

RESEARCH ARTICLE SUMMARY

PROTEIN DESIGN

Deep learning-guided design of dynamic proteins

Amy B. Guo*, Deniz Akpinaroglu, Christina A. Stephens, Michael Grabe, Colin A. Smith, Mark J. S. Kelly, Tanja Kortemme*



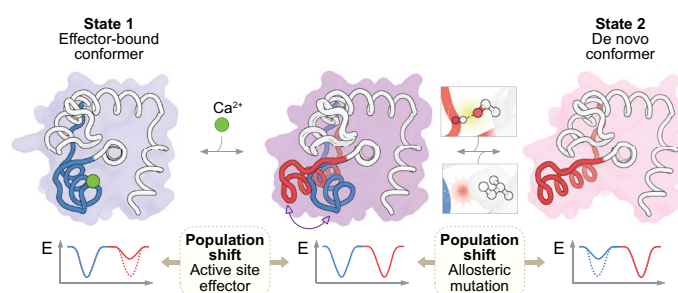
Full article and list of author affiliations:
<https://doi.org/10.1126/science.adr7094>

INTRODUCTION: Deep learning has greatly advanced the design of static proteins with new-to-nature (de novo) structures. A clear next challenge is to design the types of tightly controlled, dynamic conformational changes that underpin natural protein functions de novo. For instance, many superfamilies of protein regulators undergo functionally critical changes to the relative orientation of secondary structural elements within a domain (such as helix rotation in kinases to form active sites, kinking of helices in G protein-coupled receptors to expose binding interfaces to downstream signaling partners) and couple diverse effectors to modulate functionally distinct state populations and thereby activity. Yet, these conformational changes have been inaccessible to de novo design, despite their importance.

RATIONALE: Pioneering work to design protein conformational switches has focussed on side-chain rearrangements or large-scale hinge-like domain motions where most of the atomic-level intradomain interactions are preserved. However, no general method has been described to design the intricate controllable intradomain conformational change mechanisms prevalent in natural regulators de novo, thereby leaving many potential functionalities unexplored. In particular, it is challenging to accurately model the small energetic differences between states at this scale with traditional physics-based models, while “black box” deep learning-based models limit our biophysical understanding of the designed system. As a proof of principle, we reasoned that by combining the high performance and speed of advanced deep learning methods with the interpretability of molecular simulations, we could not only design controllable intradomain modes of motion de novo, but also understand the atomic-interaction networks underlying them.

RESULTS: We describe a general method to design dynamic proteins that uses deep learning to guide the search of sequence and structure space during multistate design. A key element of our approach is restricting the search space by first identifying the minimal set of residues required to define each user-specified conformational state through an in silico mutational scan. Specifically, we assess whether point mutations to increase sequence identity between states perturb computationally predicted structures. Focusing sampling at positions found to be important for determining state preference then allows us to generate sequences with high similarity but diverse state population distributions. This approach not only designs protein switches that exchange between our prespecified states on a functionally relevant timescale (i.e., the low-microsecond regime) but also a mechanism to tune the conformational equilibrium through minimal changes to sequence. The sequence context shared between designs helps to reveal how small sequence differences may alter the conformational ensemble on a biophysical level.

We validate our predictions by extensive experimental dynamics analyses as well as four structures solved by nuclear magnetic resonance spectroscopy, confirming atomic precision of our designs. We find that the amino acid identity at just one sequence position can



Design of dynamic proteins with controllable motions, mimicking mechanisms in natural signaling switches.

We describe a general method that searches sequence and structure space guided by deep learning to design proteins that transition between user-specified structural states (red, blue). A binding site in the dynamic region couples effector binding (Ca^{2+}) to the conformational equilibrium (left), which can in turn also be controlled by allosteric mutations (right). Experimentally determined structures show close agreement with the designed states, and physics-based simulations reveal state-specific networks that explain the observed allostery and predict new mutations that tune the equilibrium.

substantially shift the distribution of states, ranging from highly skewed to roughly equal populations. We also show regulation of the state populations through Ca^{2+} concentration—a common secondary messenger—by incorporating a Ca^{2+} binding site within the designed dynamic region, which adopts a binding-competent conformation in only one of the two states. In turn, we demonstrate the allosteric effect on Ca^{2+} binding affinity by distal mutations that tune the conformational equilibrium. Finally, we find close agreement between our deep learning-based design predictions, physics-based simulations, and experimental data, allowing us to extract and reprogram state-specific distinct atomic interaction networks.

CONCLUSION: Our results validate a general approach to designing proteins with distinct conformational states that can be specified by the designer. As in natural regulation, the equilibrium between the designed conformational states can be controlled both by ligand binding in the active site and by allosteric perturbations (mutations) to distal sites. We show that new modes of intradomain motion, inspired by those key to natural signaling, can now be realized through de novo design. This work provides a foundation for designing programmable signaling systems de novo, facilitating more complex behavior such as de novo signal integration or concerted motions coupled to energy inputs. Moreover, this approach can be applied to engineer non-native motions into natural proteins to control their activities. □

*Corresponding author. Email: amyguo1997@gmail.com (A.B.G.); tanjakortemme@gmail.com (T.K.) Cite this article as A. B. Guo *et al.*, *Science* **388**, eadr7094 (2025). DOI: 10.1126/science.adr7094

PROTEIN DESIGN

Deep learning-guided design of dynamic proteins

Amy B. Guo^{1,2*}, Deniz Akpinaroglu^{1,2}, Christina A. Stephens^{3,4}, Michael Grabe^{3,4}, Colin A. Smith⁵, Mark J. S. Kelly³, Tanja Kortemme^{1,2,6,7*}

Deep learning has advanced the design of static protein structures, but the controlled conformational changes that are hallmarks of natural signaling proteins have remained inaccessible to de novo design. Here, we describe a general deep learning-guided approach for de novo design of dynamic changes between intradomain geometries of proteins, similar to switch mechanisms prevalent in nature, with atomic-level precision. We solve four structures that validate the designed conformations, demonstrate modulation of the conformational landscape by orthosteric ligands and allosteric mutations, and show that physics-based simulations are in agreement with deep-learning predictions and experimental data. Our approach demonstrates that new modes of motion can now be realized through de novo design and provides a framework for constructing biology-inspired, tunable, and controllable protein signaling behavior de novo.

Many natural proteins can adopt multiple conformational states, and the controlled interconversion between these states often underlies biological functions and their regulation (1, 2). Binding partners, environmental conditions, posttranslational modifications, allosteric effectors, and mutations can modulate protein function by changing the populations of active and inactive conformations (3, 4). Notably, many of the conformational changes required for function and regulation do not involve large-scale structural rearrangements; rather, critical motions often consist of localized rotation, tilting, and sliding of secondary structure elements within a domain. These types of motion are hallmarks of central biological processes, such as the regulation of kinases (5) and signaling via G protein-coupled receptors (GPCRs) (6) (fig. S1).

Despite the functional importance of local conformational changes in natural proteins, predictive design of similar dynamics de novo has remained elusive (7). One key challenge is the difficulty in parameterizing sufficiently accurate physics-based energy functions to design sequences that can adopt multiple conformations [multistate design (8)], because the energy gap between user-specified folded states, off-target folded states, and disordered states is typically small, especially for intradomain conformational changes. Here, recent advances in deep learning that have enabled robust design of sequences conditioned on structure (9, 10) could provide more sensitive assessment of how changes to sequence affect conformational equilibria. A second challenge is to generate alternative conformational states de novo (not borrowed from existing proteins) that are designable, i.e., there must exist an amino acid sequence that can adopt these alternative conformations.

To date, there is no generalizable method to address this problem (11). Consequently, there are only a few examples of de novo-designed dynamic proteins and none on the local scale prevalent in natural regulators. Previous pioneering studies include the design of fold switches (12, 13), side-chain dynamics (14), and controllable changes in coiled-coil assemblies (15). Most commonly, however, design of protein switches has involved domain-level rearrangements or hinge-like motions of rigid bodies, in which conformational changes are driven by intermolecular protein-protein interactions and the multistate design problem is simplified because either most atomic interactions within the rigid bodies remain the same (16) or one of the states becomes disordered (17). As a result, most classes of conformational switch mechanisms, effectors, and their combinations have been inaccessible and hence completely unexplored by de novo design.

Design approach for dynamic proteins with tunable two-state equilibria

Inspired by the scale and modes of conformational changes prevalent in natural signaling proteins, we sought to develop a generalizable approach to design non-native controllable conformational switching in proteins de novo, providing a basis for designing tunable and complex signaling behavior beyond what exists in nature (7). Specifically, we aimed to design sequences with multiple energetic minima in structure space, each corresponding to a well-defined conformational state (Fig. 1A). We focused on conformational changes involving intradomain reorientation of secondary structure elements, mimicking a dominant mechanism in naturally occurring proteins (in contrast to the easier problem of movement of otherwise static subdomains about a hinge). As in biological regulation, we aimed to design mechanisms for modulating the conformational equilibrium by inputs such as orthosteric ligands (binding within the region of conformational change) and allosteric perturbations [acting at distal sites coupled to the active site (18)].

Our design approach (Fig. 1B) involves three general stages: The first stage identifies alternative structural states through (i) generation of a library of de novo candidate states that differ in their local geometries using systematic physics-based conformational sampling (fig. S2 and Materials and methods) (19), and (ii) assessment of the designability of these states. To assess single-state designability of the generated backbones, we designed sequences for each state, evaluated them in silico (Materials and methods), and characterized them experimentally. The second stage performs a deep learning-guided search in sequence and structure space that (i) restricts the search space by identifying the minimal set of residues required to define each state, and (ii) designs sequences at these positions that are simultaneously compatible with pairs of states (multistate design). The third stage seeks to identify perturbations that modulate the designed protein conformational landscape by (i) integrating physics-based simulations and deep learning predictions to determine state-specific interaction networks and (ii) predicting mutations leading to tunable conformational switching.

For the proof-of-concept application described here, we required the conformational landscape to be responsive to a ligand input, where the ligand-binding site (orthosteric site) changes conformations between states and thereby couples ligand binding to conformational switching. For simplicity, we used an engineered Ca^{2+} binding protein (derived from the N-terminal domain of troponin C) as our ligand binding-competent starting state ("state 1") (Fig. 1C, top row). The wild-type protein consists of two EF hand motifs (sites I and II), which both bind Ca^{2+} in the low micromolar range. We instead used a variant [Protein Data Bank (PDB) ID: 1SMG] with an E41A point mutation in site I, which weakens the affinity of site I to the millimolar range while retaining moderate micromolar affinity in site II (20). Additionally, the E41A mutant does not undergo conformational change upon Ca^{2+} binding (fig. S3). By using an existing binding state as a starting point

¹The UC Berkeley–UCSF Graduate Program in Bioengineering, University of California, San Francisco, San Francisco, CA, USA. ²Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA. ³Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA. ⁴Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, USA. ⁵Department of Chemistry, Wesleyan University, Middletown, CT, USA. ⁶Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA, USA. ⁷Chan Zuckerberg Biohub; San Francisco, CA, USA. *Corresponding author. Email: amyguo1997@gmail.com (A.B.G.); tanjakortemme@gmail.com (T.K.)

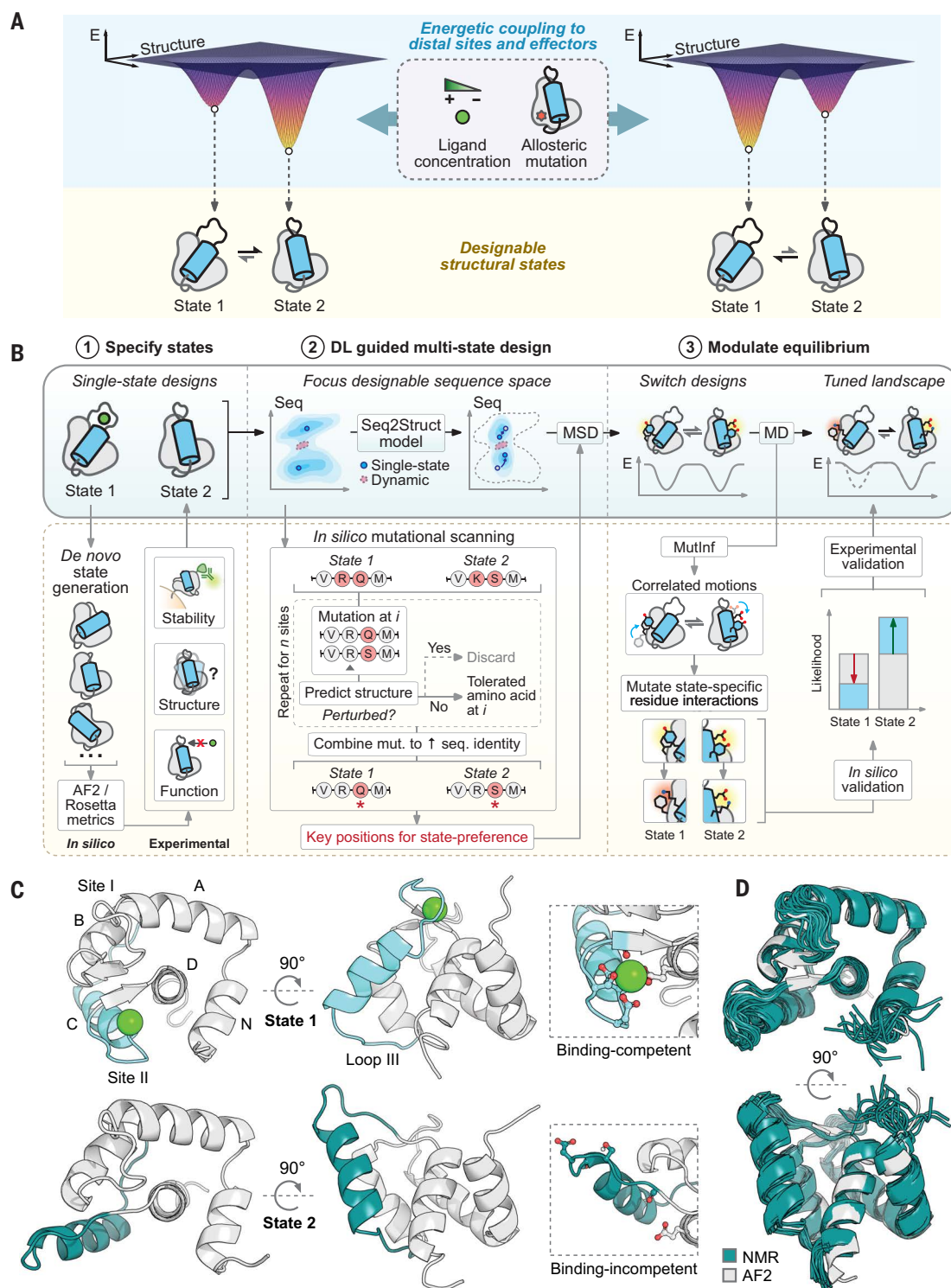


Fig. 1. Generalizable approach for the deep learning-guided design of dynamic proteins. (A) Schematic of design goal to engineer dynamic proteins in a two-state equilibrium that can be controlled by orthosteric ligands and allosteric perturbations. (B) Main stages of the approach: (1) De novo generation of alternative states that differ in their geometry (light blue region) using systematic conformational sampling (fig. S2), followed by in silico and experimental validation of single-state designs. (2) Deep learning (DL)-guided sequence-structure search to focus sampling during multistate design (MSD) at key positions for determining state preference and their neighbors. (3) Combination of physics-based molecular dynamics (MD) simulations, mutual information analysis (MutInf), and DL models to determine state-specific residue interaction networks and identify mutations capable of modulating the conformational landscape, followed by experimental validation. (C and D) Application to generate two designable states with distinct conformations coupled to ligand binding (Ca^{2+}). (C) (Top row, light blue) Binding-competent state 1 structure with Ca^{2+} binding site (inset) (PDB ID: 1SMG) shown in two orientations. (Bottom row, teal) De novo-generated alternative (binding-incompetent) state 2 model shown in two orientations. To couple Ca^{2+} binding to the designed conformational change, the Ca^{2+} binding site is highly reshaped in state 2 to disfavor Ca^{2+} binding (inset). (D) Overlay of the NMR structure of a state 2 single-state design (teal) with its AF2 prediction (gray) shows excellent agreement (Cα RMSD = 0.98 Å, excluding loops).

and designing a conformational change encompassing the binding site, we demonstrate here how our protocol can be generally applied to engineering controllable conformational changes into natural proteins; alternatively, one could generate a (static) binding-competent state de novo (21).

Generation and experimental validation of de novo designed alternative states

To generate a structurally defined alternative state (“state 2”) (Fig. 1C, bottom row), we sampled de novo orientations of a contiguous protein segment that includes loop III, helix C, and Ca^{2+} binding site II (which we refer to as the “reshaped region”) (fig. S2) using the loop-helix-loop unit combinatorial sampling algorithm (LUCS), previously shown to be capable of generating static proteins that differ in the local geometry of user-defined protein segments (Materials and methods) (19). We only kept design models with the same length as our input but allowed the secondary structure in the reshaped region to vary (i.e., some regions may be part of a loop in the input but adopt a helical structure in the output or vice versa, a common transition in natural proteins). We then used Rosetta to design sequences optimal for each of the output models (“single-state designs”) and filtered these designs computationally (Materials and methods). This procedure generated a library of $\sim 1 \times 10^3$ diverse conformations with an average $\text{C}\alpha$ root mean square deviation (RMSD) of 7.1 Å in the reshaped region, which is comparable in scale to functional conformational changes in natural signaling proteins (fig. S1). From this library, we selected 11 designs (each corresponding to a distinct backbone) for experimental testing.

To rapidly screen the de novo state 2 sequences experimentally, we displayed each design fused to a C-terminal c-Myc tag on the surface of yeast (Materials and methods) and used surface display levels (which are known to be correlated with stability) as a proxy for designability. Although we only tested a few single-state designs in this study, one could screen thousands of designs with yeast display to identify many more designable conformational states. Ten out of 11 designs (all except #615) showed high surface display levels (fig. S4). We decided to further characterize design #6306 as it had a very different conformation in the reshaped region compared with state 1 (Fig. 1C, bottom row), involving both rotation and translation of the reshaped helix C. Moreover, the Ca^{2+} binding loop was substantially restructured and partially helical, resulting in an unfavorable conformation for binding. We solved the nuclear magnetic resonance (NMR) structure of #6306 and found excellent agreement ($\text{C}\alpha$ RMSD = 0.98 Å, excluding loops) between the experimentally solved structure (teal) and the AlphaFold2 (AF2) (22) model of the design (gray) (Fig. 1D and fig. S5). In contrast to state 1 (ISMG), design #6306 did not bind Ca^{2+} in site II at Ca^{2+} concentrations up to 1 mM (fig. S6), as expected owing to notable restructuring of the binding site. These results confirm that the backbone of design #6306 is both designable and unfavorable for ligand binding, making it suitable as our binding-incompetent conformation (state 2) for two-state design.

Multistate design of dynamic proteins

We next aimed to identify sequences that were simultaneously compatible with the conformations of both state 1 and state 2. Rather than allowing all residues in the reshaped region and their contacting residue positions to be designable, we sought to generate multistate designs able to populate both states to varying degrees despite having high sequence similarity. To do so, we used deep learning-based structure predictions (AF2) to shrink the searchable sequence space and focus sampling at key positions for determining state preference. We reasoned that this approach would allow us to better interpret sequence differences predicted to shift the preference for one state versus the other, identify potential sites for allosteric mutations, and facilitate experimental characterization. To achieve this, we used AF2 to identify mutations to design #6306 (predicted to adopt state 2) that

would increase sequence similarity to ISMG (state 1) without affecting the predicted structure (Materials and methods) (Fig. 1B). We then used the resulting sequence with the highest sequence identity to the state 1 sequence (ISMG), but still predicted to fold into state 2, as input for multistate design (position-tied ProteinMPNN) (Materials and methods and fig. S7, A and B) (9). Sites where mutations caused notable structural perturbations were typically positions where there were large changes to solvent-accessible side-chain surface area, hydrogen bonding networks, or steric packing between states (table S1). The final set of multistate designable residues included these positions and their neighbors, decreasing the set of designable residues from 37 to 25 (table S2). The structures of putative switch design sequences were then predicted with AF2 to assess their compatibility with both states.

Our design approach identified a family of sequences that had AF2 structure predictions that were either in state 1, state 2, or a combination of both including structural intermediates (Fig. 2A). The designs differed from the original state 1 sequence (ISMG) by $n = 18$ mutations and from the high-sequence identity single-state state 2 design by $n = 15$ mutations, but notably differed from each other at only one residue position, 89, which was located outside the reshaped region and distal from the Ca^{2+} binding site. Position 89 was hence predicted to act as an allosteric site where perturbations, in this case mutations, were predicted to change the populations of states in the reshaped region, including the conformation of the distal Ca^{2+} binding site. In particular, smaller hydrogen bond donors and acceptors at position 89 were biased toward state 2 by forming a hydrogen bond with the backbone of loop III, bringing it closer to the central helix D. Conversely, bulky and/or hydrophobic amino acids pushed loop III outward into a conformation more consistent with state 1. The AF2 confidence metric (pLDDT) of the reshaped region distal to site 89 for this sequence family varied considerably depending on the amino acid identity at position 89 (fig. S7C).

Allosteric modulation of the conformational landscape by mutation

To test the prediction that our designs adopt two different defined conformations whose relative populations are dependent on the amino acid identity at position 89, we selected five designs that covered a range of AF2 predicted behaviors: state 2 preferred (S89, N89), state 1 preferred (I89, K89), and mixed (R89). We recorded 2D ^1H , ^{15}N -heteronuclear single-quantum coherence (HSQC) NMR spectra for each of these designs (Fig. 2B and fig. S8) and focused structural characterization on three representative designs (S89, N89, I89; analysis of R89, predicted to adopt multiple conformational states, was complicated by evidence of peak broadening, fig. S8). Notably, the 2D ^1H , ^{15}N -HSQC NMR spectra were drastically different even though the designs differed only by a single point mutation: 67 out of the 92 backbone amide peaks in the spectrum of I89 (state 1 preferred) had significantly different chemical shifts [$\Delta\delta\text{Hn} > 0.03$ parts per million (ppm) or $\Delta\delta\text{N} > 0.4$ ppm] compared with that of S89 (state 2 preferred), suggesting that these residues experience different chemical environments (Fig. 2B and table S3). Moreover, for well-separated peaks, the chemical shifts of S89 and I89 were at the two ends of a range, with N89 being intermediate. This finding was overall consistent with a simple model in which the designed proteins are in an equilibrium between two states in fast exchange on the NMR timescale where the observed chemical shifts reflect the population average of the two states (though we cannot exclude the possibility of additional exchange processes from these data). To analyze this behavior further, we assigned the backbone ^1Hn and ^{15}N chemical shifts of S89 and I89 and plotted the differences on the AF2 model of I89 (Fig. 2B). We found that the changes to chemical shifts were localized not only near the mutated position 89, but also at more distal residues throughout the reshaped region, including Ca^{2+} binding site II and its neighboring residues, consistent with a change in the ensemble-averaged conformation of the reshaped region.

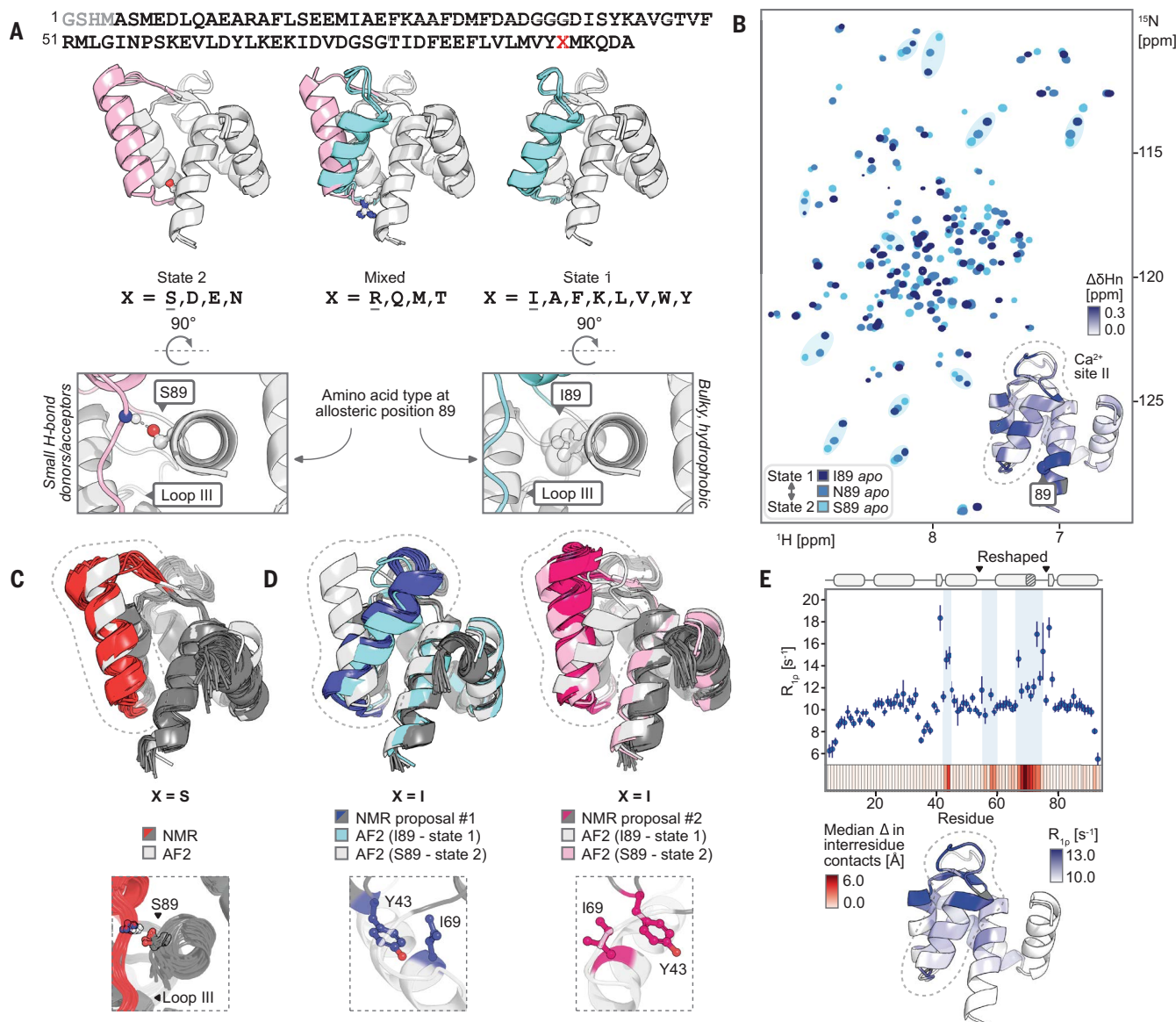


Fig. 2. Two-state equilibrium in fast exchange shifted by allosteric mutations. (A) The protocol in Fig. 1B predicted a family of sequences differing only at position 89 (X), where the amino acid identity at position X determined whether the five AF2-predicted models were entirely in state 2 (left), mixed (middle), or entirely in state 1 (right). Depicted AF2 models (gray cartoons), with the reshaped region colored by state 1 (light blue) or state 2 (pink) conformations, are for the underlined amino acid at position X (shown as sticks). Small polar residues favored state 2 by hydrogen bonding with the backbone of loop III (bottom, left), whereas bulky and/or hydrophobic residues favored state 1 by pushing loop III away from the central helix (bottom, right). (B) 2D ¹H, ¹⁵N-HSQC spectra of S89, N89, and I89, with several well-resolved peaks (shaded ovals) showing chemical shift changes consistent with a two-state equilibrium in fast exchange where I89, N89, and S89 exhibit an increasing population of state 2 relative to state 1, in that order. Inset shows ¹Hn chemical shift changes between I89 and S89 colored on the AF2 model of I89, consistent with the designed conformational change in the reshaped region [the reshaped region is circled by dashed line in panels (B) to (E)]. (C) Agreement between the NMR structure of S89 (red) and its AF2 prediction (gray) (C α RMSD = 1.31 Å excluding loops). Inset shows the hydrogen bond formed between S89 and loop III in the NMR structure consistent with AF2 predictions in (A). (D) NMR models for I89 were consistent with sampling both state 1 (blue, left) (C α RMSD = 1.67 Å excluding loops) and state 2 (pink, right) (C α RMSD = 1.31 Å excluding loops), with residue I69 buried in proposal #1 (state 1) and solvent-exposed in proposal #2 (state 2). (E) ¹⁵N on-resonance R_{1ρ} relaxation rates for design I89 plotted per residue (top) and visualized on the AF2 structure of design I89 (bottom) indicate low-microsecond exchange in the regions predicted to undergo notable conformational changes (median predicted change in C β -C β distances between states shown on the x axis colored by magnitude). Residue numbering of all designs includes the N-terminal thrombin cleavage site scar [gray in (A)]. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

To assess whether our designs indeed adopted the two specific designed conformational states, we solved the structures of S89 and I89 by NMR (Fig. 2, C and D). The structure of S89 was in excellent agreement with its AF2 prediction (state 2, C α RMSD = 1.31 Å, excluding loops), including at the side-chain level (Fig. 2C and fig. S9). By contrast,

for I89, automated structure determination using ARTINA (Materials and methods) did not converge to a single proposed conformation in the reshaped region (Fig. 2D and fig. S9). Instead, the top two structures proposed by ARTINA resembled the designed state 1 and state 2 conformations, respectively, based on assigned distance restraints

consistent with both states 1 and 2 (fig. S10). The residue predicted to undergo the largest conformational change in our computational models (I69) was buried in proposal #1 but solvent-exposed in proposal #2, hallmarks of the state 1 and 2 conformations, respectively. Taken together, our structures confirm that the designs adopt the two intended conformations but with different populations (table S4). Although S89 primarily adopts state 2 as predicted, I89 samples both state 1 and state 2 comparably, and a single-state structure cannot fully account for all ensemble-averaged distance restraints simultaneously.

To probe the suggested dynamics in design I89 more directly, we first collected a series of 2D ^1H , ^{15}N -HSQC spectra from 5° to 35°C at 5°C intervals. We observed temperature-dependent changes in peak intensity localized to residues in the reshaped region and their neighbors, consistent with changes in chemical environment due to dynamics in the reshaped region (fig. S11). Moreover, peak intensities were higher at 35°C compared to 5°C, as expected for a system being in fast exchange at higher temperatures and slowing to fast-intermediate exchange with decreasing temperature, manifesting as line broadening. Additionally, we measured $R_{1\rho}$ values for design I89 with on-resonance rotating frame relaxation dispersion at an effective field strength of 3 kHz and observed higher $R_{1\rho}$ values for residues in the reshaped region and their neighbors (Fig. 2E). This behavior is consistent with chemical exchange on the micro-to-millisecond timescale accelerating measured relaxation rates (Fig. 2E and table S5), as well as our estimate of an upper bound on the exchange time from chemical shift data $\tau_{\text{ex}} \ll 10$ ms (table S6). Finally, further NMR dynamics experiments and analyses including $R_{1\rho}$ relaxation dispersion

measurements allowed us to tighten the upper bound for exchange in the reshaped region to 42 μs (Materials and methods and figs. S12 and S13).

Orthosteric modulation of the conformational landscape by ligand binding

To probe whether Ca^{2+} binding also modulates the state populations by preferentially stabilizing state 1, we recorded 2D ^1H , ^{15}N -HSQC spectra with and without Ca^{2+} for each point mutant (fig. S14). As expected from the designed conformational change, we found that Ca^{2+} addition caused significant chemical shift perturbations ($\Delta\delta_{\text{Hn}} > 0.03$ ppm or $\Delta\delta_{\text{N}} > 0.4$ ppm) throughout the reshaped region and its neighboring residues, affecting ~30 additional peaks when compared to the single-state binding-incompetent design #6306 (Fig. 3A and table S7). Given that our NMR structural ensemble for S89 is in agreement with our binding-incompetent structure (state 2) whereas I89 has a substantial binding-competent (state 1) population (Fig. 2, C and D), the direction of chemical shift changes with Ca^{2+} was indeed consistent with a shift in the equilibrium toward state 1 (Fig. 3B and table S6). We next measured the Ca^{2+} binding affinity by monitoring Ca^{2+} concentration-dependent chemical shift changes of residues in the reshaped Ca^{2+} site and indeed found increasing affinity with an increasing estimated population of the binding-competent state 1, with an overall ~10-fold difference in Ca^{2+} binding affinity between designs I89 and S89 (Fig. 3C). These results confirm that mutations at the distal position 89—more than 15 Å away from the closest Ca^{2+} binding residue—allosterically modulate the Ca^{2+} binding site conformational equilibrium and therefore

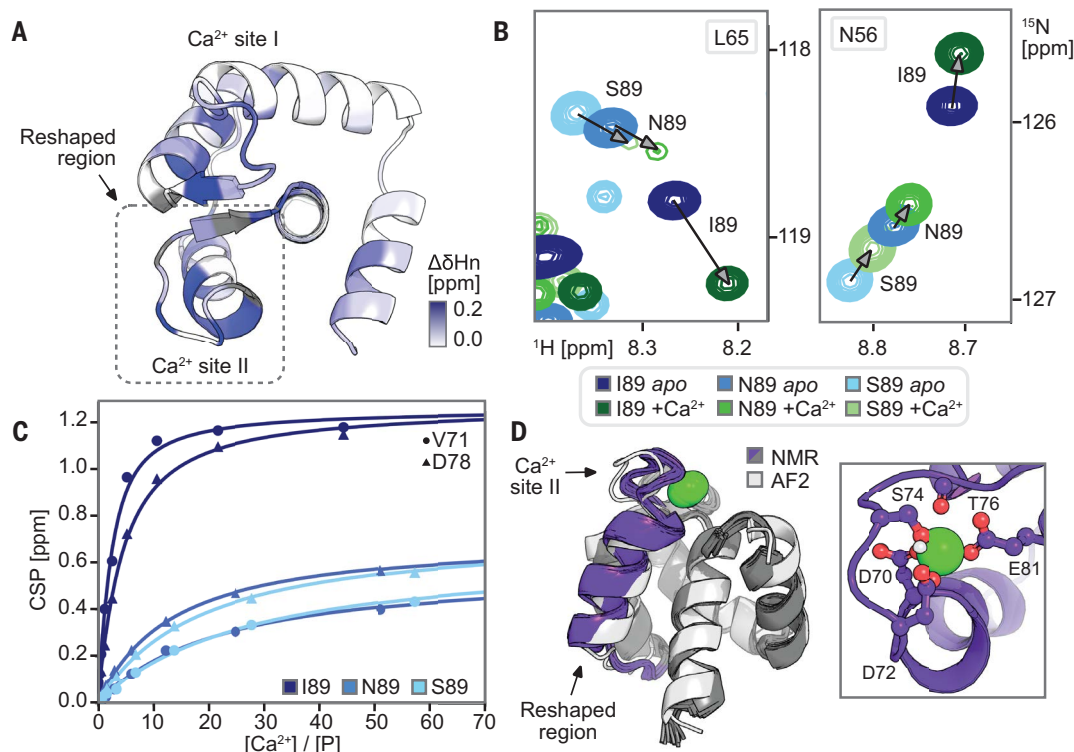


Fig. 3. Modulation of the conformational landscape by ligand binding. (A) Changes in I89 ^1Hn chemical shifts upon adding 10 equiv. of Ca^{2+} visualized on the AF2 model, showing significant chemical shift perturbations distal to the Ca^{2+} binding sites, particularly in the reshaped region and in neighboring residues. (B) 2D ^1H , ^{15}N -HSQC spectra of S89, N89, and I89 with 10 equiv. of Ca^{2+} show ligand-dependent chemical shift changes for residues in the reshaped region consistent with Ca^{2+} shifting the population distribution toward state 1 (arrows). (C) Ca^{2+} titration curves for I89 (dark blue), N89 (medium blue), and S89 (light blue) for residues V71 (circles) and D78 (triangles) in the reshaped Ca^{2+} binding site, which demonstrate increasing affinity as the estimated binding-competent state population increases. The estimated K_d values are 1.6 ± 0.2 mM (V71) and 3.9 ± 0.2 mM (D78) for I89, 20 ± 2 mM (V71) and 11.6 ± 0.4 mM (D78) for N89, and 22 ± 2 mM (V71) and 14.6 ± 0.4 mM (D78) for S89. (D) The Ca^{2+} -bound NMR structure of I89 is in excellent agreement with its AF2 prediction ($\text{C}\alpha$ RMSD = 1.34 Å, excluding loops), and the observed binding site backbone is consistent with the known EF hand binding motif with modeled-in Ca^{2+} (right).

binding affinity, consistent with previous approaches using mutagenesis to probe allosteric coupling (23, 24).

Finally, we solved the NMR structure of I89 in the presence of Ca^{2+} . The *holo* structure was in excellent agreement with our computational state 1 model ($\text{C}\alpha$ RMSD = 1.34 Å, excluding loops), and the backbone conformation of binding site II was consistent with an EF hand binding motif (Fig. 3D). Though *holo* I89 had many more distance restraints consistent with state 1 and less consistent with state 2 compared to *apo* I89, we still observed several nuclear Overhauser effect–derived distance restraints consistent with state 2 even in the presence of excess Ca^{2+} , suggesting residual dynamics (fig. S15). Taken together, these results suggest that the tested family of sequences adopted the two designed conformational states of the reshaped region in solution, where the populations of these states (table S6) can be modulated by both allosteric mutations (Fig. 2) and Ca^{2+} binding (Fig. 3).

Integration of physics-based simulations to tune conformational equilibria

To further probe the atomic-level interactions underlying the designed conformational switching, we ran molecular dynamics (MD) simulations for designs I89 and S89 with and without Ca^{2+} (Fig. 4A and figs. S16 and S17). We observed reversible transitions between states 1 and 2 for design I89 in the absence of Ca^{2+} , in excellent agreement with our design predictions and experimental data (Fig. 4A, fig. S18, and movie S1). When fitting a Markov state model (MSM) on our aggregate 36- μs Ca^{2+} -free I89 simulation data clustered by $\text{C}\alpha$ RMSD of the reshaped helix to each state, we estimated a lower bound of $\sim 3\ \mu\text{s}$ for our overall timescale of exchange (figs. S19 and S20). Combined with the upper bound from the $R_{1\rho}$ data of $\sim 42\ \mu\text{s}$ (fig. S13), these results indicate that our reshaped region undergoes exchange on a low-microsecond timescale, comparable to motions of similar scale in natural proteins (25). By contrast, we did not observe conformational switching when simulating design I89 with Ca^{2+} , and Ca^{2+} remained bound to site II during the simulation (Fig. 4A, lower plot), in agreement with the experimental data showing that Ca^{2+} stabilizes state 1. Design S89 remained in the state 2 conformation throughout the entire course of several 2- μs simulations in the Ca^{2+} -free condition (fig. S16). Though we did not observe a transition to state 1 in the presence of Ca^{2+} , we did see larger fluctuations in the reshaped region for which the RMSD to state 1 decreased compared to trajectories without Ca^{2+} (figs. S16 and S17). The lack of state transitions at this timescale may be due to insufficient sampling, though this behavior is consistent with the NMR structure of S89 (Fig. 2C) and the Ca^{2+} binding data (Fig. 3C). The MD results provide compelling support of exchange between the designed states on a low-microsecond timescale for design I89, show that Ca^{2+} preferentially stabilizes state 1 in I89, and are consistent with allosteric modulation of the designed switch at position 89.

Because of the close agreement between the design predictions (Fig. 2A), MD simulations (Fig. 4A), and NMR data (Figs. 2, B to E, and 3), we next asked whether the MD results could (i) explain the mechanism of allosteric modulation and (ii) make testable prospective predictions to further validate this mechanism. We first performed a mutual information analysis of side-chain torsional dynamics (26) in the *apo* I89 MD trajectories where a state transition was observed. We observed a network of hydrophobic core residues coupling torsional motions of the Ca^{2+} binding site II (residues 70 to 76) to distal residues in loop III (residues 50 to 58) and helix D (residues 80 to 94) (Fig. 4B). Allosteric residue 89 directly faces loop III from central helix D. Combined with our experimental results on allosteric modulation (Figs. 2 and 3C), our MD results therefore suggest a mechanism whereby the amino acid identity at residue 89 causes differences in sterics (I89) or hydrogen bonding (S89) interactions between helix D and loop III (Fig. 2A), which in turn couple through the identified correlated hydrophobic network to allosterically affect the conformation of the

distal Ca^{2+} binding site (Fig. 4B). Further analysis of the mutual information data revealed two distinct interaction networks that are specific for state 1 or state 2, respectively (Fig. 4C). State 1 appeared to be stabilized by hydrophobic interactions involving burial of I69, π - π stacking between Y64 and Y88, and less steric bulk near I89. Conversely, in state 2, I69 is pointed outward to the surface, allowing Y43 to form new contacts in a network of hydrogen-bonding and electrostatic interactions mediated by Y43, Y64, K68, and E81. Taken together, the mutual information analysis of the MD trajectories suggests two extensive state-specific interaction networks that link the reshaped region and the allosteric site 89.

To test these state-specific interactions, we predicted mutations of the identified networks that would favor state 1 and assessed them by (i) scoring with a structure-conditioned masked language model (Frame2seq) (10), (ii) predicting the mutant structures with AF2, and (iii) collecting 2D ^1H , ^{15}N -HSQC spectra. We reasoned that a Y64F mutation should be disruptive in state 2, as it is unable to hydrogen bond with Y43 and E81, but neutral in state 1, where phenylalanine can still form a π - π stacking interaction with Y88. Likewise, a K68E mutation should be disruptive in state 2, as it cannot form a stabilizing electrostatic interaction with E81, but neutral in state 1, where it is solvent-exposed. In both cases, Frame2seq predicted a higher likelihood for the mutated amino acid compared to the original amino acid given state 1 and a lower likelihood given state 2; AF2 predicted these mutants to adopt state 1 with higher confidence compared to the original I89 sequence; and the 2D ^1H , ^{15}N -HSQC spectra were consistent with the expected shift in population toward state 1 (Fig. 4, D to F). In summary, our results explain the observed allosteric modulation of the conformational landscape by mutations at the lynchpin position 89 and validate predictions to further tune the switch equilibrium.

Discussion

Our results demonstrate a general approach to design proteins with two distinct conformational states specified by the designer whose interconversion can be modulated both by ligand concentration (orthosterically) and by mutations to distal sites (allosterically). The approach generalizes beyond prior *de novo*-designed switches based primarily on domain replacement or hinge-like motions (11), where most of the atomic interactions between the rigid bodies remain constant (16, 17). By contrast, our designs interchange between distinct sets of atomic interactions and demonstrate that new modes of motions—inspired by those present in regulator superfamilies such as kinases and GPCRs—can now be realized through *de novo* design, greatly expanding accessible functional space.

One notable observation is the strong correspondence between the deep learning–based predictions, experimental data, and physics-based simulations. This agreement provided testable hypotheses on the mechanisms underlying the bistability of the switch and allowed us to modulate the conformational equilibrium at the atom level. We attribute these findings at least in part to specific features of our approach, in particular the search through sequence and structure space that narrowed designable positions to those predicted to be strong determinants of the engineered conformational change (Fig. 1B). The speed and inference offered by deep learning–based protein sequence design and structure prediction enabled this approach that, ultimately, designed distinct residue networks stabilizing the two structural states. This level of insight into a designed system is essential to advance *de novo* design of allosteric regulation.

The synergy between deep learning and physics-based simulations—demonstrated here for the *de novo* design of dynamic proteins—may be useful for developing future design methods that would enable predictive control over conformational landscapes and the timescale of exchange (we note that our current design method does not explicitly consider transition state barriers). Emerging approaches to these problems include training models on simulation and/or experimental

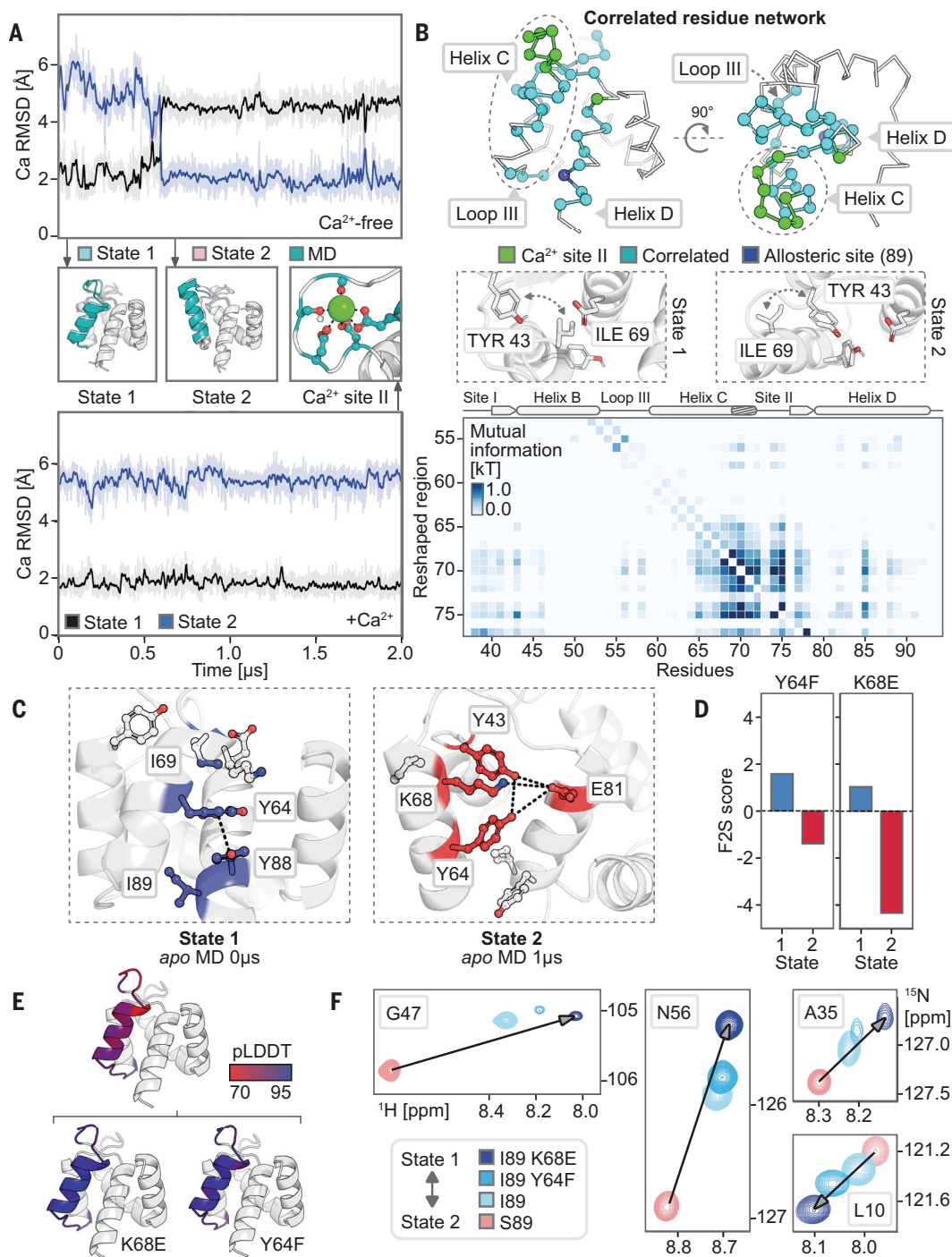


Fig. 4. Physics-based simulations reveal molecular mechanisms underlying switch behavior. (A) Ca^{2+} -free 2- μs MD trajectory of I89 (teal) showing the designed transition between a binding-competent state 1 (light blue) and binding-incompetent state 2 conformation (pink) (middle panels). The protein remains in state 1 in the presence of Ca^{2+} (bottom) and coordinates Ca^{2+} consistent with an EF hand motif (inset). α RMSD is measured for the reshaped helix compared to our computational models for states 1 (black line) and 2 (blue line) after aligning on the regions of the nonreshaped backbone with regular secondary structure. (B) Mutual informational analysis (heatmap) revealing a correlated network of residues (light blue) connecting the Ca^{2+} binding loop (green) with the allosteric mutation site (dark blue), shown in two orientations (top panels, reshaped helix circled by a dashed line). Middle panels show interaction details of a key step in which I69 becomes solvent exposed in state 2, enabling the formation of distinct contacts by Y43. (C) State-specific interactions for state 1 and state 2 (colored) observed during a 2- μs Ca^{2+} -free MD simulation for I89. (D) Frame2seq (F2S) predictions for Y64F and K68E; score refers to the negative log-likelihood difference between the mutated and original sequence, where the negative change in score predicts that the mutations disfavor a given state. (E) AF2 predictions showing higher pLDDT (greater confidence) for state 1 for mutants on the bottom compared to the original sequence on the top. (F) 2D ^1H , ^{15}N -HSQC NMR data showing chemical shift changes consistent with an increased state 1 population in the mutants [arrows denoting direction of shifts from primarily state 2 (S89) toward a larger population of state 1 (I89)].

data to accurately and rapidly predict conformational ensembles in silico, which remains a major challenge (27, 28), and using machine learning to parameterize new force fields that may allow for good estimates of energy landscapes without the resource demand of all-atom simulations (29, 30). Future development of deep learning integrated with unique dynamics data offered by experiment and simulation could allow for one-shot conformational ensemble prediction (31, 32). Such models could ultimately be applied to design entire user-defined conformational landscapes. It is encouraging that our approach can already successfully design motion involving intricate changes to the inter-residue interaction network between states, despite the individual models for sequence and structure prediction never having been explicitly trained for such a task.

Functional conformational changes in natural proteins are often complex, typically not well-understood, and hence difficult to study and manipulate using engineering approaches. By contrast, our design method both generates two-state conformational changes similar to those in nature, as well as successful predictions of how to modulate the populations of these states. The proof-of-concept application here is an exciting example of designing a de novo alternative state into a naturally occurring protein. This result suggests that our approach may be useful for controlling natural functions through tunable conformational switching, which would open many new avenues to study and manipulate the roles of natural switches in the cellular context. Moreover, our approach should be useful to generate entirely bio-orthogonal protein switches. The approach and concepts described here hence have broad relevance to both modulate and deconstruct the requirements for natural signaling and construct complex switch-like signaling systems entirely from the ground up.

Materials and methods

Residue numbering

Please note that all designs reported in this manuscript (including deposited structures/computational models, experimental data, and molecular dynamics simulation analyses) are indexed to include a four residue N-terminal thrombin cleavage scar (gray in Fig. 2A in the main manuscript). We adopted this residue numbering convention throughout (unless we are directly referring to the ISMG PDB structure) to be consistent with the design constructs characterized structurally. We note that the scar was not explicitly modeled during generation of alternative de novo states, single-state design of these states, or multi-state design, which used a scar-less sequence with numbering starting at one (see Readme note for deposited scripts).

Generating alternative backbone conformations

The starting Ca^{2+} binding-competent structure (PDB ID: 1SMG) was obtained from the Protein Data Bank. Conformer 1 of the NMR ensemble was relaxed in Rosetta with all heavy atoms in the backbone and sidechains restrained.

```
relax.linuxgccrelease -database $ROSETTA/main/database -ex1 -ex2
-use_input_sc -flip_HNQ -no_optH false -relax:constrain_relax_to_
start_coords -relax:coord_constrain_sidechains -relax:ramp_constraints
false -s 1SMG.pdb
```

The loop-helix-loop unit (“reshaped region”, residues 49 to 72) involving the higher affinity Ca^{2+} binding loop (residues 68 to 72 and 77) was reshaped with loop-helix-loop unit combinatorial sampling (19) using loop libraries of length two to five residues. The choice of loop lengths to sample were based on an analysis of loops commonly occurring in protein structures [for a detailed description, see (19)]. Note that these residue ranges follow ISMG residue numbering, i.e., these residues correspond to residues 53 to 76 for the reshaped region and 70 to 76 and 81 for Ca^{2+} binding site II in our designs and presented data.

Only models having the same length as the input were written out. Since the input structure was primarily helical, alanine was used as the clash check residue rather than valine. Any outputs that had a reshaped

helix orientation compared to the input ($\text{C}\alpha$ RMSD < 3 Å) were filtered out to generate alternative conformations that were experimentally distinguishable. Any outputs that had a reshaped helix $\text{C}\alpha$ RMSD > 10 Å compared to the input were also discarded, as these structures often lacked a well-packed hydrophobic core. This left $n = 921$ candidate alternative *apo* backbone conformations for further analysis.

Single-state sequence design

A residue i was defined as pointing toward the reshaped region if the cosine of the angle between the vector pointing from $\text{C}\alpha_i$ to $\text{C}\alpha_j$ and the vector pointing from $\text{C}\alpha_i$ to $\text{C}\beta_j$ was greater than 0.5, where j is any residue in the reshaped region. Residues in the reshaped region as well as all residues within 10 Å ($\text{C}\alpha$ - $\text{C}\alpha$ distance) from the reshaped region and pointing toward the reshaped region were designable, i.e., allowed to change amino acid identity and rotamer conformation. Residues within 8 Å ($\text{C}\alpha$ - $\text{C}\alpha$ distance) and pointing toward designable residues were repackable, i.e., constant amino acid identity but allowed to change rotamer conformation. Residues in the reshaped region that were highly conserved in EF hand motifs were held fixed during design (i.e., D66, D68, G69, S70, G71, T72, E77, following PDB: 1SMG numbering convention). Allowable amino acids at designable residue positions were determined by the extent of residue burial using the Rosetta LayerDesign task operation (33). Cysteine and histidine were disallowed at all positions to avoid disulfide bond formation and pH sensitivity issues. Rosetta FastDesign and RotamerTrial protocols using extra rotamers for χ_1 and χ_2 (enabled by the ex1 and ex2 options in the ExtraRotamersGeneric task operation) were applied during design. Relevant scripts and command line prompts can be found at https://github.com/Kortemme-Lab/local_protein_sequence_design. 20 designs per backbone were generated.

In silico validation of single-state designs

9-mer and 3-mer fragments for structure prediction were generated using the make_fragments.pl script distributed with Rosetta (34).

```
make_fragments.pl -verbose -id design_id -frag_sizes 3,9 -n_fra
200 -n_candidates 1000 sequence.fasta
```

A biased *ab initio* structure prediction simulation was run as a preliminary screen for the lowest scoring design (using the ref2015 Rosetta energy function (35)) of each backbone. This was done by using only the top three 9-mer/3-mer fragments closest in $\text{C}\alpha$ RMSD to the desired backbone and generating 30 decoys. If low-RMSD decoys did not have low Rosetta energies or the lowest energy decoy was >3 Å in $\text{C}\alpha$ RMSD to the desired backbone, the design was removed from further analysis, as it would be unlikely to fold into the desired structure in an unbiased simulation. Full Rosetta *ab initio* structure prediction simulations (20,000 decoys, 200 fragments at each sliding window) were run for the remaining designs.

```
AbinitioRelax.linuxgccrelease -abinitio:relax -use_filters true
-abinitio::increase_cycles 10 -abinitio::rg_reweight 0.5 -abinitio::rsd_
wt_helix 0.5 -abinitio::rsd_wt_loop 0.5 -relax::fast -in:file:fasta
sequence.fasta -in:file:frag3 <fragments_3mer_file> -in:file:frag9
<fragments_9mer_file> -psipred_ss2 <ss2_file_from_frag_generation>
-nstruct <num_output> -out:sf <score_file_output> -out:file:silent
<silent_file_output>
```

We selected 11 designs for experimental characterization (Fig. S4, yeast display), which had good agreement between the design model in the reshaped region and either the lowest energy Rosetta *ab initio* decoy structure or the top-ranking (by pLDDT) AF2 structure ($\text{C}\alpha$ RMSD < 2 Å). All designs selected had high RMSD to the original *holo* state in the reshaped region ($\text{C}\alpha$ RMSD > 4.5 Å).

Yeast display screen for stable single state designs

Gene fragments of selected designs flanked by homology regions (5': ggagggtcggttcgcatatg, 3': ctcgagggtggaggtccgaacaacagcttattct-gaagaggacttgta) were ordered from Integrated DNA Technologies

(table S8). Synthetic DNA (10 ng) was amplified in a 25- μ l reaction using Q5 Hot Start High-Fidelity DNA Polymerase (NEB #M0494) for 30 cycles by PCR. A strong single band at the expected size was observed for all fragments. The reaction product was cleaned using Zymo Research DNA Clean & Concentrator kits for transformation of EBY100 yeast using the protocol described in (36). We used a modified version of the pETcon vector known as pETcon v3 (RRID: Addgene_41522) that had been altered to (i) remove a long single-nucleotide stretch near the cloning region, (ii) add an N-terminal 6x-His tag to normalize intact protein level to surface display level, and (iii) include constitutively fluorescent protein Venus as the dropout sequence to aid in colony selection. Single colonies were picked, inoculated into 5 ml of SD-CAA medium, and incubated at 30°C overnight for each design. The starter culture was diluted 1:10 into SG-CAA to induce protein production. A volume corresponding to 5×10^5 cells was added to a 96-well plate and spun at 3000g for 20 min to pellet the cells. Cells were washed and resuspended in phosphate-buffered saline (PBS) then incubated with anti-c-Myc Mouse mAb (Alexa Fluor 647 conjugate) (Cell Signaling Technology #2233) and 6x-His Tag Antibody Alexa Fluor 488 conjugate (Fisher Scientific #MA1-135-A488) shaking at room temperature for 30 min. Excess unbound antibody was washed away with chilled PBS + 1% bovine serum albumin (BSA) and analyzed using a Beckman Coulter Cytoflex flow cytometer. Events were gated by forward scattering area and back scattering area to collect the live cell population then by forward scattering width and forward scattering height to select individual cells for further analysis. Cells were then gated by fluorescence intensity where the threshold separating displaying from non-displayed cells was set such that 1% of an uninduced, unstained control would pass the gate. The fraction of anti-c-Myc positive cells from single cell events was used to estimate the level of expressed intact protein level on the surface of yeast as a proxy for thermostability (37).

Protein expression and purification for experimental screening

Plasmids (pET-28a(+)) encoding the selected designed proteins were ordered from Twist Bioscience. These constructs included an N-terminal 6x-His tag and thrombin cleavage site (MGSSHHHHHGLVPRGSHM). *Escherichia coli* BL21(DE3) cells were transformed with these plasmids. Colonies were inoculated into 5 ml LB medium and cultured at 37°C overnight. Starter cultures were diluted 1:100 into 1 liter of fresh LB medium and cultured at 37°C until the OD₆₀₀ reached 0.6–0.8. Then isopropyl- β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 300 μ M to induce protein expression at 37°C overnight. Cell cultures were centrifuged at 8000g for 10 min to spin down the cells. Cell pellets were then lysed by resuspending in 4 ml/g pellet of B-PER (ThermoFisher #78243) with a dissolved cOmplete protease inhibitor cocktail EDTA-free tablet (Roche #COEDTAF-RO) and incubating at room temperature for 15 min. Cell lysate was centrifuged at 18000g for 30 min to separate the soluble and insoluble fractions. The soluble fraction was mixed with an equal volume of equilibration buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 mM imidazole) and 1 ml HisPur Ni-NTA (nitrilotriacetic acid) resin slurry volume (Thermo Scientific #88222) to pull down the His-tagged proteins. Ni-NTA resin beads were washed three times with wash buffer (50 mM Tris pH 7.5, 300 mM NaCl, 25 mM imidazole). For screening assays (e.g., size exclusion chromatography, dynamic light scattering, circular dichroism spectroscopy), the protein was then eluted three times with 0.5 ml elution buffer (50 mM Tris pH 7.5, 300 mM NaCl, 250 mM imidazole). Purity and amount of designed protein was estimated by Coomassie-stained SDS-polyacrylamide gel electrophoresis (PAGE) throughout purification (BIO-RAD #456-1095).

Circular dichroism spectroscopy

Circular dichroism (CD) data were collected on a Jasco J-710 spectrometer. Purified designs were diluted to a final salt concentration of 10 mM

KCl (pH 6.7). The concentrations of diluted samples were approximately 2 μ M determined by a Bradford assay. CD spectra were measured using a 1-mm cuvette at 25°C. Melting curves were recorded at 208 nm from 25° to 95°C using a rate of 1°C/min (fig. S21).

Analytical size exclusion chromatography

Protein samples purified by His-tag pull down were analyzed using a Superdex 75 10/300 GL (Cytiva #29148721) (switch designs 189, N89, R89, and S89) or the Superdex 200 10/300 GL (Cytiva #28990944) (single-state designs #6306 and wild-type 1SMG) size exclusion column with 100 mM KCl (pH 6.7) (figs. S21 and S22).

Dynamic light scattering

Size measurements of protein samples purified by His-tag pull down using dynamic light scattering were measured on a Zetasizer Nano S90 (fig. S22). Samples were filtered (0.22 μ m) prior to measurement to remove large particles. Number distribution of hydrodynamic diameter was plotted to provide an estimate of particle size.

Protein expression and purification for NMR structure calculation

[U-¹³C, U-¹⁵N]-labeled proteins were expressed by inoculating *E. coli* colonies into 5 ml M9 minimal medium that included 4 g/liter [U-¹³C]-glucose (99%) and 0.5 g/liter [U-¹⁵N]-NH₄Cl (99%), which were grown at 37°C overnight. Starter cultures were then diluted 1:50 into 500-ml fresh [U-¹³C, U-¹⁵N]-labeled M9 minimal medium and grown at 37°C until the OD₆₀₀ reached 0.6 to 0.8. Then IPTG was added to a final concentration of 300 μ M to induce protein expression at 37°C overnight. The expressed proteins were purified by following the Ni-NTA resin pull down protocol described in the protein purification section with the following modifications: Rather than eluting the His-tagged protein from the beads, the beads were washed with and resuspended in thrombin cleavage buffer (20 mM Tris-HCl, 150 mM NaCl, pH 8.4). Four units of biotinylated thrombin (Novagen #696022) per 1 mg of protein was added to the beads, and the immobilized His-tagged protein was cleaved from the beads while rocking at room temperature for 20 hours. To improve purity of the cleaved protein, the beads were pelleted, and the supernatant was added to fresh Ni-NTA resin. The beads were pelleted once more, and the supernatant was mixed with 32 μ l of streptavidin agarose slurry (Novagen #696022) per U of thrombin for 30 min at room temperature. The sample was then filtered (0.45 μ m) and incubated with 100 μ M EDTA for 5 min at room temperature to remove residual amounts of Ca²⁺. The sample was then buffer exchanged into 100 mM KCl pH 6.7 by dialysis overnight at 4°C and concentrated to approximately 180 μ l for NMR experiments. Ca²⁺ was then added (if applicable) to the desired concentration.

Structure determination by NMR

D₂O (5%) was added to samples. The final protein concentrations determined by Bradford assay were ~400 μ M. NMR spectra were all measured at 298.1 K. Two-dimensional (2D) ¹H, ¹⁵N-HSQC (pulse program: fhscqf3gp-ph), 2D ¹H, ¹³C-HSQC (pulse program: hsqcetgpsisp2), 36-ms 3D HCCH-TOCSY (pulse program: hccchdigp3d) and 120-ms 3D simultaneous ¹³C/¹⁵N-NOESY-HSQC (pulse program: noesyhsqcgpsismsp3d) spectra were measured using a Bruker NEO 800 MHz spectrometer with a 5-mm TCI H&F-C/N-D CryoProbe. 3D CACB(CO)NH (pulse program: hncocacbgp-wg3d), 3D CACBNH (pulse program: hncacbgp-wg3d), 3D Hcc(co)NH (pulse program: hccconhgp-wg3d2), 3D Cc(co)NH (pulse program: hccconhgp-wg3d3), 2D HBCBCGCDHDGP (pulse program: hbcbcgchdgp), and 2D HBCBCGCDCEHEGP (pulse program: hbcbcgchdhegp) were collected on a Bruker Avance 600 MHz spectrometer with an Inverse 5 mm H-C/N-D cryoprobe. Spectra were processed in TopSpin 3.6.3. For the Ca²⁺-bound structure, we also collected the following experiments on the Bruker Avance 600 MHz spectrometer: 2D TROSY (troysargp-phw-g) and 3D ¹H-¹³C NOESY-TROSY (noesytroysargp-phw-g) to aid in the assignment of aromatic nOEs.

Automated peak picking, resonance assignment, and structure calculations were performed by ARTINA (38), and dihedral angle restraints were generated by TALOS (39). Manual inspection of ARTINA chemical shift assignments was done in CCPNMR v. 3.1.0 and corrections were made as necessary (40). The manually curated chemical shift list was then included as input for a subsequent structure calculation run in ARTINA. The output candidate structure (an ensemble of $n = 20$ conformers) with the lowest CYANA target function value was refined in XPLOR-NIH-3.7 with the refine.py script included in the distribution (eginput/gb1_rdc/refine.py) (41). The 20 lowest-scoring structures out of 100 were then refined in explicit water with the wrefine.py script (eginput/gb1_rdc/wrefine.py). The ensemble of the refined structures was validated using the PDB validation server (42). For the *apo* I89 structures (i.e., proposals #1 and #2 from ARTINA), we recalculated the structures in CYANA after excluding a small subset of distance restraints that could be unambiguously assigned to the other state prior to refinement. For the Ca^{2+} -bound I89 structure, the structure was first calculated with no Ca^{2+} restraints using ARTINA as described above. The Ca^{2+} restraints were then included during refinement in XPLOR-NIH-3.7, which were set to 2.8 Å between the Ca^{2+} ion and the oxygens involved in coordinating Ca^{2+} (43). The structural statistics from the PDB validation server can be found in table S9.

Two-state sequence design

Initially, we attempted to use Rosetta-based methods for multi-state design. We first tried a restrained convergence algorithm, which allowed each state to explore sequence space independently and encouraged convergence to a single sequence by incrementally increasing an energy bonus if the same amino acid was sampled at corresponding positions in both states (44). This method did not appear to work well for states varying in conformation, as it was unlikely for positions that differed greatly in solvent accessible surface area or secondary structure to converge on the same amino acid, even with an energy bonus. We also attempted to use a genetic algorithm that allowed users to input a custom fitness function as input (45). However, this method performed Monte Carlo optimization of the Rosetta energy function over discrete rotamer space and was consequently computationally expensive compared to rotamer-free design methods. Due to what appeared to be low in silico success rates, we decided to use ProteinMPNN, a deep learning-based design method capable of quickly generating tens of thousands of multi-state designs (9). This method has since been shown to successfully generate proteins where domains can be hinged open and closed upon binding a peptide effector (16).

In general, a PDB file containing both the single-state alternative state 2 conformation and the natural Ca^{2+} binding protein state 1 conformation separated by a distance much larger than the approximate diameter of either state was generated by PyMOL. State 1 was conformer 1 of the PDB ID: 1SMG NMR ensemble relaxed using Rosetta with restraints on heavy-atom positions. Ca^{2+} was removed from the structure, as this implementation of ProteinMPNN did not support ligand atoms.

The single-state 2 input was generated with the following design task in mind: beyond individual bistable designs, we wanted to design a family of sequences that had different state population distributions in the absence of ligand despite having high sequence identity. We reasoned it would be more likely to design such a set of sequences if we first restricted the allowable sequence space during design. Initially, the set of potential designable residues for two-state design included the reshaped region and its neighbors in both states as defined for single-state design ($n = 37$ residues). To exclude positions that did not change in environment significantly between states or were otherwise not critical in determining the preferred structure, we individually mutated each designable position in the state 2 design to the corresponding state 1 amino acid and predicted the structural impact of each “reversion” using AF2 (though one could also use an in silico deep

mutational scanning approach to evaluate other potential amino acids at each position). If the predicted structure was still consistent with the state 2 conformation ($\text{C}\alpha$ RMSD < 1.5 Å), this mutation was considered to be in the “tolerated” sequence space of state 1. We then input a state 2 sequence containing all “tolerated” amino acid reversions into AlphaFold2 and used the best-ranked structure by pLDDT as the state 2 state during two-state design. Of note, this prediction was still highly similar to the original single-state design in structure ($\text{C}\alpha$ RMSD = 1.15 Å in the reshaped helix (residues 59-69) and 1.21 Å over the entire backbone) and had an average pLDDT of 84.6 in the reshaped helix. The final set of designable residues included (i) all positions that were not reverted to the corresponding state 1 amino acid through this process and (ii) their neighbors (defined as in single-state design) ($n = 25$ residues) (table S2).

All corresponding positions between states were tied together across chains during design with equal weighting of each state. During preliminary rounds of design, we observed the last turn of the reshaped helix (residues 68 to 72) fraying at the C-terminal region in some of the AF2 structure predictions. To generate designs more closely matching our desired target state, we performed a sequence analysis of the designs closest in RMSD to our input structure for state 2 and found that there was a strong preference for positively charged residues and those capable of hydrogen bonding at position 62 (i.e., K, R, Q), which could then interact with D66. In subsequent rounds of two-state design, we biased this position toward those amino acid types. This kind of iterative in silico refinement of the design process to increase sequence and structure compatibility is common and should be generalizable across input structures. ProteinMPNN was used to generate 10^4 sequences, which were input without MSA generation into ColabFold for structure prediction (46). Designs selected for experimental characterization can be found in table S10.

NMR experiments to characterize dynamics

[^{15}N]-labeled protein was prepared as described above for structure calculation except unlabeled glucose was used in place of [^{13}C]-glucose (99%) in culture media. All experiments were done on the Bruker Avance NEO 800 MHz spectrometer with the cryoprobe specifications mentioned previously. The concentration of samples ranged from 100 to 200 μM in 100 mM KCl, pH 6.7, 5% v/v D_2O . An initial screen for Ca^{2+} modulation of conformational dynamics was done by collecting 2D ^1H , ^{15}N -HSQC spectra of all two-state designs with and without 10 equiv. of Ca^{2+} (pulse program: flhsqcf3gpph). To further probe which residues may be dynamic in the absence of Ca^{2+} , a 2D ^1H , ^{15}N -HSQC temperature series from 5° to 35°C in 5°C increments was collected for design I89. To characterize motions occurring on a micro-to-millisecond timescale, we acquired ^{15}N rotating frame ($R_{1\rho}$) relaxation dispersion experiments using spin-lock field strengths (ω_{SL}) of 1.0 kHz, 1.25 kHz, 2.0 kHz, 2.5 kHz, and 3.0 kHz (pulse program: hsqctretf3gpsitc3d) at $T = 288.1\text{K}$ and 298.1K with delay times of 2, 4, 6, 8, 10, 20, 30, 50, 70, and 100 ms, where the 30-ms experiment was repeated to assess reproducibility and estimate error (47, 48). $R_{1\rho}$ rates were calculated using the Bruker Dynamics Center 2.8.4. We additionally measured longitudinal relaxation (R_1) rates in the laboratory frame at both temperatures using delay times of 20, 40, 80, 200, 500, 900, and 1400 ms (pulse program: hsqc1etf3gpsitc3d) (49) and then converted the $R_{1\rho}$ rates to observed transverse relaxation (R_2) rates, accounting for differences in the site-specific R_1 rates and tilt angles. Heterogeneity in the temperature dependence of R_2 suggested there may be multiple exchange processes occurring on the micro-to-millisecond timescale (fig. S12A). To further probe this possibility, we acquired a two-point ^{15}N CPMG experiment (pulse program: hsqctretx-etc3gpsitc3d) at $\nu_{\text{CPMG}} = 2000$ Hz and 25 Hz at $T = 298.1\text{K}$ (fig. S12B, upper panel) (50). These data again indicated there were at least two exchange processes occurring at distinct timescales: a slower process detectable by ^{15}N CPMG in helix B/site I and a faster process detectable

by ^{15}N $R_{1\rho}$ at high ω_{SL} in the reshaped region (see fig. S12 for details). The slower process may have originated as a feature of our single-state design for state 2, as we observed lower peak intensity in that region consistent with line broadening due to exchange (fig. S12B, lower panel).

As detailed in Fig. S13, we fit the ^{15}N $R_{1\rho}$ relaxation dispersion data globally across all residues to a simple model of the sum of two independent exchange processes occurring at separate timescales in the fast-exchange limit (57) as follows:

$$R_2 = \Phi_{\text{ex,fast}} \tau_{\text{ex,fast}} / (1 + \tau_{\text{ex,fast}} \omega_{\text{eff}}) + \Phi_{\text{ex,slow}} \tau_{\text{ex,slow}} / (1 + \tau_{\text{ex,slow}} \omega_{\text{eff}}) + R_{2,0} \quad (1)$$

where $\Phi_{\text{ex},i}$ is the product of the state populations and chemical shift difference for a given spin (i.e., $p_{A,i} p_{B,i} \Delta\omega_i^2$) and $R_{2,0}$ is the intrinsic transverse relaxation rate in the absence of exchange. The two exchange timescale (τ_{ex}) values were each sampled on a 2D grid with values ranging from 42 to 199 μs (1.25-fold lower and higher, respectively, than the lowest and highest measured $1/\omega_{\text{eff}}$ values). Given each pair of fixed τ_{ex} values, the parameters $\Phi_{\text{ex,fast}}$, $\Phi_{\text{ex,slow}}$ and $R_{2,0}$ were determined using linear least-squares fitting for all residues. Φ_{ex} values were constrained to be greater than or equal to zero by removing the respective term from the fit if it produced a negative nonphysical value.

Approximation of state populations using chemical shifts

To approximate the chemical shifts of pure state 1 or state 2, we used the chemical shifts of switch design I89 in the presence of Ca^{2+} and design S89 without Ca^{2+} , respectively, and assumed a 5% minor state population for each. Peaks that were well-resolved in the 2D ^1H , ^{15}N -HSQC spectrum and corresponded to residues facing or in the reshaped region but distal from position 89 and Ca^{2+} site II were selected for analysis. To estimate state populations, we assumed that each peak position was a population-weighted average of the two endpoints.

Determining Ca^{2+} binding affinity

The expression and purification of $[\text{U-}^{15}\text{N}]$ -labeled protein were as described previously to yield ~ 1 mM protein. Stock solutions of 10 mM, 100 mM, and 1 M CaCl_2 in 100 mM KCl, 95% $\text{H}_2\text{O}/5\%$ D_2O were prepared and adjusted to pH 6.7 with HCl. For each titration point, a volume of stock solution was added such that the $[\text{Ca}^{2+}]/[\text{L}]$ ratio approximately doubled, ranging from $[\text{Ca}^{2+}] = 0.025$ to 51.2 mM. The titration was monitored by collecting a series of 2D ^1H , ^{15}N -HSQC spectra and tracking the chemical shift of residues in Ca^{2+} binding site II that were well-resolved at all titration points (V71, D78). We fit the data to the following equation to calculate the apparent dissociation constant (K_d) for each residue:

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \left\{ \left(\frac{n[\text{P}] + [\text{L}] + K_d}{n[\text{P}] + [\text{L}] + K_d} \right) - \left[\frac{n[\text{P}] + [\text{L}] + K_d}{n[\text{P}] + [\text{L}] + K_d} - 4n[\text{P}][\text{L}] \right]^{1/2} \right\} / 2n[\text{P}] \quad (2)$$

where $\Delta\delta_{\text{obs}}$ is the change in the observed shift from the Ca^{2+} -free state, $\Delta\delta_{\text{max}}$ is the maximum shift change on saturation (treated as a free variable during fitting), $[\text{P}]$ and $[\text{L}]$ are the total protein and ligand concentrations, respectively, and n is the number of sites with approximately (within an order of magnitude) similar K_d values. To account for binding to site I, which is of comparable affinity, we set n equal to 2, as described previously (52).

Molecular dynamics simulations

Simulations of designs were performed using GROMACS 2022.5 (53). The initial structures used for the triplicate 2- μs simulations of I89 and S89 in the presence/absence of Ca^{2+} were the AMBER-relaxed AF2 predictions for the design with the N-terminal thrombin cleavage site scar included to be consistent with experimentally characterized

designs. For simulations with Ca^{2+} , we initiated the trajectory with the Ca^{2+} ion in site II. To more extensively sample the conformational space of design I89, we collected seven additional 1- μs -long trajectories starting from the AMBER-relaxed AF2 prediction. Furthermore, we sampled design I89 initiated from different starting conformations using snapshots from previously collected trajectories that had (i) low $\text{C}\alpha$ RMSD of the reshaped helix to state 2 to initiate 10 additional 1- μs trajectories (three of which were extended to 2 μs) and (ii) intermediate $\text{C}\alpha$ RMSD of the reshaped helix to both states 1 and 2 to initiate 10 additional 1- μs trajectories from an intermediate conformation (frames designated by triangles in fig. S18). The proteins were parameterized using the a99SB-disp force field and water molecules were parameterized using the a99SB-disp water model (54). The systems were solvated and neutralized as in table S11. Energy minimization of the system was performed with the steepest descent minimization algorithm to a tolerance of 1000.0 $\text{kJ mol}^{-1} \text{nm}^{-1}$. Equilibration was performed in the NVT ensemble for 1000 ps at 300 K using the Berendsen thermostat. Systems were then equilibrated in the NPT ensemble for 100 ps at a target pressure of 1 bar at 300 K maintained by the Berendsen thermostat with position restraints on all heavy atoms. Bond lengths and angles of protein atoms were constrained with the LINCS algorithm and water constraints were applied using the SETTLE algorithm. The PME algorithm was used for electrostatics with a grid spacing of 1.8 nm. Van der Waals forces were calculated with the Verlet cut-off scheme using a 1.2-nm cut-off distance. Snapshots were saved every 80 ps. Initial structures were AlphaFold2 structure predictions of two-state designs (including the residues GSHM at the N terminus corresponding to a scar from the thrombin cleavage site) relaxed in AMBER. The $\text{C}\alpha$ RMSD of the reshaped region (residues 53-76) were calculated over the course of a 1- μs simulation using GROMACS 2022.5 command line tools.

Mutual information analysis

To quantify correlations between residue dihedral angles during the *apo* I89 simulation, the first and third replicates (in which a state 1 to state 2 transition was observed) were split into 200ns blocks, discarding the first 100 ns. For each block, dihedral angle data was extracted with:

```
gmx chi -f md.xtc -s md_1_us.tpr -phi -psi -omega -rama -all -maxchi 4 -HChi -b 5000
```

We then calculated the mutual information matrix using the MutInf method (23): `dihedral_mutent -x <base dir> -d / -o 6 -n 7 -w 30 -p 0 -c "yes" -a "yes" I89.reslist > I89_mutinf.out` where the *.reslist file was a list of all residues. The file containing mutual information between pairs of residues, with zero diagonal (*bootstrap_avg_mutinf_res_sum_0diag.txt) was visualized as a heatmap.

Markov state modeling

For each simulation frame, we first aligned the $\text{C}\alpha$ atoms of the non-reshaped region with regular secondary structure (residues 7 to 17, 20 to 33, 40 to 53, 79 to 90) to our computational models for state 1 or state 2 and calculated the $\text{C}\alpha$ RMSD of the reshaped helix (residues 59 to 69) to each reference state. We then used mini-batch k-means clustering on these two RMSD values, assessing models fit to 2, 5, and 10 clusters using all aggregate simulation data from I89 simulations in the absence of Ca^{2+} (fig. S18). Markov state models (MSMs) were then constructed for each number of total states using a range of lag times. For each model, mean first passage times (MFPTs) were computed between the clusters most representative of our designed state 1 and 2. MFPT values appeared to converge as the number of states increased where the absolute difference between MFPTs computed from a 5-state versus 10-state MSM model was less than 10 μs for transitions from state 2 to state 1 and less than 0.5 μs from state 1 to state 2 (fig. S20A). As both cluster sizes provided similar results, we chose five states for further analysis. We did not sample lag times greater than 300 ns as that would result in discarding over one-third of our simulation data. The implied timescales for the four slowest

modes of the five-state MSM were then plotted (fig. S20B), and a lag time of 200 ns was chosen for further analysis as the timescales begin to plateau at this point, suggesting that this MSM is a faithful representation of the underlying dynamics (55). Clusters, MSMs, and MFPTs were calculated using MSMBuilder v3.8 (56). MFPTs from the MSM were converted to implied radian units (i.e., divided by a factor of 2π) for direct comparison to NMR experimental data in the main text. The complete list of MFPTs (as output by MSMBuilder without implied radian units) and stationary probabilities can be found in table S12. Alignment, RMSD calculations, and each cluster medoid structure file were performed and generated using MDAnalysis v2.8 (57).

Frame2seq scoring of point mutants

To predict the effects of single point mutations on the conformational equilibrium of switch design I89, we computed sequence likelihoods given structure using Frame2seq (10). Frame2seq models learn to approximate $P(\text{sequence}|\text{structure})$ via a masked language modeling objective. Pseudo log-likelihood (PLL) has been explored for scoring sequences (58), which we adapted for a structure-conditioned ranking task. For each conformational state, we separately output Frame2seq model PLL by providing the structure and the sequence as input with a mask introduced at the mutated position. We then subtracted the single point mutant negative PLL from the wild-type negative PLL (the reference point) to compute the score as follows:

$$\text{score} = \log(p(x_i = x_i^{\text{wt}} | x_i^{\text{wt}}, Y)) - \log(p(x_i = x_i^{\text{mt}} | x_i^{\text{mt}}, Y)) \quad (3)$$

where i is the mutated position, x^{wt} is the wild-type sequence, x^{mt} is the mutant sequence, x_i is the sequence x with a mask introduced at position i , and Y is the structure.

REFERENCES AND NOTES

- D. D. Boehr, R. Nussinov, P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009). doi: [10.1038/nchembio.232](https://doi.org/10.1038/nchembio.232); pmid: [19841628](https://pubmed.ncbi.nlm.nih.gov/19841628/)
- K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007). doi: [10.1038/nature06522](https://doi.org/10.1038/nature06522); pmid: [18075575](https://pubmed.ncbi.nlm.nih.gov/18075575/)
- K. M. Chen, D. Keri, P. Barth, Computational design of G Protein-Coupled Receptor allosteric signal transductions. *Nat. Chem. Biol.* **16**, 77–86 (2020). doi: [10.1038/s41589-019-0407-2](https://doi.org/10.1038/s41589-019-0407-2); pmid: [31792443](https://pubmed.ncbi.nlm.nih.gov/31792443/)
- T. Perica *et al.*, Systems-level effects of allosteric perturbations to a model molecular switch. *Nature* **599**, 152–157 (2021). doi: [10.1038/s41586-021-03982-6](https://doi.org/10.1038/s41586-021-03982-6); pmid: [34646016](https://pubmed.ncbi.nlm.nih.gov/34646016/)
- M. Huse, J. Kuriyan, The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002). doi: [10.1016/S0092-8674\(02\)00741-9](https://doi.org/10.1016/S0092-8674(02)00741-9); pmid: [12015977](https://pubmed.ncbi.nlm.nih.gov/12015977/)
- A. S. Hauser *et al.*, GPCR activation mechanisms across classes and macro/microscales. *Nat. Struct. Mol. Biol.* **28**, 879–888 (2021). doi: [10.1038/s41594-021-00674-7](https://doi.org/10.1038/s41594-021-00674-7); pmid: [34759375](https://pubmed.ncbi.nlm.nih.gov/34759375/)
- T. Kortemme, De novo protein design-From new structures to programmable functions. *Cell* **187**, 526–544 (2024). doi: [10.1016/j.cell.2023.12.028](https://doi.org/10.1016/j.cell.2023.12.028); pmid: [38306980](https://pubmed.ncbi.nlm.nih.gov/38306980/)
- J. A. Davey, R. A. Chica, Multistate approaches in computational protein design. *Protein Sci.* **21**, 1241–1252 (2012). doi: [10.1002/pro.2128](https://doi.org/10.1002/pro.2128); pmid: [22811394](https://pubmed.ncbi.nlm.nih.gov/22811394/)
- J. Dauparas *et al.*, Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022). doi: [10.1126/science.add2187](https://doi.org/10.1126/science.add2187); pmid: [36108050](https://pubmed.ncbi.nlm.nih.gov/36108050/)
- D. Akpinaroglu *et al.*, Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space. *bioRxiv* 2023.12.15.571823 [Preprint] (2023). <https://doi.org/10.1101/2023.12.15.571823>
- R. G. Alberstein, A. B. Guo, T. Kortemme, Design principles of protein switches. *Curr. Opin. Struct. Biol.* **72**, 71–78 (2022). doi: [10.1016/j.sbi.2021.08.004](https://doi.org/10.1016/j.sbi.2021.08.004); pmid: [34537489](https://pubmed.ncbi.nlm.nih.gov/34537489/)
- X. I. Ambroggio, B. Kuhlman, Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154–1161 (2006). doi: [10.1021/ja054718w](https://doi.org/10.1021/ja054718w); pmid: [16433531](https://pubmed.ncbi.nlm.nih.gov/16433531/)
- B. Ruan *et al.*, Design and characterization of a protein fold switching network. *Nat. Commun.* **14**, 431 (2023). doi: [10.1038/s41467-023-36065-3](https://doi.org/10.1038/s41467-023-36065-3); pmid: [36702827](https://pubmed.ncbi.nlm.nih.gov/36702827/)
- J. A. Davey, A. M. Damry, N. K. Goto, R. A. Chica, Rational design of proteins that exchange on functional timescales. *Nat. Chem. Biol.* **13**, 1280–1285 (2017). doi: [10.1038/nchembio.2503](https://doi.org/10.1038/nchembio.2503); pmid: [29058725](https://pubmed.ncbi.nlm.nih.gov/29058725/)
- J. A. Cross *et al.*, A de novo designed coiled coil-based switch regulates the microtubule motor kinesin-1. *Nat. Chem. Biol.* **20**, 916–923 (2024). doi: [10.1038/s41589-024-01640-2](https://doi.org/10.1038/s41589-024-01640-2); pmid: [38849529](https://pubmed.ncbi.nlm.nih.gov/38849529/)
- F. Praetorius *et al.*, Design of stimulus-responsive two-state hinge proteins. *Science* **381**, 754–760 (2023). doi: [10.1126/science.adg7731](https://doi.org/10.1126/science.adg7731); pmid: [37590357](https://pubmed.ncbi.nlm.nih.gov/37590357/)
- R. A. Langan *et al.*, De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019). doi: [10.1038/s41586-019-1432-8](https://doi.org/10.1038/s41586-019-1432-8); pmid: [31341284](https://pubmed.ncbi.nlm.nih.gov/31341284/)
- N. M. Goodey, S. J. Benkovic, Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4**, 474–482 (2008). doi: [10.1038/nchembio.98](https://doi.org/10.1038/nchembio.98); pmid: [18641628](https://pubmed.ncbi.nlm.nih.gov/18641628/)
- X. Pan *et al.*, Expanding the space of protein geometries by computational design of de novo fold families. *Science* **369**, 1132–1136 (2020). doi: [10.1126/science.abc0881](https://doi.org/10.1126/science.abc0881); pmid: [32855341](https://pubmed.ncbi.nlm.nih.gov/32855341/)
- S. M. Gagné, M. X. Li, B. D. Sykes, Mechanism of direct coupling between binding and induced structural change in regulatory calcium binding proteins. *Biochemistry* **36**, 4386–4392 (1997). doi: [10.1021/bi963076+](https://doi.org/10.1021/bi963076+); pmid: [9109645](https://pubmed.ncbi.nlm.nih.gov/9109645/)
- R. Krishna *et al.*, Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, ead12528 (2024). doi: [10.1126/science.adl2528](https://doi.org/10.1126/science.adl2528); pmid: [38452047](https://pubmed.ncbi.nlm.nih.gov/38452047/)
- J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2); pmid: [34265844](https://pubmed.ncbi.nlm.nih.gov/34265844/)
- A. J. Faure *et al.*, Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022). doi: [10.1038/s41586-022-04586-4](https://doi.org/10.1038/s41586-022-04586-4); pmid: [35388192](https://pubmed.ncbi.nlm.nih.gov/35388192/)
- C. J. P. Mathy, T. Kortemme, Emerging maps of allosteric regulation in cellular networks. *Curr. Opin. Struct. Biol.* **80**, 102602 (2023). doi: [10.1016/j.sbi.2023.102602](https://doi.org/10.1016/j.sbi.2023.102602); pmid: [37150039](https://pubmed.ncbi.nlm.nih.gov/37150039/)
- D. Kern, E. R. Zuiderweg, The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757 (2003). doi: [10.1016/j.sbi.2003.10.008](https://doi.org/10.1016/j.sbi.2003.10.008); pmid: [14675554](https://pubmed.ncbi.nlm.nih.gov/14675554/)
- C. L. McClelland, G. Friedland, D. L. Mobley, H. Amirkhani, M. P. Jacobson, Quantifying correlations between allosteric sites in thermodynamic ensembles. *J. Chem. Theory Comput.* **5**, 2486–2502 (2009). doi: [10.1021/ct9001812](https://doi.org/10.1021/ct9001812); pmid: [20161451](https://pubmed.ncbi.nlm.nih.gov/20161451/)
- F. Noé, S. Olsson, J. Köhler, H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019). doi: [10.1126/science.aaw1147](https://doi.org/10.1126/science.aaw1147); pmid: [31488660](https://pubmed.ncbi.nlm.nih.gov/31488660/)
- S. Mansoor, M. Baek, H. Park, G. R. Lee, D. Baker, Protein Ensemble Generation Through Variational Autoencoder Latent Space Sampling. *J. Chem. Theory Comput.* **20**, 2689–2695 (2024). doi: [10.1021/acs.jctc.3c01057](https://doi.org/10.1021/acs.jctc.3c01057); pmid: [38547871](https://pubmed.ncbi.nlm.nih.gov/38547871/)
- A. E. P. Durumeric *et al.*, Machine learned coarse-grained protein force-fields: Are we there yet? *Curr. Opin. Struct. Biol.* **79**, 102533 (2023). doi: [10.1016/j.sbi.2023.102533](https://doi.org/10.1016/j.sbi.2023.102533); pmid: [36731338](https://pubmed.ncbi.nlm.nih.gov/36731338/)
- M. Majewski *et al.*, Machine learning coarse-grained potentials of protein thermodynamics. *Nat. Commun.* **14**, 5739 (2023). doi: [10.1038/s41467-023-41343-1](https://doi.org/10.1038/s41467-023-41343-1); pmid: [37714883](https://pubmed.ncbi.nlm.nih.gov/37714883/)
- G. Janson, G. Valdes-Garcia, L. Heo, M. Feig, Direct generation of protein conformational ensembles via machine learning. *Nat. Commun.* **14**, 774 (2023). doi: [10.1038/s41467-023-36443-x](https://doi.org/10.1038/s41467-023-36443-x); pmid: [36774359](https://pubmed.ncbi.nlm.nih.gov/36774359/)
- S. Zheng *et al.*, Predicting equilibrium distributions for molecular systems with deep learning. *Nat. Mach. Intell.* **6**, 558–567 (2024). doi: [10.1038/s42256-024-00837-3](https://doi.org/10.1038/s42256-024-00837-3)
- S. J. Fleishman *et al.*, RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLOS ONE* **6**, e20161 (2011). doi: [10.1371/journal.pone.0020161](https://doi.org/10.1371/journal.pone.0020161); pmid: [21731610](https://pubmed.ncbi.nlm.nih.gov/21731610/)
- D. Gront, D. W. Kulp, R. M. Vernon, C. E. M. Strauss, D. Baker, Generalized fragment picking in Rosetta: Design, protocols and applications. *PLOS ONE* **6**, e23294 (2011). doi: [10.1371/journal.pone.0023294](https://doi.org/10.1371/journal.pone.0023294); pmid: [21887241](https://pubmed.ncbi.nlm.nih.gov/21887241/)
- R. F. Alford *et al.*, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017). doi: [10.1021/acs.jctc.7b00125](https://doi.org/10.1021/acs.jctc.7b00125); pmid: [28430426](https://pubmed.ncbi.nlm.nih.gov/28430426/)
- R. D. Gietz, R. H. Schiestl, High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007). doi: [10.1038/nprot.2007.13](https://doi.org/10.1038/nprot.2007.13); pmid: [17401334](https://pubmed.ncbi.nlm.nih.gov/17401334/)
- E. V. Shusta, M. C. Kieke, E. Parke, D. M. Kranz, K. D. Wittrup, Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J. Mol. Biol.* **292**, 949–956 (1999). doi: [10.1006/jmbi.1999.3130](https://doi.org/10.1006/jmbi.1999.3130); pmid: [10512694](https://pubmed.ncbi.nlm.nih.gov/10512694/)
- P. Klukowski, R. Riek, P. Güntert, Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA. *Nat. Commun.* **13**, 6151 (2022). doi: [10.1038/s41467-022-33879-5](https://doi.org/10.1038/s41467-022-33879-5); pmid: [36257955](https://pubmed.ncbi.nlm.nih.gov/36257955/)
- Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009). doi: [10.1007/s10858-009-9333-z](https://doi.org/10.1007/s10858-009-9333-z); pmid: [19548092](https://pubmed.ncbi.nlm.nih.gov/19548092/)
- S. P. Skinner *et al.*, CcpNmr AnalysisAssign: A flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124 (2016). doi: [10.1007/s10858-016-0060-y](https://doi.org/10.1007/s10858-016-0060-y); pmid: [27663422](https://pubmed.ncbi.nlm.nih.gov/27663422/)
- C. D. Schwieters, G. A. Bermejo, G. M. Clore, Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci.* **27**, 26–40 (2018). doi: [10.1002/pro.3248](https://doi.org/10.1002/pro.3248); pmid: [28766807](https://pubmed.ncbi.nlm.nih.gov/28766807/)
- H. Berman, K. Henrick, H. Nakamura, J. L. Markley, The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **35** (Database), D301–D303 (2007). doi: [10.1093/nar/gkl971](https://doi.org/10.1093/nar/gkl971); pmid: [17142228](https://pubmed.ncbi.nlm.nih.gov/17142228/)

43. N. C. J. Strynadka, M. N. G. James, Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem.* **58**, 951–999 (1989). doi: [10.1146/annurev.bi.58.070189.004511](https://doi.org/10.1146/annurev.bi.58.070189.004511); pmid: 2673026
44. A. M. Sevy, T. M. Jacobs, J. E. Crowe Jr., J. Meiler, Design of Protein Multi-specificity Using an Independent Sequence Search Reduces the Barrier to Low Energy Sequences. *PLoS Comput. Biol.* **11**, e1004300 (2015). doi: [10.1371/journal.pcbi.1004300](https://doi.org/10.1371/journal.pcbi.1004300); pmid: 26147100
45. P. Löffler, S. Schmitz, E. Hupfeld, R. Sterner, R. Merkl, Rosetta:MSF: a modular framework for multi-state computational protein design. *PLoS Comput. Biol.* **13**, e1005600 (2017). doi: [10.1371/journal.pcbi.1005600](https://doi.org/10.1371/journal.pcbi.1005600); pmid: 28604768
46. M. Mirdita *et al.*, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022). doi: [10.1038/s41592-022-01488-1](https://doi.org/10.1038/s41592-022-01488-1); pmid: 35637307
47. N.-A. Lakomek, J. Ying, A. Bax, Measurement of ¹⁵N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* **53**, 209–221 (2012). doi: [10.1007/s10858-012-9626-5](https://doi.org/10.1007/s10858-012-9626-5); pmid: 22689066
48. F. A. A. Mulder, R. Kaptein, R. Boelens, R. Boelens, de Graaf RA, An Off-resonance Rotating Frame Relaxation Experiment for the Investigation of Macromolecular Dynamics Using Adiabatic Rotations. *J. Magn. Reson.* **131**, 351–357 (1998). doi: [10.1006/jmre.1998.1380](https://doi.org/10.1006/jmre.1998.1380); pmid: 9571112
49. N. A. Farrow *et al.*, Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994). doi: [10.1021/bi00185a040](https://doi.org/10.1021/bi00185a040); pmid: 7514039
50. M. Tollinger, N. R. Skrynnikov, F. A. Mulder, J. D. Forman-Kay, L. E. Kay, Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001). doi: [10.1021/ja011300z](https://doi.org/10.1021/ja011300z); pmid: 11707108
51. D. G. Davis, M. E. Perlman, R. E. London, Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1 rho and T2 (CPMG) methods. *J. Magn. Reson. B* **104**, 266–275 (1994). doi: [10.1006/jmrb.1994.1084](https://doi.org/10.1006/jmrb.1994.1084); pmid: 8069484
52. M. P. Williamson, Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013). doi: [10.1016/j.pnmrs.2013.02.001](https://doi.org/10.1016/j.pnmrs.2013.02.001); pmid: 23962882
53. D. Van Der Spoel *et al.*, GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005). doi: [10.1002/jcc.20291](https://doi.org/10.1002/jcc.20291); pmid: 16211538
54. P. Robustelli, S. Piana, D. E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4758–E4766 (2018). doi: [10.1073/pnas.1800690115](https://doi.org/10.1073/pnas.1800690115); pmid: 29735687
55. W. C. Swope *et al.*, Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a β -Hairpin Peptide. *J. Phys. Chem. B* **108**, 6582–6594 (2004). doi: [10.1021/jp037422q](https://doi.org/10.1021/jp037422q)
56. M. P. Harrigan *et al.*, MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys. J.* **112**, 10–15 (2017). doi: [10.1016/j.bpj.2016.10.042](https://doi.org/10.1016/j.bpj.2016.10.042); pmid: 28076801
57. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011). doi: [10.1002/jcc.21787](https://doi.org/10.1002/jcc.21787); pmid: 21500218
58. J. Meier *et al.*, Language models enable zero-shot prediction of the effects of mutations on protein function. *bioRxiv* 2021.07.09.450648 [Preprint] (2021). <https://doi.org/10.1101/2021.07.09.450648>.
59. A. B. Guo, T. Kortemme, Deep learning guided design of dynamic proteins, Dryad; <https://doi.org/10.1101/2021.07.09.450648>.

ACKNOWLEDGMENTS

We thank members of the Kortemme lab for discussion and P. Huettemann, S. Crilly, and R. G. Alberstein for comments on the manuscript. **Funding:** Supported by NIH grant R35 GM145236 (T.K.); a grant from the Alfred P. Sloan Foundation (G-2021-16899); a UCSF Discovery Fellowship (A.B.G.); NSF GRFP fellowships (A.B.G. and D.A.); NIH grant S10 OD023455; NIH grant R01 GM137109 (M.G.); and a PBBR TMC award (UCSF NMR core facility). T.K. is a Chan Zuckerberg Biohub Investigator. **Author contributions:** A.B.G. and T.K. developed the conceptual approach for designing dynamic proteins. A.B.G. designed, screened, and characterized de novo state 2 designs. A.B.G. performed multi-state design and conceptualized, designed, screened, and characterized allosteric point mutants. A.B.G. collected and processed NMR data with aid and supervision from M.J.S.K. C. A. Smith performed the two-timescale fit of NMR relaxation data. A.B.G. ran MD simulations and mutual information analysis. A.B.G., C. A. Stephens, and M.G. performed Markov state modeling. D.A. performed Frame2seq scoring. T.K. provided guidance and resources. A.B.G. and T.K. wrote the manuscript with contributions from the other authors. All authors read and commented on the manuscript. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Expression plasmids for single-state and switch designs have been deposited to Addgene with accession codes 231958, 231959, 231960, 231961, 231962, 231963, 231964, 231965, and 231966. Structures have been deposited to the Protein Data Bank (PDB) with accession codes 9CIC, 9CID, 9CIE, 9CIF, and 9CIG. NMR data have been deposited to the Biological Magnetic Resonance Data Bank with accession codes 31182, 31183, 31184, 31185, and 31186. The data and scripts used in this study have been deposited in Dryad (59) and GitHub (https://github.com/amyguo1997/dynamic_protein_design). All other relevant data are available in the main text or supplementary materials. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.sciencemag.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adr7094

Figs. S1 to S22; Tables S1 to S12; References (60, 61); MDAR Reproducibility Checklist; Movie S1

Submitted 17 July 2024; accepted 5 March 2025

10.1126/science.adr7094