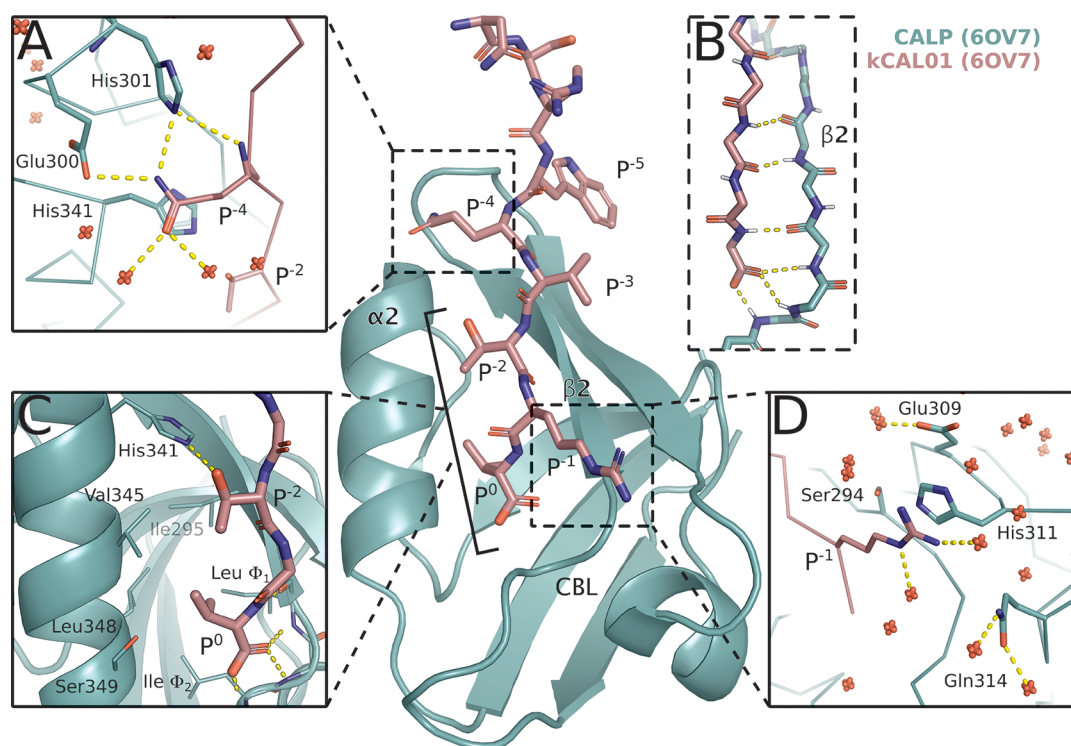
**Table 1. Data Collection and Refinement Statistics for CALP:kCAL01 Complex (PDB ID: 6OV7)**

Data Collection	
space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
unit cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	43.6, 60.8, 81.2
resolution <sup>a</sup> (Å)	48.6–1.71 (1.83–1.71)
R <sub>sym</sub> <sup>b</sup> (%)	10.8 (95.9)
I/σ <sub>I</sub>	14.0 (2.0)
completeness (%)	99.6 (99.8)
Refinement	
total number of reflections	23681
reflections in test set	1160
R <sub>work</sub> <sup>c</sup> /R <sub>free</sub> <sup>d</sup> (%)	18.2/21.4 (28.1/30.7)
no. atoms protein	1470
no. atoms water	175
Ramachandran plot <sup>e</sup> (%)	98.4, 1.6, 0, 0
B <sub>av</sub> (Å <sup>2</sup> )	
protein	22.3
solvent	33.1
bond length RMSD (Å)	0.01
bond angle RMSD (deg)	1.18

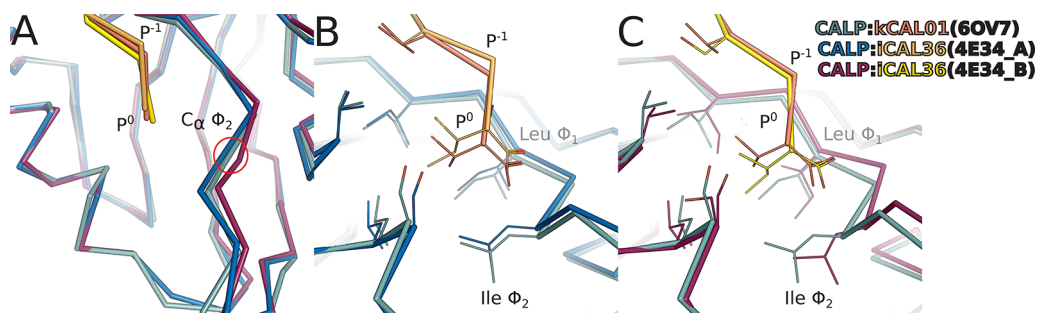
<sup>a</sup>Values in parentheses are for data in the highest-resolution shell. <sup>b</sup>R<sub>sym</sub> =  $\sum_h \sum_i |I_i(h) - I_i(h)| / \sum_h \sum_i I_i(h)$ , where  $I_i(h)$  and  $I(h)$  values are the *i*th and mean measurements of the intensity of reflection *h*. <sup>c</sup>R<sub>work</sub> =  $\sum |F_{obs,h} - F_{calc,h}| / \sum |F_{obs,h}|$ , *h* ∈ {working set}. <sup>d</sup>R<sub>free</sub> is calculated as R<sub>work</sub> for the reflections *h* ∈ {test set}. <sup>e</sup>Core, allowed, generously allowed, disallowed.

and linker by HRV-3C protease cleavage, CALP was recovered in the flow-through fraction of a NiNTA affinity column and further purified by SEC. To facilitate crystallization, kCAL01 was synthesized as a decapeptide (ANSRWQVTRV) containing four N-terminal residues (in italics) that form lattice contacts in other CALP:peptide co-crystals.<sup>26,27,64,65</sup> The kCAL01 decapeptide was synthesized using standard Fmoc solid-phase peptide synthesis and purified by reverse-phase HPLC. Peptide mass was confirmed using liquid chromatography/mass spectrometry (LC/MS). Using the hanging-drop method, CALP:kCAL01 co-crystals were obtained by mixing 1 mM kCAL01 and 6 mg/mL CALP with reservoir solution containing 25% (*w/v*) PEG 8000, 5% (*v/v*) PEG 400, 150 mM sodium chloride, and 100 mM Tris pH 8.5. Crystals were transferred into cryoprotectant solution (25% [*w/v*] PEG 8000, 15% [*v/v*] PEG 400, 150 mM sodium chloride, and 100 mM Tris pH 8.5) and flash-cooled in a liquid-nitrogen bath. Oscillation diffraction data were recorded at 100 K on beamline BL9-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) over a 180° range with 0.5 s, 0.2° exposures. Reflection intensities were integrated and scaled using the XDS package<sup>66</sup> (version 20190315). Initial phase estimates were obtained by molecular replacement using Phaser<sup>67</sup> within the Phenix package<sup>68</sup> (version 1.15.2) and using PDB ID 4E34<sup>26</sup> as the search model (containing chains A and C only). Subsequent model building and refinement were performed using Phenix and Coot<sup>69</sup> (version 0.8.9.2) to generate the final model of CALP in a complex with the kCAL01 decapeptide at 1.71 Å resolution. Data quality and refinement statistics are reported in Table 1. The coordinates and structure factors have been deposited in the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) with ID 6OV7.

**2.2. Computational Methods.** The new crystal structure of CALP:kCAL01 (PDB ID: 6OV7, protomer A) and the crystal structure of CALP:iCAL36 (PDB ID: 4E34, protomer



**Figure 1.** Crystal structure of CALP:kCAL01 (PDB ID: 6OV7) displays canonical class 1 PDZ binding and favorable interactions at  $P^{-1}$  and  $P^{-4}$ . The CALP and kCAL01 crystal structures are shown in green and pink, respectively (protomer A). Hydrogen bonds predicted by the Probe<sup>78</sup> software are represented as dashed yellow lines. The kCAL01 peptide binds in the groove defined by helix  $\alpha 2$  and strand  $\beta 2$ . (A) Gln  $P^{-4}$  makes favorable hydrogen bonds with Glu300 or His301 and forms van der Waals interactions with His341. Additionally, Gln  $P^{-4}$  can coordinate with several water molecules, shown as red, nonbonded oxygens. (B) kCAL01 binds in the groove defined by helix  $\alpha 2$  and strand  $\beta 2$  and forms an antiparallel  $\beta$ -sheet interaction with strand  $\beta 2$ . The C-terminus of kCAL01 forms favorable hydrogen bonds with the carboxylate-binding loop (CBL). (C) kCAL01 displays features of class 1 PDZ binding,<sup>28</sup> including the conserved hydrogen bond between Thr  $P^{-2}$  and His341, as well as the interaction between the Val  $P^0$  side chain and the hydrophobic pocket. (D) Arg  $P^{-1}$  appears to form favorable  $\pi$ -interactions with His311 and also coordinates with several waters.

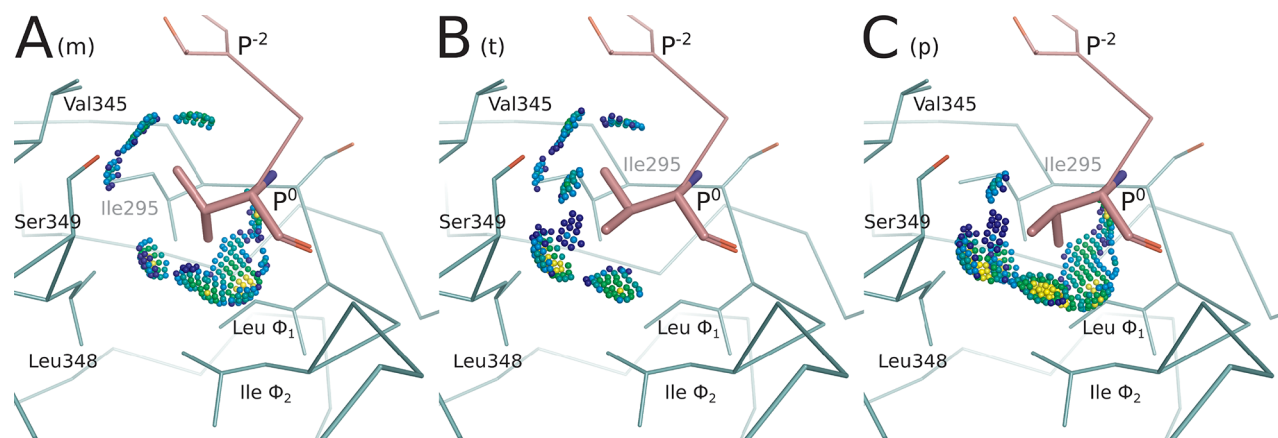


**Figure 2.** Binding geometry of kCAL01  $P^0$  and the carboxylate-binding loop. Superimposed views of the  $P^0$  interaction with the carboxylate-binding loop (CBL) and hydrophobic pocket (side chains shown as lines) are shown for CALP:kCAL01 protomer A (green: pink), CALP:iCAL36 protomer A (blue: orange), and CALP:iCAL36 protomer B (purple: yellow). (A) Superimposed  $C_\alpha$  traces show that the CALP conformation at the Ile  $\Phi_2$   $C_\alpha$  is more similar to the CALP:iCAL36 protomer A conformation than the CALP:iCAL36 protomer B conformation. (B) A pairwise comparison shows that the CALP:kCAL01 CBL geometry matches most closely with CALP:iCAL36 protomer A, seen at the side chains at CBL positions  $\Phi_1$  and  $\Phi_2$ . However, the kCAL01 peptide  $P^0$  shifts toward the CBL by 0.7 Å relative to the CALP:iCAL36 structure. (C) A pairwise comparison shows that the CALP:kCAL01 peptide orientation matches most closely with CALP:iCAL36 protomer B, seen at position  $P^0$ . However, the CALP:iCAL36 CBL shifts outward by 1.3 Å relative to the CALP:kCAL01 structure, and the hydrophobic pocket expands due to changes in rotamer at CBL position  $\Phi_1$ .

A) were used to model energy landscapes of CALP binding to kCAL01 and iCAL36, respectively. For each structure, we computed energy landscapes for three states: the bound CALP:peptide complex, the unbound CALP, and the unbound peptide. This was accomplished by first defining a set of accessible conformations for each state and then using the MARK\* algorithm<sup>63</sup> in OSPREY 3.0<sup>37</sup> to compute both a

provable approximation to the partition function value and an approximation to the energy landscape.

Sets of accessible conformations, or *conformation spaces*, were defined as follows for each state. These conformation spaces are an approximation to the ensemble of conformations available to each state *in vivo*. First, hydrogens were added to each crystal structure using the MolProbity server<sup>70</sup> in order to



**Figure 3.** Energy landscape analysis reveals conformational heterogeneity at Val  $P^0$  for CALP:kCAL01. Energy landscape analysis of bound kCAL01 indicates three rotamers at peptide  $P^0$  that contribute significantly to the partition function. We refer to these rotamers as **m**, **t**, or **p**, which describe the valine  $N-C_{\alpha}-C_{\beta}-C_{\gamma1}$  dihedral angle as *minus 60* ( $\sim -60^\circ$ ), *trans* ( $\sim 180^\circ$ ), or *plus 60* ( $\sim 60^\circ$ ), respectively, conforming to the convention defined in ref 50. This landscape analysis (see Figure 4C, outermost ring) suggests that the complex can sample any of these rotamers with relatively high probability, but that the **m** rotamer will be most occupied, and the **p** rotamer will be least occupied. Conformations containing each of these three rotamers were selected from the bound kCAL01 ensemble. Interactions between the  $P^0$  side-chain atoms and the CALP structure are shown using Probe dots,<sup>78</sup> where green and blue dots indicate favorable interactions, yellow dots indicate small overlaps, and red and pink lines show steric clashes. (A) The **m** rotamer forms favorable interactions with Thr  $P^{-2}$ , Val345, Ser349, Leu348, and Ile  $\Phi_2$ . The **t** rotamer forms favorable interactions with Thr  $P^{-2}$ , Val345, Ser349, Leu348, and Ile  $\Phi_2$ . The slight overlaps (yellow dots) generated due to the interaction with Leu348, along with the lack of interaction with Leu  $\Phi_1$ , suggest that this conformation is slightly less favorable than the **m** rotamer. (C) The **p** rotamer forms favorable interactions with Ser349, Leu348, Ile  $\Phi_2$ , and Leu  $\Phi_1$ . Slight overlaps (yellow dots) can be seen in interactions with Leu348, Ile  $\Phi_2$ , and Leu  $\Phi_1$ , and there is no interaction with Val345, suggesting that this rotamer may be slightly less favorable than either the **m** or **t** rotamers. Nevertheless, all three rotamers are well-sampled in the ensemble and contribute significantly to the partition function.

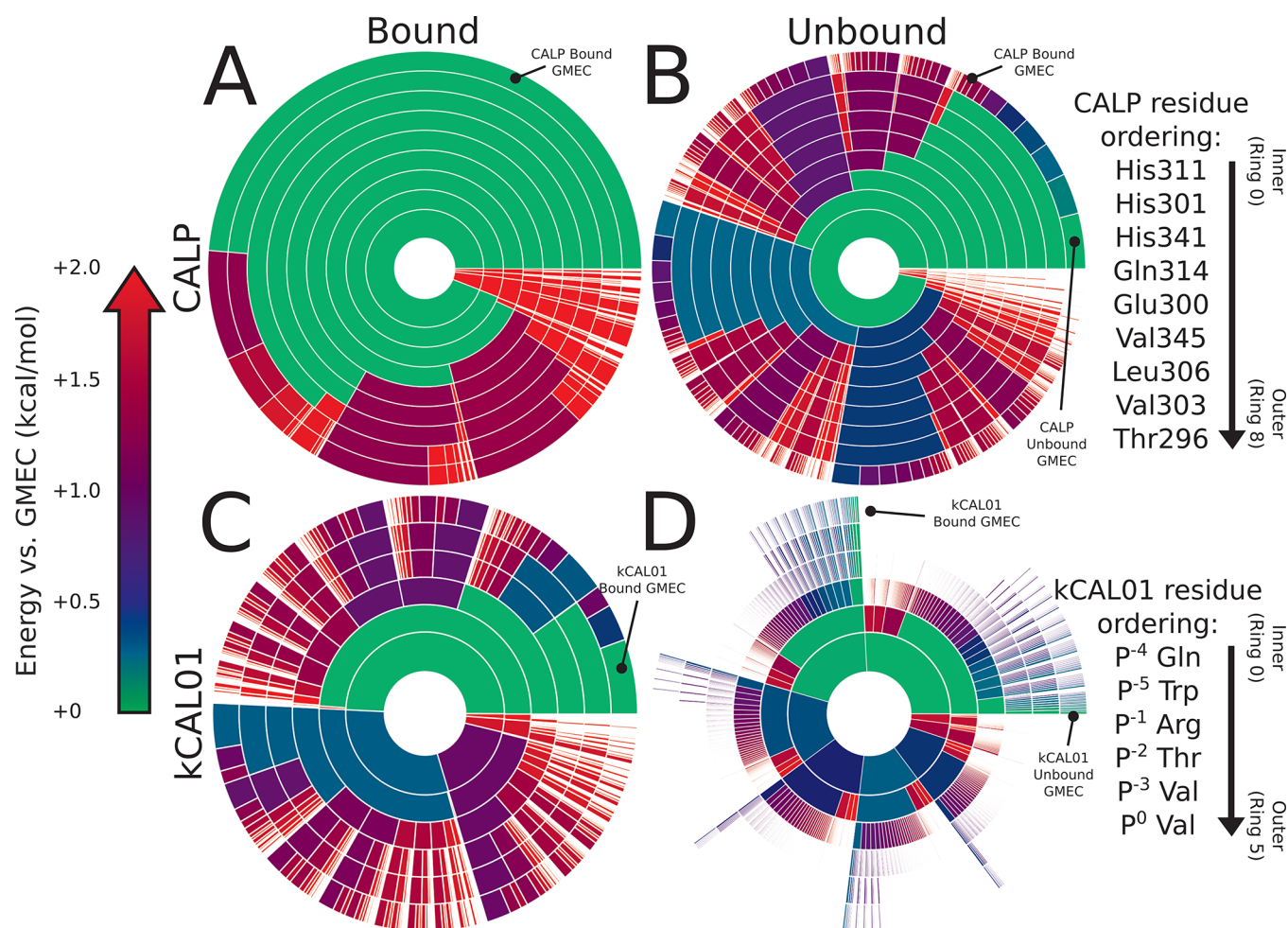
generate *protonated crystal structures*. Backbone atom coordinates for the bound complex state were obtained directly from protonated crystal structures 6OV7 (for the CALP:kCAL01 state) and 4E34 (for the CALP:iCAL36 state). Nine residues for CALP and the six most C-terminal residues for kCAL01 or iCAL36 (for a total of 15 residues in each bound complex, see Table S2) were modeled as continuously flexible using continuous rotamers<sup>71,72</sup> in OSPREY. As in refs 46, 63, 71, and 73, rotamers from the Penultimate Rotamer Library<sup>50</sup> were allowed to adopt any side-chain conformation such that all  $\chi$ -angles are within  $\pm 9^\circ$  of their modal  $\chi$ -angles. For all other residues, side-chain coordinates were obtained from protonated crystal structures. Models for unbound CALP and peptide states were obtained by removing all atoms of the peptide or CALP structure, respectively, from the complex state. Thus, we defined approximations to the conformational ensembles for bound and unbound states, herein referred to as *models*, for CALP:kCAL01 and CALP:iCAL36.

For each model, we computed  $\varepsilon$ -approximate bounds on the value of the partition function to a deterministic, guaranteed accuracy of  $\varepsilon < 0.01$  using the MARK\* algorithm<sup>63</sup> in OSPREY. All computations were run on 40–48 core Intel Xeon nodes with up to 500 GB of memory. As proved previously,<sup>63</sup> not only does MARK\* compute a provable  $\varepsilon$ -approximation to the partition function, it also bounds the energy landscape by provably approximating the energy and therefore statistical weight of all model conformations in the conformation space.

**2.3. Entropy, Internal Energy, and Helmholtz Free Energy Calculation.** Aggregate values for the ensembles in each state were computed by bounding the energy for each conformation in the ensemble, and combining these energy bounds. For each bound and unbound state, we first computed bounds on the energy of each conformation in the conformational ensemble defined by that state, as was done in refs 35,

37, 44, 63, 74, and 75 described in Section 2.2. Using these energy bounds, we then computed bounds on the corresponding Boltzmann-weighted partition function  $Z_C = \sum_{c \in C} \exp(-E(c)/RT)$ , where  $E$  is a function that returns the energy of conformation  $c$ , by computing and summing bounds on the Boltzmann weights for conformations  $c$  in that state  $C$  (see ref 63 for details). We then divided the upper bound on Boltzmann weight of  $c$  by the upper bound on the partition function to compute the probability  $p_c$  for each conformation  $c$  within the ensemble. Using these probabilities, we then calculated the entropy  $S = -R \sum_{c \in C} p_c \ln p_c$  and internal energy  $U = \sum_{c \in C} p_c E(c)$  of the ensemble and combined these two values to compute the Helmholtz free energy  $F = U - TS$  at a temperature of 298.15 K. Here,  $E$  is a function that returns the lower bound on the energy of conformation  $c$ . We also used the upper bounds on Boltzmann weight for each conformation to compute energy landscape diagrams,<sup>63</sup> which are explained in Section 2.4.

Although the  $K^*$  score exhibits good Spearman's rank correlation with experimental  $K_a$  values,<sup>37,74</sup> the correlation between  $K^*$  scores and  $K_a$  is not yet quantitative. First, most physics-based energy functions are based on small-molecule energetics, which can overestimate van der Waals terms and thereby overestimate internal energy. Additionally, the input models used in the current computation model only a subset of biologically available flexibility; in this case, flexibility was restricted to up to 4 side-chain  $\chi$  angles per residue. We allowed side-chain  $\chi$  angles to minimize continuously within  $\pm 9^\circ$  of modal  $\chi$ -angles but did not model backbone flexibility. Furthermore, we did not model explicit waters, instead relying on the EEF1 implicit solvation model<sup>76</sup> in OSPREY. As a result, our models likely underestimate entropy and overestimate internal energy. Therefore, we scaled our thermodynamic



**Figure 4.** Energy landscape analysis reveals components of binding thermodynamics for CALP:kCAL01. Upper bounds on the Boltzmann-weighted partition function computed using the MARK\* algorithm<sup>63</sup> in OSPREY<sup>37</sup> for a 15-residue design at the protein–protein interface of CALP:kCAL01 are shown as colored ring charts. A brief explanation of the ring chart diagram can be found in Section 2.4. (A, B) Energy landscapes for CALP in the bound (A) and unbound (B) states show the change in conformation distribution induced by binding. (A) Bound CALP has a narrow distribution, with the GMEC accounting for nearly 50% of the partition function. (B) In contrast, unbound CALP shows a wide conformation distribution, with the unbound GMEC accounting for roughly 5% of the partition function, with conformational entropy generated largely by residues Thr296 and His301, indicated by a large number of similarly sized arcs at their corresponding rings. Importantly, the bound and unbound GMECs are not the same conformation and do not have the same energy. (C, D) Energy landscapes for kCAL01 in the bound (C) and unbound (D) states show a similar change in conformation distribution upon binding. (C) The bound kCAL01 energy landscape shows that the GMEC accounts for roughly 5% of the partition function. Even so, the landscape suggests considerable entropy, driven by residues at P<sup>-4</sup>, P<sup>-1</sup>, and P<sup>-2</sup>, and P<sup>0</sup>. Prediction of conformational heterogeneity at P<sup>0</sup> is particularly interesting given its buried location. (D) Conversely, unbound kCAL01 shows a very high-entropy conformation distribution, with many conformations that contribute to the partition function, as seen by the presence of many small arcs in the outer ring. This is consistent with our expectations for an extended peptide backbone. Thermodynamic parameters calculated from these partition functions (see Section 3.2.2) indicate that binding of kCAL01 to CALP results in a decrease in internal energy and a decrease in entropy, which is represented visually in these energy landscapes.

values, decreasing internal energy  $U$  by a factor of 4, similar to the method described in ref 63.

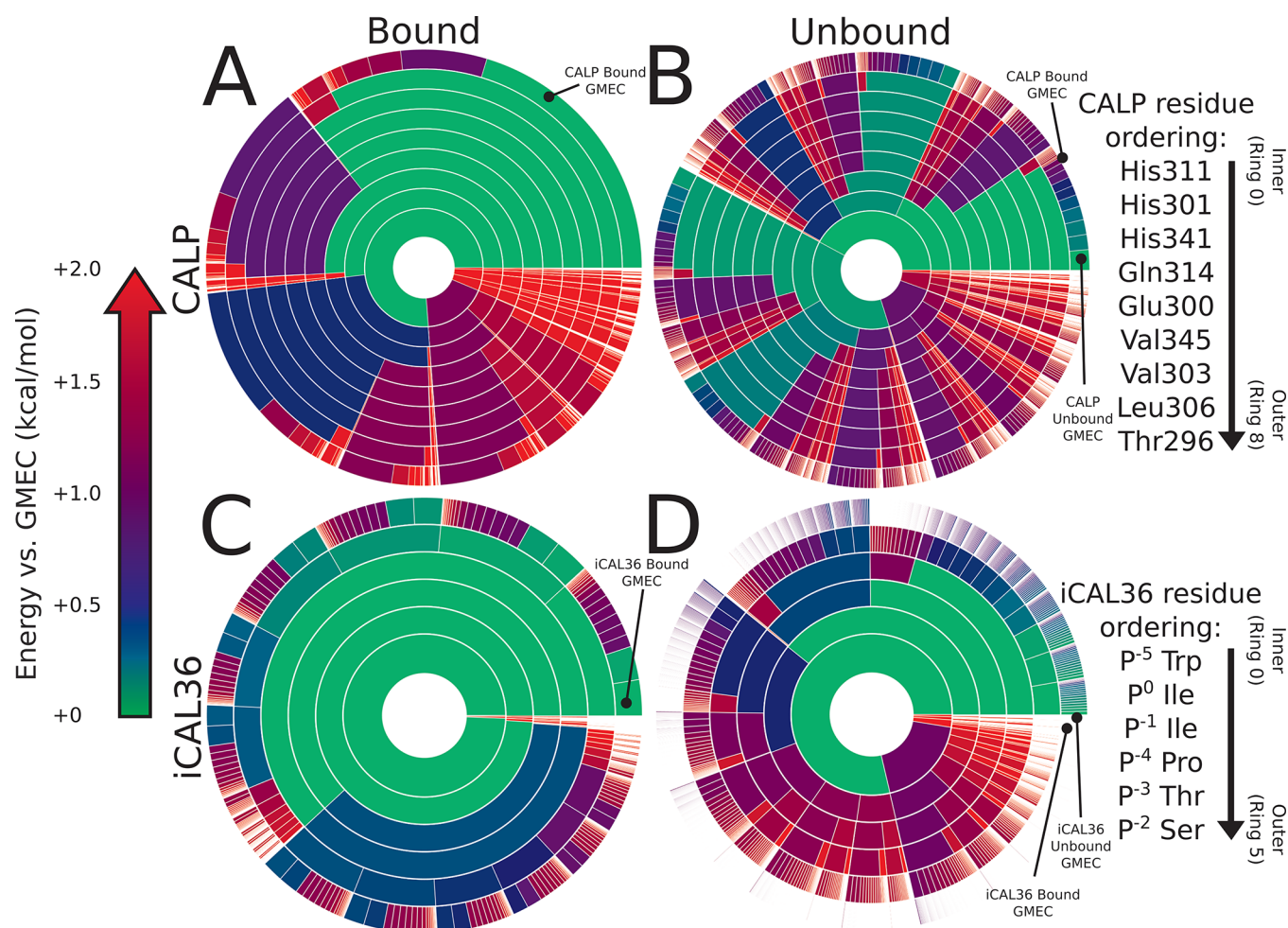
#### 2.4. Interpretation of Energy Landscape Diagrams.

Visual representations of computed energy landscapes can be found in Figures 4 and 5. For a full description of ring diagram visualizations, see ref 63. Briefly, each concentric ring represents a design amino acid residue, with each ring arc representing a single rotamer assignment to that residue given the residue assignments of the inner arcs, or “partial conformation”. Therefore, any arc in the outermost ring represents a “full conformation”, where all amino acid positions are each assigned a single rotamer. The angle of any given arc corresponds to the partition function contribution of all conformations that contain the given partial conformation. The

color of any given arc corresponds to the smallest energy difference between the GMEC and the lowest-energy conformation that contains the given partial conformation, with small energy differences colored green, and larger energy differences colored red. Notably, white gaps are indicative of relatively high-energy conformations that individually contribute less than 0.1% of the partition function value. Therefore, a ring diagram visually represents the entire energy landscape for a design problem, showing the distribution of conformations according to their probability.

### 3. RESULTS AND DISCUSSION

**3.1. Structural Analysis of CALP:kCAL01.** Co-crystals were formed with recombinant CALP and a decapeptide



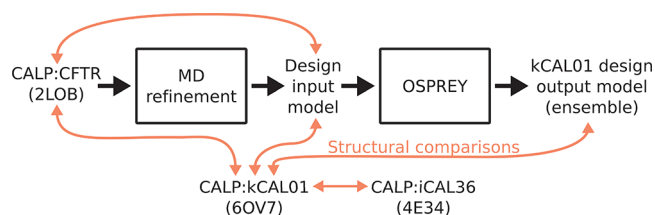
**Figure 5.** Energy landscape analysis reveals components of binding thermodynamics for CALP:iCAL36. Upper bounds on the Boltzmann-weighted partition function computed using the MARK\* algorithm<sup>63</sup> in OSPREY<sup>37</sup> for a 15-residue design at the protein–protein interface of CALP:iCAL36 are shown as colored ring charts. A brief explanation of the ring chart diagram can be found in Section 2.4. (A, B) Energy landscapes for CALP in the bound and unbound states show the change in conformation distribution induced by binding. (A) Bound CALP has a narrower distribution, with the GMEC accounting for roughly 20% of the partition function. (B) Unbound CALP shows a wide conformation distribution, with the unbound GMEC accounting for roughly 2% of the partition function, with conformational variation in multiple residues, indicated by a large number of similarly sized arcs at multiple rings. (C, D) Energy landscapes for iCAL36 in the bound and unbound states show a change in conformation distribution upon binding. (C) The bound iCAL36 energy landscape exhibits a lower-entropy distribution, with the GMEC accounting for roughly 3% of the partition function. Even so, the landscape suggests considerable heterogeneity, much of which is attributable to variation at P<sup>-2</sup>, and P<sup>-3</sup>. Interestingly, in contrast to the bound kCAL01 landscape, little heterogeneity is predicted at P<sup>0</sup>. (D) Unbound iCAL36 has a high-entropy conformation distribution, with many conformations that contribute to the partition function, as seen by the presence of many small arcs in the outer ring. This is consistent with our expectations for an extended peptide backbone. Thermodynamic parameters calculated from these partition functions (see Section 3.2.2) indicate that binding of iCAL36 to CALP results in a decrease in internal energy and a decrease in entropy, which is represented visually in these energy landscapes.

variant of kCAL01 with a four-residue N-terminal extension (ANSRWQVTRV, extension in italics). The refined model exhibits an excellent fit to the density, with  $R_{\text{work}}$  and  $R_{\text{free}}$  of 0.182 and 0.214, respectively, and lies within typical peptide geometry constraints (Table 1). The asymmetric unit consists of two protomers (A and B) of CALP complexed with kCAL01 (Figure S1A). The 9 or 8 C-terminal residues of the peptide ligand are well-resolved in protomers A and B, respectively. The crystal structure was deposited as PDB ID: 6OV7.

Alignment of protomers A and B of 6OV7 by CALP main-chain atoms using PyMOL<sup>77</sup> results in good overlap, with 281 of the 348 total backbone atoms aligning with an RMSD of 0.32 Å. Notable differences can be seen between the two protomer structures at two sites: helix  $\alpha 1$  and adjacent to the carboxylate-binding loop (CBL). Significant distortion of the

protomer B  $\alpha 1$  helix results from an interprotomer disulfide bond between CALP residues 319 (protomer A) and 319 (protomer B) (Figure S1A,B). We hypothesize that this disulfide bond and resulting helix  $\alpha 1$  distortion are artifacts of crystallization. Pronounced conformational differences adjacent to the protomer A and B CBLs, which connect the  $\beta$ -strands  $\beta 1$  and  $\beta 2$ , are evident between CALP residues 284 and 289 (Figure S1B). However, these differences occur upstream of the CBL sequence motif residues 291 ( $\Phi_1$ ) and 293 ( $\Phi_2$ ) and do not appear to affect peptide binding. Due to the distortion of protomer B helix  $\alpha 1$ , the following analysis focuses on the protomer A CALP:kCAL01 structure.

**3.1.1. Gross Structural Analysis of CALP:kCAL01 Reveals Canonical PDZ:Peptide Binding.** The overall topology of the protomer A CALP fold, composed of 5  $\beta$ -strands and 2  $\alpha$ -

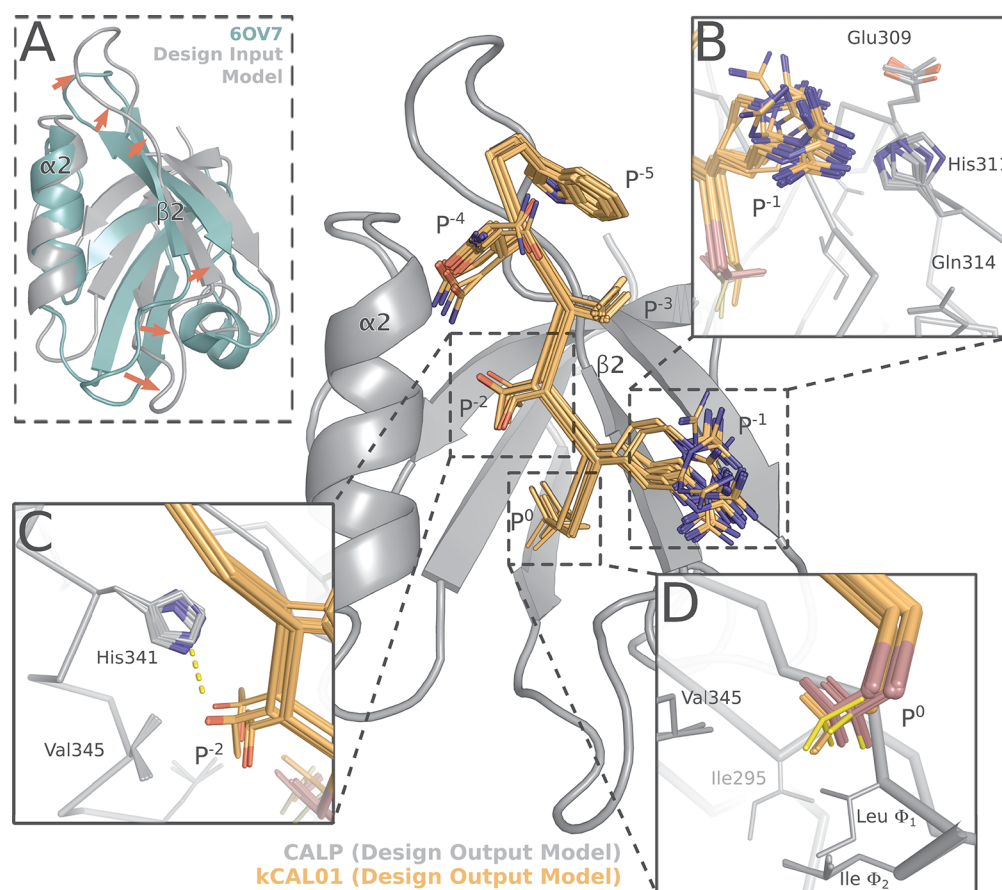


**Figure 6.** Schematic diagram of design process reported in ref 35 and structural comparisons presented in this work. A flowchart of design work performed in ref 35 is shown with black arrows. First, a design input model was generated by performing MD refinement of an NMR structure of CALP:CFTR.<sup>24</sup> Using OSPREY, this input model was used for the  $K^*$  algorithm's<sup>38</sup> ensemble-based design of peptide inhibitors of CALP, resulting in a design output model (ensemble) of CALP:kCAL01.<sup>35</sup> Finally, in this work we perform detailed structural comparisons between several CALP:peptide structures and models, indicated by red arrows.

helices, matches well with previous CALP structures and represents a canonical PDZ fold.<sup>2,3</sup> Class 1 PDZ domains bind to peptides containing a C-terminal S/T-X- $\Phi$  binding motif, where  $\Phi$  is hydrophobic and X is any residue, and form an antiparallel  $\beta$ -sheet with the  $\beta 2$  strand.<sup>28</sup> kCAL01 binds in a

manner consistent with typical class 1 PDZ domains, occupying the groove defined by helix  $\alpha 2$  and strand  $\beta 2$  (Figure 1). Four main-chain hydrogen bonds are formed between CALP  $\beta 2$  and the 3 C-terminal kCAL01 residues  $P^{-2}$ – $P^0$ , forming an antiparallel  $\beta$ -strand interaction (Figure 1B). This positions the most C-terminal peptide residue ( $P^0$ ) such that the main-chain carboxyl terminus interacts with the CBL, defined by an  $X\Phi_1G\Phi_2$  sequence motif, where  $\Phi_i$  is a hydrophobic amino acid.<sup>3,26</sup> Additionally, the hydrophobic  $P^0$  side chain is buried in the pocket defined by CALP residues Leu291, Ile293, Ile295, Val345, and Leu348 (Figure 1C). The CALP:kCAL01 structure also contains the critical hydrogen bond between Thr  $P^{-2}$  and His341, which plays a significant role in defining PDZ domain class 1 specificity<sup>28</sup> (Figure 1C). Notably, only the six C-terminal residues of the extended kCAL01 peptide form any direct contacts with CALP; the four additional N-terminal residues make only lattice contacts and were added to facilitate crystallization. Overall, this structure depicts a binding interaction that is consistent with the structural characteristics observed for canonical class 1 PDZ domains.<sup>3,28</sup>

To evaluate the basis for the enhanced efficiency of the CALP:kCAL01 binding interaction, we compared this crystal



**Figure 7.** Structural analysis of the kCAL01 design models<sup>35</sup> reveals similarities to the CALP:kCAL01 crystal structure. The design output model ensemble<sup>35</sup> of CALP (gray) bound to kCAL01 (orange) closely resembles the bound CALP:kCAL01 crystal structure (Figure 1). (A) Comparison of the design input model (gray) and 6OV7 crystal structure (green) CALP conformations shows significant shifts (red arrows) in strand  $\beta 2$  and the  $\beta 2$ – $\beta 3$  loop. These shifts greatly expand the binding cleft between helix  $\alpha 2$  and strand  $\beta 2$ . The shift in  $\beta 2$ – $\beta 3$  loop conformation is a result of MD refinement, as it is not present in the structure of 2LOB before refinement (see Figure S2). The 100 conformations in the design output model ensemble capture interactions between (B) Arg  $P^{-1}$  and His311, as well as (C) the interaction between Thr  $P^{-2}$  and His341. (D) The design output ensemble models conformational heterogeneity at  $P^0$ , suggesting that modeling of entropy at this site was important for this design's success. The t, p, or m rotamers of  $P^0$  are shown in red, orange, and yellow, respectively.



structure (PDB ID: 6OV7, protomer A) to the structure of CALP:iCAL36 (PDB ID: 4E34, protomers A and B), a previously developed inhibitor of CALP that also exhibits in-cell activity,<sup>29</sup> but binds less tightly to CALP. We note that CALP residue numbering differs between these two structures, with numbering for 4E34 +8 relative to 6OV7. Unless otherwise noted, all residue numbering refers to the 6OV7 numbering convention.

**3.1.2. Comparison of CALP:kCAL01 to CALP:iCAL36 Reveals Differences in Carboxylate-Binding Loop Conformation.** First, we analyzed the CBL conformation and peptide orientation, because previous work<sup>26</sup> demonstrated that these features play a role in modulating CALP specificity for peptide residue P<sup>0</sup>. In particular, through analysis of 4E34, ref 26 presented two structural mechanisms by which CALP accommodates a Ile P<sup>0</sup> side chain: (1) a CBL conformation that narrows the entrance to the hydrophobic binding pocket concomitant with an N-terminal peptide shift (4E34, protomer A), and (2) a CBL conformation that widens the entrance to the hydrophobic binding pocket concomitant with a change in rotamer at Leu  $\Phi_1$ , thus expanding the hydrophobic binding pocket (4E34, protomer B). kCAL01, in contrast, has a valine at position P<sup>0</sup>. To investigate the structural consequences of this substitution, we aligned each protomer of CALP:iCAL36 to CALP:kCAL01 by the main-chain atoms of CALP secondary-structure elements  $\beta_2$  and  $\alpha_2$ , which flank the peptide-binding groove. Structures of CALP:kCAL01 protomer A and CALP:iCAL36 protomer A showed good correspondence at the binding pocket, with an RMSD of 0.24 Å, and CALP:kCAL01 protomer A and CALP:iCAL36 protomer B aligned with an RMSD of 0.41 Å (RMSD calculated using backbone heavy atoms of secondary-structure elements  $\alpha_2$  and  $\beta_2$ ).

The overall binding geometry of the CBL and peptide for CALP:kCAL01 contains distinct features of both CALP:iCAL36 protomers A and B (Figure 2). The CBL conformation of kCAL01 at residue  $\Phi_2$  is more similar to that of the iCAL36 protomer A than protomer B, with  $C_\alpha$  deviations of 0.5 and 1.3 Å, respectively (see Figure 2A). Additionally, the rotamer at loop residue  $\Phi_1$  matches with iCAL36 protomer A (Figure 2B). Overall, the CALP CBL when bound to kCAL01 adopts a conformation that narrows the entrance of the hydrophobic binding pocket, which is similar to the previously reported CALP:iCAL36 protomer A structure (Figure 2B). This suggests that kCAL01 binding to CALP does not require hydrophobic pocket expansion to accommodate the Val P<sup>0</sup>.

However, the bound kCAL01 peptide shifts toward the CBL, similar to the CALP:iCAL36 protomer B structure (Figure 2C). kCAL01 shifts toward the CBL relative to iCAL36 protomer A by 0.7 Å measured at the P<sup>0</sup>  $C_\alpha$  (Figure 2B). This results in side-chain positioning at kCAL01 Val P<sup>0</sup> that is intermediate to CALP:iCAL36 protomers A and B. This shift propagates up the backbone of the peptide, as the P<sup>-1</sup>  $C_\alpha$  also shifts by 0.8 Å. These results suggest that the presence of a valine at P<sup>0</sup>, rather than a sterically larger leucine or isoleucine, allows shifts in the peptide backbone that accommodate the less common C-terminal side chain within a high-affinity interaction.

On the basis of this structural analysis, it is unclear how the changes in P<sup>0</sup> binding mode between CALP:kCAL01 and CALP:iCAL36 affect binding affinity. On one hand, the Val P<sup>0</sup> present in kCAL01 appears to allow a peptide C-terminal shift

without requiring a shift in the CBL and hydrophobic pocket expansion. On the other hand, it is not clear whether this C-terminal peptide shift is either favorable or unfavorable for binding. Indeed, the interactions formed by Val and Ile P<sup>0</sup> in 6OV7 and 4E34, respectively, appear qualitatively similar, and the inclusion of the sterically larger Ile could more effectively fill the pocket. Overall, more analysis is needed to clarify the effects of structural variation at this site. This analysis is provided in Section 3.2.1, where investigation of CALP:kCAL01 and CALP:iCAL36 energy landscapes suggests that these structural shifts allow the kCAL01 Val P<sup>0</sup> to sample three favorable rotamers, which we predict to be favorable for binding.

**3.1.3. Comparison of CALP:kCAL01 and CALP:iCAL36 at Modulator Residues Reveals Interactions That Favor kCAL01 Binding.** Previous work<sup>25,27,65</sup> pinpointed “modulator” residues at P<sup>-1</sup>, P<sup>-3</sup>, and P<sup>-4</sup>–P<sup>-9</sup> that show individually modest effects on binding and specificity but together can create significant effects. We compared the CALP:kCAL01 and CALP:iCAL36 protomer A structures in order to determine the effect of these modulator residues on inhibitor binding.

kCAL01 contains an arginine residue at P<sup>-1</sup> that interacts with His311 in an apparent  $\pi$ –cation interaction (Figure 1D). In contrast, iCAL36 contains an isoleucine at this position, which forms minor van der Waals interactions with His311 and Ser294. These interactions appear to be much less extensive than those formed by the Arg P<sup>-1</sup> (see Figure S3A). Favorable interactions between CALP and kCAL01 are indicated at Gln P<sup>-4</sup>, which hydrogen bonds with Glu300 or His301, in addition to forming van der Waals interactions with His341 and interacting with several waters (Figure 1A). While the Pro P<sup>-4</sup> found in the CALP:iCAL36 structure does interact with His301, His341, and a single water molecule, these interactions appear to be less favorable (see Figure S3B).

kCAL01 and iCAL36 differ only slightly at P<sup>-3</sup>, containing a valine and threonine residue, respectively. Both residues form van der Waals interactions between a methyl group and Ser308, but only the kCAL01 Val P<sup>-3</sup> forms interactions with Thr296 due to the larger steric volume of the methyl group. These differences, while minor, suggest slightly more favorable interactions for kCAL01 at this position (see Figure S3C).

Overall, the most notable structural differences in binding stereochemistry between kCAL01 and iCAL36 occur at residues P<sup>-1</sup> and P<sup>-4</sup>, where mutations to long polar and charged residues likely result in an increase in favorable energetic interactions. These results suggest that sequence differences shift the thermodynamic balance: the more hydrophobic iCAL36 peptide may have higher energy alone in solvent, whereas the more polar kCAL01 sequence is preferentially stabilized in the bound state.

**3.2. Energy Landscape Analysis of CALP:kCAL01.** Conformational entropy can play a significant role in defining protein structure and function.<sup>79–81</sup> For this reason, when modeling binding of protein:ligand complexes, it is useful to compute partition functions over protein ensembles to better model and understand binding thermodynamics.<sup>35,38,40–46</sup> To approximate the conformational ensembles involved in CALP:kCAL01 and CALP:iCAL36 binding, we computed partition functions and energy landscapes using OSPREY for bound and unbound models of CALP:kCAL01 (PDB ID: 6OV7, protomer A) and CALP:iCAL36 (PDB ID: 4E34, protomer A) as described in Section 2.2. We compared energy landscape features of CALP:kCAL01 and CALP:iCAL36 to

reveal dynamic features that contribute to CALP:kCAL01 binding. Furthermore, we used these energy landscapes to compute approximations to thermodynamic components of binding (described in Section 2.3) to analyze differences in binding of CALP:kCAL01 and CALP:iCAL36. A discussion on how to interpret energy landscape diagrams presented herein can be found in Section 2.4.

**3.2.1. Energy Landscape Comparison of CALP:kCAL01 and CALP:iCAL36 Foregrounds Structural and Dynamic Features That Explain Differences in Binding.** Detailed comparison of CALP:kCAL01 (Figure 4) and CALP:iCAL36 (Figure 5) landscapes reveals local differences in side-chain conformational distributions for each structure. This is most notable when comparing the bound inhibitor landscapes of kCAL01 and iCAL36 (Figures 4C and 5C, respectively). Note that for ease of comparison we have decomposed the bound complex CALP:kCAL01 landscape into bound CALP and kCAL01 landscapes. The original convolved landscapes can be found in Figure S4.

The bound kCAL01 landscape (Figure 4C) indicates that residue Val P<sup>0</sup> adopts three significant rotamers, shown by subdivision into three arcs at the outermost ring (ring 5). In contrast, the bound iCAL36 landscape (Figure 5C) indicates that residue Ile P<sup>0</sup> adopts only one significant rotamer, shown in the second ring from the center (ring 1). Structural analysis of the rotamer distribution for this residue (Figure 3) suggests that Val P<sup>0</sup> forms favorable interactions with the CALP hydrophobic pocket in each of three rotamers (Figure 3A–C) defined by a rotation of  $\sim 60^\circ$  around the N–C <sub>$\alpha$</sub> –C <sub>$\beta$</sub> –C <sub>$\gamma$ 1</sub> dihedral angle. As a result, the landscape analysis of bound kCAL01 suggests that residue Val P<sup>0</sup> interacts with low energy and locally high entropy. Conversely, the iCAL36 Ile P<sup>0</sup> is likely too large to interact favorably in multiple rotameric states and is predicted to occupy only one significant rotamer. These predicted differences in conformational heterogeneity at P<sup>0</sup> could help explain the improved binding efficiency of kCAL01. Ensemble features are difficult to visualize when examining only a static crystal structure, but are now made clear by the energy landscape analysis.

Comparison of the energy landscapes of bound CALP for the CALP:kCAL01 and CALP:iCAL36 models (Figures 4A and 5A) also reveals differences in conformational heterogeneity. CALP bound to kCAL01 appears to be heavily conformationally restricted, with the GMEC occupying nearly 50% of the landscape (Figure 4A), while CALP bound to iCAL36 is less conformationally restricted, with the GMEC occupying roughly 20% of the landscape (Figure 5A). These differences appear to be driven in large part by differences in residue conformational heterogeneity at His311 and His301 (CALP:kCAL01 residue numbering, CALP:iCAL36 numbering is +8 relative), which can be seen by comparing the innermost two rings (ring 0 and 1) in the two bound CALP landscapes (Figures 4A and 5A). For each of ring 0 (the innermost) or 1 (the second innermost), the bound CALP landscape in the iCAL36 structure shows an additional minor rotamer population, shown by a purple or blue arc, respectively. This indicates that the rotamer distribution for His311 and His301 in the bound CALP:iCAL36 model has more entropy than that in the CALP:kCAL01 model.

The greater calculated side-chain entropy for His301 and His311 in the iCAL36-bound state might appear counter-intuitive, given the better affinity of the kCAL01 complex. However, these entropic contributions may be offset by other

changes in the protein and by a loss of favorable energetic interactions locally. Indeed, examination of structural interactions between His311 and His301 and the peptide inhibitor for CALP:kCAL01 and CALP:iCAL36 structures suggests that this relative increase in entropy for CALP:iCAL36 can be explained by a loss of interactions between His311 and peptide P<sup>-1</sup>, and between His301 and peptide P<sup>-4</sup>. Specifically, kCAL01 Arg P<sup>-1</sup> forms strong  $\pi$ -stacking interactions with His311, while iCAL36 Ile P<sup>-1</sup> forms weaker interactions with the analogous His319. Similarly, kCAL01 Gln P<sup>-4</sup> forms a hydrogen bond with His301, while iCAL36 Pro P<sup>-4</sup> forms van der Waals interactions. As a result, we expect these histidines to form more energetically favorable interactions with kCAL01 than iCAL36. Our models predict that these favorable interactions are sensitive to the rotamer choice at His311 and His301, resulting in a less conformationally heterogeneous ensemble at these positions.

Thus, energy landscape analysis both reveals conformational heterogeneity at Val P<sup>0</sup> and draws attention to important modulator residue interactions in the bound state. These ensemble features are not clearly evident from electron density or B-factor analysis, indicating that our models capture information that is missed by traditional structural analysis. This information could likely be captured by measuring 3-bond scalar couplings by NMR,<sup>82</sup> analysis of a higher-resolution structure with Ringer<sup>83</sup> or qFit,<sup>84–86</sup> or use of advanced crystallography techniques including room-temperature<sup>85,87</sup> and multitemperature<sup>88</sup> crystallography.

**3.2.2. Thermodynamics of CALP:kCAL01 and CALP:iCAL36 Energy Landscapes Reveal Decreases in Internal Energy and Entropy upon Binding.** Energy landscapes of CALP:kCAL01 binding visualize the loss of entropy upon binding. Figure 4 depicts energy landscapes of the unbound CALP (Figure 4B), unbound kCAL01 (Figure 4D), and bound CALP and kCAL01 (Figure 4A,C, respectively). Comparison of the unbound and bound ensemble landscapes for CALP (Figure 4A,B) reveals the significant loss of entropy due to conformational rearrangement upon binding. The unbound CALP landscape (Figure 4B) shows many low-energy conformations that contribute to the partition function, indicated by the many green-blue arcs in the outermost ring. Conversely, the bound ensemble of CALP (Figure 4A) is dominated by a single low-energy conformation, indicated by the large green arc, with the rest of the landscape occupied by higher-energy minor conformations. This indicates that the conformational rearrangement due to binding of kCAL01 imposes a significant entropic cost that must be compensated for by the gain of favorable intermolecular interactions.

Comparison of the unbound and bound ensemble landscapes for kCAL01 (Figure 4C,D, respectively) reveals a similar picture, with the decrease in entropy upon binding illustrated by the decrease in number and increase in size of the arcs in the outermost ring. Notably, for the unbound kCAL01 landscape, the GMEC occupies less than 0.2% of the partition function and outer rings are characterized by extensive white-space, indicating the presence of many conformations that occupy individually less than 0.1% of the partition function. Together, these features are indicative of a high-entropy landscape. In contrast, the bound kCAL01 landscape is characterized by fewer conformations that contribute relatively more to the partition function, and a GMEC that occupies roughly 5% of the partition function. These landscape

representations depict the loss of entropy upon binding of CALP:kCAL01.

Using these energy landscapes, we calculated approximations to the ensemble-weighted internal energy and entropy for the bound and unbound states of CALP:kCAL01 as described in Section 2.3. Additionally, we computed the same approximations for the binding-competent ensemble<sup>89</sup>—an alchemical state defined by the conformations and occupancies found in the bound protein or ligand modeled with the energy field of the unbound state—to deconvolve the changes in entropy vs internal energy upon binding. Conveniently, as shown previously<sup>89</sup> this construction allows us to decompose binding into an “induced fit” step, involving a change in conformation distribution, and a “lock and key” step, in which protein:ligand interactions are formed, without regard to actual mechanism. At the chosen temperature of ~298 K, both CALP and kCAL01 exhibit a change in conformation distribution upon binding. This change in distribution results in a change in internal energy and entropy for CALP of 0.113 and -1.97 kcal/mol, respectively, and for kCAL01 of 0.234 and -2.23 kcal/mol, respectively, quantifying the large decrease in entropy due to binding, where the total contribution of the entropy change to the Helmholtz free energy  $-T\Delta S_{\text{binding}}$  is +4.19 kcal/mol. Complex formation—the “lock and key” step—results in a decrease in internal energy: whereas the combined internal energy of the unbound models is -18.84 kcal/mol, the bound model internal energy is -28.49 kcal/mol, resulting in a  $\Delta U_{\text{binding}}$  of -9.65 kcal/mol. As a result, the approximated change in Helmholtz free energy due to binding  $\Delta F_{\text{binding}}$  is -5.46 kcal/mol. These models suggest that both binding partners incur penalties to entropy and internal energy when adopting the binding-competent ensemble, which are compensated for by a large decrease in internal energy upon complex formation.

Energy landscapes of CALP:iCAL36 binding reveal a loss of entropy upon binding that is similar to that of CALP:kCAL01. Figure 5 depicts energy landscapes of the unbound CALP (Figure 5B), unbound iCAL36 (Figure 5D), and bound CALP and iCAL36 (Figure 5A,C, respectively). Comparison of bound and unbound states for CALP and iCAL36 also reveal a decrease in entropy upon binding for both binding partners, indicated by CALP and iCAL36 energy landscapes exhibiting a reduction in the number of arcs and an increase in arc size upon binding.

Approximations of ensemble-weighted internal energy and entropy for bound, binding-competent ensemble, and unbound models of CALP:iCAL36 revealed a smaller decrease in entropy upon binding compared to CALP:kCAL01, but also showed a smaller decrease in internal energy upon binding. At a temperature of ~298 K, our models indicate that both CALP and iCAL36 undergo a change in conformation distribution upon binding. This results in a change in internal energy and entropy for CALP of 0.004 and -1.60 kcal/mol, respectively, and for iCAL36 of 0.558 and -1.89 kcal/mol, respectively, illustrating a decrease in entropy due to binding, where the total contribution of the entropy change to the Helmholtz free energy  $-T\Delta S_{\text{binding}}$  is +3.48 kcal/mol. Complex formation results in a decrease in internal energy: whereas the combined internal energy of the unbound models is -17.7 kcal/mol, the bound model internal energy is -25.6 kcal/mol, resulting in a  $\Delta U_{\text{binding}}$  of -7.84 kcal/mol. As a result, the approximated change in Helmholtz free energy due to binding  $\Delta F_{\text{binding}}$  is -4.36 kcal/mol. Similar to CALP:kCAL01, both binding

partners undergo a loss of entropy and reduction in internal energy upon binding.

Overall, these landscapes and thermodynamic calculations suggest that although both CALP:kCAL01 and CALP:iCAL36 undergo a decrease in both entropy and internal energy upon binding, the energetic interactions gained upon binding are less favorable for CALP:iCAL36 than for CALP:kCAL01. This is reflected in the change in internal energy due to binding for each model, with CALP:kCAL01 and CALP:iCAL36 exhibiting  $\Delta U_{\text{binding}}$  of -9.65 kcal/mol and  $\Delta U_{\text{binding}}$  of -7.84 kcal/mol, respectively. Although the entropic penalty due to binding is less for CALP:iCAL36, these models predict that kCAL01 binds more tightly to CALP than does iCAL36, with  $\Delta F_{\text{binding}}$  values of -5.46 kcal/mol and -4.36 kcal/mol, respectively. Despite the fact that our models account for only a subset of biologically relevant flexibility (see Section 2.3), this predicted 1.1 kcal/mol difference in free energy of binding is semi-quantitatively in line with experimentally determined inhibition constants<sup>25,26,35</sup> (see Table S1), suggesting that these models are capturing biologically relevant features.

These results suggest that side-chain conformational entropy played a role in the design for improved binding efficiency of kCAL01, an observation that supports related work by Head-Gordon and co-workers.<sup>81</sup> Previous studies<sup>24</sup> suggested that backbone flexibility may play a role in CALP:peptide binding, but backbone conformational entropy was not addressed in this study. Investigation into the effects of backbone flexibility on predicted energy landscapes and binding thermodynamics for both kCAL01 and CALP is a promising avenue for future work. In particular, future directions include modeling backbone flexibility with the CATS<sup>59</sup> and DEEPER<sup>90</sup> algorithms in OSPREY, focusing on the CALP  $\alpha 2$  helix,  $\beta 1$ - $\beta 2$  loop, and  $\beta 2$ - $\beta 3$  loop. Investigation of the backbone flexibility of iCAL36 and kCAL01 would also be valuable, especially given that iCAL36 contains a Pro P<sup>-4</sup>, whereas kCAL01 contains a Gln P<sup>-4</sup>.

**3.3. Design Model Corresponds Closely with Bound Crystal Structure.** In this section, we briefly compare the design model reported in ref 35 with 6OV7 to determine which key structural features contributed to a successful kCAL01 design. We perform structural comparisons for both the CALP:kCAL01 *design output model* (defined as the ensemble of structures output from the  $K^*$  algorithm<sup>38</sup> in OSPREY) and the CALP:CFTR *design input model* (defined as the structural input to OSPREY). A schematic diagram depicting these definitions is shown in Figure 6. We begin by comparing gross features, and then we proceed to identify shared side-chain energetic and dynamic interactions.

To determine the accuracy of the structural design model reported in ref 35, we compared the crystal structure of CALP:kCAL01 to the ensemble of 100 low-energy structures that comprise the CALP:kCAL01 design output model<sup>35</sup> (Figure 7). We aligned members of the design output ensemble to the CALP:kCAL01 crystal structure using the main chain of secondary-structure elements  $\alpha 2$  and  $\beta 2$  and obtained good alignment quality, with one representative structure aligning with an RMSD of 0.94 Å. Deviations were primarily a result of a more relaxed hydrophobic binding pocket in the design model relative to the CALP:kCAL01 crystal structure 6OV7, involving an outward shift of the  $\beta 2$  strand (Figure 7A). Additionally the loop connecting  $\beta 2$  and  $\beta 3$  adopts a different conformation, resulting in a large loop

shift and a significant change in orientation of the  $\beta 2$  strand (Figure 7A).

We hypothesized that these differences were inherited from the CALP:CFTR design input model, generated during MD refinement<sup>24,35</sup> of the bound NMR structure of CALP:CFTR (PDB ID: 2LOB).<sup>24</sup> This MD-refined structure was chosen as the design input<sup>35</sup> due to the optimization of the  $\beta$ -strand interactions between the CFTR peptide and strand  $\beta 2$ .<sup>24</sup> In order to test this hypothesis, we aligned both 2LOB NMR model 1 and the design input model to 6OV7 by the main-chain atoms of secondary-structure elements  $\alpha 2$  and  $\beta 2$  which revealed good alignment quality, with an RMSD of 0.69 and 0.94 Å, respectively. Overall, 2LOB is more relaxed than the bound CALP:kCAL01 structure, with slight outward shifts and changes in angle in both helix  $\alpha 2$  and strand  $\beta 2$  that result in an apparent expansion of the hydrophobic pocket that interacts with peptide residue P<sup>0</sup> (Figure S2). However, the loop connecting  $\beta 2$  and  $\beta 3$  shows good correspondence between the CALP:CFTR NMR structure and the CALP:kCAL01 crystal structure (Figure S2). The change in loop conformation and significant reorientation of strand  $\beta 2$  is a result of MD refinement and does not appear in either the unrefined NMR structure (2LOB), the CALP:kCAL01 crystal structure (6OV7), or indeed in the CALP:iCAL36 structure (4E34). Therefore, we conclude that deviations in CALP conformation observed in the design output model of ref 35 were inherited from MD refinement of a relaxed NMR structure.

Nonetheless, this NMR-based design model captured key structural and ensemble properties of the CALP:kCAL01 complex,<sup>91</sup> which allowed OSPREY to design kCAL01, the most binding-efficient inhibitor of CALP to date. Almost all members of the design ensemble predict favorable interactions between Arg P<sup>-1</sup> and His311 (Figure 7B). Additionally, all three rotamers of Val P<sup>0</sup> (t, p, and m) appear in the design output ensemble, indicating that the K\* algorithm successfully identified sequences with multiple low-energy states (Figure 7D). A significant subset of the design ensemble captures the important hydrogen bond between His341 and Thr P<sup>-2</sup>, indicating that the design of kCAL01 captured key components of the class 1 PDZ binding geometry<sup>28</sup> (Figure 7C). Interestingly, some members of the design output ensemble do not contain the hydrogen bond between histidine 341 and threonine P<sup>-2</sup>, consistent with observations of hydrogen bond breaking and reformation from experimental<sup>92</sup> and MD simulation<sup>93–96</sup> studies. Indeed, solution NMR studies of ubiquitin showed that threonine residues occupy multiple rotameric states by measuring 3-bond scalar couplings.<sup>82</sup> Overall, the crystal structure and design ensemble are quite quantitatively and qualitatively similar, despite differences in CALP structure, indicating that the design presented in ref 35 succeeded in capturing important structural and ensemble interactions. We conclude that these features allowed OSPREY to design kCAL01, the most binding-efficient inhibitor of CALP to date with rescue activity for F508del-CFTR, a disease-associated variant present in approximately 90% of CF patients. Key interactions and entropic effects predicted by the OSPREY design model are supported by the new crystal structure and landscape analysis presented herein.

**3.4. PDZ Domain Energy Landscapes.** Modeling of energy landscapes complements traditional structural analysis of CALP:peptide crystal structures and provides a novel way to probe the conformational distribution available to the protein complex. We submit that these tools may prove useful for

analyzing PDZ:peptide complexes in general. To investigate differences in conformational distributions for PDZ domains, we computed partition functions and energy landscapes for 10 structures of bound PDZ:peptide complexes. A preliminary analysis of these landscapes reveals a general trend of loss of entropy upon binding for both PDZ and peptide ligands, similar to that observed for CALP:kCAL01 and CALP:iCAL36. Additionally, these supplementary energy landscapes predict no significant conformational heterogeneity for any studied system at the peptide position P<sup>0</sup> in the bound state, contrasting with the heterogeneity we observed for CALP:kCAL01 at Val P<sup>0</sup>. This raises the intriguing possibility that, similarly to kCAL01 for CALP, more binding-efficient inhibitors for other PDZ domains could be designed by maximizing relative entropy at P<sup>0</sup>. We include these energy landscapes for the scientific community in SI Section S1.3 in the hope that these insights and data may be of further benefit.

#### 4. CONCLUSION

In this work we investigated the basis for the binding efficiency of kCAL01, an OSPREY-designed peptide inhibitor of CALP that rescued functional CFTR activity as assessed by *in vitro* Ussing chamber assays.<sup>35</sup> On the basis of structure and energy landscape analysis of the new crystal structure of CALP:kCAL01, we conclude that the comparative binding efficiency of kCAL01 stems from entropic effects at P<sup>0</sup> and substitutions that result in more favorable energetic interactions at modulator residues. This conclusion is supported not only by comparative analysis of the CALP:kCAL01 and CALP:iCAL36 crystal structure conformations, but also by investigating energy landscapes for each ensemble model. We used energy landscape analysis enabled by the MARK\* algorithm<sup>63</sup> in OSPREY to provably approximate the energies of all conformations in each ensemble model, generated by assigning flexibility to residues in each crystal structure. These landscapes probed local residue conformational heterogeneity and enabled us to approximate binding thermodynamics to correctly predict that kCAL01 binds more tightly to CALP than does iCAL36. We conclude that modeling of energy landscapes complemented traditional structural analysis of CALP:peptide crystal structures and provided a novel way to probe the conformational distribution available to the protein complex. Modeling energy landscapes may prove useful for analyzing PDZ:peptide complexes in general, and hence, we provide energy landscapes for 10 additional bound PDZ:peptide complexes. Finally, we show that our successful design of kCAL01<sup>35</sup> was a result of effective modeling of both energetic and ensemble properties of CALP:peptide binding.

#### ■ ASSOCIATED CONTENT

##### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.9b07278>.

Supplementary figures and tables of structures (Figures S1–S3), previous biochemical characterization (Table S1), computational design flexibility (Table S2), and energy landscapes (Figures S4–S15) (PDF)

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*E-mail: [brd+jpc19@cs.duke.edu](mailto:brd+jpc19@cs.duke.edu). Phone: 919-660-6583.

ORCID 

Graham T. Holt: 0000-0003-3354-9748

## Author Contributions

<sup>†</sup>G.T.H. and J.D.J. contributed equally to the work.

## Notes

The authors declare the following competing financial interest(s): B.R.D. and J.D.J. are founders of Gavilán Biodesign, Inc. All other authors declare no conflicts of interest.

## ACKNOWLEDGMENTS

The authors thank Carrie Ann Davison and Dr. Mark Spaller for kCAL01 peptide synthesis; Dr. Terrence Oas, Drs. Jane and Dave Richardson, Nathan Guerin, Hong Niu, and all members of the Donald lab for helpful discussion and comments; the NIH (R01-GM078031, R01-GM118543 to B.R.D.; R01-DK101541, P20-GM113132, P30-DK117469, T32-GM008704 to D.R.M.) for funding; and the Stanford Synchrotron Radiation Lightsource (SSRL) staff, especially Irimpan I. Mathews. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences.

## REFERENCES

- (1) Diella, F.; Haslam, N.; Chica, C.; Budd, A.; Michael, S.; Brown, N. P.; Trave, G.; Gibson, T. J. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* **2008**, *13*, 6580.
- (2) Fanning, A. S.; Anderson, J. M. Protein–protein interactions: PDZ domain networks. *Curr. Biol.* **1996**, *6*, 1385–1388.
- (3) Lee, H.-J.; Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signal.* **2010**, *8*, 8.
- (4) Harris, B. Z.; Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **2001**, *114*, 3219–3231.
- (5) Kim, E.; Sheng, M. PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* **2004**, *5*, 771–781.
- (6) Piserchio, A.; Salinas, G. D.; Li, T.; Marshall, J.; Spaller, M. R.; Mierke, D. F. Targeting specific PDZ domains of PSD-95: structural basis for enhanced affinity and enzymatic stability of a cyclic peptide. *Chem. Biol.* **2004**, *11*, 469–473.
- (7) Lee, H.-J.; Wang, N. X.; Shi, D.-L.; Zheng, J. J. Sulindac inhibits canonical Wnt signaling by blocking the PDZ domain of the protein Dishevelled. *Angew. Chem., Int. Ed.* **2009**, *48*, 6448–52.
- (8) Babault, N.; Cordier, F.; Lafage, M.; Cockburn, J.; Haouz, A.; Prehaud, C.; Rey, F.; Delepierre, M.; Buc, H.; Lafon, M.; et al. Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death. *Structure* **2011**, *19*, 1518–1524.
- (9) Reiners, J.; Nagel-Wolfrum, K.; Jürgens, K.; Märker, T.; Wolfrum, U. Molecular basis of human Usher syndrome: Deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. *Exp. Eye Res.* **2006**, *83*, 97–119.
- (10) Liu, W.; Wen, W.; Wei, Z.; Yu, J.; Ye, F.; Liu, C.-H.; Hardie, R. C.; Zhang, M. The INAD scaffold is a dynamic, redox-regulated modulator of signaling in the drosophila eye. *Cell* **2011**, *145*, 1088–1101.
- (11) Thorsen, T. S.; Madsen, K. L.; Rebola, N.; Rathje, M.; Anggono, V.; Bach, A.; Moreira, I. S.; Stühr-Hansen, N.; Dyhring, T.; Peters, D.; et al. Identification of a small-molecule inhibitor of the

PICK1 PDZ domain that inhibits hippocampal LTP and LTD. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 413–8.

(12) Humbert, P.; Russell, S.; Richardson, H. Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer. *BioEssays* **2003**, *25*, 542–553.

(13) Shepherd, T. R.; Klaus, S. M.; Liu, X.; Ramaswamy, S.; DeMali, K. A.; Fuentes, E. J. The Tiam1 PDZ domain couples to Syndecan1 and promotes cell–matrix adhesion. *J. Mol. Biol.* **2010**, *398*, 730–746.

(14) Guggino, W. B.; Stanton, B. A. New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 426–436.

(15) Cheng, J.; Moyer, B. D.; Milewski, M.; Loffing, J.; Ikeda, M.; Mickle, J. E.; Cutting, G. R.; Li, M.; Stanton, B. A.; Guggino, W. B. A golgi-associated PDZ domain protein modulates cystic fibrosis transmembrane regulator plasma membrane expression. *J. Biol. Chem.* **2002**, *277*, 3520–3529.

(16) Cheng, J.; Wang, H.; Guggino, W. B. Modulation of mature cystic fibrosis transmembrane regulator protein by the PDZ domain protein CAL. *J. Biol. Chem.* **2004**, *279*, 1892–1898.

(17) Bergbower, E.; Boinot, C.; Sabirzhanova, I.; Guggino, W.; Cebotaru, L. The CFTR-associated ligand arrests the trafficking of the mutant  $\Delta$ F508 CFTR channel in the ER contributing to cystic fibrosis. *Cell. Physiol. Biochem.* **2018**, *45*, 639–655.

(18) *Worldwide Survey of the  $\Delta$ F508 Mutation—Report from the Cystic Fibrosis Genetic Analysis Consortium*; 1990; Vol. 47, pp 354–359.

(19) Population variation of common cystic fibrosis mutations. *Hum. Mutat.* **1994**, *4*, 167–177.

(20) Cheng, S. H.; Gregory, R. J.; Marshall, J.; Paul, S.; Souza, D. W.; White, G. A.; O’Riordan, C. R.; Smith, A. E. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **1990**, *63*, 827–834.

(21) Welch, W. J. Role of quality control pathways in human diseases involving protein misfolding. *Semin. Cell Dev. Biol.* **2004**, *15*, 31–38.

(22) Swiatecka-Urban, A.; Brown, A.; Moreau-Marquis, S.; Renuka, J.; Coutermarsh, B.; Barnaby, R.; Karlson, K. H.; Flotte, T. R.; Fukuda, M.; Langford, G. M.; et al. The short apical membrane half-life of rescued  $\Delta$ F508-cystic fibrosis transmembrane conductance regulator (CFTR) results from accelerated endocytosis of  $\Delta$ F508-CFTR in polarized human airway epithelial cells. *J. Biol. Chem.* **2005**, *280*, 36762–36772.

(23) Wolde, M.; Fellows, A.; Cheng, J.; Kivenson, A.; Coutermarsh, B.; Talebian, L.; Karlson, K.; Piserchio, A.; Mierke, D. F.; Stanton, B. A.; et al. Targeting CAL as a negative regulator of F508-CFTR cell-surface expression: an RNA interference and structure-based mutagenic approach. *J. Biol. Chem.* **2007**, *282*, 8099–8109.

(24) Piserchio, A.; Fellows, A.; Madden, D. R.; Mierke, D. F. Association of the cystic fibrosis transmembrane regulator with CAL: structural features and molecular dynamics. *Biochemistry* **2005**, *44*, 16158–16166.

(25) Vouilleme, L.; Cushing, P. R.; Volkmer, R.; Madden, D. R.; Boisguerin, P. Engineering peptide inhibitors to overcome PDZ binding promiscuity. *Angew. Chem., Int. Ed.* **2010**, *49*, 9912–9916.

(26) Amacher, J. F.; Cushing, P. R.; Bahl, C. D.; Beck, T.; Madden, D. R. Stereochemical determinants of C-terminal specificity in PDZ peptide-binding domains: a novel contribution of the carboxylate-binding loop. *J. Biol. Chem.* **2013**, *288*, 5114–26.

(27) Amacher, J. F.; Cushing, P. R.; Brooks, L.; Boisguerin, P.; Madden, D. R. Stereochemical preferences modulate affinity and selectivity among five pdz domains that bind CFTR: comparative structural and sequence analyses. *Structure* **2014**, *22*, 82–93.

(28) Songyang, Z.; Fanning, A. S.; Fu, C.; Xu, J.; Marfatia, S. M.; Chishti, A. H.; Crompton, A.; Chan, A. C.; Andersen, J. M.; Cantley, L. C. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **1997**, *275*, 73–77.

(29) Cushing, P. R.; Vouilleme, L.; Pellegrini, M.; Boisguerin, P.; Madden, D. R. A stabilizing influence: CAL PDZ inhibition extends

the half-life of  $\Delta F508$ -CFTR. *Angew. Chem., Int. Ed.* **2010**, *49*, 9907–9911.

(30) Cushing, P. R.; Fellows, A.; Villone, D.; Boisguérin, P.; Madden, D. R. The relative binding affinities of PDZ partners for CFTR: a biochemical basis for efficient endocytic recycling. *Biochemistry* **2008**, *47*, 10084–98.

(31) Basdevant, N.; Weinstein, H.; Ceruso, M. Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *J. Am. Chem. Soc.* **2006**, *128*, 12766–12777.

(32) Gerek, Z. N.; Keskin, O.; Ozkan, S. B. Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 796–811.

(33) Petit, C. M.; Zhang, J.; Sapienza, P. J.; Fuentes, E. J.; Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18249–54.

(34) Lee, J. H.; Park, H.; Park, S. J.; Kim, H. J.; Eom, S. H. The structural flexibility of the shank1 PDZ domain is important for its binding to different ligands. *Biochem. Biophys. Res. Commun.* **2011**, *407*, 207–212.

(35) Roberts, K. E.; Cushing, P. R.; Boisguerin, P.; Madden, D. R.; Donald, B. R. Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity. *PLoS Comput. Biol.* **2012**, *8*, No. e1002477.

(36) Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005**, *10*, 464–469.

(37) Hallen, M. A.; Martin, J. W.; Ojewole, A.; Jou, J. D.; Lowegard, A. U.; Frenkel, M. S.; Gainza, P.; Nisonoff, H. M.; Mukund, A.; Wang, S.; et al. OSPREY 3.0: open-source protein redesign for you, with powerful new features. *J. Comput. Chem.* **2018**, *39*, 2494–2507.

(38) Lilien, R. H.; Stevens, B. W.; Anderson, A. C.; Donald, B. R. Algorithm for protein redesign and its application synthetase a phenylalanine adenylation enzyme. *J. Comput. Biol.* **2005**, *12*, 740–761.

(39) Qian, Z.; Xu, X.; Amacher, J. F.; Madden, D. R.; Cormet-Boyaka, E.; Pei, D. Intracellular delivery of peptidyl ligands by reversible cyclization: discovery of a PDZ domain inhibitor that rescues CFTR activity. *Angew. Chem., Int. Ed.* **2015**, *54*, 5874–5878.

(40) Tzeng, S. R.; Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* **2012**, *488*, 236–240.

(41) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047–1069.

(42) Sciretti, D.; Bruscolini, P.; Pelizzola, A.; Pretti, M.; Jaramillo, A. Computational protein design with side-chain conformational entropy. *Proteins: Struct., Funct., Genet.* **2009**, *74*, 176–191.

(43) Georgiev, I.; Keedy, D.; Richardson, J. S.; Richardson, D. C.; Donald, B. R. Algorithm for backrub motions in protein design. *Bioinformatics* **2008**, *24*, i196–i204.

(44) Chen, C.-Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3764–3769.

(45) Silver, N. W.; King, B. M.; Nalam, M. N. L.; Cao, H.; Ali, A.; Reddy, G. S. K. K.; Rana, T. M.; Schiffer, C. A.; Tidor, B. Efficient computation of small-molecule configurational binding entropy and free energy changes by ensemble enumeration. *J. Chem. Theory Comput.* **2013**, *9*, 5098–5115.

(46) Georgiev, I.; Lilien, R. H.; Donald, B. R. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *J. Comput. Chem.* **2008**, *29*, 1527–1542.

(47) Donald, B. R. *Algorithms in Structural Molecular Biology*; MIT Press: Cambridge, MA, 2011.

(48) Karplus, M.; Ichiye, T.; Pettitt, B. M. Configurational entropy of native proteins. *Biophys. J.* **1987**, *52*, 1083–5.

(49) Dill, K.; Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*; CRC Press, 2012.

(50) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389–408.

(51) Traoré, S.; Allouche, D.; André, I.; de Givry, S.; Katsirelos, G.; Schiex, T.; Barbe, S. A new framework for computational protein design through cost function network optimization. *Bioinformatics* **2013**, *29*, 2129–36.

(52) Viricel, C.; Simoncini, D.; Schiex, T.; Barbe, S. Guaranteed Weighted Counting for Affinity Computation: Beyond Determinism and Structure. *The 22nd International Conference on Principles and Practice of Constraint Programming*; 2016.

(53) Simoncini, D.; Allouche, D.; de Givry, S.; Delmas, C.; Barbe, S.; Schiex, T. Guaranteed discrete energy optimization on large protein design problems. *J. Chem. Theory Comput.* **2015**, *11*, 5980–9.

(54) Georgiev, I.; Lilien, R. H.; Donald, B. R. Improved pruning algorithms and divide-and-conquer strategies for dead-end elimination, with application to protein design. *Bioinformatics* **2006**, *22*, e174–e183.

(55) Dahiyat, B. I.; Mayo, S. L. De novo protein design: fully automated sequence selection. *Science* **1997**, *278*, 82–7.

(56) Leach, A. R.; Lemon, A. P. Exploring the conformational space of protein side chains using dead-end elimination and the A\* algorithm. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 227–39.

(57) Chazelle, B.; Kingsford, C.; Singh, M. A semidefinite programming approach to side chain positioning with new rounding strategies. *INFORMS J. Comput.* **2004**, *16*, 380–392.

(58) Hallen, M. A.; Gainza, P.; Donald, B. R. Compact representation of continuous energy surfaces for more efficient protein design. *J. Chem. Theory Comput.* **2015**, *11*, 2292–2306.

(59) Hallen, M. A.; Donald, B. R. CATS (coordinates of atoms by taylor series): protein design with backbone flexibility in all locally feasible directions. *Bioinformatics* **2017**, *33*, i5–i12.

(60) Lee, C.; Subbiah, S. Prediction of protein side-chain conformation by packing optimization. *J. Mol. Biol.* **1991**, *217*, 373–388.

(61) Kuhlman, B.; Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10383–8.

(62) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **2011**, *487*, 545–74.

(63) Jou, J. D.; Holt, G. T.; Lowegard, A. U.; Donald, B. R. Minimization-aware recursive K\* (MARK\*): a novel, provable algorithm that accelerates ensemble-based protein design and provably approximates the energy landscape. *J. Comput. Biol.* In press.

(64) Amacher, J. F.; Cushing, P. R.; Weiner, J. A.; Madden, D. R. Crystallization and preliminary diffraction analysis of the CAL PDZ domain in complex with a selective peptide inhibitor. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2011**, *67*, 600–3.

(65) Amacher, J. F.; Zhao, R.; Spaller, M. R.; Madden, D. R. Chemically modified peptide scaffolds target the CFTR-associated ligand PDZ domain. *PLoS One* **2014**, *9*, No. e103650.

(66) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125–32.

(67) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

(68) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–21.

(69) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486–501.

(70) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.;

Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **2010**, *66*, 12–21.

(71) Gainza, P.; Roberts, K. E.; Donald, B. R. Protein design using continuous rotamers. *PLoS Comput. Biol.* **2012**, *8*, e1002335.

(72) Roberts, K. E.; Donald, B. R. Improved energy bound accuracy enhances the efficiency of continuous protein design. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 1151–1164.

(73) Ojewole, A. A.; Jou, J. D.; Fowler, V. G.; Donald, B. R. BBK\* (branch and bound over K\*): a provable and efficient ensemble-based protein design algorithm to optimize stability and binding affinity over large sequence spaces. *J. Comput. Biol.* **2018**, *25*, 726–739.

(74) Reeve, S. M.; Gainza, P.; Frey, K. M.; Georgiev, I.; Donald, B. R.; Anderson, A. C. Protein design algorithms predict viable resistance to an experimental antifolate. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 749–54.

(75) Gorczyński, M. J.; Grembecka, J.; Zhou, Y.; Kong, Y.; Roudaia, L.; Douvas, M. G.; Newman, M.; Bielnicka, I.; Baber, G.; Corpora, T.; et al. Allosteric inhibition of the protein-protein interaction between the leukemia-associated proteins runx1 and CBF $\beta$ . *Chem. Biol.* **2007**, *14*, 1186–1197.

(76) Lazaridis, T.; Karplus, M. Effective energy function for proteins in solution. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.

(77) Schrödinger, L. L. C. *The PyMOL Molecular Graphics System*, version 1.8; 2015.

(78) Word, J. M.; Lovell, S. C.; LaBean, T. H.; Taylor, H. C.; Zalis, M. E.; Presley, B. K.; Richardson, J. S.; Richardson, D. C. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **1999**, *285*, 1711–33.

(79) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448*, 325–329.

(80) Fleishman, S. J.; Khare, S. D.; Koga, N.; Baker, D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **2011**, *20*, 753–7.

(81) Bhowmick, A.; Sharma, S. C.; Honma, H.; Head-Gordon, T. The role of side chain entropy and mutual information for improving the de novo design of Kemp eliminases KE07 and KE70. *Phys. Chem. Chem. Phys.* **2016**, *18*, 19386–19396.

(82) Chou, J. J.; Case, D. A.; Bax, A. Insights into the mobility of methyl-bearing side chains in proteins from 3 J CC and 3 J CN couplings. *J. Am. Chem. Soc.* **2003**, *125*, 8959–8966.

(83) Lang, P. T.; Ng, H. L.; Fraser, J. S.; Corn, J. E.; Echols, N.; Sales, M.; Holton, J. M.; Alber, T. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* **2010**, *19*, 1420–1431.

(84) Van Den Bedem, H.; Dhanik, A.; Latombe, J. C.; Deacon, A. M. Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 1107–1117.

(85) Fraser, J. S.; Van Den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16247–16252.

(86) Keedy, D. A.; Fraser, J. S.; van den Bedem, H. Exposing hidden alternative backbone conformations in X-ray crystallography using qFit. *PLoS Comput. Biol.* **2015**, *11*, e1004507.

(87) Woldeyes, R. A.; Sivak, D. A.; Fraser, J. S. E pluribus unum, no more: From one crystal, many conformations. *Curr. Opin. Struct. Biol.* **2014**, *28*, 56–62.

(88) 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.

(89) Qi, Y.; Martin, J. W.; Barb, A. W.; Thelot, F.; Yan, A.; Donald, B. R.; Oas, T. G. Continuous interdomain orientation distributions reveal components of binding thermodynamics. *J. Mol. Biol.* **2018**, *430*, 3412–3426.

(90) Hallen, M. A.; Keedy, D. A.; Donald, B. R. Dead-End Elimination with Perturbations (“DEEPer”): A provable protein design algorithm with continuous sidechain and backbone flexibility. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 18–39.

(91) Schneider, M.; Fu, X.; Keating, A. E. X-ray vs. NMR structures as templates for computational protein design. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 97–110.

(92) Gaffney, K. J.; Piletic, I. R.; Fayer, M. D. Hydrogen bond breaking and reformation in alcohol oligomers following vibrational relaxation of a non-hydrogen-bond donating hydroxyl stretch. *J. Phys. Chem. A* **2002**, *106*, 9428–9435.

(93) Sessions, R. B.; Gibbs, N.; Dempsey, C. E. Hydrogen bonding in helical polypeptides from molecular dynamics simulations and amide hydrogen exchange analysis: alamethicin and melittin in methanol. *Biophys. J.* **1998**, *74*, 138–152.

(94) Bondar, A.-N.; White, S. H. Hydrogen bond dynamics in membrane protein function. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 942–50.

(95) Kaiser, A.; Ismailova, O.; Koskela, A.; Huber, S. E.; Ritter, M.; Cosenza, B.; Bengler, W.; Nazmutdinov, R.; Probst, M. Ethylene glycol revisited: Molecular dynamics simulations and visualization of the liquid and its hydrogen-bond network. *J. Mol. Liq.* **2014**, *189*, 20–29.

(96) Guardia, E.; Skarmoutsos, I.; Masia, M. Hydrogen bonding and related properties in liquid water: a Car–Parrinello molecular dynamics simulation study. *J. Phys. Chem. B* **2015**, *119*, 8926–8938.