Computational design of conformation-biasing mutations to alter protein functions

Peter E. Cavanagh[#1], Andrew G. Xue[#2], Shizhong Dai[3], Albert Qiang[3], Tsutomu Matsui[4], Alice Y. Ting[3,5,6*].

[1]Department of Biochemistry, Stanford University
[2]Biophysics Program, Stanford University
[3]Department of Genetics, Stanford University
[4]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Stanford University
[5]Departments of Biology, and by courtesy, Chemistry
[6]Chan Zuckerberg Biohub – San Francisco
[#]These authors contributed equally.
[*]Correspondence to: ayting@stanford.edu

Abstract: Nearly all proteins undergo conformational changes to perform their biological functions. Intentional manipulation of conformational occupancy could produce novel or altered protein functions. We developed "Conformational Biasing" (CB), a streamlined computational workflow that utilizes an inverse folding model to predict mutations that bias proteins toward desired conformational states. By applying CB to diverse deep mutational scanning datasets, we could predict mutations which, through their conformational effects, enhance K-Ras effector binding, increase SARS-CoV-2 spike protein interactions with ACE-2, and improve activation of the β2 adrenergic receptor. Additionally, CB predicted activity-increasing and -decreasing mutations for the conformationally-gated enzymes Src, B-Raf, MurA, and FabZ. We then used CB to design novel variants of *E. coli* lipoic acid ligase (LplA) biased toward "open" or "closed" adenylate ester-bound conformations, which were characterized through tryptophan fluorescence and small angle X-ray scattering measurements. Closed-biased LplA variants exhibited greater sequence-specificity than wild-type, enabling targeted fluorophore labeling of proteins in living cells, while open-based LplA variants were far more promiscuous, enabling proximity-dependent proteome tagging in cells. Due to its straightforward and rapid implementation (https://github.com/alicetinglab/ConformationalBiasing/), CB should be a versatile tool for engineering conformational bias into a wide range of proteins, resulting in altered or novel functions with utility for basic science, biotechnology, or medicine.

Introduction

Computational protein design has excelled in the generation of novel proteins with static conformations, producing potent binders(1, 2), high-order complexes(3), and even enzymes(4). However, most natural proteins transition between multiple conformational states, which are critical to their functions. For example, receptors change conformation across the membrane in response to ligand binding, G proteins toggle between a GTP-bound "on" conformation and a GDP-bound "off" conformation, and enzymes undergo structural rearrangements during catalysis to drive substrate binding, chemical bond formation or breakage, and product release. An important step in the direction of designing more complex, functional proteins for biology, biotechnology, and medicine involves addressing the challenge of modeling, predicting, and designing proteins that undergo conformational changes.

Recent studies have addressed the challenge of designing multi-conformation proteins by using AlphaFold alignment-based sub-clustering (5-7), molecular dynamics (8-11), physics-based deep learning models(12), and multi-state design (13-18). The clustering approach subsamples

multiple sequence alignments (MSAs) to predict, for example, mutations that switch the ground state of a metamorphic protein KaiB(5). However, this approach is limited to proteins with a deep MSA in which different evolutionary branches adopt alternative ground states. Molecular dynamics and other physics-based approaches for studying(19, 20) or predicting(21, 22) alternate conformations are compute-intensive, can be difficult to scale, and are unlikely to capture the longer timescales needed for many conformational changes.

Another strategy, multi-state design, uses multiple structures or a sampled conformational ensemble to design favorable sequences that can exist in all required conformations(13, 15, 23, 24). The multi-state design objective can also include negative selection against unfavorable conformations, for example to create proteins that preferentially assemble into heteromers instead of homomers (14, 18).

We envisioned an alternative and potentially simpler approach leveraging inverse folding models (IFMs) to evaluate sequences against alternative backbone conformations. Sequence-to-structure models, such as AlphaFold2, often exhibit limited sensitivity to the subtle conformational effects of single point mutations(20). In contrast, structure-to-sequence "inverse folding" models such as ProteinMPNN(25) and ESM-IF(26), are trained to recover the wild-type sequence based solely on a protein's backbone fold. Therefore, inherent to the training task is the differentiation between structurally similar amino acids at a given position in a sequence. Indeed, recent examples of protein engineering leverage the ability of IFMs to suggest structure- and complex-stabilizing point mutations, supporting the mutation-sensitivity of IFMs(25, 27-29). For this reason, we hypothesized that IFMs might be well-suited to score variants of the same sequence against alternative backbone conformations, and thereby identify mutations that preferentially stabilize for one conformation over another.

Here we describe the development of "Conformational Biasing" (CB) methodology, and validate it using several published deep mutational scanning datasets (30-35). Our analysis shows that CB can predict mutants of the small GTPase K-Ras, several enzymes (Src, B-Raf, MurA, FabZ), a GPCR (β2 adrenergic receptor), and the SARS CoV-2 spike protein that possess conformation-specific functionality, such as the ability to bind to downstream effectors or catalyze kinase-dependent cell growth. We then applied CB to the design of *E. coli* lipoic acid ligase (LplA), an enzyme that catalyzes the covalent ligation of small molecule probes onto specific acceptor proteins and peptides via an activated adenylate ester (probe-AMP) intermediate(36). By using CB to predict LplA variants biased towards "open" versus "closed" probe-AMP bound conformations, we discovered a novel mechanism for conformational-gating of LplA's specificity, and produced variants with vastly altered protein sequence specificity, useful for either site-specific or promiscuous bio-orthogonal protein labeling in living cells. Our results show that CB may be a generally useful tool for designing protein variants with biased conformational states.

Results

In the CB workflow shown in **Figure 1**, all possible point mutants of a protein of interest are paired with two or more alternative backbone conformations. Each pairing is scored, relative to wild-type, using an inverse folding model (IFM) such as ProteinMPNN(25). The resulting plot of inverse folding scores can be filtered for point mutants with predicted conformational bias towards one backbone structure versus another (**Figure 1C**). To test the validity of CB's predictions, we used a recently-reported deep mutational scanning (DMS) dataset on the small GTPase K-Ras(33), which alternates between two distinct conformational states, "State 1" and "State 2," which exhibit different binding affinities to downstream effectors (**Figure 2A**). Oncogenic mutations in K-Ras can alter the State 1: State 2 equilibrium, leading to uncontrolled cell proliferation(37). Weng et al.(33) experimentally evaluated >30,000 single and double mutants of K-Ras that show changes in binding to a panel of

State 2 and State 1-specific effectors. We used the CB workflow in **Figure 1** to predict State 1 or State 2-biased K-Ras mutations (**Figure 2B**). We then colored the mutants based on their experimentally-determined binding scores to RALGDS, a State 2-specific effector, or DARPin K27, a State 1-specific binder(33, 38). **Figure 2B** shows that State 2-biased mutants predicted by CB (lower right) exhibit higher RALGDS binding activity, while State1-biased mutants predicted by CB (upper left) exhibit higher DARPin K27 binding. We performed a similar analysis of binding data for five other State 2-specific effectors reported in Weng et al. and observed, in each case, a significant ($p < 0.0001$) increase in binding scores for CB-predicted top 2.5% State 2-biased versus top 2.5% State 1-biased K-Ras variants (**Figures 2C**).

To test the generality of CB, we then applied it to six other experimental datasets, encompassing protein binders, receptors, and enzymes. For SARS-CoV2 spike protein, which binds to the ACE2 receptor on human cells in the receptor binding domain (RBD) "up" conformation but not "down" conformation (**Figure 2D**), we crossed our CB predictions with DMS data from Starr et al., which measured ACE2 binding affinities via yeast surface display(32). **Figure 2E** shows that our CB-predicted up-biased mutants have significantly higher ACE2 binding than a set of "neutral" variants (predicted by CB to be neither up-biased nor down-biased). On the other hand, CB-predicted down-biased variants have reduced ACE2 binding compared to the neutral variant population.

We then applied CB to the β2 adrenergic receptor (β2AR), a G-protein coupled receptor that signals to G proteins in its active, agonist-bound conformation but not in its inactive, apo conformation (**Figure 3A)**. A study by Jones et al. used DMS to identify mutations in β2AR that increase receptor activity upon agonist treatment, using a transcriptional reporter as readout (30). After performing CB using experimental structures of the inactive(39) and active(40) receptor, we found that the top CB-predicted active conformation-biased mutants of β2AR have significantly greater receptor activity than inactive conformation-biased mutants or a set of neutral mutants ($p < 0.0001$) (**Figure 3B**). **Figure 3A** also highlights how CB-predicted active-biased mutations are distributed across the β2AR structure, not just in the sixth transmembrane helix that undergoes the largest conformational change upon receptor activation.

As a further test of CB's ability to discriminate between binding-competent conformations, we examined a set of 10 highly similar synthetic protein binder pairs, whose affinities, cross-affinities, and crystal structures have been determined (**Figure S1**)(41). Following the same CB workflow, we scored the sequences for each binder pair across all 10 binder-pair structures. Though the binder pairs differ only slightly in sequence (**Figure S1D**) and structure (**Figure S1E**), we observed that most scored better on their experimentally-validated binding conformations than on the conformations of other, similar sequences ($p < 0.01$) (**Figures S1C, F**). This difference was greater when structural/sequence differences were larger (e.g. L1.6/L1.1 vs. L2.17). Collectively, our results suggest that CB can predict conformation-biasing mutations on diverse protein scaffolds, leading to functional shifts in binding activities.

CB tunes the enzymatic activities of Src, B-Raf, FabZ, and MurA

We next turned our attention to enzyme datasets. The human kinases Src and B-Raf (34, 35), as well as the *E. coli* biosynthetic enzymes FabZ and MurA(31), all interconvert between active and inactive conformations. The inactive state of Src is auto-inhibited, while auto-phosphorylation of Tyr527 turns on the enzyme through displacement of its auto-inhibitory domain (**Figure 3C**)(42, 43). Using CB and two experimentally-solved structures of Src(44, 45), we predicted active-biased and inactive-biased mutations. **Figure 3D** overlays experimentally-measured Src activities (via cell growth from Nguyen et al.(35)). CB-predicted active-biased mutants at the bottom right are enriched

for higher activity variants, while inactive-biased mutants at the top left have reduced activity. This effect is quantified in **Figure 3E**.

B-Raf kinase, downstream of Ras, interconverts between a monomeric, autoinhibited conformation, and a dimeric, uninhibited form (**Figure 3F**) (46, 47). Using CB, we predicted dimer-biased and monomer-biased mutations. Although the DMS data from Simon et al.(34) focuses primarily on mutations in the kinase domain of B-Raf, we found that predicted dimer-biased mutants showed a significant increase in kinase activity, as measured by phospho-Erk immunoprecipitation (**Figure 3G**).

FabZ and MurA also interconvert between active and inactive conformations (**Figures 3H, I, L**)(48, 49). Because a complete set of experimentally-determined structures is not available, we used two AlphaFold2-templated structures of FabZ (see Methods)(49, 50), a crystal structure for the closed (active) form of MurA (PDB: 1UAE (48)), and an AlphaFold2-predicted structure for the open (inactive) form of MurA (51). For both enzymes, we predicted a set of active-biased and inactive-biased variants. Based on experimental growth assay data, the CB-predicted active variants were significantly more active, while CB-predicted inactive variants were less active than a set of neutral variants (**Figures 3J, K**).

Benchmarking CB

Though CB was able to predict conformation-specific functionality for a diverse set of proteins, we wished to benchmark it against previous methodologies and examine the nature of CB-predicted variants more closely. First, we compared CB to IFM-based scoring on a single backbone conformation, because this approach has previously been used to dramatically increase antibody affinity(28) through stabilization of the antigen-bound antibody state. Using the SARS-CoV2 spike protein DMS dataset (32), we scored all variants against "up" or "down" structures alone, and took the highest scoring variants in each group. When we examined their experimentally-determined ACE2 binding scores, we found that, interestingly, both up- and down-biased variants both showed increases in ACE2 binding (**Figure 2H**, left). However, if we normalized the binding scores by variant expression level (which was also experimentally determined (32)), all differences between groups vanished (**Figure 2H**, right). Thus, single structure scores may correlate more with expression level or stability than with conformational preference. By contrast, the top up-biased and down-biased variants predicted by CB showed significant differences in ACE2 binding in the expected directions, even when binding scores were normalized by variant expression level (**Figure 2G**, right).

We performed a similar comparison for K-Ras, B-Raf, and β2AR datasets, and found that for all of them, single structure scoring using ProteinMPNN was either unable to or less able to predict differences in measured activities, while CB-based contrastive scoring generated variants with robust activity differences (**Figure S2**).

Next, we benchmarked CB using BioEmu (52), a state-of-the-art model for sampling equilibrium conformational ensembles of proteins. We selected a subset of CB-predicted "up" and "down"-biased SARS-CoV2-2 RBD mutations, with high and low ACE2 binding scores, respectively, avoiding the binding interface with ACE2 (**Figure 2I**). Using BioEmu, we generated an ensemble (n=4000) of conformational states for each variant, and classified them by structural topology as "up"-like or "down"-like. We found that 4 out of 6 CB-predicted "up" variants showed increased "up" conformational occupancy by BioEmu, while 4 out of 5 "down" predicted variants showed increased "down" conformational occupancy (**Figure 2J**). The general agreement between CB and BioEmu supports the idea that CB may be enriching variants with altered conformational occupancy, as opposed to altering protein behavior through other mechanisms.

We were curious how CB would perform if we used other inverse folding models besides ProteinMPNN to score variants, as our workflow is in principle agnostic as to the specific IFM used.

Using the β2AR dataset, we applied two masked language models - ESM-IF1(26) and Frame2Seq(53) - as well as ThermoMPNN(54), a version of ProteinMPNN fine-tuned on thermal stability data. Since ThermoMPNN is specifically designed for predicting ΔΔG, we reasoned that it may be especially sensitive to the state-specific stabilizing effects of specific variants. In **Figure S3A**, we show that ProteinMPNN and ThermoMPNN perform the best on this dataset, while Frame2Seq predicted the most unique (not predicted by other models) variants (**Figure S3C**). Interestingly, the 18 active-biased variants predicted by all four models had even higher receptor activation scores than the original CB prediction group (**Figure S3D**).

Finally, we wondered if the effect demonstrated by CB is largely dependent on destabilizing one conformation, or if merely stabilizing one conformation more than another is sufficient. To test this, we compared all of the K-Ras single point mutants (excluding double mutants) to just the subset of quadrant 1 (Q1) variants that have above-average score on both backbone conformations (**Figure S4**). While the significance of the effect is reduced for Q1 variants only, an effect is still observed for 4/5 effector datasets where an effect was previously seen (all single mutants). When we performed the same analysis on K-Ras double mutants, the Q1 subset showed an effect across all effector datasets, only with reduced significance compared to the full double mutant set (**Figures S4C-D**). Utilizing the experimentally-determined variant abundance data included in the same study(33), we observed that variants that destabilize one conformation and stabilize the other (Pos/Neg in **Figure S4E**) have on average near-WT expression, and thus do not necessarily compromise the overall stability of the protein, which would reduce the utility of CB for engineering purposes.

We also analyzed the Q1 subset of SARS-CoV2 S1, β2AR, and Src variants (**Figure S4F-I**). This subset has generally decreased CB bias scores, and we observed a corresponding decrease in effect strength and significance (possibly mediated by decrease in number of variants), while still maintaining some evidence of enriching for conformationally-biased mutations (**Figure S4F-I**). We conclude that for point mutants, the greatest effect is demonstrated by variants that destabilize one conformation while stabilizing the other; however, when introducing multiple mutations, a strong effect may be seen even when selecting for only stabilizing mutations.

CB predicts conformationally-biased variants of lipoic acid ligase

The *E. coli* enzyme lipoic acid ligase (LplA) plays an essential role in energy metabolism(55) and has been harnessed for biotechnological applications, including site-specific protein labeling with fluorophores in living cells(56, 57). LplA undergoes a large conformational change over the course of its catalytic cycle, which has been characterized by several experimentally-solved crystal structures (**Figure 4A**). During the first half-reaction, formation of lipoyl-AMP from lipoic acid and ATP, the adenylate binding loop closes over the active site, leading to a major rotation of the 10 kDa C-terminal domain from a downward "closed" conformation to an extended "open" conformation (58). This rotation exposes a docking site for acceptor proteins, enabling LplA's second half reaction: lipoyl transfer onto the lysine sidechain of one of LplA's three acceptor proteins (56, 59).

Given the essential role of this conformational change in LplA's mechanism, we used CB to predict conformationally-biased variants that we hypothesized might have altered enzymatic properties (**Figure 4** and **S5**). **Figure 4B** shows CB scoring of all LplA point mutants, with closed-biased, open-biased, and neutral variants highlighted. We selected a representative set of biased mutants, avoiding mutations lining the substrate-binding pocket, and cloned them into the template W37V LplA. The W37 sidechain is a previously-identified gatekeeper in the lipoic acid binding pocket that, when mutated to a smaller valine or alanine, allows LplA to ligate a range of unnatural probes including alkyl azides and alkynes, trans-cyclooctene, and coumarin, instead of lipoic acid(56, 59). To examine the conformational bias of our LplA variants, we first recorded the intrinsic tryptophan fluorescence of the purified proteins before and after addition of a non-hydrolyzable lipoyl-AMP ester

analog, lipoyl-AMS (**Figure 4C** and Methods). The template W37V showed a drop in Trp fluorescence upon lipoyl-AMS addition, consistent with a shift towards the open conformation in which more Trp sidechains become solvent-exposed (**Figure S5F**). Interestingly, most of the CB-predicted open-biased variants showed a greater drop than W37V in Trp fluorescence, suggesting greater occupancy of the open conformation upon lipoyl-AMS addition (**Figure 4D**). By contrast, 8 out of 11 CB-predicted closed-biased variants showed a smaller drop in Trp fluorescence than W37V. These results provide experimental support for CB's ability to predict changes in conformational occupancy of adenylate ester-bound LplA.

To measure LplA conformation more directly, we used small angle X-ray scattering (SAXS) to study W37V-LplA, two CB-predicted closed-biased variants ($C55M^V$ and $T57I^V$; the V superscript indicates that the mutation is on the W37V background) and two predicted open-biased variants ($E265R^V$ and $A48N^V$) (**Figure 4E**). We found that all five enzymes were predominantly in the closed conformation before substrate addition, as expected (only 1.9-5.6% open). After incubation with lipoyl-AMS, a 22% of W37V became open, while the open-biased mutants $265R^V$ and $A48N^V$ shifted much more, to 89% and 99% in the open conformation, respectively (**Figures 4E-F** and **S6**). Interestingly, our closed-biased variants exhibited the opposite trend, opening *less* than W37V upon lipoyl-AMS addition, to just 11% and 18% open. The SAXS measurements showed excellent correlation with our Trp fluorescence measurements on the same LplA variants (**Figure 4G**). Together, our experimental characterizations of open and closed-biased LplA designs suggest that CB bias score is a strong predictor of protein conformational occupancy.

Conformation-biasing mutations alter LplA's sequence specificity in cells

Next we explored the functional consequences of biasing LplA's conformation. Because the open conformation is required for acceptor protein engagement and transfer of lipoate, and wild-type LplA is predominantly closed in its lipoyl-AMP bound form, we hypothesized that the closed-to-open conformational transition may function as a selectivity filter to ensure that only correct, cognate acceptor proteins receive lipoate from LplA. In other words, open-biased mutants of LplA may exhibit more promiscuous labeling than WT.

To test this hypothesis, we developed a flow cytometry-based assay in mammalian cells to rapidly evaluate the promiscuous activity of LplA variants (**Figure 5A**). We expressed each mCherry-tagged LplA variant in the cytosol of HEK 293T cells, and treated with a bicyclo-nonyne substrate (BCN, **Figure 5A**), a Click handle known to be used by W37V mutants of LplA (60). After 5 min of labeling, BCN tagged proteins were detected in live cells by Click with fluorogenic tetrazine-BODIPY. Under these conditions, in which no LplA acceptor protein or peptide(33) are co-expressed, any appearance of BODIPY fluorescence represents promiscuous rather than site-specific labeling (**Figure S7**). **Figure S8** shows the correspondence between BODIPY signal, measured by imaging or flow cytometry, and promiscuous labeling extent, as quantified by SDS-PAGE and in-gel fluorescence detection of many endogenous BODIPY-tagged proteins (see **Supporting Text** for details).

Flow cytometry data plotting expression-normalized promiscuous activity (BODIPY/mCherry) for single cells show that the W37V-LplA template has a moderate degree of promiscuous activity at high expression levels in HEK 293T cells (**Figures 5C** and **S7A-C**). For the 32 closed-biased mutants that we evaluated, all but one had lower promiscuous activity than W37V, despite clearly detectable expression (**Figures 5B-C** and **S9**). By contrast, 21 of the 31 open-biased mutants that we tested had promiscuous activity greater or much greater than the W37V template. We also tested 36 neutral mutants, which score similarly high for both open and closed structures

of LplA; the promiscuity profiles of these mutants fall in between those of the open and closed-biased mutants (**Figures 5B-D**).

These measurements could reflect a change in LplA specificity, from more specific to more promiscuous for open-biased variants – or they could represent changes in the total activity of LplA variants, against both specific (cognate) and non-specific (non-cognate) substrates. To distinguish between these possibilities, we selected 6 open-biased variants and 6 closed-biased variants, and tested them in HEK 293T cells with the LplA acceptor protein E2p co-expressed(61). We performed the BCN and tetrazine-BODIPY labeling as in **Figure 5A** and then lysed the cells, ran SDS-PAGE, and visualized labeled proteins by in-gel fluorescence. When examining the E2p-YFP band at 39 kDa in **Figure S7F**, we observed robust labeling for all 12 LplA mutants, whether open-biased or closed-biased. After quantification (**Figures 5G** and **S7F-H**), the closed-biased variants exhibit a higher ratio of specific to promiscuous labeling activity than W37V template, while open-biased mutants exhibit the opposite trend.

A mechanistic model is presented in **Figure 5H**. The lipoyl-AMP bound form of wild-type LplA equilibrates between closed and open conformations, favoring the closed state. Since the C-terminal domain (CTD), in the closed state, blocks substrate protein access, it can be modeled as a tethered competitive inhibitor of the probe transfer step. As such, its effect is to tune (upward) the observed $K_{M\text{-}Obs}$ of LplA for both cognate and non-cognate substrates. In the crowded cell interior, labeling of cognate substrate is likely to be K_{cat}-limited ([cognate substrate] > $K_{M(cog)}$), while labeling of non-cognate substrate is K_M-limited. Thus, the shift in observed $K_{M\text{-}Obs}$ will primarily slow *non-cognate,* K_M-limited protein labeling, resulting in greater apparent sequence-specificity. By contrast, open-biased LplA variants, with higher K_{conf}, lack a competitive inhibition mechanism and the active site remains accessible to both cognate and non-cognate protein substrates alike. The net result is higher promiscuous labeling for open-biased LplA variants, while closed-biased variants exhibit greater protein substrate specificity than even wild-type.

Harnessing conformationally-biased LplA variants for protein labeling technologies in living cells

W37V LplA has previously been harnessed for site-specific protein labeling with fluorophores in living cells(56) and in vitro(62). The combination of LplA's malleable small-molecule binding pocket and its high natural sequence specificity has provided a powerful platform for protein labeling technologies that work in complex biological settings. Because CB was able to design LplA variants with both greater sequence-specificity and greater sequence-promiscuity than W37V LplA, we wondered if these could be harnessed for new and improved protein labeling technologies.

For site-specific protein labeling applications in cells, greater sequence-specificity is highly advantageous, for minimizing background and improving the detection of target proteins of interest. We co-expressed W37V-LplA with a target protein fused to a 13-amino acid LplA acceptor peptide (LAP) in HEK 293T cells (**Figure 5I**). After 10 minute labeling with BCN and 15 minute Click with methyltetrazine-BODIPY, we performed confocal imaging. **Figure 5J** shows a significant degree of non-specific labeling in the cytosol, not matching the localization of the LAP-BFP fusion at the plasma membrane. When we repeated the experiment with a closed-biased LplA variant, that showed greater sequence-specificity in our E2p assay in **Figure S7F-G**, we observed substantially improved labeling specificity, indicated by strong overlap between BODIPY and LAP-BFP channels. This improvement in S/N was also apparent when we took the labeled cells, lysed, and analyzed the labeled proteins by SDS-PAGE and in-gel fluorescence (**Figure 5K**).

Another important use for engineered ligase enzymes is promiscuous proximity-dependent tagging (**Figure 5I**). The promiscuous labeling enzymes APEX, TurboID, and BioID have been broadly deployed in cell biology and organismal biology to map cell-type specific and organelle-specific

proteomes, as well as protein and RNA interactomes(63). Yet improved methods are needed that bypass the toxicity of APEX, the biotin-dependent background of TurboID, and the slow speed of BioID. CB produced LplA variants with vastly increased promiscuous activities (**Figure 5L**). In separate work (unpublished), we show that these variants provide the foundation for a new proximity labeling technology called "FlexID," which can be deployed for high-resolution interactome and organelle mapping in cells and *in vivo*.

Benchmarking CB's predictions on LplA

One measure of CB's accuracy is its ability to predict experimental data. We plotted CB-predicted bias scores for LplA against three experimental datasets: Trp fluorescence change for 24 variants (**Figure S10A**), SAXS measurement of conformational occupancy for 6 variants (**Figure S10B**), and mean promiscuous labeling activities for 99 variants (**Figure 5E**). All three measurements show good correlation with CB bias score, which suggests that CB is predicting conformational change rather than altering LplA function by other mechanisms. In **Figure 5E**, 9 open-biased mutants have predicted steric clashes with BCN in the substrate binding pocket, providing a likely explanation for their low promiscuous activities (independent of their conformational effect).

Using the LplA dataset, we tested the use of other inverse folding models in the CB workflow. We found that CB bias scores generated by ProteinMPNN, ESM-IF1(26), Frame2Seq(53), and ThermoMPNN(54) all showed good correlation with LplA experimental data (spearman correlations between 0.63-0.7, excluding substrate clash variants) (**Figure S10**). Thus, similar to our observations above with B2AR (**Figure S3**), all models seem sensitive to potential biasing mutations, especially the variants with highest and lowest promiscuous activities.

We also benchmarked CB against the use of inverse-folding models for single structure scoring, analogous to previous studies(28, 29). 27 out of 32 top-scoring variants on the closed LplA backbone alone, and 24 out of 31 top-scoring variants on the open LplA backbone alone, were present in our promiscuous labeling dataset. **Figure S2C** shows that these groups do not exhibit a significant difference in promiscuous labeling activities. By contrast, CB's contrastive scoring method produces robust and bidirectional differences in promiscuous labeling activities among the top open-biased and closed-biased groups (**Figure 5D**).

Next we compared CB predictions to BioEmu(52), which naturally lends itself to prediction of conformationally-biasing mutations. We generated structural samples using BioEmu (n=500) for each of the 24 LplA variants that we analyzed by Trp fluorescence in **Figure 4D**, and analyzed the conformational distribution generated by BioEmu (**Figures S11A, S11B**) (see Methods). Overall, we observe some agreement between BioEmu and our experimental results, with open-biased variants generally having a greater proportion of samples in the open state. Combining open-biasing mutations with closed-biasing mutations for sampling (n=100) resulted in double mutants with predictions between the original two mutants, suggesting that introduced mutations are acting on the conformational equilibrium rather than altering other properties (**Figure S11C**). A similar proportion of samples are assigned to open/closed states for each variant, suggesting that variants are not acting through creation of new conformational states or destabilization of native states. While BioEmu is able to resolve some differences in open-biased and closed-biased variants, in comparison to CB scoring, however, the correlation of BioEmu samples to our Trp fluorescence data is much weaker (**Figure S11D**). BioEmu is also orders of magnitude more computationally intensive, and requires a different evaluation method per system for classifying samples.

Currently, no gold standard method for predicting conformation-specific mutations has been established. As a baseline, we attempted to use AFCluster(5) to predict conformationally-biased variants of LplA. To do so, we used AFCluster to generate clusters of similar sequences from LplA's multiple sequence alignment (MSA) and scored each cluster's AlphaFold2-predicted

structure for similarity to LplA's open or closed conformations (**Figure S12A**). Scoring the enrichment of individual mutations in open- or closed-like clusters, we determined conformational bias scores (see Methods) and found that they correlated much more poorly with our experimental data (**Figure S12C,** r = 0.26 compared to r = 0.67 for CB method in **Figure 5E**). Furthermore, 1250 LplA point mutants (20%) could not be scored by AFCluster at all because they are absent from LplA's MSA. AFCluster requires evolutionary branches in which the protein of interest adopts alternate ground-state conformations, a feature not present in the majority of conformation-changing proteins (including LplA). Overall, CB may have a stronger ability to predict conformationally-biased variants than both AFCluster and BioEmu.

Discussion

Conformational Biasing (CB) offers a broadly applicable method for predicting mutations that bias protein conformational state, enabling both mechanistic insight and the engineering of new capabilities, as we demonstrated with the enzyme lipoic acid ligase (LplA). Structural modeling suggests that the CB-predicted mutations in LplA work in diverse ways, by introducing steric collisions that disfavor one conformation, or beneficial hydrophobic packing that favors one conformation over the other (**Figure S13**). However, the mechanistic basis of some CB-predicted mutations is unclear, highlighting the ability of the method to identify non-obvious solutions to conformational biasing.

Protein design methods that consider more than one target structure for a given protein sequence are just starting to come into focus(3, 14, 16, 17, 24). However, most multi-state design objectives target high occupancy of both (or more) backbone structures, whereas our design approach identifies mutations that bias multi-state proteins towards one conformation over another, producing sometimes dramatic effects on protein function. We showed (**Figure 5C-D**) that mutations in LplA that scored highly in both conformations were not as active for promiscuous labeling as open-biased mutations. This suggests that, at least in some cases, designing conformational occupancy rather than absolute fitness may be more appropriate to achieve specific design goals.

Inverse folding models have mostly been used in the context of de novo design(64), for generating sequences that fold into a desired backbone structure. However, their role in design and evaluation of targeted mutations may be equally important. For example, inverse folding has been applied to antibody-antigen complexes for unsupervised prediction of higher affinity mutants(28). Our CB method could potentially enable extensions on that task, for example by designing antibodies that are selectively stabilized by complexation with antigen versus un-complexed.

Other machine learning methods for zero-shot or low-shot engineering of proteins have often suffered from black-box effects(65). For example, protein language models can vastly improve the hit rate of proposed designs(27, 28, 66), but mutations predicted on the basis of "evolutionary fitness" are agnostic as to whether they improve stability, activity, trafficking, or some other characteristic of the protein in question. Here, a mechanistic hypothesis for LplA, tying backbone structure to enzyme specificity, enabled us to use machine learning tools in a more directed and hypothesis-driven manner.

Another significant consideration is method accessibility. The approach outlined here is biologically intuitive, and can be run on short time scales on commonly available workstation resources (<1 min to evaluate ~3000 point mutations using a RTX 4090 GPU) or using free compute like Google Colaboratory. Our workflow requires far fewer resources than methods based on AlphaFold or molecular dynamics simulations, as well as a lower expertise barrier to entry. While BioEmu requires fewer resources than MD-type experiments, it is significantly more compute-intensive than CB, and does not outperform CB in predicting LplA experimental data.

While CB currently has multiple limitations, such as the requirement for high-quality experimental structures or predicted structures in multiple states, and limited predictive power throughout all point mutants, we anticipate that these will improve as both inverse folding models and structure prediction methods improve. For most of the mutational datasets that we analyzed, we did not have a direct measurement of conformational occupancy, which is an added layer of noise in our CB workflow. Despite this, we still found that CB consistently produced signal across diverse protein types (binders, receptors, enzymes), even when structural changes are small. Extensions of our CB method, whether with a larger inverse folding model, more complex training data, inclusion of side-chain and ligand modelling, or integration of protein language models, could further improve performance and unlock new applications.

Acknowledgements: We are grateful for funding from the National Institutes of Health (RC2DK129964), the Chan Zuckerberg Biohub – San Francisco, the GPCR collaborative of the St. Jude Children's Research Hospital, the Phil and Penny Knight Initiative for Brain Resilience, the Wu Tsai Neurosciences Institute, and Stanford Bio-X. P.E.C. is supported by an NSF Graduate Research Fellowship. Dan Herschlag, Brian Trippe, Brian Hie, Chris Garcia, Matthew Ravalin, Jessica Simon, Zhaoyang Li, and Nicholas Kalogriopoulos provided valuable feedback. We thank Reika Tei for assistance with LplA microscopy, Pratik Kumar and Luke Lavis for sharing JF-646 tetrazine fluorophore.

Conflict of interest statement: The authors declare no conflict of interest.

Author contributions: P.E.C., A.G.X., and A.Y.T. conceived of the project and developed the CB methodology. A.G.X. wrote the code for CB implementation. A.G.X. and P.E.C analyzed deep mutational scan datasets, with input from A.Y.T. and S.D.. A.G.X. performed benchmarking analyses. P.E.C., A.G.X., and A.Y.T. designed the LplA library. P.E.C. and A.Q. designed and executed the cell-based characterization of LplA variants. S.D. developed the tryptophan fluorescence assay, LplA purification, and synthesized lipoyl-AMS. P.E.C. purified LplA variants and performed the tryptophan fluorescence assay. P.E.C., T.M., and A.G.X. performed SEC-SAXS analysis. P.E.C., A.G.X., and A.Y.T. wrote the manuscript with input from all authors.

Supplementary Materials
Figs. S1 to S13
Supplementary Table 1 and Table 2 (as .xlsx file)
Supplementary Text
Materials and Methods

Figures

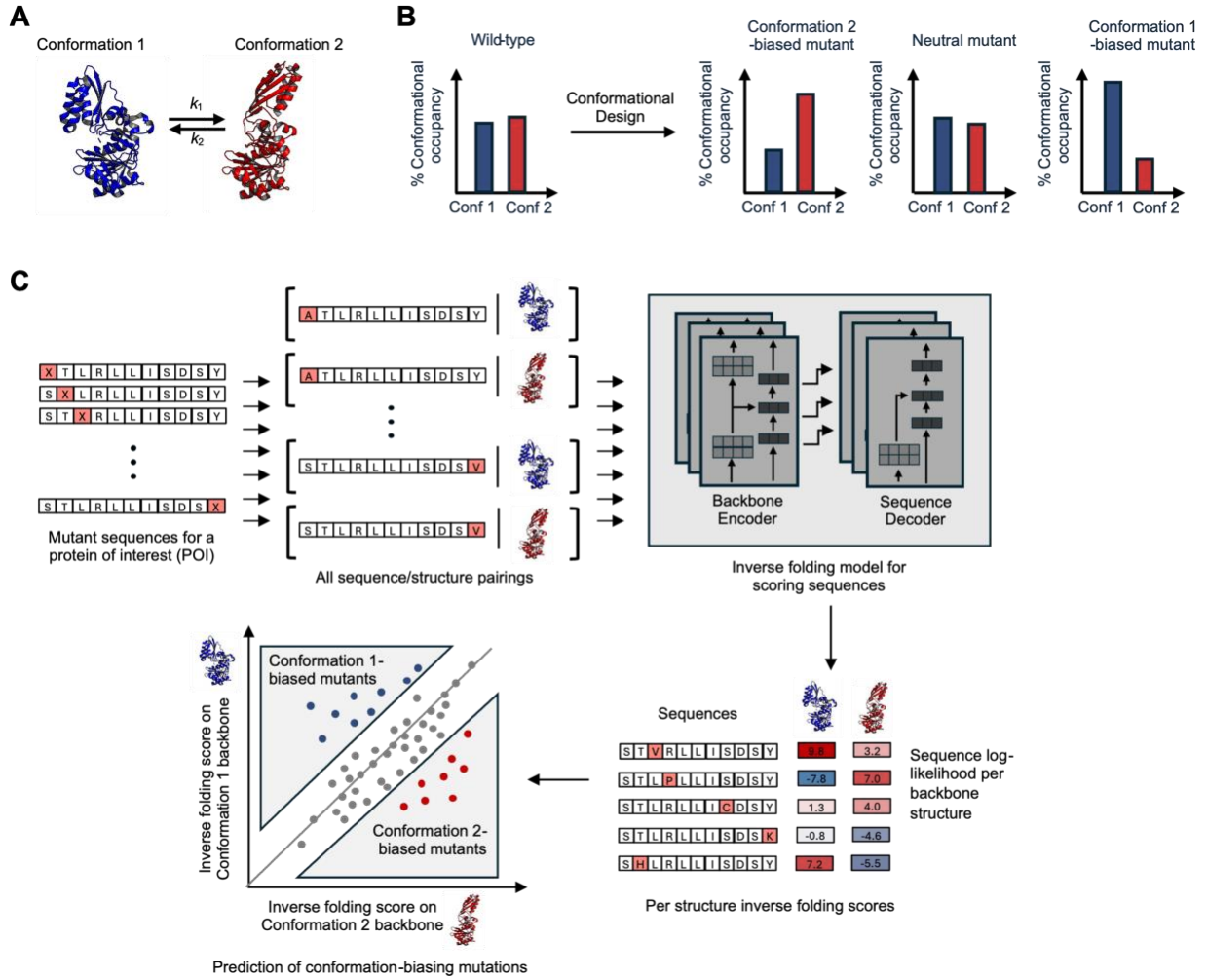


Figure 1. Conformational Biasing (CB) approach for predicting mutations that shift protein conformational occupancy. (**A**) Many proteins switch between alternative conformations. Specific conformations may exhibit unique binding, catalytic, and functional properties. (**B**) CB predicts mutations that shift conformational occupancy. (**C**) Computational workflow for CB. Mutant sequences for a protein of interest are paired with the protein's structure in two or more conformational states. The structures can be experimentally-solved or computationally-predicted structures. The sequence-structure pairings are scored using an inverse folding model such as ProteinMPNN and plotted as shown. Conformation-biasing mutations maximize the difference in conformation-specific inverse folding scores.

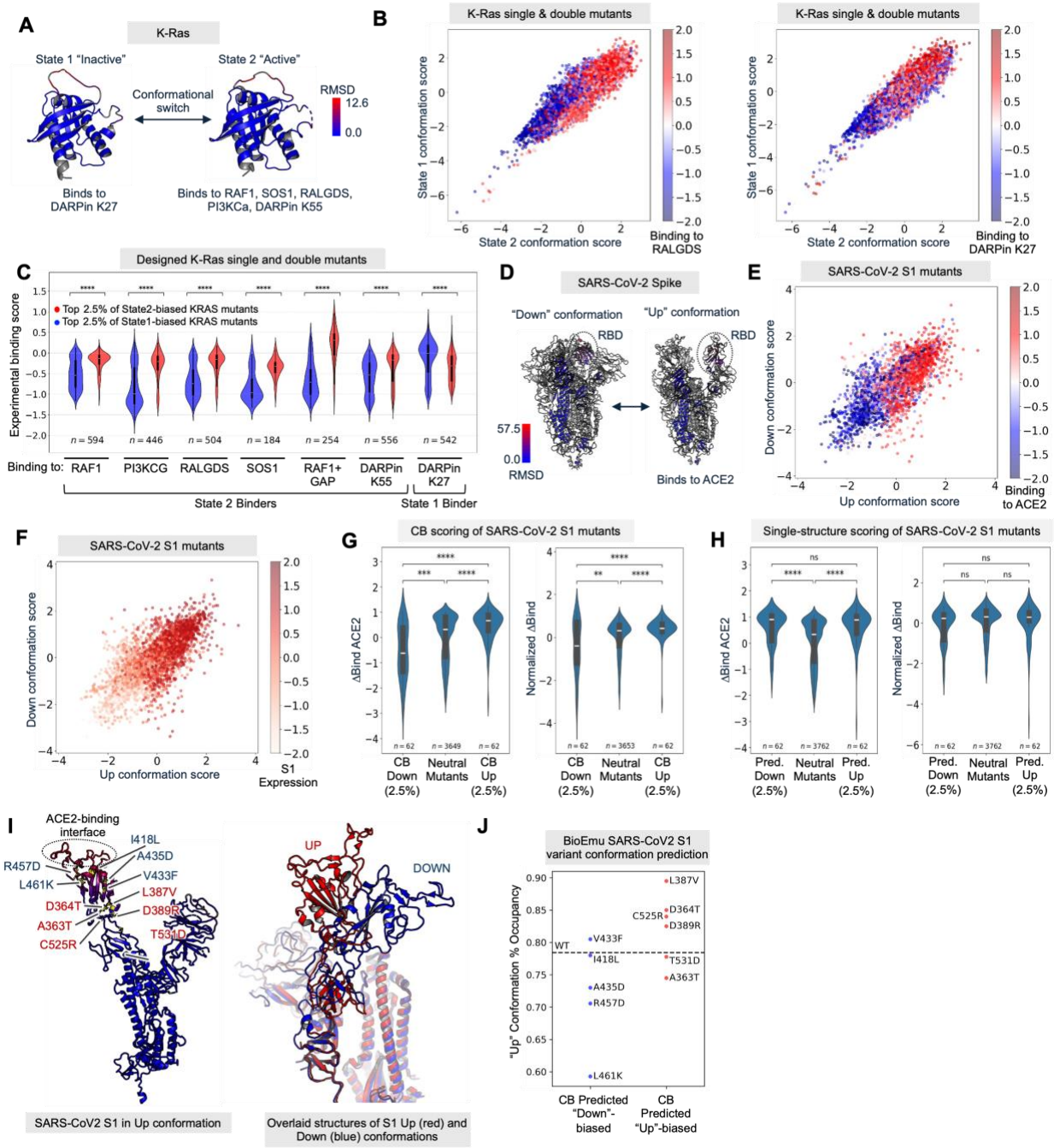


Figure 2. CB analysis of K-Ras and SARS-CoV2 spike protein predicts binding-competent and -incompetent variants. (**A**) The G protein K-Ras switches between an inactive State 1 conformation (PBD: 8T71, bound to GDP) and an active State 2 conformation (PDB: 6XHB, bound to GTP-analogue). State 2 binds to the effectors listed. RMSD, root mean square deviation. (**B**) 3320 Single mutants and 26761 designed K-Ras double mutants were scored using CB against State 1 and State 2 experimental structures in (A). Points are colored according to experimental binding data from Weng et al.(33), for RALGDS, a State 2-specific effector (left) or DARPin K27, a State 1-specific binder (right). (**C**) Violin plot showing distributions of experimentally-determined binding scores from Weng et al.(33) for top CB-predicted State 1-biased mutants (blue) and State 2-biased mutants (red). Mutations at the KRAS-effector binding interface that are likely to directly affect effector binding are

excluded from analysis (**Figure S2B** and Methods). ****p < 0.0001. (**D**) SARS-CoV-2 spike protein with ACE2 receptor binding domain (RBD) in "down" (PDB: 7XIX(67)) versus "up" (PDB: 7XO8(68)) conformations. (**E**) SARS-CoV-2 spike protein S1 mutants were scored against RBD up/down structures in (D). Points are colored by experimentally-determined ACE2 binding scores from Starr et al.(32). (**F**) Same scores in (E) are colored by S1 expression level (32). (**G**) Violin plot showing distributions of experimentally-determined ACE2 binding scores (left) and expression-normalized ACE2 binding scores (right), for CB-predicted up-biased and down-biased S1 mutants. **p<0.01, ***p < 0.001, ****p<0.0001. (**H**) Same as (G) except groups are based on single conformation scoring using ProteinMPNN instead of contrastive scoring with CB. ****p<0.0001. (**I**) CB-predicted down and up-biased S1 RBD mutants selected for further validation using BioEmu(52). Red-colored mutants show increased ACE2 binding in Starr et al.(32), and blue mutants show decreased ACE2 binding. Right: Overlaid structures of S1 in up (red) and down (blue) conformations. (**J**) BioEmu was used to generate a conformational ensemble for each variant from (I). Sampled conformations were assigned to up/down conformations, and percent occupancy of "up" conformation is shown for each variant. The horizontal dashed line shows wild-type occupancy predicted by BioEmu.

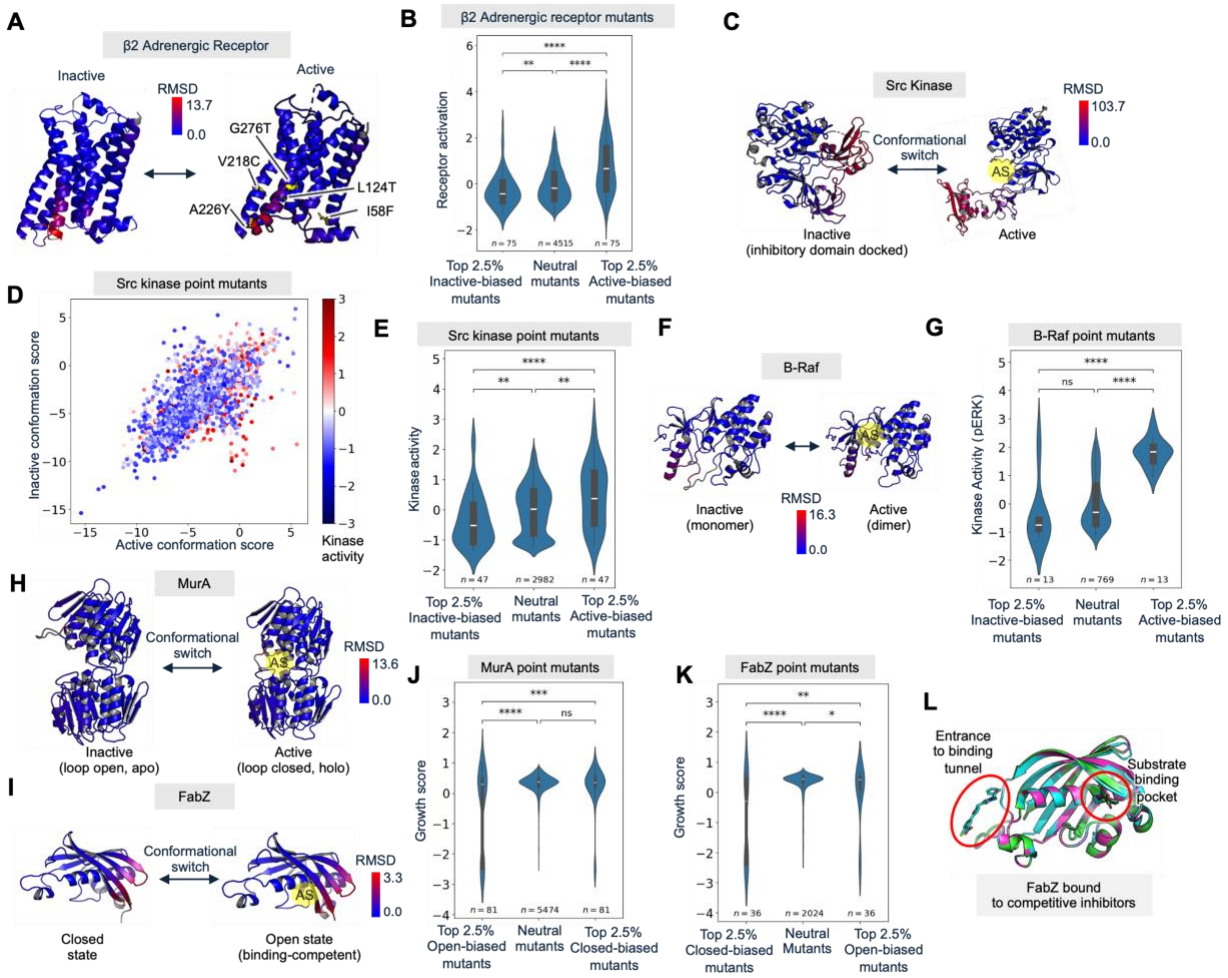


Figure 3. CB predicts conformationally-biased variants of a GPCR and four enzymes. (**A**) Human β2 adrenergic receptor (β2AR) in inactive and active conformations (PDB: 3SN6 and 2RH1, respectively). The change in transmembrane helix 6 (red) enables the active form to bind to G proteins and arrestin. Positions of five CB-predicted active-biased mutations that are corroborated by experimental data from Jones et al (30) (Z > 1) are labeled. (**B**) 5263 β2AR variants were scored using CB, and top active-biased, inactive-biased, and neutral variants are plotted by their experimentally-determined receptor basal activation scores (30). ****p < 0.0001. (**C**) Src kinase repositions its auto-inhibitory SH2/SH3 domains upon autophosphorylation, becoming active (PDB: 2SRC and 1Y57, respectively). AS, active site. (**D**) 3372 Src kinase variants were scored using CB and the inactive and active structures shown in (C). Points are colored according to experimentally-determined kinase activity data (based on inhibition of cell growth) from Nguyen et al.(35). (**E**) Distributions of Src kinase activities for inactive-biased, active-biased, and neutral variant populations. **p < 0.01, ****p < 0.0001. (**F**) B-Raf is a human kinase downstream of Ras, which has an inactive (autoinhibited, monomeric) state, and an active (phosphorylated, dimerized) state (PDB: 7MFD and 7MFF, respectively). AS, active site. (**G**) Distributions of experimentally-determined B-Raf activities from Simon et al.(34), based on quantification of phospho-ERK levels. Categories are based on CB analysis of active and inactive B-Raf structures in (F). ****p < 0.0001. (**H**) MurA is an *E. coli* transferase required for peptidoglycan synthesis that interconverts between an inactive "loop open" state and an active "loop closed" conformation in which substrates are re-positioned for

catalysis (PBD: 1UAE). The “loop open” structure was generated by AlphaFold2, templating on the open structure of *E. cloacae* MurA (PBD: 1NAW). AS, active site. (**I**) Distributions of MurA activity scores from Dewachter et al.(54) for open-biased, closed-biased, and neutral populations. ***$p < 0.001$; ****$p < 0.0001$. (**J**) FabZ is an *E. coli* dehydratase involved in peptidoglycan synthesis that switches between a closed conformation and an open, substrate-bound conformation. Structures were generated using AlphaFold2, templating on open and closed structures of *H. pylori* FabZ (PDB: 4ZJB and 2GLL). AS, active site. (**K**) Distributions of FabZ activity scores from Dewachter et al.(54) for closed-biased, open-biased, and neutral populations. *$p<0.05$; **$p<0.01$; ****$p < 0.0001$. (**L**) Overlaid crystal structures of *H. Pylori* FabZ bound to the competitive inhibitors quercetin, apigenin, and sakuranetin (PDB: 3CF8, 3CF9, 3D04)(71) showing positions of substrate binding pocket and entrance to binding tunnel.

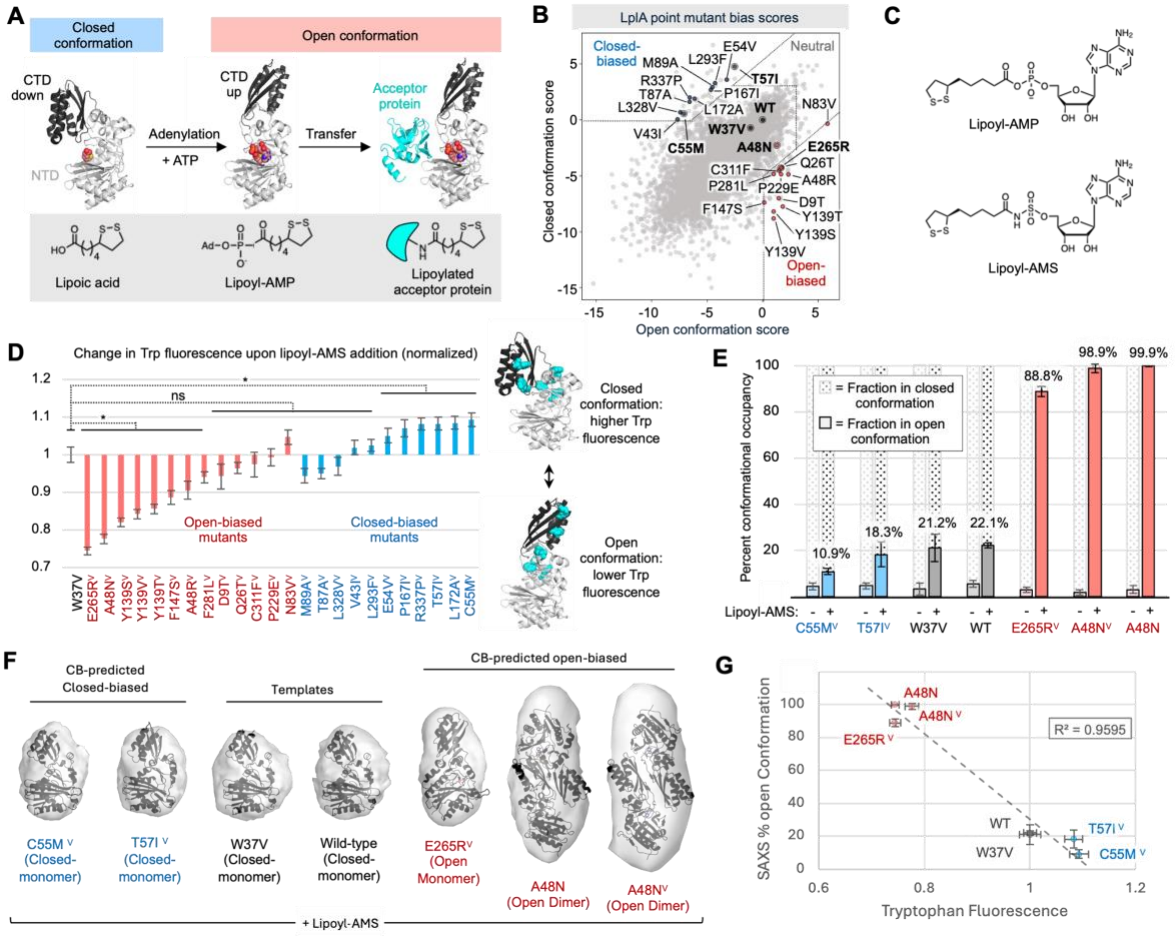
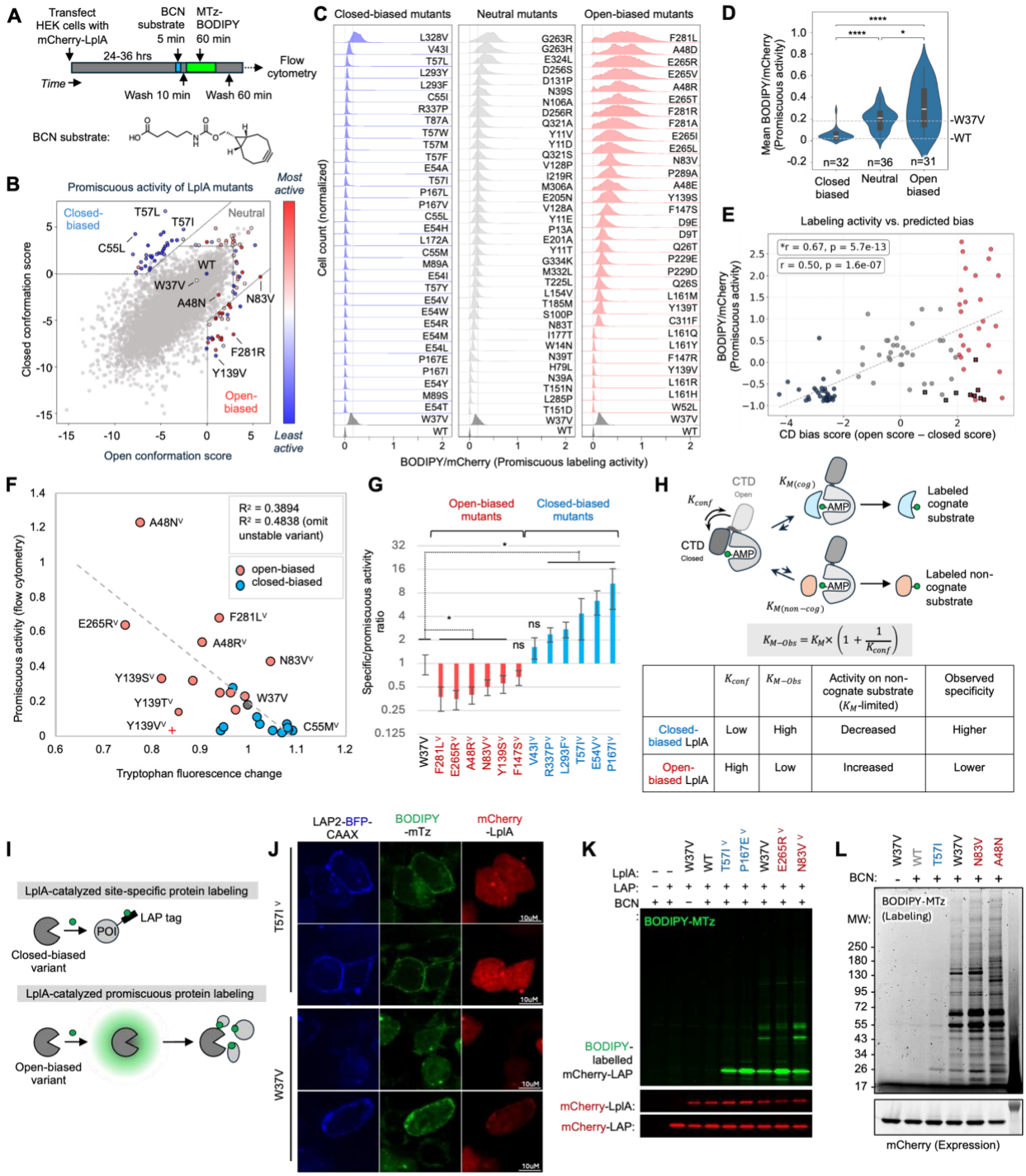


Figure 4. CB biases *E. coli* LplA towards open or closed adenylate ester-bound conformations. (**A**) Two half-reactions catalyzed by LplA: adenylation of lipoic acid by ATP to generate lipoyl-AMP, and transfer of lipoyl onto lysine sidechain of an acceptor protein. LplA switches from a closed to open conformation between the first and second half reactions. Ad, adenosine. CTD, C-terminal domain of LplA. NTD, N-terminal domain. Structures from PBD IDs 1X2H, 3A7R, and 3A7A (left to right). Lipoic acid, lipoyl-AMP, and octyl-AMP shown as red spheres (left to right). (**B**) Conformational design of LplA, using open and closed backbone structures from PDB: 1X2G (closed) and 3A7R (open). Closed-biased, neutral, and open-biased mutants selected for tryptophan fluorescence analysis in (D) are labeled. The 6 variants analyzed by SAXS in (E) are in bold. (**C**) Structure of non-hydrolyzable lipoyl-AMS analog. (**D**) Fold-change in tryptophan (Trp) fluorescence for purified LplA variants upon addition of lipoyl-AMS. All variants were tested on W37V background (notated as mutation^{V}). This experiment was performed in triplicate. Open and closed LplA ribbon structures show positions of Trp sidechains in light blue. *p < 0.01. (**E**) SEC-SAXS analysis of W37V LplA and four biased variants (on W37V background), in apo state or in complex with lipoyl-AMS. Fractional occupancy predicted using Oligomer (see Methods and **Figure S6**), from 2-5 independent measurements per variant. Errors, ±1 std. dev. (**F**) DENSS *ab initio* modeling of protein envelopes, fit with major LplA conformer detected under +lipoyl-AMS condition. At the high protein concentrations

required for SAXS, A48N appears as a dimer (model generated by AlphaFold3). (**G**) Correlation of Trp fluorescence data from (D) to fractional occupancy predicted by SAXS Oligomer analysis in (E).



	K_{conf}	K_{M-Obs}	Activity on non-cognate substrate (K_M-limited)	Observed specificity
Closed-biased LplA	Low	High	Decreased	Higher
Open-biased LplA	High	Low	Increased	Lower

Figure 5. Biased LplA variants exhibit changes in sequence-specific versus promiscuous protein labeling. (**A**) Flow cytometry assay for measuring the promiscuous labeling activity of mCherry-tagged LplA variants. LplA acceptor proteins/peptides are not co-expressed in HEK cells, and thus labeling represents promiscuous tagging of endogenous proteins. After conjugation to proteins, BCN is derivatized via Click with fluorogenic methyltetrazine(MTz)-BODIPY. See **Supplementary Text** for further details about this assay. (**B**) CB scoring of LplA (same as **Figure 4B**) with closed-biased, neutral, and open-biased mutants colored by their experimentally-measured

promiscuous activities in (C). (**C**) Promiscuous activity histograms for closed, neutral, and open-biased LplA variants (on W37V background). 2D BODIPY vs. mCherry flow cytometry plots for all variants shown in **Figure S9**. This experiment was performed 2 times with similar results (correlation between replicates shown in **Figure S7D**). (**D**) Mean promiscuous activity scores for each category of CB-designed LplA variant in (C). * $p<0.05$, **** $p<0.0001$. (**E**) Correlation between experimentally-measured promiscuous activity from (C) and CB-predicted bias score for each LplA variant tested. Variants marked with an X are likely to have a steric clash with BCN substrate in the active site (AS), explaining their low activities. r* = spearman correlation calculated without steric clash variants, r = spearman correlation calculated for all variants. (**F**) Correlation between experimentally-measured promiscuous activity from (C) and Trp fluorescence change measured in **Figure 4D**. Y139V (red cross) is an unstable variant that shows strong aggregation. (**G**) Ratio of specific to promiscuous activity for 12 purified LplA variants, based on E2p labeling and SDS-PAGE analysis in **Figure S7F-G**. Errors, 1 std. dev. *$p < 0.05$. (**H**) Proposed mechanistic model for how conformational occupancy (K_{conf}) tunes substrate specificity in LplA. The C-terminal domain (CTD) may act as a tethered competitive inhibitor for substrate binding, so that K_{conf} tunes the effective $K_{M\text{-}Obs}$ of LplA for its protein substrates by the equation shown. $K_{M(cog)}$ and $K_{M(non\text{-}cog)}$ are K_M values for labeling of cognate and non-cognate substrates, respectively. (**I**) Schematic showing how LplA variants can be used for either site-specific (top) or promiscuous (bottom) protein labeling in cells. POI, protein of interest. LAP is a 13-amino acid engineered LplA Acceptor Peptide(69). (**J**) Imaging of HEK293T cells expressing membrane-localized LAP (LAP-BFP-CAAX), labeled with W37V-LplA or a closed-biased variant ($T57I^{V}$) and BCN for 10 minutes. BCN was Clicked with mTz-BODIPY (middle column). (**K**) In-gel fluorescence for cell lysates labeled as in (J). (**L**) The open-biased LplA variants A48N and N83V catalyze promiscuous labeling of endogenous cytosolic proteins in HEK293T cells. Cells were treated with BCN for 10 minutes before lysis, Click with mTz-BODIPY, and SDS-PAGE. No LAP fusion was expressed in these samples.

References

1. Bryan CM, Rocklin GJ, Bick MJ, Ford A, Majri-Morrison S, Kroll AV, et al. Computational design of a synthetic PD-1 agonist. Proc Natl Acad Sci U S A. 2021 2021/7/20;118(29):e2102164118. PMCID: PMC8307378.
2. Torres SV, Leung PJ, Venkatesh P, Lutz ID, Hink F, Huynh H, et al. De novo design of high-affinity binders of bioactive helical peptides. Nature. 2023 2023/12/18;626:435-42.
3. Divine R, Dang HV, Ueda G, Fallas JA, Vulovic I, Sheffler W, et al. Designed proteins assemble antibodies into modular nanocages. Science. 2021 2021/4/2;372(6537):eabd9994. PMCID: PMC8592034.
4. Hsien-Wei A, Yeh H-W, Norn CH, Kipnis Y, Tischer D, Pellock S, et al. De novo design of luciferases using deep learning. Nature. 2023 2023/2/1;614:774-80.
5. Wayment-Steele HK, Ojoawo AM, Otten R, Apitz JM, Pitsawong W, Hömberger M, et al. Predicting multiple conformations via sequence clustering and AlphaFold2. Nature. 2023 2023/11/13;625:832-9.
6. Stein RA, McHaourab HS. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. PLoS Comput Biol. 2022 2022/8;18(8):e1010483. PMCID: PMC9436118.
7. del Alamo D, Sala D, McHaourab H, Meiler J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. Elife. 2022 2022/3/3;11.
8. Chen J, Wang W, Pang L, Zhu W. Unveiling conformational dynamics changes of H-Ras induced by mutations based on accelerated molecular dynamics. Phys Chem Chem Phys. 2020 2020/9/30;22(37):21238-50.
9. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol. 2009 2009/4;19(2):120-7.
10. Weinkam P, Chen YC, Pons J, Sali A. Impact of mutations on the allosteric conformational equilibrium. J Mol Biol. 2013 2013/2/8;425(3):647-61. PMCID: PMC3557769.
11. Sala D, Del Alamo D, McHaourab HS, Meiler J. Modeling of protein conformational changes with Rosetta guided by limited experimental data. Structure. 2022 2022/8/4;30(8):1157-68.e3. PMCID: PMC9357069.
12. Qiao Z, Nie W, Vahdat A, Miller TF, III, Anandkumar A. State-specific protein-ligand complex structure prediction with a multi-scale deep generative model. arXiv [q-bioQM]. 2022 2022/9/29.
13. Norn C, Wicky BIM, Juergens D, Liu S, Kim D, Tischer D, et al. Protein sequence design by conformational landscape optimization. Proc Natl Acad Sci U S A. 2021 2021/3/16;118(11):e2017228118. PMCID: PMC7980421.
14. Dessaux D, Buchet S, Barthe L, Defresne M, Cioci G, de Givry S, et al. Designing symmetrical multi-component proteins using a hybrid generative AI approach. bioRxiv. 2024 2024/6/15.
15. Sauer MF, Sevy AM, Crowe JE, Jr., Meiler J. Multi-state design of flexible proteins predicts sequences optimal for conformational change. PLoS Comput Biol. 2020 2020/2;16(2):e1007339. PMCID: PMC7032724.
16. Guo AB, Akpinaroglu D, Kelly MJS, Kortemme T. Deep learning guided design of dynamic proteins. bioRxiv. 2024 2024/7/19.

17. Kortemme T. De novo protein design-From new structures to programmable functions. Cell. 2024 2024/2/1;187(3):526-44. PMCID: PMC10990048.
18. Havranek JJ, Harbury PB. Automated design of specificity in molecular recognition. Nat Struct Biol. 2003 2003/1;10(1):45-52.
19. Garrido-Rodríguez P, Carmena-Bargueño M, de la Morena-Barrio ME, Bravo-Pérez C, de la Morena-Barrio B, Cifuentes-Riquelme R, et al. Analysis of AlphaFold and molecular dynamics structure predictions of mutations in serpins. PLoS One. 2024 2024/7/5;19(7):e0304451. PMCID: PMC11226102.
20. Pak MA, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, Maksimova ES, et al. Using AlphaFold to predict the impact of single mutations on protein stability and function. PLoS One. 2023 2023/3/16;18(3):e0282689. PMCID: PMC10019719.
21. Porter LL, Chakravarty D, Schafer JW, Chen EA. ColabFold predicts alternative protein structures from single sequences, coevolution unnecessary for AF-cluster. bioRxivorg. 2023 2023/11/25.
22. Chakravarty D, Schafer JW, Chen EA, Thole JF, Ronish LA, Lee M, et al. AlphaFold predictions of fold-switched conformations are driven by structure memorization. Nat Commun. 2024 2024/8/24;15(1):7296. PMCID: PMC11344769.
23. Praetorius F, Leung PJY, Tessmer MH, Broerman A, Demakis C, Dishman AF, et al. Design of stimulus-responsive two-state hinge proteins. Science. 2023;381(6659):754-60.
24. Pillai A, Idris A, Philomin A, Weidle C, Skotheim R, Leung PJY, et al. De novo design of allosterically switchable protein assemblies. Nature. 2024 2024/8;632(8026):911-20. PMCID: PMC11338832.
25. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, et al. Robust deep learning-based protein sequence design using ProteinMPNN. Science. 2022 2022/10/7;378(6615):49-56. PMCID: PMC9997061.
26. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al. Learning inverse folding from millions of predicted structures. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, editors. Proceedings of the 39th International Conference on Machine Learning; 2022: PMLR; 2022. p. 8946-70.
27. Hie BL, Shanker VR, Xu D, Bruun TUJ, Weidenbacher PA, Tang S, et al. Efficient evolution of human antibodies from general protein language models. Nat Biotechnol. 2023 2023/4/24:1-9.
28. Shanker VR, Bruun TUJ, Hie BL, Kim PS. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. Science. 2024;385(6704):46-53.
29. Sumida KH, Núñez-Franco R, Kalvet I, Pellock SJ, Wicky BIM, Milles LF, et al. Improving protein expression, stability, and function with ProteinMPNN. J Am Chem Soc. 2024 2024/1/24;146(3):2054-61. PMCID: PMC10811672.
30. Jones E, Lubock N, Venkatakrishnan AJ, Wang J, Tseng A, Paggi JM, et al. Structural and functional characterization of G protein–coupled receptors with deep mutational scanning. eLife. 2019 2019/4/30;9.
31. Dewachter L, Brooks AN, Noon K, Cialek C, Clark-ElSayed A, Schalck T, et al. Deep mutational scanning of essential bacterial proteins can guide antibiotic development. Nat Commun. 2023 2023/1/16;14(1):241. PMCID: PMC9842644.

32. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. Cell. 2020 2020/9/3;182(5):1295-310.e20. PMCID: PMC7418704.
33. Weng C, Faure AJ, Escobedo A, Lehner B. The energetic and allosteric landscape for KRAS inhibition. Nature. 2024 2024.
34. Simon JJ, Fowler DM, Maly D. Multiplexed profiling of intracellular protein abundance, activity, interactions and druggability with LABEL-seq. Nat Methods. 2024 2024/10/21.
35. Nguyen V, Ahler E, Sitko KA, Stephany JJ, Maly DJ, Fowler DM. Molecular determinants of Hsp90 dependence of Src kinase revealed by deep mutational scanning. Protein Sci. 2023 2023/7;32(7):e4656. PMCID: PMC10273359.
36. Morris TW, Reed KE, Cronan JE, Jr. Lipoic acid metabolism in Escherichia coli: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein. J Bacteriol. 1995 1995/1;177(1):1-10. PMCID: PMC176549.
37. Pantsar T. The current understanding of KRAS protein structure and dynamics. Comput Struct Biotechnol J. 2020 2020;18:189-98. PMCID: PMC6965201.
38. Guillard S, Kolasinska-Zwierz P, Debreczeni J, Breed J, Zhang J, Bery N, et al. Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange. Nat Commun. 2017 2017/7/14;8(1):16111. PMCID: PMC5519984.
39. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science. 2007 2007/11/23;318(5854):1258-65. PMCID: PMC2583103.
40. Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the β2 adrenergic receptor-Gs protein complex. Nature. 2011 2011/7/19;477(7366):549-55. PMCID: PMC3184188.
41. Yang A, Jude KM, Lai B, Minot M, Kocyla AM, Glassman CR, et al. Deploying synthetic coevolution and machine learning to engineer protein-protein interactions. Science. 2023 2023/7/28;381(6656):eadh1720. PMCID: PMC10403280.
42. Boggon T, Eck M. Structure and regulation of Src family kinases. Oncogene. 2004 2004/10/18;23:7918-27.
43. Roskoski R, Jr. Src kinase regulation by phosphorylation and dephosphorylation. Biochem Biophys Res Commun. 2005 2005/5/27;331(1):1-14.
44. Cowan-Jacob SW, Fendrich G, Manley PW, Jahnke W, Fabbro D, Liebetanz J, et al. The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. Structure. 2005 2005/6;13(6):861-71.
45. Xu W, Doshi A, Lei M, Eck MJ, Harrison SC. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. Mol Cell. 1999 1999/5;3(5):629-38.
46. Freeman AK, Ritt DA, Morrison DK. Effects of Raf dimerization and its inhibition on normal and disease-associated Raf signaling. Mol Cell. 2013 2013/2/21;49(4):751-8. PMCID: PMC3582845.
47. Martinez Fiesco JA, Durrant DE, Morrison DK, Zhang P. Structural insights into the BRAF monomer-to-dimer transition mediated by RAS binding. Nat Commun. 2022 2022/1/25;13(1):486. PMCID: PMC8789793.

48. Mihalovits LM, Ferenczy GG, Keserű GM. Catalytic mechanism and covalent inhibition of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA): Implications to the design of novel antibacterials. J Chem Inf Model. 2019 2019/12/23;59(12):5161-73.
49. Zhang L, Xiao J, Xu J, Fu T, Cao Z, Zhu L, et al. Crystal structure of FabZ-ACP complex reveals a dynamic seesaw-like catalytic mechanism of dehydratase in fatty acid biosynthesis. Cell Res. 2016 2016/11/22;26:1330-44.
50. Zhang L, Liu W, Hu T, Du L, Luo C, Chen K, et al. Structural basis for catalytic and inhibitory mechanisms of beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). J Biol Chem. 2008 2008/2/29;283(9):5370-9.
51. Schönbrunn E, Sack S, Eschenburg S, Perrakis A, Krekel F, Amrhein N, et al. Crystal structure of UDP-N-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomycin. Structure. 1996 1996/9/15;4(9):1065-75.
52. Lewis S, Hempel T, Jiménez-Luna J, Gastegger M, Xie Y, Foong AYK, et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. bioRxiv. 2024:2024-12.
53. Akpinaroglu D, Seki K, Guo A, Zhu E, Kelly MJS, Kortemme T. Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space. BioRxiv. 2023:2023-12.
54. Dieckhaus H, Brocidiacono M, Randolph NZ, Kuhlman B. Transfer learning to leverage larger datasets for improved prediction of protein stability changes. Proceedings of the national academy of sciences. 2024;121(6):e2314853121.
55. Cronan JE. Lipoic acid attachment to proteins: stimulating new developments. Microbiol Mol Biol Rev. 2024 2024/6/27;88(2):e0000524. PMCID: PMC11332335.
56. Uttamapinant C, White KA, Baruah H, Thompson S, Fernández-Suárez M, Puthenveetil S, et al. A fluorophore ligase for site-specific protein labeling inside living cells. Proc Natl Acad Sci U S A. 2010 2010/6/15;107(24):10914-9. PMCID: PMC2890758.
57. Uttamapinant C, Sanchez MI, Liu DS, Yao JZ, White KA, Grecian S, et al. Site-specific protein labeling using PRIME and chelation-assisted click chemistry. Nature protocols. 2013;8(8):1620-34.
58. Fujiwara K, Maita N, Hosaka H, Okamura-Ikeda K, Nakagawa A, Taniguchi H. Global conformational change associated with the two-step reaction catalyzed by Escherichia coli lipoate-protein ligase A. J Biol Chem. 2010 2010/3/26;285(13):9971-80. PMCID: PMC2843243.
59. Fernández-Suárez M, Baruah H, Martínez-Hernández L, Xie KT, Baskin JM, Bertozzi CR, et al. Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. Nature biotechnology. 2007;25(12):1483-7.
60. Baalmann M, Neises L, Bitsch S, Schneider H, Deweid L, Werther P, et al. A Bioorthogonal Click Chemistry Toolbox for Targeted Synthesis of Branched and Well‐Defined Protein‐Protein Conjugates. Angewandte Chemie. 2020;132(31):12985-93.
61. Zhao X, Miller JR, Jiang Y, Marletta MA, Cronan JE. Assembly of the covalent linkage between lipoic acid and its cognate enzymes. Chemistry & biology. 2003;10(12):1293-302.
62. Slavoff SA, Liu DS, Cohen JD, Ting AY. Imaging protein–protein interactions inside living cells via interaction-dependent fluorophore ligation. Journal of the American Chemical Society. 2011;133(49):19769-76.

63. Qin W, Cho K, Cavanagh PE, Ting A. Deciphering molecular interactions by proximity labeling. Nat Methods. 2021 2021/1/11;18:133-43.
64. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. De novo design of protein structure and function with RFdiffusion. Nature. 2023;620(7976):1089-100.
65. Ferruz N, Höcker B. Controllable protein design with language models. Nature Machine Intelligence. 2022;4(6):521-32.
66. Jiang K, Yan Z, Di Bernardo M, Sgrizzi SR, Villiger L, Kayabolen A, et al. Rapid in silico directed evolution by a protein language model with EVOLVEpro. Science. 2024:eadr6006.
67. Cao Y, Yisimayi A, Jian F, Song W, Xiao T, Wang L, et al. BA. 2.12. 1, BA. 4 and BA. 5 escape antibodies elicited by Omicron infection. Nature. 2022;608(7923):593-602.
68. Xu Y, Wu C, Cao X, Gu C, Liu H, Jiang M, et al. Structural and biochemical mechanism for increased infectivity and immune evasion of Omicron BA. 2 variant compared to BA. 1 and their possible mouse origins. Cell research. 2022;32(7):609-20.
69. Puthenveetil S, Liu DS, White KA, Thompson S, Ting AY. Yeast display evolution of a kinetically efficient 13-amino acid substrate for lipoic acid ligase. J Am Chem Soc. 2009 2009/11/18;131(45):16430-8. PMCID: PMC2799336.