De Novo Design of Protein Conformational Changes and Protein-Peptide Interactions

Philip Leung

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

David Baker, Chair

Jeffrey Nivala

Georg Seelig

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

De Novo Design of Protein Conformational Changes and Protein-Peptide Interactions

Philip Leung

Chair of the Supervisory Committee:

David Baker

Department of Biochemistry

This work addresses two difficult challenges in protein design: designing proteins with two distinct, interconvertible structural conformations and designing proteins to bind tightly to biologically active, flexible helical peptides. First, through the design of “hinge” proteins that can switch between two fully structured conformations in response to an effector-binding event, we demonstrate the modular transformation of biochemical information. Through detailed structural and biophysical characterization, we show tight coupling between the conformational change and the effector binding. Second, we present methods for designing proteins to bind helical peptides, which we show bind specifically to therapeutically relevant targets with nanomolar and picomolar affinities. These two approaches go beyond the single-structure paradigm in protein design and enable new possibilities for therapeutics and bioengineering.

Table of Contents

Acknowledgments

In no particular order or organization. Scott Schmelze and Nisha Kanwar for kindling my interest for the research path. Justin, Luke, Colin, Aspen, Sara, and Kaitlin, for fanning the research flames. Amin, Gen, Drew and Sarah for being my Seattle family. Alex, for teaching me the finer points of discrete maths with ravioli cans. Bo, Tim, Susan, and Dave, for material and familial support. Ellie and Anna, and our friend Jim for keeping me physically and mentally in good condition. Nate, Caci and Erin, for showing me how to do dumb things with smart people. Stacey and Harley, for hanging out. Flo, for friendship, mentorship, and emotional support. Susana, for making me a better scientist and showing me how to be fearless. Ryan and Tim, for Rosetta-handholding and everything else. Basile and Lukas, for being cowboys (and PyMol presets). The crispy-shifties, for being the best scientific subgroup in the lab hands down. Sam for teaching me structural biology and PyMol finesse. Adam, for showing me how to use github copilot and making me learn git. Cullen, and Jeremy, for making sure I understand Levinthal's Paradox and Anfinsen's Dogma. MISL labmates, for helping me get really solid at scientific presentations. The Rosettacommons, for helping me find and fix random Rosetta problems, and having the best research meetings. The IPD admins, Luki, Kandise et al. for keeping the lights (and the digs) on. DB, for giving me a chance to build things out of proteins. Jeff, for countless slack conversations at odd hours. NERSC, for giving me nearly unfettered access to what was at the time the world's 5th largest supercomputer. The National Science Foundation, for giving me money. Many others.

Chapter I: Theoretical basis for designing protein conformational changes and protein-peptide interactions

Biochemistry, Central Dogma, and Protein Structure-Based Design with Computational Energy-based Methods

Proteins are genetically encoded macromolecules that facilitate myriad functions in nature, including structuring cells and tissues(*1*), fighting disease(*2*), and catalyzing the reactions that allow plants and animals to respirate(*3*, *4*). Thus, genetic variations and genetic engineering frequently realize their impact at the protein level, due to the fact that protein-encoding genes map to functional proteins. While nature has sampled a vast array of protein structures and functions to solve innumerable challenges over geologic timescales, solving modern engineering problems may require unique and specific functionalities that have not yet been met by evolution. *De novo* protein design presents a pathway towards achieving solutions for these problems, through the rational design of entirely novel protein structures and functions, purely based on the knowledge of biophysics, and often only loosely inspired by natural proteins(*5*). The is is possible because, while the primary encoding of proteins is a one-dimensional sequence of amino acid residues specified at the genetic level, in structured proteins, this macromolecular sequence usually folds surprisingly rapidly(*6*), through physical forces(*7*), from a one-dimensional polymer into a three dimensional shape(*8*), whose specific arrangement confers function(*9*). Thus it can be said that the structures of proteins is determined by their sequence(*10*), and since the functions of proteins are determined by their structures, the *de novo* design task then is to work backwards from desired function, by first designing structures that confer such function, then by designing sequences that form those structures with extremely favorable physical energies(*11*). Once those sequences are obtained, one can then work in a forwards direction, by encoding them in synthetic genes, producing the proteins, and screening for the desired function(*12*).

Design of Protein Shapes and Functions

Design of Complex Shapes and Structures with De Novo Helical Proteins

As a field, protein design is fairly mature in its capability for designing arbitrary protein shapes(*13*). Proteins have a structural hierarchy with four levels. The primary structure

is the sequence of the protein. Secondary structure comprises local repetitive patterns in the polymer chain backbone. Tertiary structure is the three dimensional arrangement of the primary and secondary structures. Finally, quaternary structure describes the arrangement of multiple associated tertiary structures across disconnected polymers, which occurs in a subset of proteins. Many tertiary and quaternary structures are arbitrarily accessible using fairly simple patterns - one example at the tertiary structure level is the vast geometric space accessible using only combinations of repeats of the alpha-helix secondary structure(*14*, *15*). Taking advantage of structural symmetry, all-helical proteins can also make a variety of quaternary structures, including oligomers(*16*), fibers(*17*), and cages(*18*). Elegant asymmetric quaternary structures are possible as well, including orthogonal heterodimers(*19*) and branched heterotrimers(*20*).

Design of Dynamic and Functional Systems using Helical Proteins

Going beyond structure into functional properties, designed all-helical proteins have achieved complex functions such as digital logic(*21*), pH-driven disassembly/unfolding(*22*), coupled binding and conformational changes(*23*), sensitivity to phosphorylation(*24*), and sensitive detection of molecules(*25*).

Deep Learning as an Efficient and Effective Tool in Protein Design

The Sampling Problem in De Novo Design

Modern protein design can be conceptualized as a process that requires two stages of problem solving, the sampling problem and the decoy discrimination problem. The exact details of each stage depends on the desired result, but in general the sampling problem takes two stages: First, one must arrive at realistic, “designable” protein backbones that confer a specific desired functionality. Second, one must sample sequences that robustly encode proteins that will reliably fold to the protein backbones(*5*). There are a variety of ways to sample protein backbones effectively, including fragment assembly(*26*), parametric sampling(*27*, *28*), alignment-based docking(*29*, *30*), hallucination(*31–33*), and denoising diffusion probabilistic models(*34*, *35*). For sequence sampling, conventional paradigms use Monte Carlo simulations with a scorefunction that provides a concise approximation of biophysical forces(*36*), often with some biasing based on structural data(*37*) gleaned from the rich statistics of the protein databank's (PDB) structure datasets(*38*). Recent methods combine or

circumvent the physics-based approaches with deep learning(*39*), which has been shown to implicitly learn physical properties of proteins(*40*). These deep learning approaches have been shown to be highly general(*41*), and avoid a particular pathology of physics-based optimization - they maximize the probability of a sequence given a structure, instead of minimizing the energy of a sequence given a structure(*39*, *42*). Maximizing probability implicitly minimizes the chances that a given sampled sequence might be even lower in energy for some off-target structure; this is not the case for physics-based methods, which minimize the energy of a sequence on a particular structure but do not check the energy of that sequence against all possible off target structures(*42*). This property becomes particularly important when trying to design a sequence that can adopt multiple structural states, optimizing the energy of such a sequence against multiple structures is not only computationally more challenging, but also is much more likely to result in a sequence that has a degenerate energetic landscape(*39*).

The Decoy Discrimination Problem and Structure Prediction

Given sets of amino acid sequences that supposedly encode a target protein structure, the next stage of problem solving is the decoy discrimination problem. Decoy discrimination, in this context, is the evaluation of a suitability of a sequence for a given structure. Historically, one way this could be done is with “forward folding,” simulating folding of the sequence *ab initio(43)*. Typically, several thousand computational trajectories per sequence would be launched. Fragment assembly(*44*), Monte Carlo sampling, and energy-based minimization would be used to sample likely folded structures for the given sequence(*45*), with resulting “decoys” ranked by a physically-based and statistically-biased scorefunction(*36*), with the goal of obtaining highly ranked decoys that are structurally very close to the intended target. This process leaves a lot to be desired in terms of computational efficiency and real world accuracy. Recently, Alphafold2 demonstrated accuracy competitive with experimental approaches for predicting protein structures from their sequence(*46*). This capability is extremely valuable for reasons beyond its accuracy: it is considerably more computationally efficient than forward folding (a few CPU hours per sequence vs thousands of hours), and it also results in useful confidence metrics(*47*), generated internally by the network, that can be used to easily rank and interpret the decoys resulting from the prediction trajectories.

Chapter II: Multistate design of stimulus-responsive protein hinges

Adapted from Florian Praetorius, Philip J. Y. Leung, Maxx H. Tessmer, Adam Broerman, Cullen Demakis, Acacia F. Dishman, Arvind Pillai, Abbas Idris, David Juergens, Justas Dauparas, Xinting Li, Paul M. Levine, Mila Lamb, Ryanne K. Ballard, Stacey R. Gerben, Hannah Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Brian F. Volkman, Jeff Nivala, Stefan Stoll, and David Baker. (2023). Design of stimulus-responsive two-state hinge proteins. biorxiv

Natural and Designed Protein Conformational Changes

Proteins Conformational Changes are Widespread in Nature

While many naturally occurring proteins adopt single folded states, conformational changes between distinct protein states are crucial to the functions of enzymes(*48*, *49*), cell receptors(*50*), and molecular motors(*51*). The extent of these changes ranges from small rearrangements of secondary structure elements(*52*, *53*) over domain movement(*54*) to fold-switching or metamorphic proteins(*55*) that adopt completely different structures. In many cases, these conformational changes are triggered by “input” stimuli such as the binding of a target molecule, post-translational modification, or change in pH. These changes in conformation can in turn result in “output" actions such as enzyme activation, target binding, or oligomerization(*56*); protein conformational changes can thus couple a specific input to a specific output. The generation of proteins that can switch between two quite different structural states is a difficult challenge for computational protein design, which usually aims to optimize a single, very stable conformation to be the global minimum of the folding energy landscape(*27*, *57*); design of proteins that can undergo controlled, major conformational changes requires reframing this paradigm towards optimizing for more than one minimum on the energy landscape, while simultaneously avoiding undesired off-target minima(*58*).

Designing Protein Conformational Changes

Previously, multi-state design has been used to design proteins that undergo very subtle conformational changes(*59*, *60*), cyclic peptides that switch conformations based on the presence of metal ions(*61*), and closely related sequences that fold into dramatically different conformations(*62*). Stimulus-responsive proteins have been designed to undergo conformational changes upon binding to a target peptide or protein(*22*, *23*); however, while the “closed” unbound state of these “switch” proteins is a well-defined

and fully structured conformation, the “open” bound state is a broad distribution of conformations. These proteins have found use as biosensors(*25*, *63*), but the lack of a defined second state makes them not well suited for mechanical coupling in a molecular machine or discrete state-based computing systems.
We set out to design proteins that can switch between two well-defined and fully structured conformations, with one conformational state being capable of highly favorable binding to an effector molecule. To facilitate experimental characterization of the conformational change and to ensure compatibility with downstream applications, we imposed several additional requirements. First, the conformational change between the two states should be large, with some inter-residue distances changing by tens of angstroms between the two states. Second, conformational change should not require global unfolding, which can be very slow. Third, neither of the two states should have substantially exposed patches of hydrophobic residues, which can compromise solubility. Fourth, the conformational change should be readily coupled to a range of inputs and outputs. Given that proteins are stabilized by hydrophobic cores, collectively achieving all of these properties in one protein system is challenging: protein conformations that differ considerably typically will have different sets of buried hydrophobic residues and require substantial structural rearrangements for interconversion.

We reasoned that these goals can collectively be achieved with a “hinge”-like design in which two rigid domains move relative to each other while remaining individually folded. The hinge amplifies small local structural and chemical changes to achieve large global changes while the chemical environment for most residues remains similar throughout the conformational change, avoiding the need for global unfolding. Provided that the two states of the hinge bury similar sets of hydrophobic residues, the amount of exposed hydrophobic surface area can be kept low in both states. Designing one of the resulting conformations to bind to a target effector couples the conformational equilibrium with target binding (Figure 1A). This design concept has precedent in nature; for example, bacterial two-component systems utilize binding proteins that undergo hinging between two discrete conformations in response to ligand binding(*64*).

Effectively Sampling Pairs of Designable Backbones

To implement this two-state hinge design concept, we took advantage of designed helical repeat proteins (DHRs, (*14*); Figure 1B, C left) or DHR-based junction proteins(*15*). The backbone conformation of the DHR serves as the first conformational state of our hinge protein (“state X”). To generate a second conformation, a copy of the parent protein is rotated around a “pivot helix” by aligning the copy to the original DHR shifted by N residues, where -7 < N < 7 (Figure 1B, C). A new backbone conformation is then created by combining the first half of the original protein (domain 1), the second half of the copy (domain 2), and either the helix following the pivot helix from the original protein or the helix preceding the pivot helix from the rotated copy (“peptide”). Backbone arrangements with large backbone clashes, as evaluated by the Lennard-Jones potential in Rosetta(*36*), are discarded. Rosetta FastDesign with backbone

movement(*65*, *66*) is used to re-design the interface between the three parts. Using fragment-based loop closure(*13*, *14*, *67*), domains 1 and 2 are connected into a single chain that serves as the second conformational state of the hinge protein (“state Y”, Figure 1B, C right). The loop between domains 1 and 2 is rebuilt in state X, yielding pairs of state X and state Y backbones with matching loop lengths and secondary structures.

Approaches to the Sequence Sampling Problem

Using a combination of Rosetta two-state design (see methods section for details) and proteinMPNN(*41*) with linked residue identities, a single amino acid sequence is generated that is compatible with the state X hinge as well as with the state Y hinge-peptide complex.

Decoy Discrimination using Deep Learning and Physics-Based Approaches

AlphaFold2 (AF2)(*46*) with an initial guess(*47*) can then be used to predict the structure of the hinge with and without the effector peptide, allowing for the selection of designs that are predicted in the correct state X in absence of the peptide and in the correct state Y complex in presence of the peptide. The designs are additionally filtered based on the Rosetta energies and spatial aggregation propensity (SAP)(*68*) for state X and for state Y with and without peptide. To favor designs that are predominantly in the closed state in absence of the peptide (Figure 1A, D), designs are selected only if state X has lower energy than state Y in absence of the peptide, and if the state Y complex has lower energy than state X plus the peptide, spatially separated. Designs are also filtered on standard interface design metrics for the bound conformation (see methods for details on filtering)(*69*).

Effector Binding Controls the Conformational State

Screening for Monomericity and Binding using Size Exclusion Chromatography

We used our hinge design approach to generate hinge-peptide pairs that cover a large structural space with diverse changes in terminal angle and globularity (Figures 1D, 2A, S1, S3). We experimentally tested multiple rounds of designs, using both DHRs(*14*) and helical junctions(*15*) as input scaffolds, and improving individual steps of the design pipeline between iterations (see Methods). To facilitate expression and improve solubility, the peptides were fused to superfolder GFP (sfGFP) via a flexible linker.

Hinges and sfGFP-peptide fusions were expressed in *Escherichia coli* (E. coli) and purified using immobilized metal affinity chromatography (IMAC) followed by size exclusion chromatography (SEC).

Quantifying Binding Affinities using Fluorescence Polarization

For hinge designs that were predominantly monomeric by SEC at high expression levels, we evaluated peptide binding by running hinge, the corresponding peptide-sfGFP fusion, and a mixture of both on SEC (Figures S2, S3). Designs that showed significant peak shifts of the mixture compared to individual components, indicating hinge-peptide binding, were selected for further characterization using a fluorescence polarization (FP) assay in which a chemically synthesized 5-carboxy-tetramethylrhodamine (TAMRA)-labeled peptide was incubated with different concentrations of purified hinge protein (Figures 2B, S4). The measured polarization is proportional to the fraction of bound peptide. Hinge-peptide binding affinities obtained from titration experiments with constant peptide concentration and varying hinge concentrations ranged from 1 nM to the low μM range (Figures 2B, S4). To circumvent the bottleneck of finding soluble peptide sequences we sought to design hinges that bind to a given target peptide. Starting from hinge cs201, we used a modified version of our design pipeline to redesign the hinge to bind the peptide from designs cs074 or cs221, respectively. This one-sided two-state design approach yielded hinge designs that showed strong binding to their new target peptide while showing no or only weak off-target binding (Figure S5).

Probing Hinge Conformational Changes with DEER

To characterize the conformational equilibrium of the designed hinges, we introduced two surface cysteine residues into the hinge protein and covalently labeled them with the nitroxide spin label MTSL(*70*). We then used double electron-electron resonance spectroscopy (DEER) to determine distance distributions between the two spin labels and compared these to simulated(*71*) distance distributions based on the state X and state Y design models. This experiment was performed on two different labeling site pairs for each design: one pair where the distance is predicted to decrease in the presence of peptide (Figures 2C, S4) and the other where it is predicted to increase (Figures 2D, S4). In the absence of the peptide, the observed distance distributions closely matched the state X simulations. In all cases the distances between the two pairs of probes shifted upon the addition of peptide to better match the state Y simulations, suggesting that the addition of effector peptide causes the conformational equilibrium to shift towards state Y as designed. For example, cs074 (site pair 1) showed a clear peak between 40 and 50 Å in absence of the peptide, and a peak between 30 and 40 Å in presence of the peptide, and both peaks agree well with the corresponding simulations (Figure 2C, top row). In a control experiment using the static parent DHR protein of design cs074, the distance distributions with and without peptide

were identical and matched both the simulation for the parent design model, which closely resembles state X and the experimental distance distribution for state X of cs074 (Figure S4). Design cs094 (site pair 1) showed the expected state X peak around 50 Å in absence of the peptide. In presence of peptide, this design showed a new peak around 30 Å matching the state Y simulation but still showed a residual state X peak around 50 Å, indicating incomplete binding either due to weak binding affinity or to insufficient peptide concentration. For all hinge designs tested using DEER, the observed conformational changes were consistent with the designed state X and Y models.

Structural Confirmation of the Designed Conformations

X-ray Crystallography of Hinges Confirms their Atomic Accuracy

We solved crystal structures for two designs, cs207 and cs074. For design cs207, crystals were obtained from two separate crystallization screens: one screen for the hinge alone (Figure 3A), and one screen for the hinge in complex with the target peptide (Figure 3B). In the absence of peptide, the experimental structure agrees well with the state X design model, with an RMSD (root mean square deviation) of 1.03 Å between the model and experimental structure at a resolution of 2.50 Å. The structure of the hinge-peptide complex agrees well with the state Y design model with an RMSD of 0.69 Å between the model and experimental structure at a resolution of 2.66 Å. For comparison, the RMSD between design models for states X and Y of cs207 is 4.59 Å. The electron densities for the side chains involved in the interdomain interface in state X and in the hinge-peptide interface in state Y closely match the design models, providing high-resolution validation of our two-state design procedure (Figure 3A-C, right). The two crystal structures of cs207 with and without peptide, combined with our DEER distance distributions, corroborate the fundamental hinge design concept: In the absence of the peptide the hinge is predominantly in state X, and the addition of the peptide shifts the equilibrium towards the state Y-peptide complex. For another design, cs074, the crystal structure of the hinge-peptide complex (Figure 3C) agrees well with the corresponding state Y design model (RMSD 1.46 Å to the original Rosetta model, 0.67 Å to the closest AF2 prediction, at a resolution of 2.75 Å).

Electron Microscopy Confirms Structural Accuracy and Modularity

To test whether our hinges can be incorporated as components of more complex protein assemblies without affecting their ability to undergo conformational changes, we designed a fully structured C3-symmetric protein with three hinge arms (Figure 3D). We used inpainting(*33*) with RoseTTAFold(*72*) to rigidly connect one end of hinge cs221 to a previously validated homotrimer(*73*) and the other end of the hinge to a previously validated monomeric protein(*29*). Negative-stain electron microscopy (nsEM) with

reference-free class averaging shows straight arms in absence of peptide and bent arms in presence of peptide cs221B, corroborating the designed conformational change (Figures 3D, S6).

The Conformational State Modulates Effector Binding

Kinetic Analysis of Effector Binding and Orthogonality with FRET

A critical feature of two-state switches in biology and technology is the coupling between the state control mechanism and the populations of the two states. To quantitatively investigate the thermodynamics and kinetics of the effector-induced switching between the two states of our designed hinges, we used Förster resonance energy transfer (FRET). To increase both the absolute distance from N- to C- terminus and the change in termini distance between the two conformational states, we took advantage of the extensibility of repeat proteins and extended hinges cs074, cs221, and cs201 by 1-2 helices on their N and C termini, yielding cs074F, cs221F, and cs201F, respectively (Figure 4A, first column). Single cysteines were introduced in helical regions near the termini of the extended hinges and stochastically labeled with an equal mixture of donor and acceptor dyes. For hinges cs074F and cs221F the distance between the label sites is above the R_0 of the dye pair in state X and below R_0 in state Y, and hence, acceptor emission upon donor excitation increases upon the addition of the corresponding peptides cs074B and cs221B, respectively (Figure 4A, second column). We used labeled, extended DHR82, the parent protein for cs074F, as a static control, and observed fluorescence spectra comparable to cs074F but no change in fluorescence upon the addition of the peptide (Figure S7). For cs201F, the dye distance is above R_0 in state X and below R_0 in state Y, and donor emission decreases upon the addition of peptide cs201B (Figure 4A, second column). To test the specificity of our hinge-peptide pairs, we performed pairwise titrations of all three labeled hinges at 2 nM with all three target peptides at varying concentrations. The on-target titrations had sigmoidal transitions that can be fitted with standard binding isotherms (Figures 4A, third column; S7), while the off-target titrations for cs201F and cs221F show flat lines, indicating no conformational change of these hinges upon the addition of off-target peptides at μM concentrations. cs074F showed weak off-target binding that was three orders of magnitude weaker for cs201B and two orders of magnitude weaker for cs221B compared to the on-target interaction for cs074B. cs201F and cs221F are thus orthogonal in the nM and μM range, and the set of cs201F, cs221F, and cs074F is orthogonal over two orders of magnitude of effector concentration.

Association kinetics for the on-target interactions measured using constant concentrations of the labeled hinge and varying excess concentrations of the peptide are well fit by single exponentials (Figures 4A, fourth column; S8). The apparent rate constants increase linearly with increasing peptide concentration, exhibiting standard pseudo-first-order kinetics for bimolecular reactions (Figures 4A, fifth column; S8). We analyze these data using a model comprising the three states (X, Y, Y+peptide) and four

rate constants (Figure 4B). The kinetic measurements using the FRET system follow the decrease in state X over time (d[X]/dt) upon the addition of peptide. The observed pseudo-first-order behavior (Figure 4A, fifth column) indicates that the conformational change happens on a timescale that is faster than that of the observed binding and can be treated as a fast pre-equilibrium (Supplementary Note 1). The slopes of the linear pseudo-first order fits (k_{on}) can thus be interpreted as the product of the microscopic association rate k_2 and the fractional population of state Y in absence of the peptide (F_Y = [Y]/([X]+[Y]), see Supplementary Note 1).

Comparison to Kinetic Analysis of Binding with Fluorescence Polarization

FP-based titrations and kinetic characterization using the unlabeled extended hinge cs074F in excess over the TAMRA-labeled peptide cs074B agree well with the corresponding FRET experiments, further supporting the pre-equilibrium model (Figures 4C, S8). FP kinetics experiments for other hinge designs also follow pseudo-first order behavior with k_{on} values ranging from $2.5x10^3$ $M^{-1}s^{-1}$ to $7.8x10^4$ $M^{-1}s^{-1}$ (Figures S4, S9).

Probing Reversibility of Switching with FRET

To study the reversibility of hinge conformational changes, we started with 30 nM of FRET-labeled hinge cs201F (Figure 4D), added 200 nM peptide to drive the conformational change, and then added excess unlabeled hinge cs201 to compete away the peptide. The FRET signal decreased upon the addition of the peptide, consistent with a conformational change from state X to state Y, and then returned to nearly the original level upon the addition of an unlabeled hinge, indicating that the hinge conformational change is fully reversible.

Using Proteins Effectors Instead of Peptide Effectors

To explore whether peptide-responsive hinges could be turned into protein-responsive hinges, we used inpainting with RoseTTAFold to add two additional helices to a validated effector peptide, resulting in fully structured 3-helix bundles (3hb). For nine of our validated hinges, we designed and experimentally characterized these effector proteins using SEC (Figures 4E, S10, S11). Hinge-3hb binding was tested qualitatively by SEC and, for hinges that had a corresponding FRET construct, quantitatively with the FRET-labeled variant, and DEER was used in addition to FRET to confirm that 3hb binding caused the same conformational change as effector peptide binding (Figures 4E, bottom; S10). The affinity of 3hb05 to cs074F was similar to the affinity observed for the original peptide cs074B (Figure 4E), whereas 3hb21 bound its target hinge cs221F significantly tighter than the original peptide cs221B (Figure S12). The 3hb approach was able to rescue designs for which the peptide alone or the hinge-peptide complex

had shown the tendency to form higher-order oligomers (Figure S11). For two designs, 3hb05 and 3hb12, we obtained crystal structures that agreed well with the design models, indicating that the three-helix bundles are fully structured in isolation (Figures 4E top right, S13).

Making Hinges Red/Ox Responsive with Disulfide “Stapling”

To test the effect of the conformational pre-equilibrium on effector binding, we introduced disulfide “staples” that lock the hinge in one conformation. We used a 6D hashing approach(*74*) to identify pairs of cysteine positions that would be compatible with disulfide formation in one state but not the other, and used AF2 prediction as a computational filter, selecting only sequences for which cysteine side chains were predicted to have side chain distances and relative orientations compatible with disulfide formation in the intended structural state. Using FP we analyzed peptide binding to stapled versions of hinge cs221 (Figure 5A, B). The variant that forms a disulfide bond in state X (“locked X”) showed only weak residual binding, likely due to a small fraction of hinges not forming the disulfide (Figure 5A). Upon addition of the reducing agent dithiothreitol (DTT) to break the disulfide, peptide binding was fully restored, making this hinge variant a red/ox-dependent peptide binder that binds the effector peptide under reducing but not under oxidizing conditions. The variant that forms a disulfide bond in state Y (“locked Y”) showed fast peptide binding following pseudo-first-order kinetics with an observed association rate that was 200-fold higher than for the original hinge without disulfides (Figures 5B, S14). Using the pre-equilibrium model described above, the observed association rates provide an estimate of the fraction of hinge that is in state Y in absence of the peptide: a 200-fold higher observed on rate for the locked Y variant indicates a 200-fold higher fraction of hinge in state Y compared to the original hinge. Assuming that the locked Y variant is 100% in state Y and assuming that the microscopic rate constant k_2 is identical for the locked Y hinge and state Y of the original hinge, this would indicate that the original hinge is 99.5% in state X and 0.5% in state Y at equilibrium.

The conformational pre-equilibrium controls effector binding kinetics

Having established the edge cases of locked state X and locked state Y, we sought to tune the pre-equilibrium by introducing single point mutations expected to specifically stabilize one state over the other while not directly affecting the peptide-binding interface. We used proteinMPNN to generate consensus sequences(*75*) for each state and identified non-interface positions with distinct residue preferences that were different between both states (Figure 5C). We experimentally tested individual protein variants carrying mutations expected to stabilize one state over the other without disrupting either conformation, as evaluated by AF2 predictions. Several mutations expected to stabilize state X led to weaker peptide binding, and mutations expected to stabilize state Y led to stronger binding (Figures 5C,D, S14). Mutations expected to

stabilize state X had little effect on k_{on}, suggesting that these mutations primarily caused destabilization of the state Y-peptide complex. Mutations that stabilized state Y, on the other hand, strongly affected k_{on}, indicating that they effectively shift the pre-equilibrium between both states (Figures 5C,D, S15).

The double mutant cs221_V111L_A114T has a 22-fold higher on rate than the original cs221, suggesting the occupancy of state Y in cs221_V111L_A114T is 22x higher in the absence of peptide. Distance distributions obtained from DEER measurements on site pair 2 of the double mutant cs221_V111L_A114T in absence of the peptide indeed showed an additional peak at a distance closely matching state Y (Figures 5E, S16). DEER measurements on site pair 1 of the double mutant showed a broader distribution with occupancy in the region corresponding to state Y (Figures 5E, S16). Measurements in the presence of the peptide were virtually indistinguishable from the original cs221 (Figure S16). The double mutant thus populates two distinct states in the absence of the effector and collapses to one state upon effector addition (Figures 5E, S16). The observation of a significant state Y population at equilibrium in the absence of the peptide as predicted based on the kinetic measurements further corroborates that the mutations affect the conformational pre-equilibrium, and provides strong support for our quantitative two-state model of the kinetics and thermodynamics of the designed hinge-effector systems.

Conclusion

Extensions and Future Directions of the Hinge Design Concept

Taken together, our structural, kinetic and thermodynamic data strongly suggest that our hinge design method generates proteins that populate two well-defined and structured conformational states. The crystal structures, nsEM class averages, and one-dimensional distance distributions obtained from DEER measurements suggest that the designed hinges preferentially populate the two designed states, rather than adopting a heterogenous mixture of structures. The DEER and FRET experiments show that a peptide or protein effector can drive the conformational equilibrium from state X to state Y. The kinetics of the conformational change measured by FRET and the kinetics of peptide binding measured by FP agree well, corroborating the two-state model. DEER measurements on a hinge variant that populates both states in absence of the peptide further confirm the absence of additional states and agree well with population estimates based on association kinetics. To our knowledge, these are the first designed proteins that can reversibly switch between two substantially different structured conformations. Our design strategy for creating protein two-state systems without exposing substantial hydrophobic surface area in either state should be quite broadly applicable.

Like transistors in electronic circuits, our designed two-state switches can now be coupled to external outputs and inputs to create sensing devices and incorporated into

larger protein systems to address a wide range of outstanding design challenges. Hinges containing a disulfide that locks them in state X couple the input “red/ox state” to the output “target binding,” where the target can be a peptide or a protein, and our FRET-labeled hinges couple the input “target binding” to the output “FRET signal.” Our approach can be readily extended to have state switching driven by naturally occurring rather than designed peptides; hinge-like proteins have recently been designed to target peptides such as glucagon, secretin, or neuropeptide Y(*76*), enabling new routes to sensing and detection. Stimuli-responsive protein assemblies that change shape or oligomeric state in the presence of an effector can now be built by incorporating the hinges as modular building blocks. Installing enzymatic sites in hinges such that substrate binding favors one state and product release favors the other state should enable fuel-driven conformational cycling, a crucial step towards the *de novo* design of molecular motors. More generally, the ability to design two-state systems, and the designed two-state switches presented here, should enable protein design to go beyond static structures to more complex multistate assemblies and machines.

Figures for Multistate Design of Stimulus-Responsive Protein Hinges

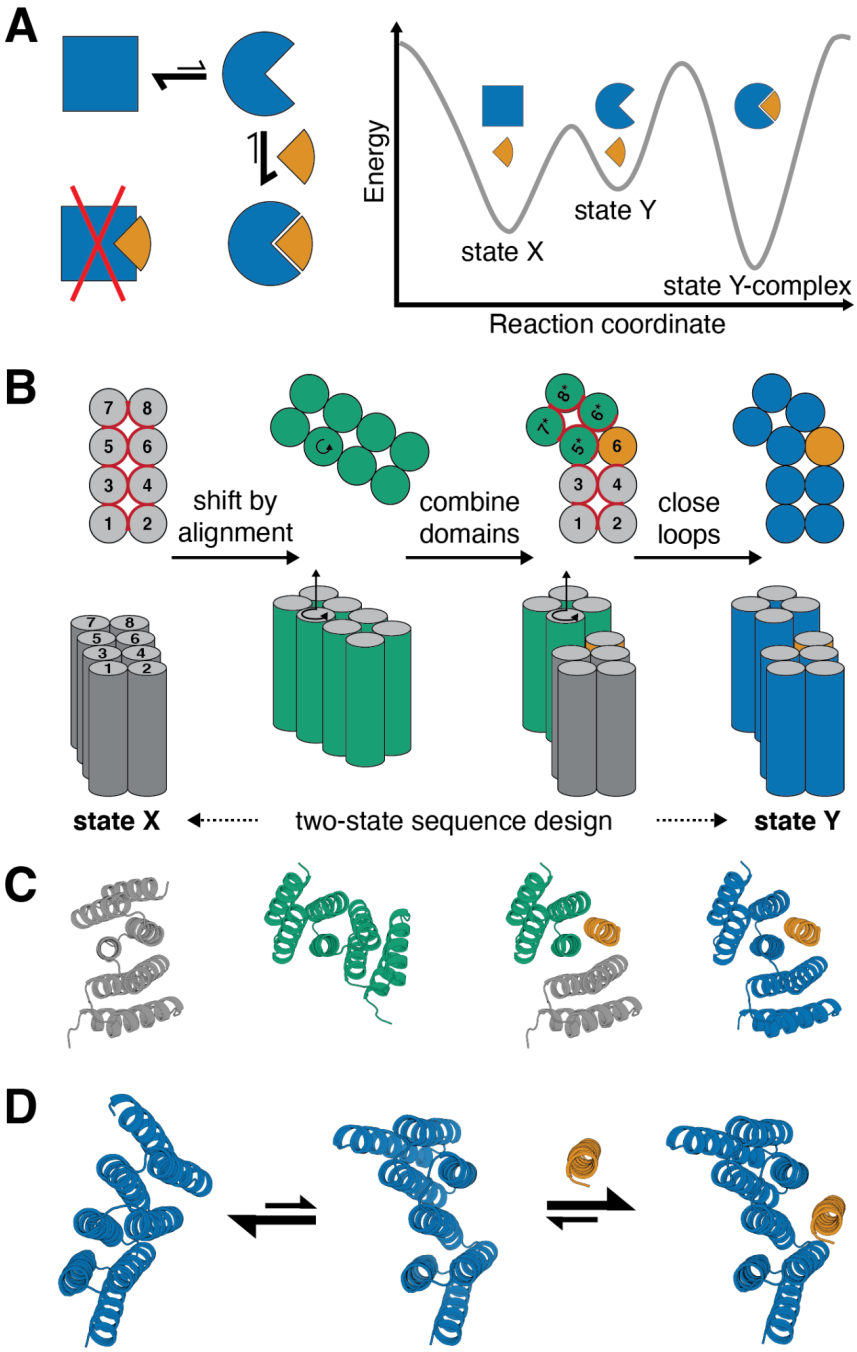

Figure 1: Strategy for designing proteins that can switch between different conformations.

A) Left: reaction scheme for a protein (blue) that undergoes a conformational change and can bind an effector (orange) in one (circle) but not in the other conformational state (square). Right: Energy landscape for the system shown on the left. **B)** Schematic representation of the hinge design approach. Alpha-helices are represented as circles (top view, top) or cylinders (side view, bottom). From left to right: A previously designed repeat protein (gray) serves as the first conformation of the hinge. To generate the second conformation a copy of the repeat protein (green) is moved by shifted alignment along a pivot helix, causing a rotation (top and bottom, indicated by circular arrow) and a translation along the helix axis (bottom). The first 4 helices of the original protein form domain 1 of the hinge, the last 4 helices of the rotated copy form domain 2, and an additional helix is copied over from the original protein to serve as an effector peptide (orange) that can bind to this second conformation of the hinge. Both domains of the hinge are connected into one continuous chain (blue) using fragment-based loop closure, and a single amino acid sequence is designed to be compatible with both conformations. **C)** Design steps from B illustrated using cartoon representations of an exemplary design trajectory. **D)** Exemplary design models of a designed hinge protein in state X (left), state Y (center), and in state Y bound to an effector peptide (right). Hinge is shown in blue, peptide is in orange.

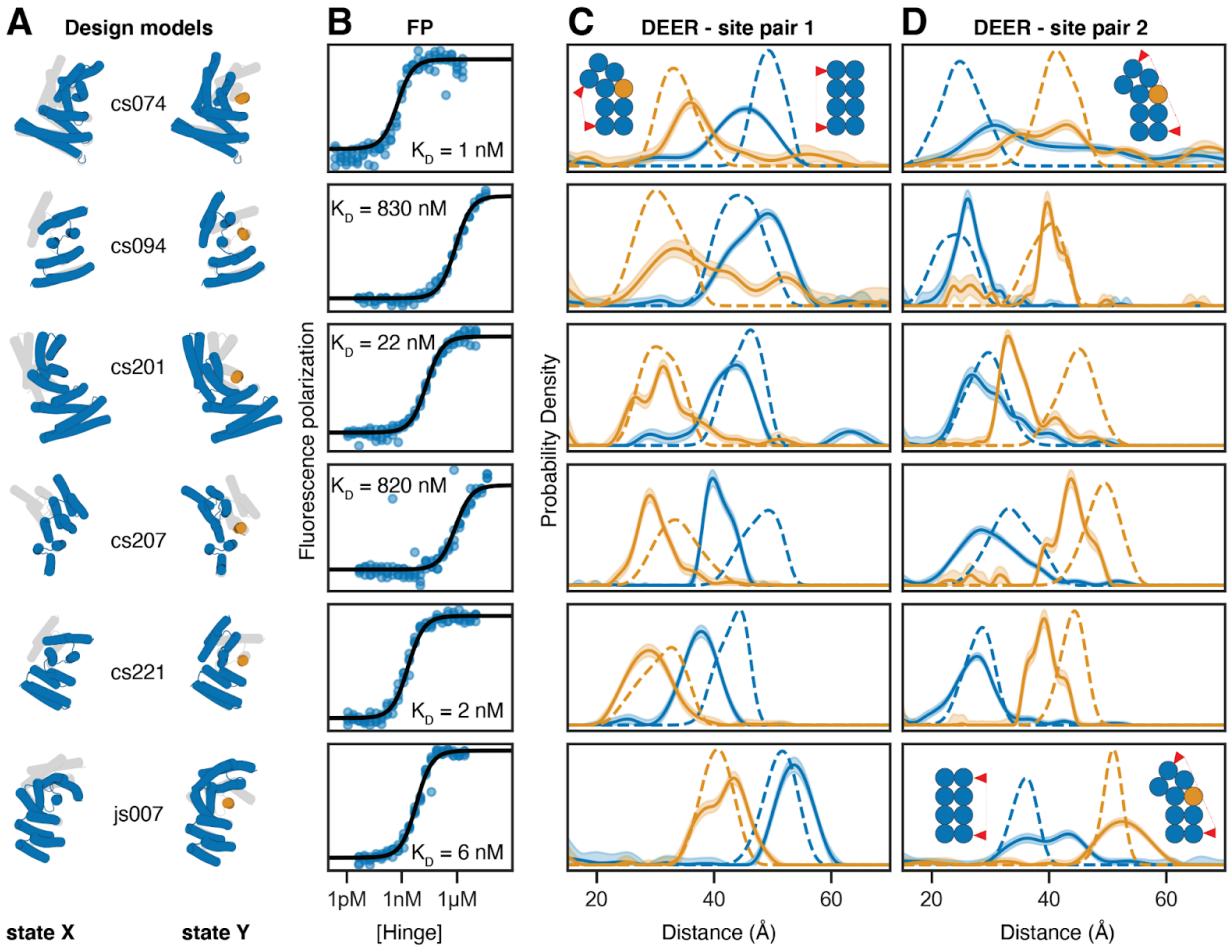

Figure 2: Experimental validation of peptide-binding hinges.

A) Design models of hinges (blue) and peptides (orange) in state X (left model) and state Y bound to the peptide (right model). Gray shades behind models in state X and Y indicate the corresponding states Y and X, respectively. **B)** Fluorescence Polarization (FP) titrations with a constant concentration of TAMRA-labeled peptide (0.1 nM for cs074 and cs221; 0.5 nM for cs201; 1 nM for cs094, cs207, and js007) and varying hinge concentrations. Circles represent data points from four independent measurements, lines are fits of standard binding isotherms to all data points, and dissociation constants (K_D) are obtained from those fits. **C, D)** Distance distributions between spin labels covalently attached to cysteine side chains. Solid lines are obtained from DEER experiments without (blue) or with (orange) an excess of peptide, shaded areas are 95% confidence intervals, and dashed lines are simulated based on the design models for state X (blue) or the state Y complex (orange). For each hinge, two different label site pairs were tested, one in which the distance was expected to decrease with peptide binding (C) and one in which the distance was expected to

increase upon peptide binding (D). Chemically synthesized peptides were used for all measurements except for cs074 site pair 1, for which sfGFP-peptide fusion was used.

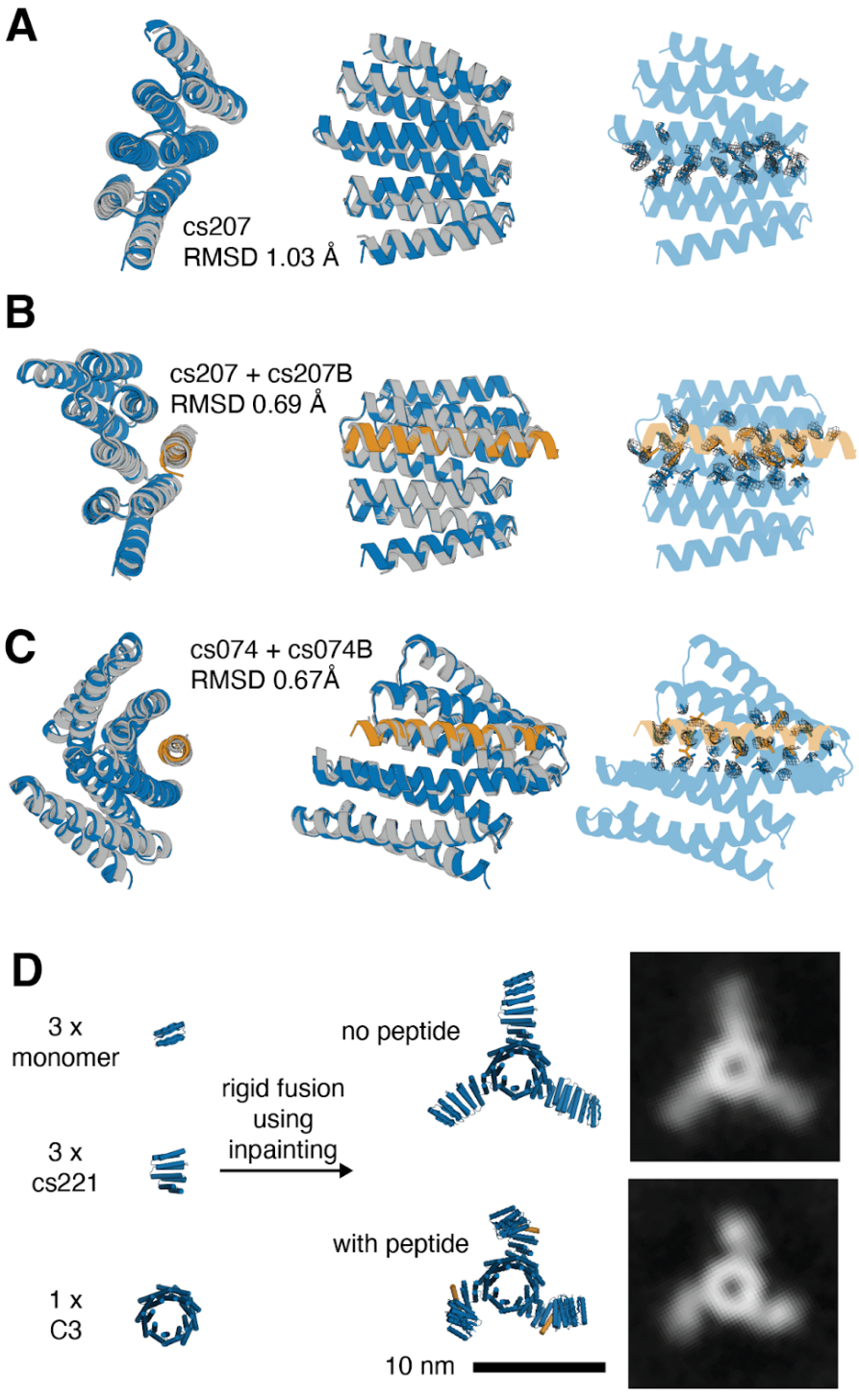

Figure 3: Structural validation of the conformational change in peptide-binding hinges.

A, B, C) Design models (hinge in blue, peptide in orange) overlaid with crystal structures (gray) in the top view (left) and side view (center) and design models with electron density for selected side chains (right). **A)** Design model of hinge cs207 in state X overlaid with the crystal structure of hinge cs207 crystallized without peptide. The right panel highlights electron densities for side chains in the interface between the two hinge domains. **B)** Design model of hinge cs207 in state Y overlaid with the crystal structure of hinge cs207 crystallized with peptide cs207B. The right panel highlights electron densities for the same side chains as A; these side chains now contribute to the hinge-peptide interface. **C)** Design model of hinge cs074 in state Y overlaid with the crystal structure of hinge cs074 crystallized with peptide cs207B. The right panel highlights electron densities for side chains in the interface between the hinge and peptide. **D)** Left: Components for the design of a C3-symmetric homotrimer with three cs221 hinge arms. Center: Design model of the hinge-armed trimer in state X (top) and in state Y (bottom). Right: nsEM class averages of the trimer in absence of peptide (top) and in the presence (bottom) of peptide cs221B.

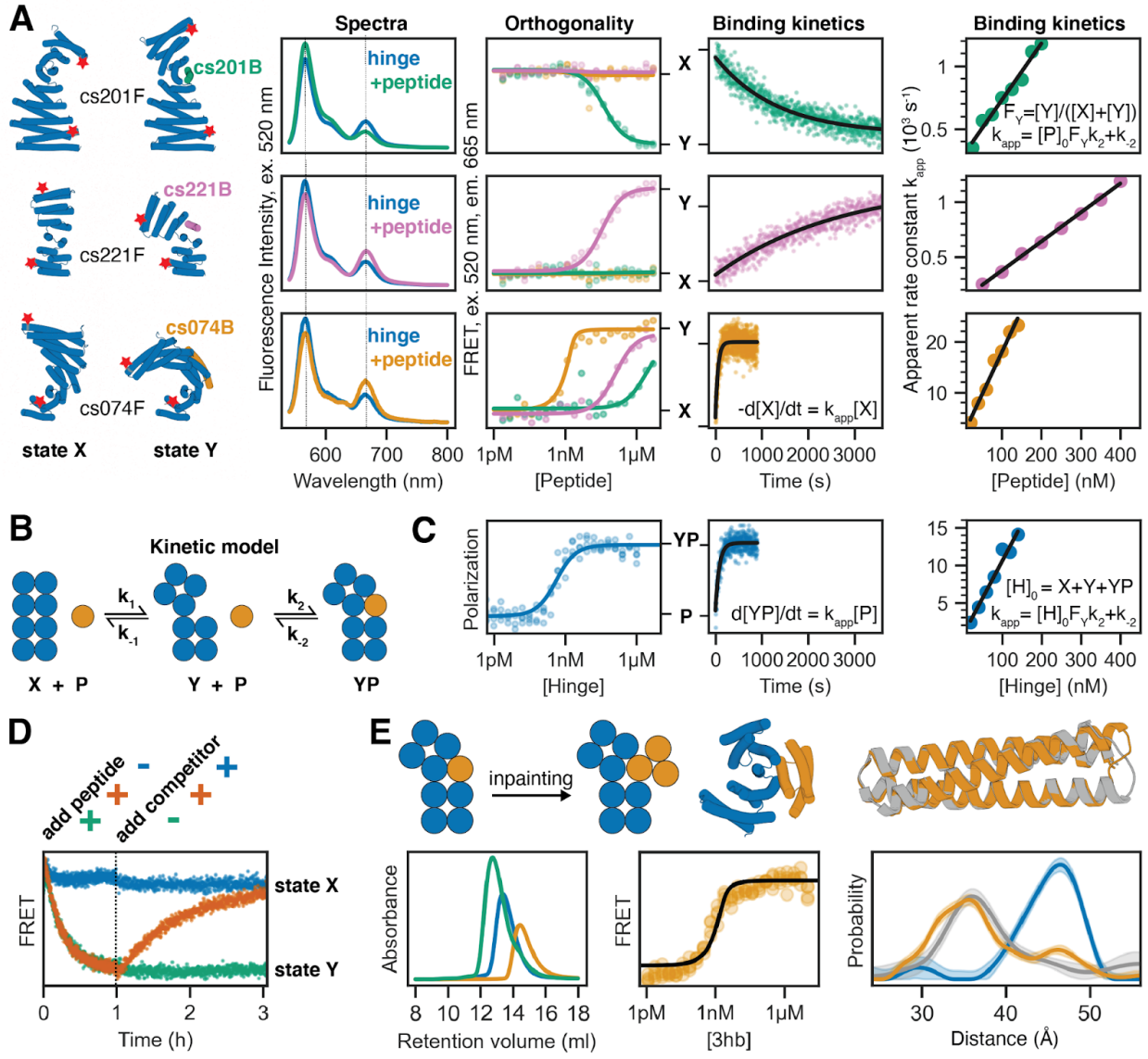

Figure 4: Quantitative analysis of conformational changes in designed hinge proteins

A) FRET-based characterization of three extended hinges. From left to right: cylindrical representation of extended hinges (blue) and their corresponding target peptides (green: cs201B, pink: cs221B, orange: cs074B) with red stars indicating attachment sites for fluorescent dyes; fluorescence spectra (excitation at 520 nm) of labeled hinge without (blue) or with (green/pink/orange) target peptide; FRET-based binding titrations (excitation 520 nm, emission 665 nm) at 2 nM labeled hinge and varying peptide concentrations fitted with standard binding isotherms (solid lines); time course after mixing 2 nM (cs201F, cs074F) or 5 nM (cs221F) labeled hinge and 100 nM peptide fitted with a single-exponential equation (black line); apparent rate constants obtained from single-exponential kinetic fits plotted against absolute peptide concentrations

(circles) and fitted with a linear equation (black line). Dotted lines in spectra indicate acceptor and donor emission peaks. **B)** Kinetic model describing the coupling of the conformational equilibrium to the binding equilibrium. X and Y: hinge in state X and Y, respectively; P: peptide; YP: peptide bound to hinge in state Y. k_1, k_{-1}, k_2, and k_{-2} are the microscopic rate constants. **C)** FP characterization of unlabeled extended hinge cs074F. From left to right: binding titration at 0.1 nM TAMRA-labeled peptide and varying hinge concentrations; time course after mixing 2 nM TAMRA-labeled peptide and 100 nM hinge fitted with a single-exponential equation (black line); apparent rate constants obtained from single-exponential kinetic fits plotted against absolute hinge concentrations (circles) and fitted with a linear equation (black line). **D)** FRET-based reversibility experiment using the labeled extended hinge cs201F introduced in C). Hinge concentration is 30 nM for all traces; 1 μM peptide is added at t=0 (green/orange), and 3 μM unlabeled competitor hinge is added after 1 h (blue/orange). **E)** Top from left to right: schematic representation of the inpainting procedure that adds two helices to the peptide cs074B yielding a three-helix bundle (3hb); cylindrical representation of 3hb_05(orange) bound to hinge cs074 (blue); overlay of design model (orange) and crystal structure (gray) of 3hb_05. Bottom from left to right: SEC traces for hinge cs074 (blue), 3hb_05 (orange), and a mixture of both (green); FRET-based titration of 2 nM extended labeled hinge cs074F and varying concentrations of 3hb_05 fitted with a standard binding isotherm (back line); Distance distributions obtained from DEER experiments as described in Figure 2 (blue: cs074, gray: cs074 + peptide cs074B, orange: cs074 + 3hb_05).

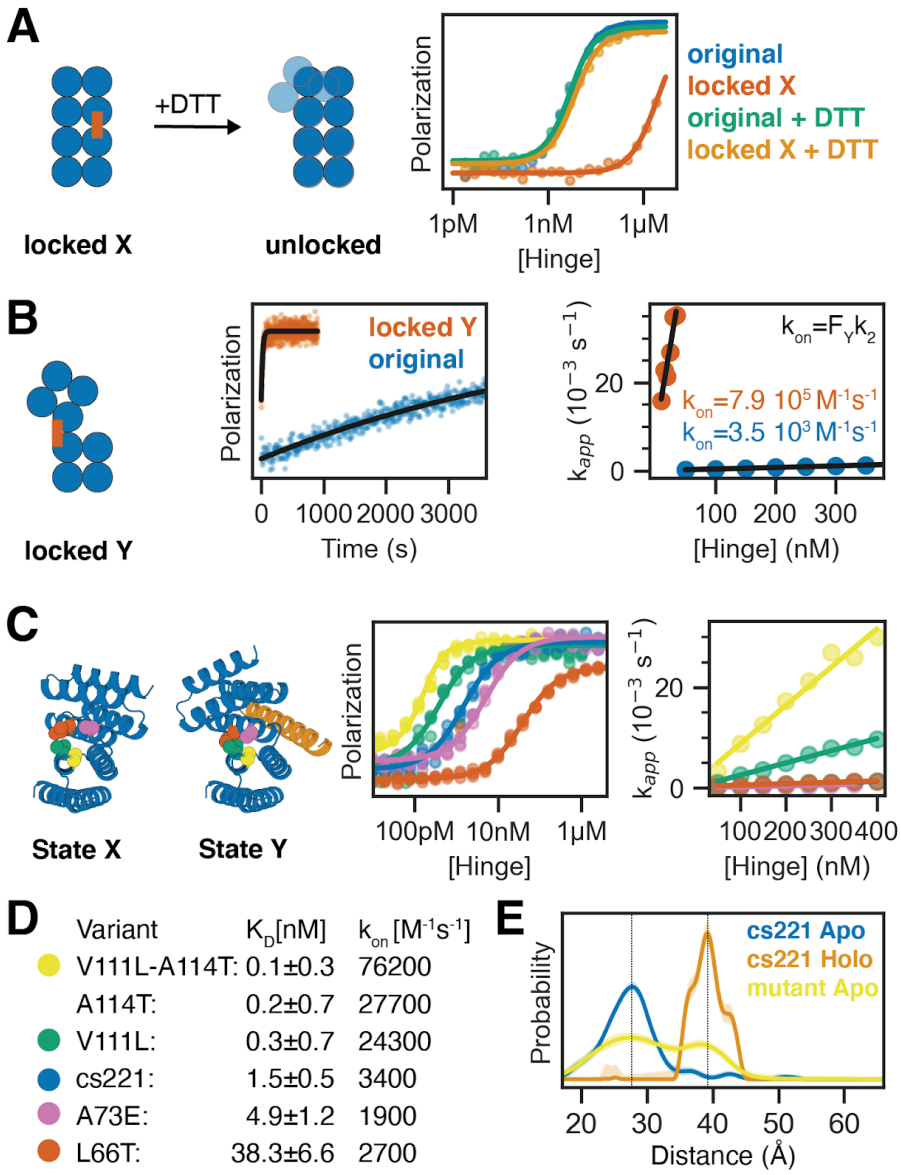

Figure 5: Controlling the conformational pre-equilibrium affects peptide binding.

A) Left: Schematic representation of a hinge containing two cysteine residues that can form a disulfide bond in state X but not in state Y, effectively locking the hinge in state X under oxidizing conditions. Upon addition of reducing agent DTT the disulfide bond is broken and the conformational equilibrium is restored. Right: FP-based titration of 1 nM TAMRA-labeled peptide and a hinge with state X disulfide (red, orange) or the parent hinge without cysteines (blue, green) under oxidizing (blue, red) or reducing (green, orange) conditions. **B)** From left to right: schematic representation of a hinge that is disulfide-locked in state Y; time course after mixing 2 nM TAMRA-labeled peptide and 50 nM locked hinge (red) or original hinge without cysteines (blue) fitted with a single-exponential equation (black line); apparent rate constants obtained from

single-exponential kinetic fits plotted against absolute hinge concentrations (circles) and fitted with a linear equation (black line). **C)** Tuning the pre-equilibrium with point mutations. Left: Cartoon representation of hinge cs221 highlighting positions of point mutations. Center: FP-based titration of 0.1 nM (yellow, green, blue) or 1 nM (pink, red) TAMRA-labeled peptide cs221B and varying concentrations of hinge variants containing one or two point mutations. Right: Apparent rate constants obtained from single-exponential kinetic fits plotted against absolute hinge concentrations (circles) and fitted with a linear equation (black line). **D)** Dissociation constants (K_D) and observed binding rate constants (k_{on}) for the hinge variants shown in C. **E)** DEER distance distribution for the double mutant cs221-V111L-A114T in absence of peptide (yellow) in comparison to the original cs221 with (orange) and without (blue) peptide. Gray lines serve as guides to the eye indicating state X and state Y distances.

Authors and Acknowledgements for Multistate Design of Stimulus-Responsive Protein Hinges

Authors and affiliations

Florian Praetorius[†*1,2], Philip J. Y. Leung[†1,2,3], Maxx H. Tessmer[4], Adam Broerman[1,2,5], Cullen Demakis[1,2,6], Acacia F. Dishman[1,2,7,8], Arvind Pillai[1,2], Abbas Idris[1,2,9], David Juergens[1,2,3], Justas Dauparas[1,2], Xinting Li[1,2], Paul M. Levine[1,2], Mila Lamb[1,2], Ryanne K. Ballard[1,2], Stacey R. Gerben[1,2], Hannah Nguyen[1,2], Alex Kang[1,2], Banumathi Sankaran[10], Asim K. Bera[1,2], Brian F. Volkman[7], Jeff Nivala[11,12], Stefan Stoll[4], and David Baker[*1,2,13]

1. Department of Biochemistry, University of Washington, Seattle, WA, USA.
2. Institute for Protein Design, University of Washington, Seattle, WA, USA.
3. Graduate Program in Molecular Engineering, University of Washington, Seattle, WA, USA.
4. Department of Chemistry, University of Washington, Seattle, WA, USA.
5. Department of Chemical Engineering, University of Washington, Seattle, WA, USA.
6. Graduate Program in Biological Physics, Structure, and Design, University of Washington, Seattle, Washington, USA.
7. Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI, USA.
8. Medical Scientist Training Program, Medical College of Wisconsin, Milwaukee, WI, USA.
9. Department of Bioengineering, University of Washington, Seattle, WA, USA.
10. Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

11. Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.
12. Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA.
13. Howard Hughes Medical Institute, University of Washington, Seattle, WA,USA.

*Corresponding authors. Email: flop@uw.edu, dabaker@uw.edu
†These authors contributed equally to this work

Acknowledgments and funding

We thank Basile I. M. Wicky, Lukas F. Milles, and Danny D. Sahtoe for helpful discussions and technical support, Alexis Courbet for inspiring discussions, Annika Philomin and Andrew Borst for EM support, and Kandise VanWormer and Luki Goldschmidt for technical support.
We want to thank the Advanced Light Source (ALS) beamline 8.2.2/8.2.1 at Lawrence Berkeley National Laboratory for X-ray crystallography data collection. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health (NIH), the National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences and US Department of Energy (DOE) (DE-AC02-05CH11231). This work was supported by a Human Frontier Science Program Long Term Fellowship (LT000880/2019, F.P.), the Open Philanthropy Project Improving Protein Design Fund (P.J.Y.L., C.W.D., H.N., D.B.), an NSF Graduate Research Fellowship (DGE-2140004; P.J.Y.L), NERSC award BER-ERCAP0022018 (P.J.Y.L, D.B.), the Audacious Project at the Institute for Protein Design (A.B., A.P., A.I., M.L., R.K.B., S.R.G., A.K., D.B.), a gift from Microsoft (D.J., J.D., D.B.), a grant from DARPA supporting the Harnessing Enzymatic Activity for Lifesaving Remedies (HEALR) program (HR001120S0052 contract HR0011-21-2-0012, X.L., A.K.B., D.B.), the Defense Threat Reduction Agency (DTRA) grant # HDTRA1-19-1-0003 (P.M.L.), NSF Award #2006864 (J.N.), and the Howard Hughes Medical Institute (D.B.). DEER measurements were supported by R01 GM125753 (to S.S.). The spectrometer used was funded by NIH grant S10 OD021557 (S.S.).

Author contributions

F.P. developed the hinge design concept. F.P. and P.J.Y.L. developed the computational hinge design pipeline. F.P. and P.J.Y.L. designed, screened, and characterized most hinges with help from C.D. and A.P.. M.H.T. designed, performed, and analyzed DEER experiments. S.S. analyzed DEER data and supervised research. C.D. developed the one-sided two-state design protocol and designed and tested hinges with swapped targets. A.B. designed and characterized 3-helix bundles with support from D.C.J.. A.I. designed and characterized the hinge-armed trimer with support from A.B.. A.I. performed electron microscopy and image processing with support from A.P.. J.D. provided conceptual support for two-state sequence design. X.L., P.M.L., M.L., and

R.K.B. synthesized and purified peptides. X.L., M.L., and R.K.B. performed LC-MS validation of proteins and peptides. S.R.G. performed additional protein purification. H.N., A.K., B.S., and A.K.B. determined crystal structures. A.F.D. and B.V. contributed conceptual support. D.B. and J.N. supervised research. F.P., P.J.Y.L., and D.B. wrote the manuscript. M.H.T. contributed to the manuscript. All authors read and commented on the manuscript.

Competing interests

A provisional patent application will be filed prior to publication, listing F.P., P.J.Y.L., M.H.T., A.B., C.D., A.F.D., A.P., A.I., B.V., S.S., and D.B. as inventors or contributors.

Data and materials availability

Crystallographic datasets have been deposited in the Protein Data Bank (PDB) (accession codes 8FIH, 8FVT, 8FIT, 8FIN and 8FIQ).
Code can be found at https://github.com/proleu/hinge_paper

Supplementary Material for Multistate Design of Stimulus-Responsive Protein Hinges

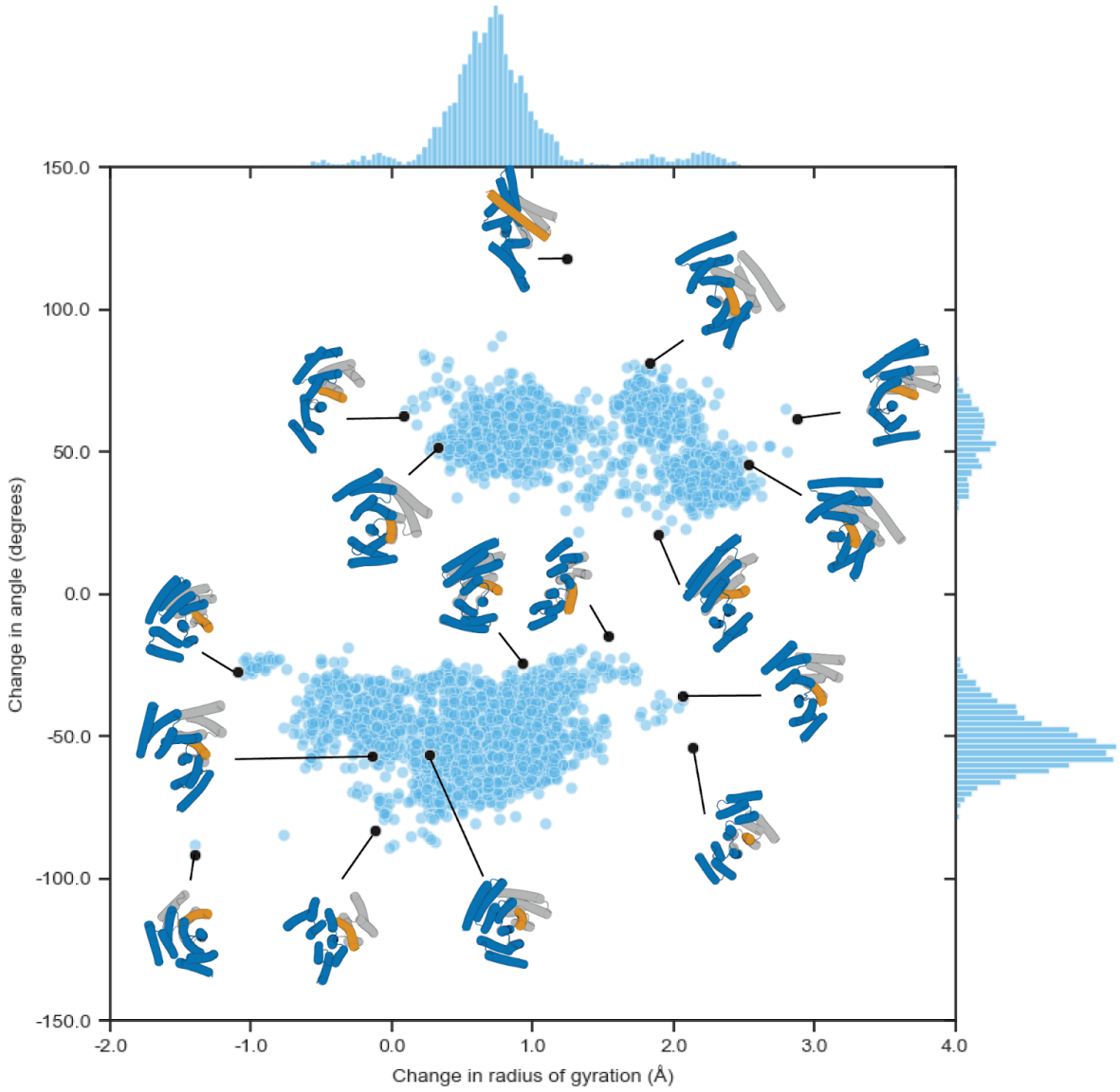

Figure S1: Diversity of hinge structures and conformational changes.

Light blue points represent individual designs, with positions as a function of the angle change of the hinge (state Y - state X, N-terminus-midpoint-C-terminus angle), measured in degrees, and the change in radius of gyration of the hinge (state Y - state X) in angstrom as computed by PyRosetta. Black points are representative examples and are depicted as cartoon models (state X in gray, state Y in blue and orange).

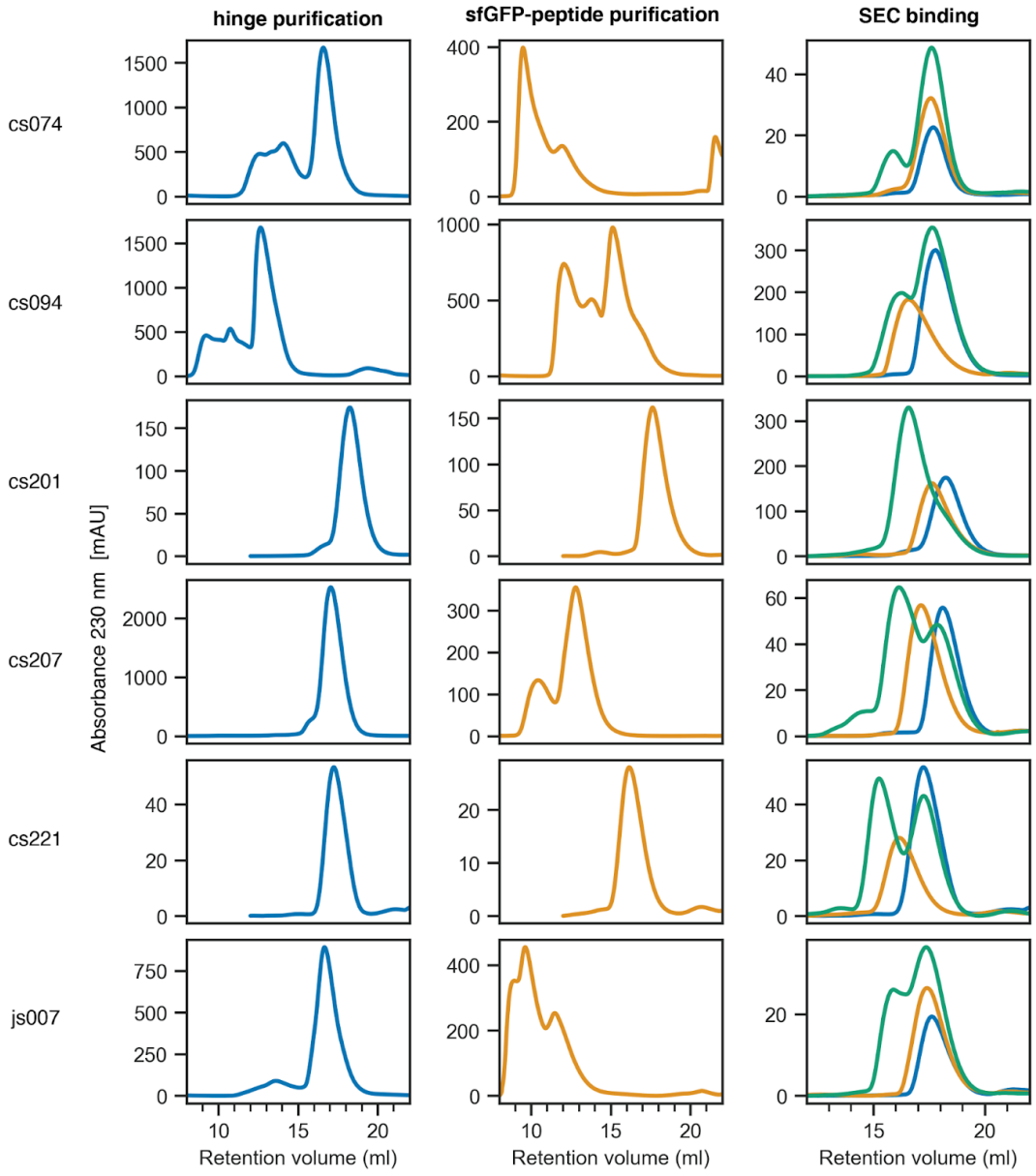

Figure S2: Size exclusion chromatography (SEC) of the hinges shown in Figure 2.

Purification runs of hinges (left, blue) and sfGFP-peptide fusions (center, orange) were performed on Superdex 75 Increase 10/300 GL columns (Cytiva). SEC binding experiments (blue: hinge, orange: sfGFP-peptide fusion, green: a mixture of both) were performed on Superdex 200 Increase 10/300 GL columns (Cytiva).

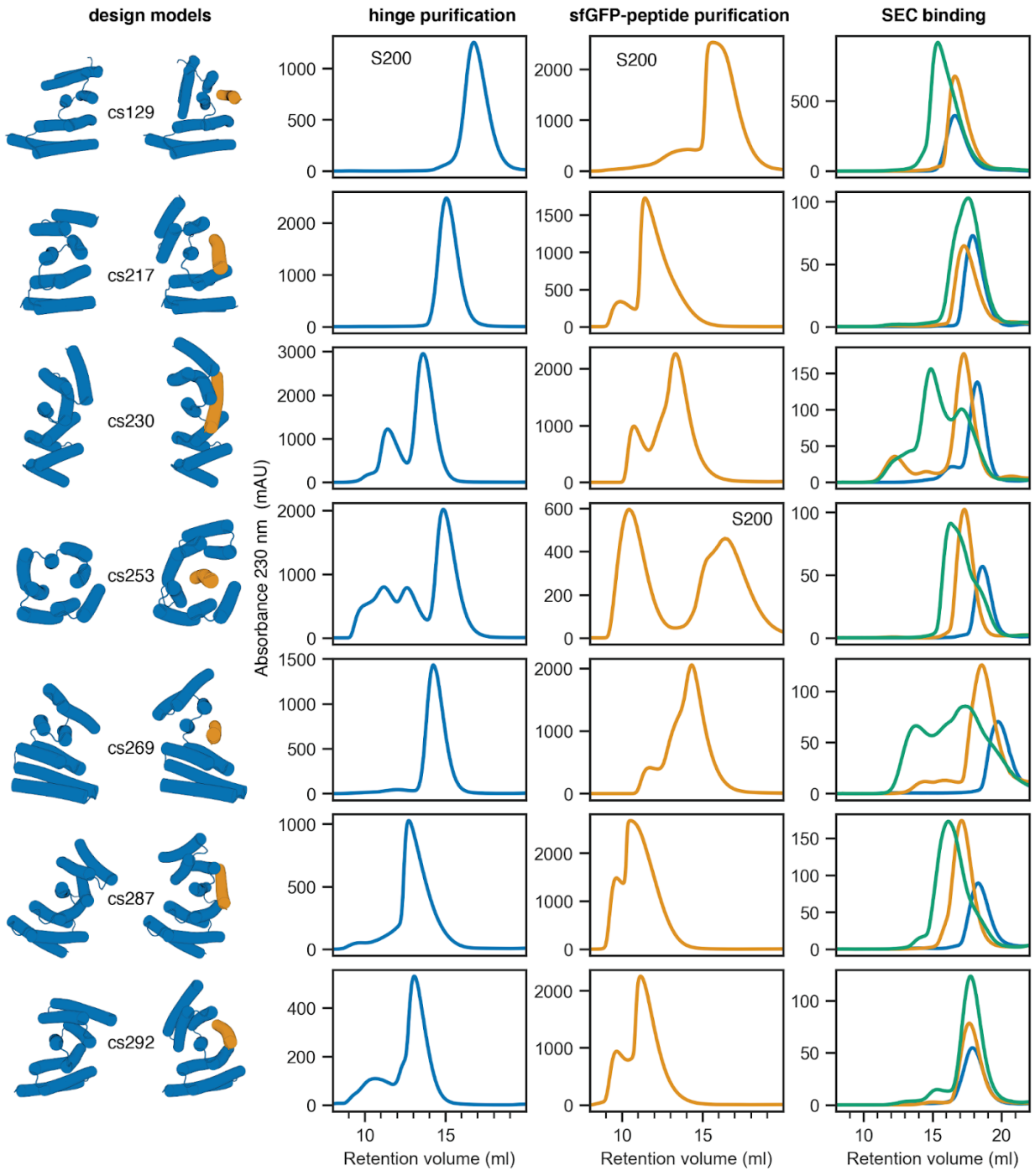

Figure S3: SEC Characterization of additional hinges not shown in Figure 2.

Purification runs of hinges (left, blue) and sfGFP-peptide fusions (center, orange) were performed on Superdex 75 Increase 10/300 GL columns (Cytiva) except for traces with label “S200” that were run on a Superdex 200 Increase 10/300 GL column. SEC binding

experiments (blue: hinge, orange: sfGFP-peptide fusion, green: a mixture of both) were performed on Superdex 200 Increase 10/300 GL columns (Cytiva).

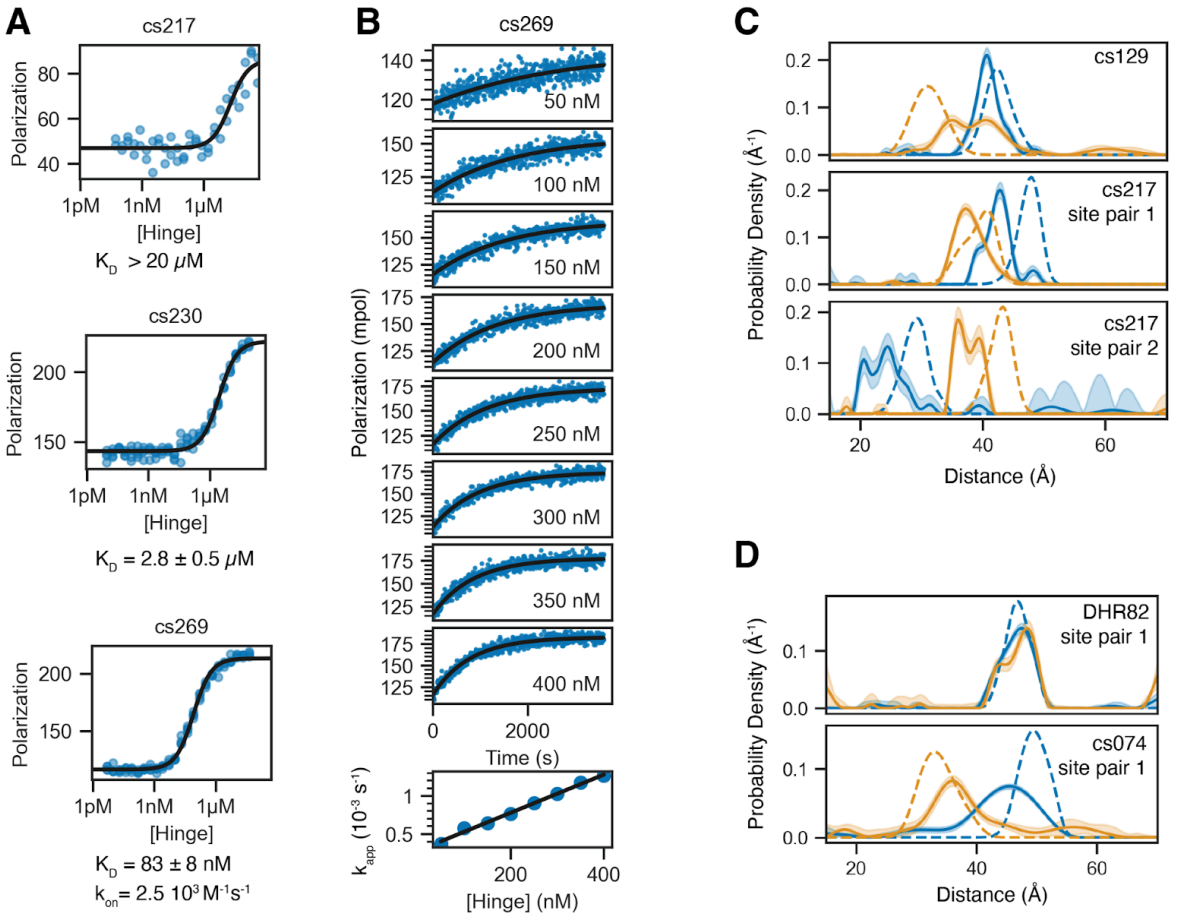

Figure S4: Additional characterization experiments.

A) FP-based titration using TAMRA-labeled peptide at 1 nM and varying concentrations of the hinge. For cs217 the signal never reached a plateau, thus only an approximate lower bound for the K_D can be estimated. **B)** FP-based kinetics experiment using 5 nM TAMRA-labeled peptide cs269B and varying concentrations of hinge cs269 as indicated by plot labels. Top: All kinetic traces were fitted using a single-exponential equation (black lines). Bottom: Apparent rate constants (blue points) from the single exponential fits plotted against the hinge concentration and fitted as linear (black lines). The slope of the linear fit gives the observed on rate k_{on}. **C)** DEER distance distributions for hinges cs129 (one site pair) and cs217 (two site pairs) in absence of peptide (blue) and with excess peptide (orange). Dashed lines are simulated distributions based on design models, solid lines are fitted to the experimental data, and shaded areas are confidence intervals of these fits. **D)** DEER control experiment using DHR82 (the parent of hinge

cs074) labeled at the same sites as cs074. Top: DHR82 shows a sharp peak that does not shift upon the addition of the peptide cs074B. Bottom: The distance distribution for cs074 that is shown in Figure 2 is shown again as a comparison to the DHR82 distribution. The hinge cs074 in absence of peptide shows a slightly broader peak than the parent DHR, suggesting increased flexibility and conformational breathing.

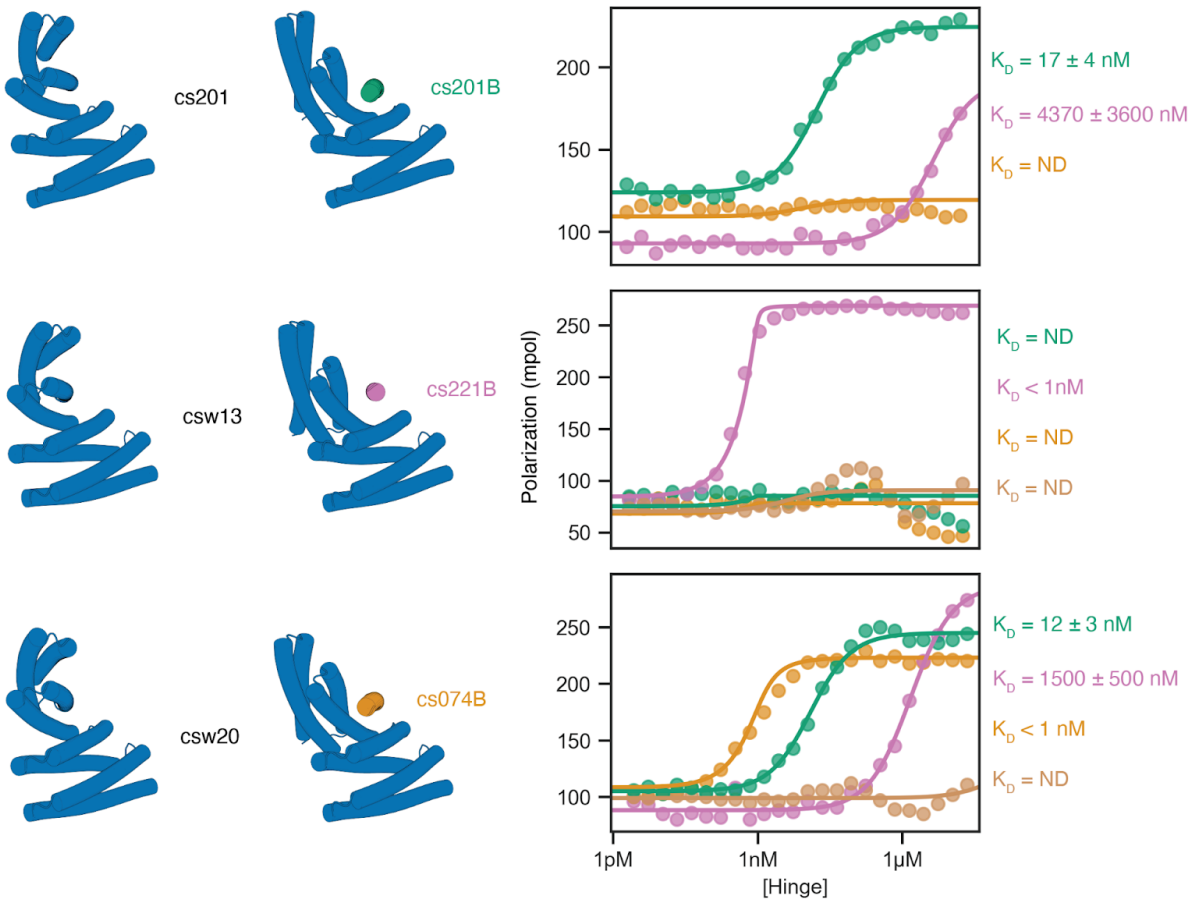

Figure S5: Redesign of hinges to bind other target peptides.

An experimentally tested hinge (top, cs201) is redesigned using a one-sided two-state design approach to bind to different target peptides. Hinges csw13 and csw20 are designed to bind to peptides cs221B and cs074B, respectively while having a backbone conformation that is similar to the parent cs201 that binds cs201B. Left: design models, right: FP titrations using 1 nM peptide (green: cs201B, pink: cs221B, orange: cs074B, brown: js007B) and varying concentrations of hinge (from top to bottom: cs201, csw13, csw20). csw13 is an example of a successful orthogonal redesign that specifically binds the new target peptide cs221B while not binding the parent target peptide cs201B or the off-target peptides cs074B or js007B. csw20 is a less orthogonal example that binds the

new target peptide cs074B most strongly but still binds to the parent-target peptide cs201B and to the off-target peptide cs221B albeit weaker than the new target.

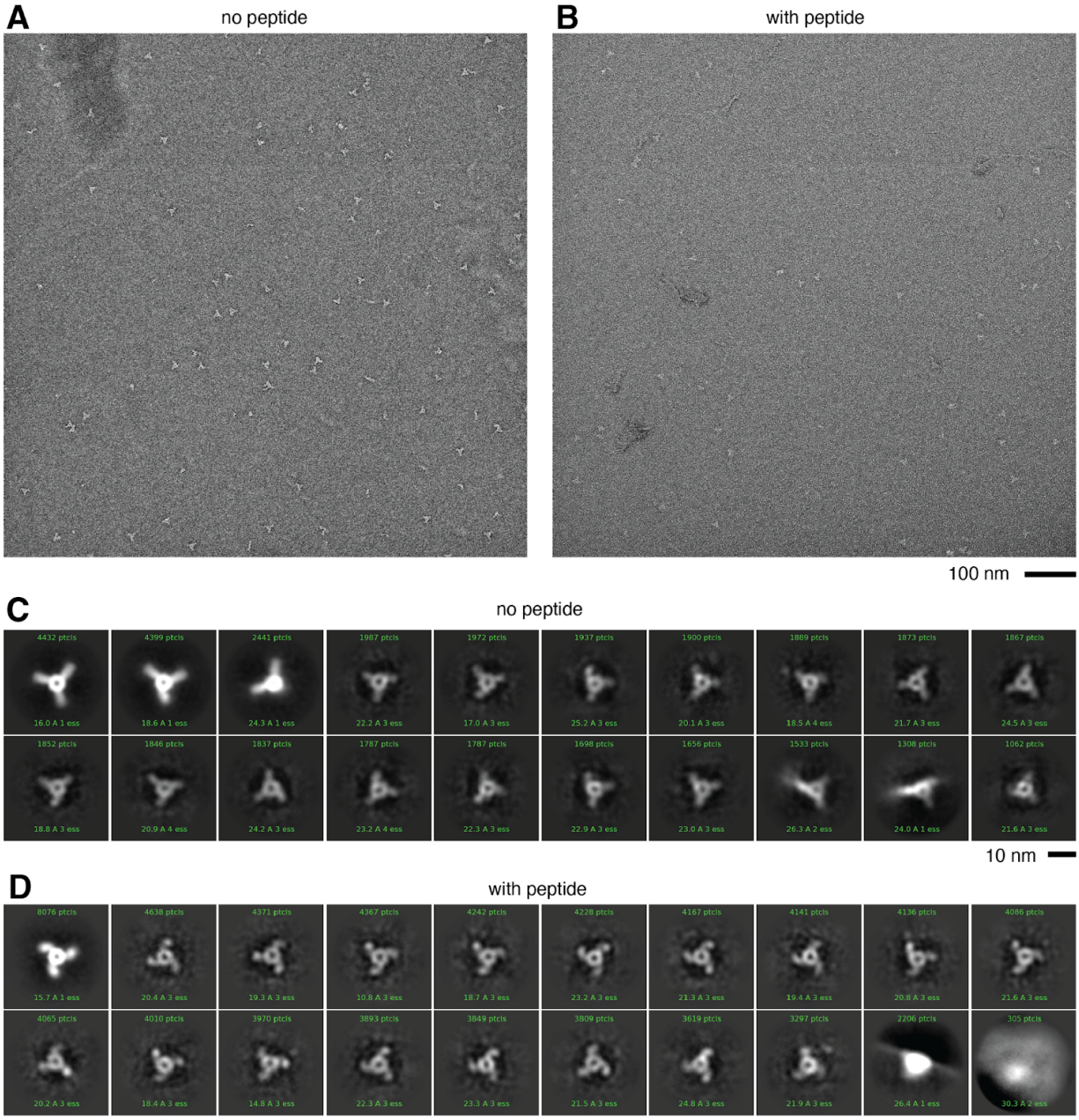

Figure S6: negative stain electron microscopy on hinge-armed trimers.

A, B) Field-of-view electron micrographs of hinge-armed trimers in the absence (A) or presence (B) of peptide cs221B. **C, D)** Class averages from one round of classification (20 classes) using particles obtained from hinge-armed trimers in the absence (C) or presence (D) of peptide cs221B.

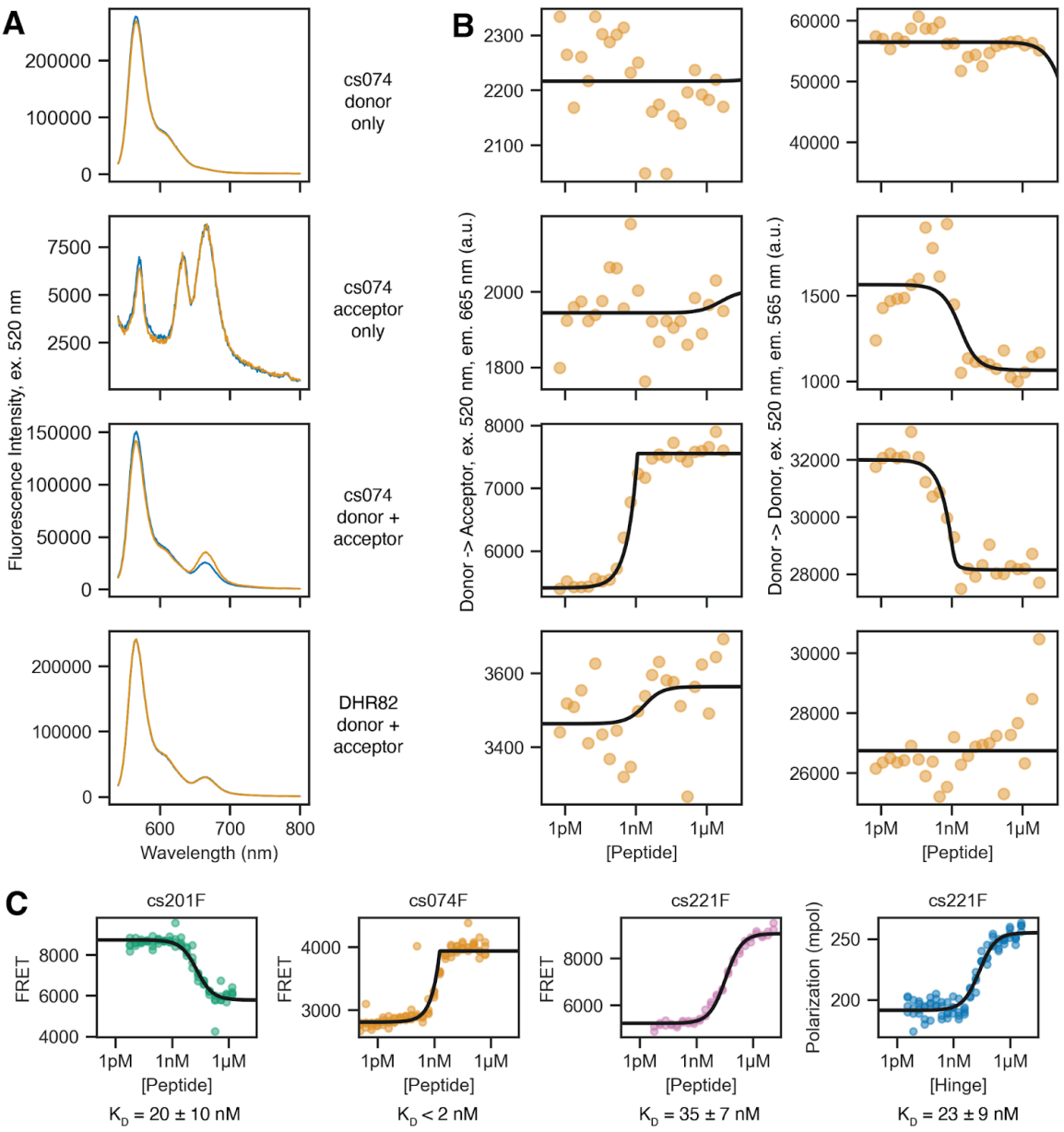

Figure S7: Additional FRET experiments.

A) Fluorescence spectra of proteins (2nM) without (blue) or with (orange) peptide cs074B (5 nM for cs074 variants, 2 µM for DHR82). Proteins from top to bottom: cs074F labeled with AlexaFluor 555 (= donor only), cs074F labeled with AlexaFluor 647 (= acceptor only), cs074F labeled with a 1:1 mixture of AlexaFluor 555 and AlexaFluor 647 (= donor+acceptor), DHR82 labeled with a 1:1 mixture of both dyes. DHR82 shows no significant change in FRET upon the addition of the peptide. **B)** Titrations of the same proteins as in A at 1.2 nM and peptide cs074B at varying concentrations. Left: Acceptor emission upon donor excitation, right: donor emission upon donor excitation. **C)**

Replicate titrations of the FRET-labeled extended hinges shown in Figure 4 (2 nM hinge), and FP titration of the unlabeled extended hinge cs221F (right, 1 nM TAMRA-peptide cs221B).

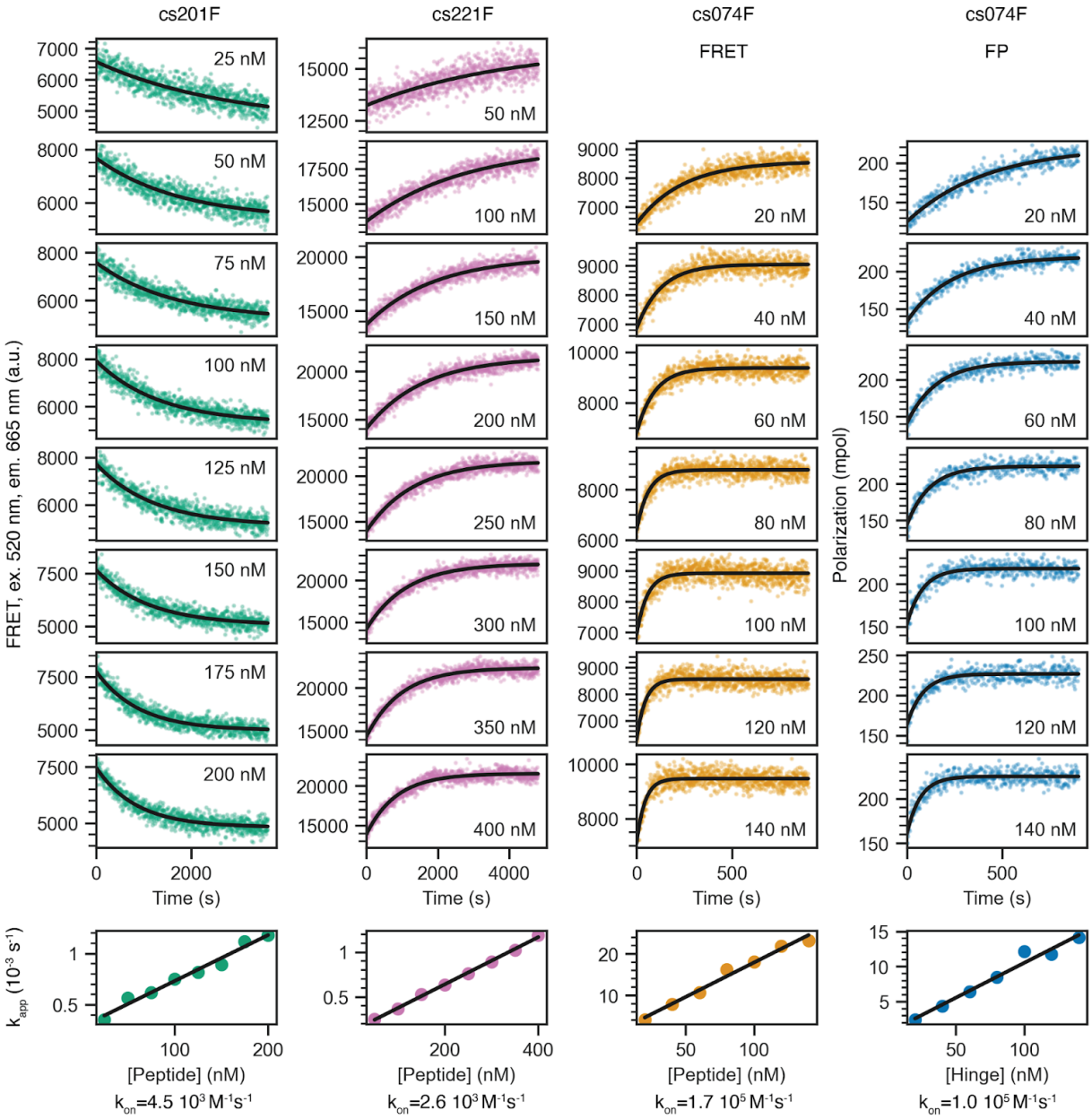

Figure S8: Full kinetics measurements of the extended hinges shown in Figure 4.

Columns 1-3: FRET kinetics using extended hinges labeled with AlexaFluor 555 and AlexaFluor 647 at a constant concentration (2 nM for cs201F and cs074F, 5 nM for

cs221F) and corresponding peptides at varying concentrations. Column 4: FP kinetics using TAMRA-labeled peptide cs074B at 2nM and extended hinge cs074F at varying concentrations. All kinetic traces (rows 1-8) were fitted using a single-exponential equation (black lines). Row 9 shows apparent rate constants from the single exponential fits plotted against the hinge concentration and fitted as linear (black lines). The slope of the linear fit gives the observed on rate k_{on}.

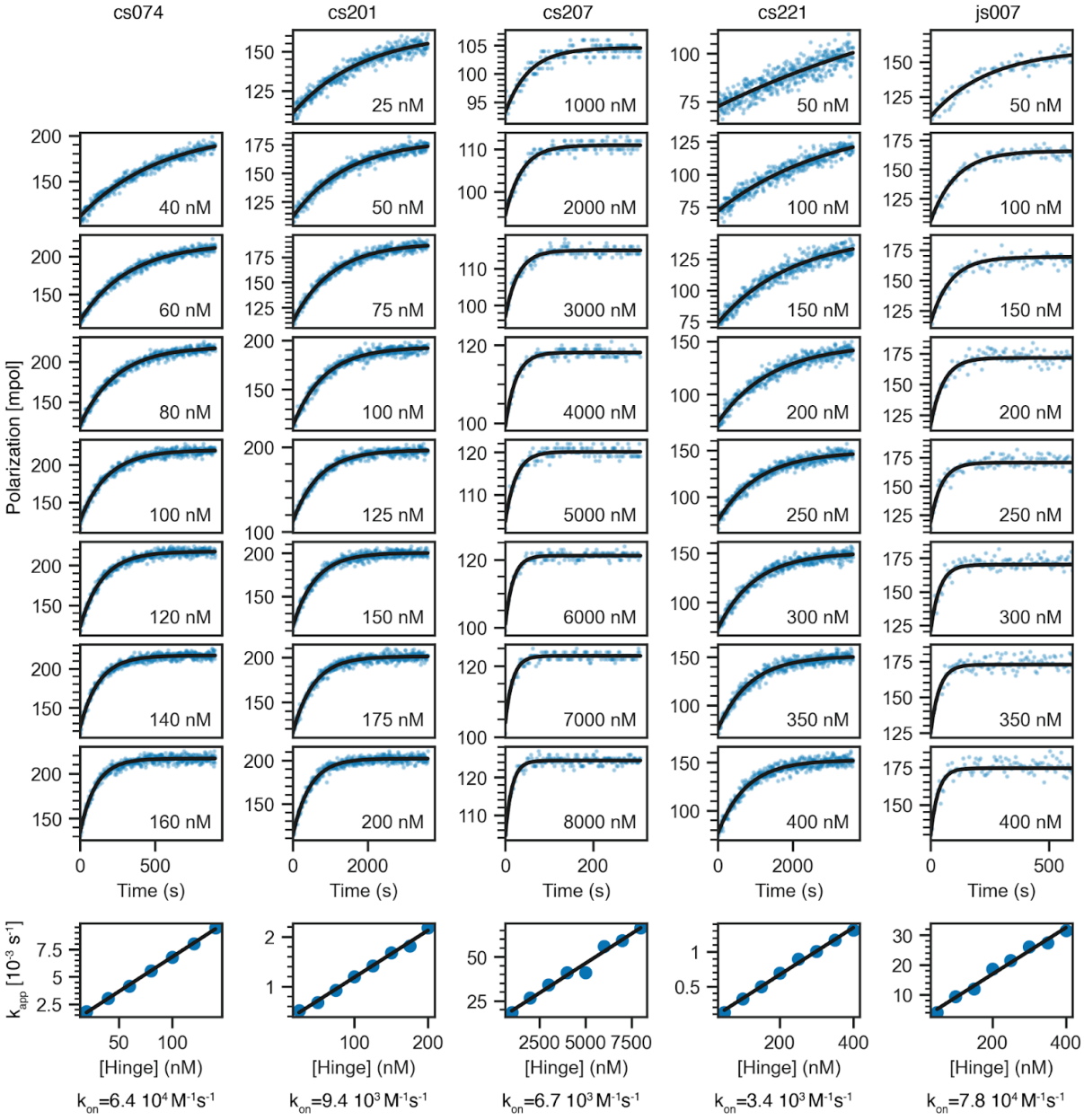

Figure S9: FP kinetics measurements of hinges shown in Figure 2.

Rows 1-8: TAMRA-labeled peptide at a constant concentration (2 nM for cs074, 50 nM for cs207, 5 nM for cs201, cs221, and js007) was mixed with the hinge at varying

concentrations (labels in each plot indicate the hinge concentration for the corresponding experiment). All kinetic traces were fitted using a single-exponential equation (black lines). Row 9: Apparent rate constants from the single exponential fits plotted against the hinge concentration and fitted as linear (black lines). The slope of the linear fit gives the observed on rate k_{on}.

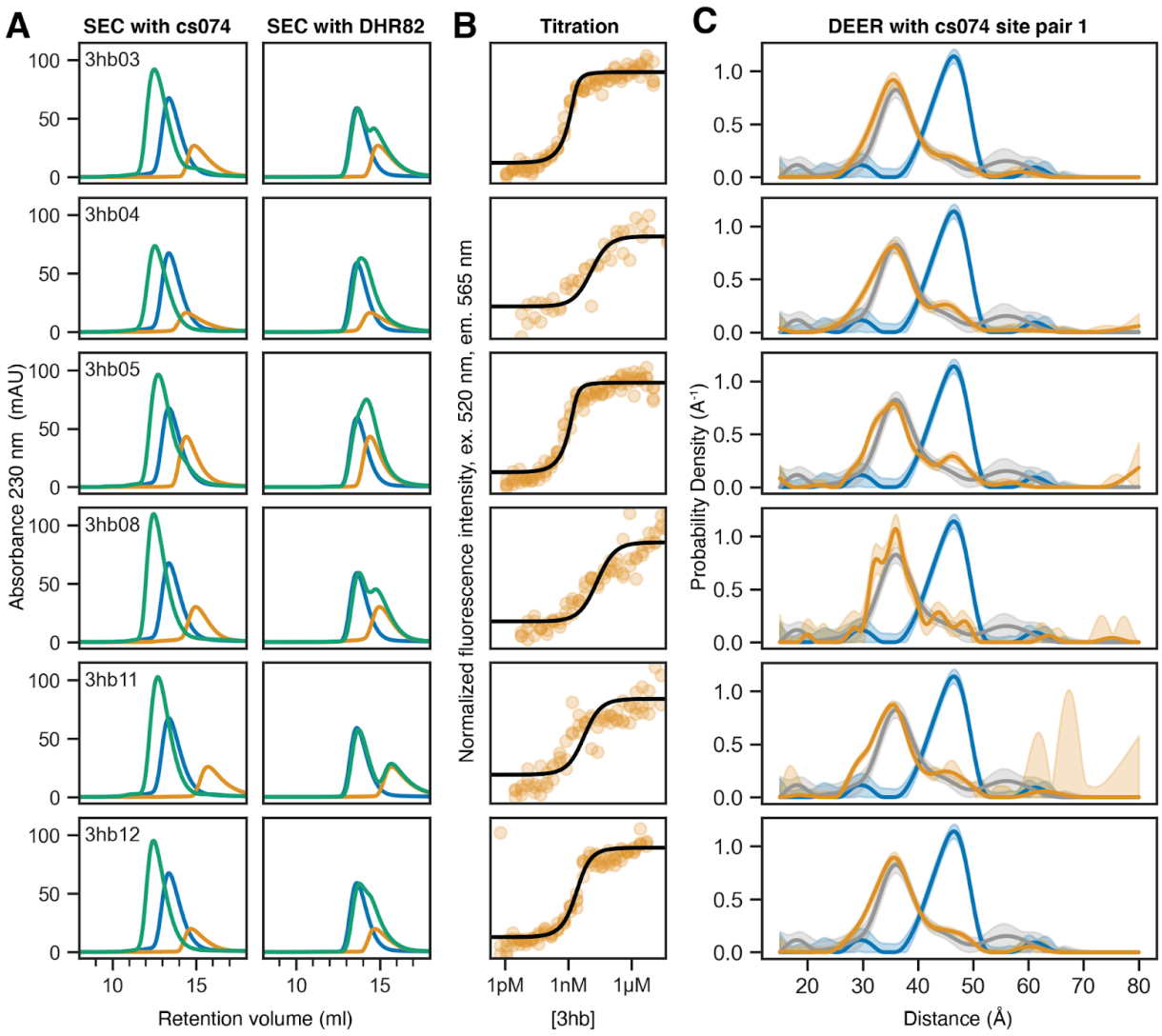

Figure S10: Additional three-helix bundles that bind cs074.

A) SEC binding experiments. Left: Overlaid chromatograms of three-helix bundles (orange), hinge cs074 (blue), and mixtures of both (green) show clear binding. Right: Overlaid chromatograms of three-helix bundles (orange), parent DHR82 (blue), and mixtures of both (green) show no significant binding. **B)** FRET-based titration

experiments using 2 nM labeled cs074F and varying concentrations of 3hb show that the 3hb designs bind to the target hinge with nanomolar affinities and cause a conformational change. **C)** DEER experiments with MTSL-labeled cs074 show that the 3hb designs cause the same conformational change as the original peptide cs074B. Blue: hinge only, gray: hinge + peptide cs074B, orange: hinge + 3hb. Shaded areas indicate 95% confidence intervals.

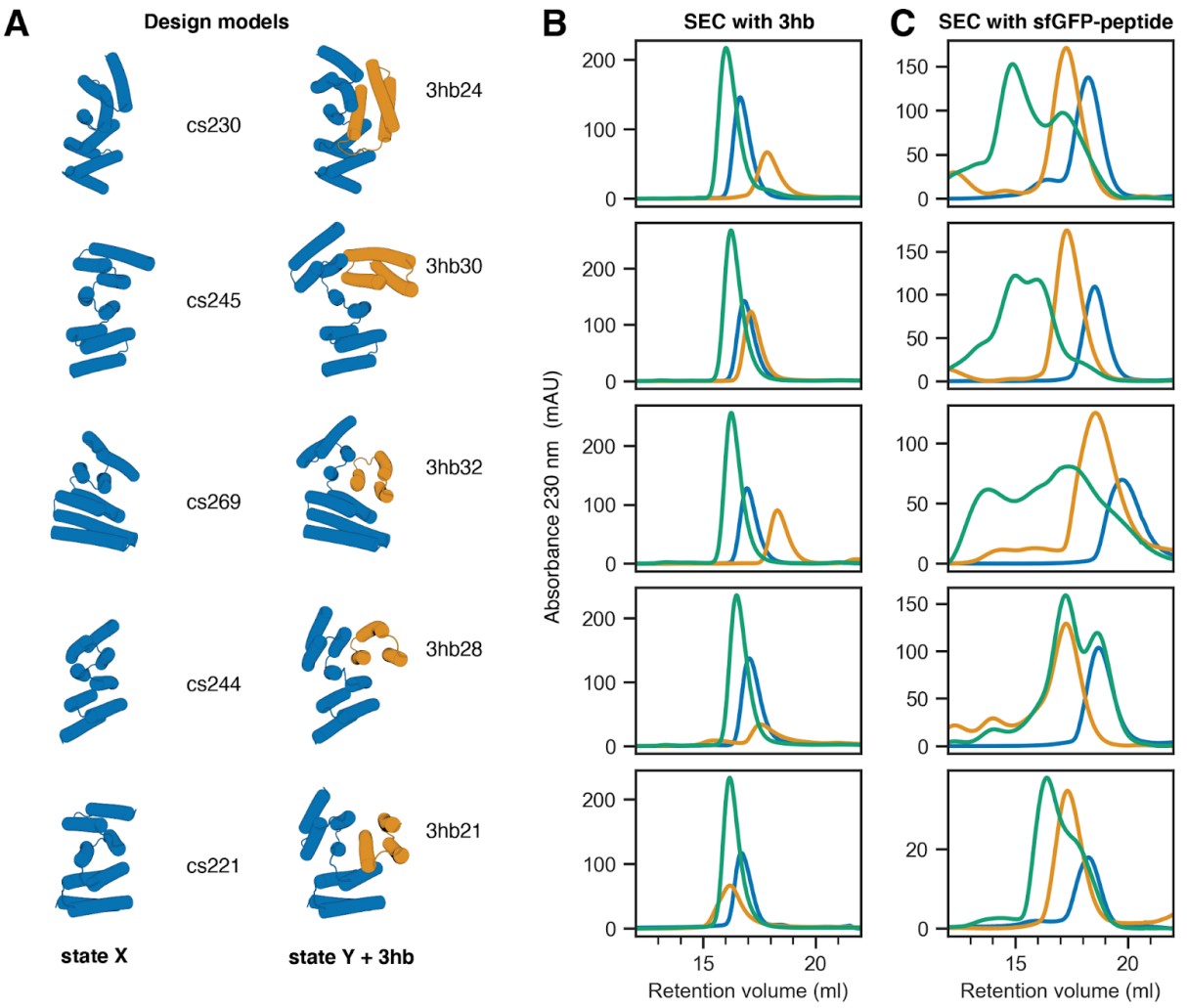

Figure S11: Additional three-helix bundles.

A) Models of hinges (blue) in state X and in state Y bound to a three-helix bundle (3hb, orange). **B)** SEC binding experiments of hinge (blue), 3hb (orange), and a mixture of both (green) show clear monodisperse complex peaks. **C)** SEC binding experiments of the same hinges with the original peptides fused to superfolder green fluorescent protein (sfGFP). For cs230, cs245, and cs269 the hinge-peptide complex shows higher-order peaks while the corresponding hinge-3hb peaks look much cleaner. For cs244, the original peptide showed no clear binding in the SEC experiment, while 3hb28 showed clear binding.

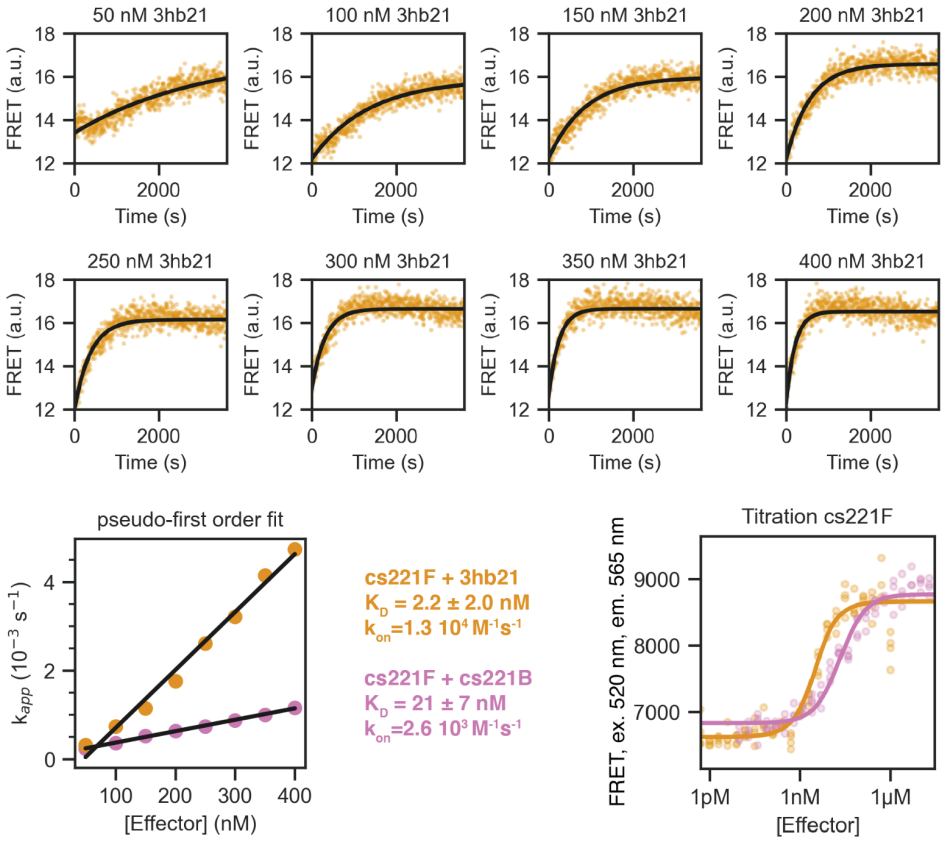

Figure S12: FRET-based quantitative analysis of the interaction between cs221F and 3hb21.

Individual kinetic traces were obtained using a constant hinge concentration of 5 nM and varying 3hb concentrations as indicated by plot labels. Single exponential fits give apparent rate constants that increase linearly with the total 3hb concentration (orange points in pseudo-first order plot). The linear fit of k_{app} against 3hb concentration gives an observed on rate that is 5 times faster than the observed on rate of the original peptide (pink points). FRET titration of 2 nM hinge cs221F and varying concentrations of 3hb (orange) gives a KD below 2 nM which is at least 20 times stronger than the KD of the original peptide(pink).

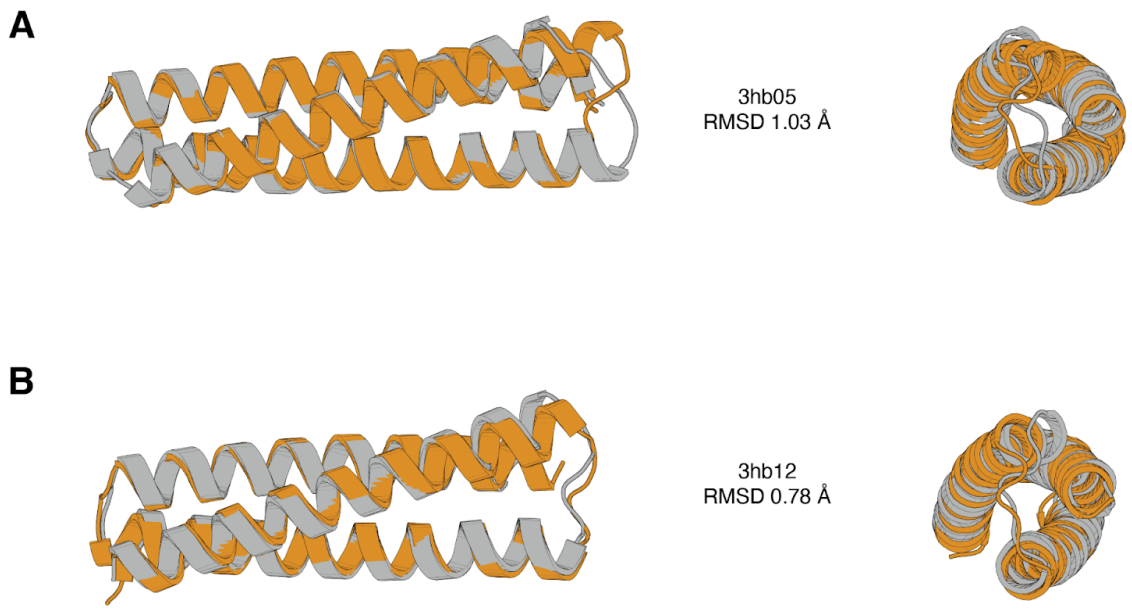

Figure S13: Structural validation of three-helix bundles.

A, B) Overlay of design model (orange) and crystal structure (gray) in side view (left) and top view (right) for designs 3hb05 (A, also shown in Figure 4E) and 3hb12 (B).

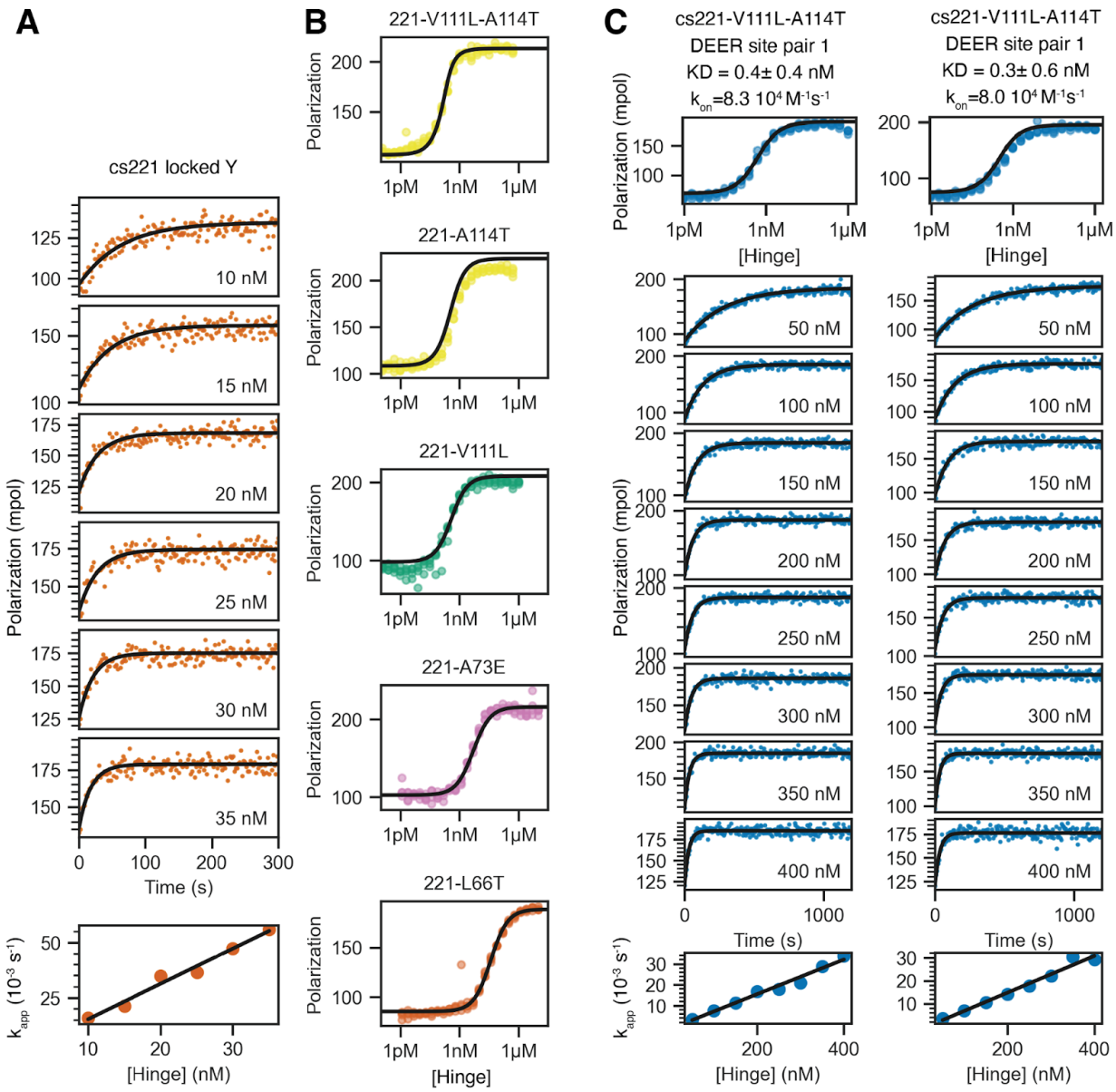

Figure S14: Additional data supplementing Figure 5.

A) Full FP kinetics experiment for the cs221 locked Y hinge variant shown in Figure 5B. Individual kinetic traces were obtained using 1 nM TAMRA-labeled peptide cs221B and varying hinge concentrations as indicated by plot labels. Single exponential fits give apparent rate constants that increase linearly with the total hinge concentration (orange points in pseudo-first order plot). **B)** Full-range individual plots of the FP titrations shown in Figure 5C. **C)** FP titrations and kinetics for the MTSL-labeled variants of hinge cs221-V111L-A114T. The spin labels have no measurable effect on affinity or association kinetics.

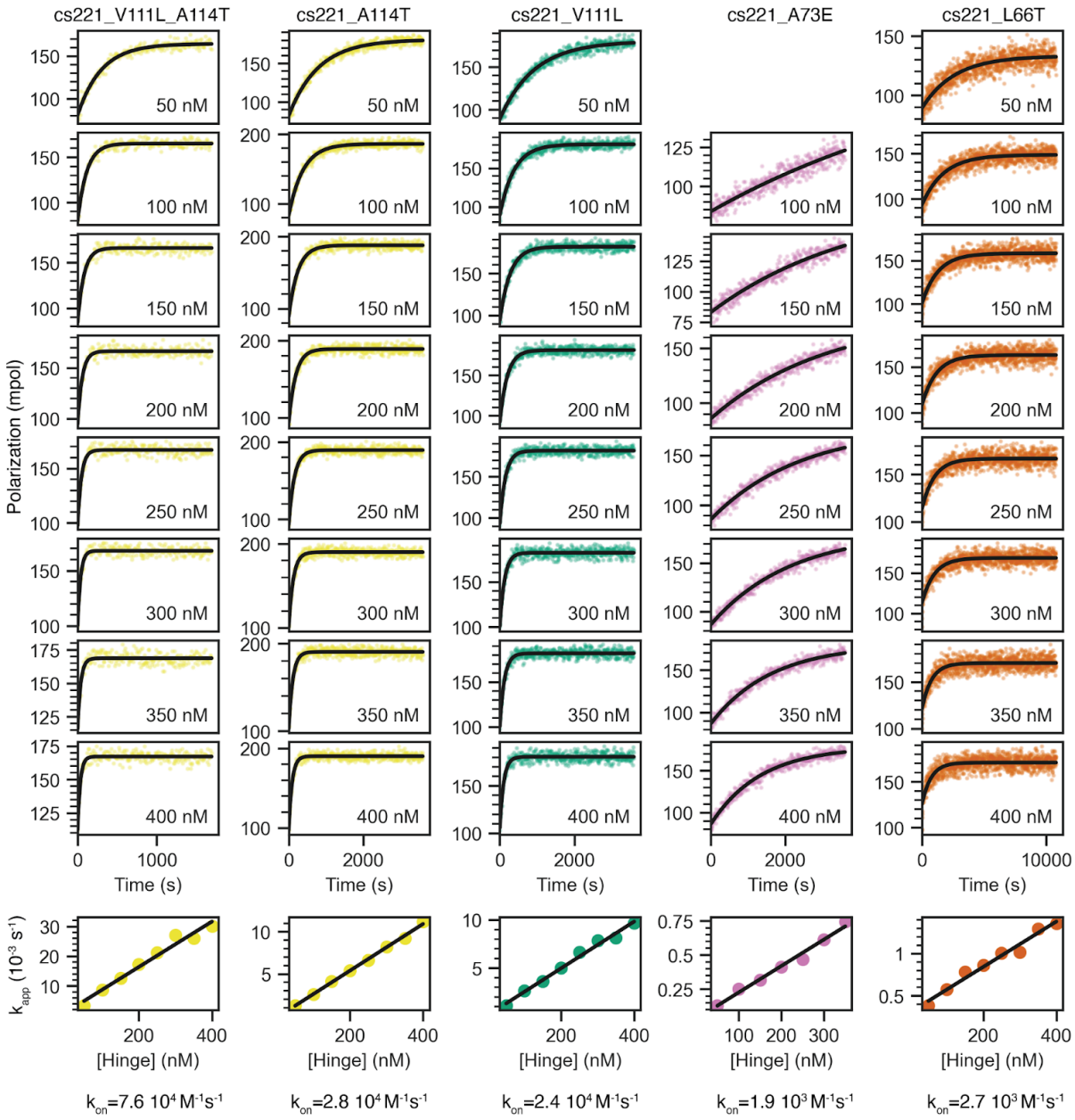

Figure S15: Full FP kinetics experiments for the cs221 mutants shown in Figure 5C.

Rows 1-8: TAMRA-labeled peptide cs221B at a constant concentration of 5 nM was mixed with the hinge at varying concentrations (labels in each plot indicate the hinge concentration for the corresponding experiment). All kinetic traces were fitted using a single-exponential equation (black lines). Row 9: Apparent rate constants from the single exponential fits plotted against the hinge concentration and fitted as linear (black lines). The slope of the linear fit gives the observed on rate k_{on}.

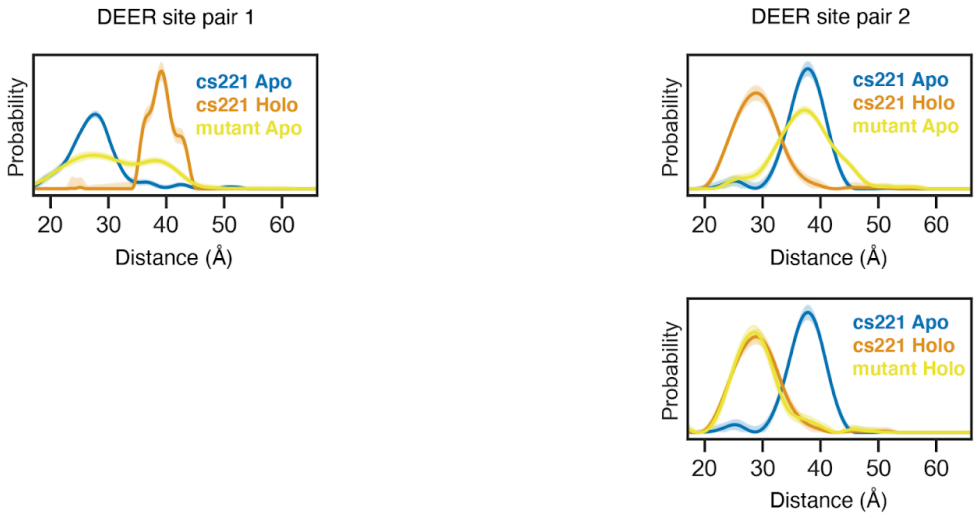

Figure S16: DEER data on hinge variants that populate both states in absence of peptide.

Distance distributions obtained from DEER experiments with the original hinge cs221 are shown in absence of peptide (blue) and in presence of peptide (orange) as well as of the double mutant cs221-V111L-A114T in absence of peptide (mutant Apo, yellow) and in presence of peptide (mutant Holo, yellow).

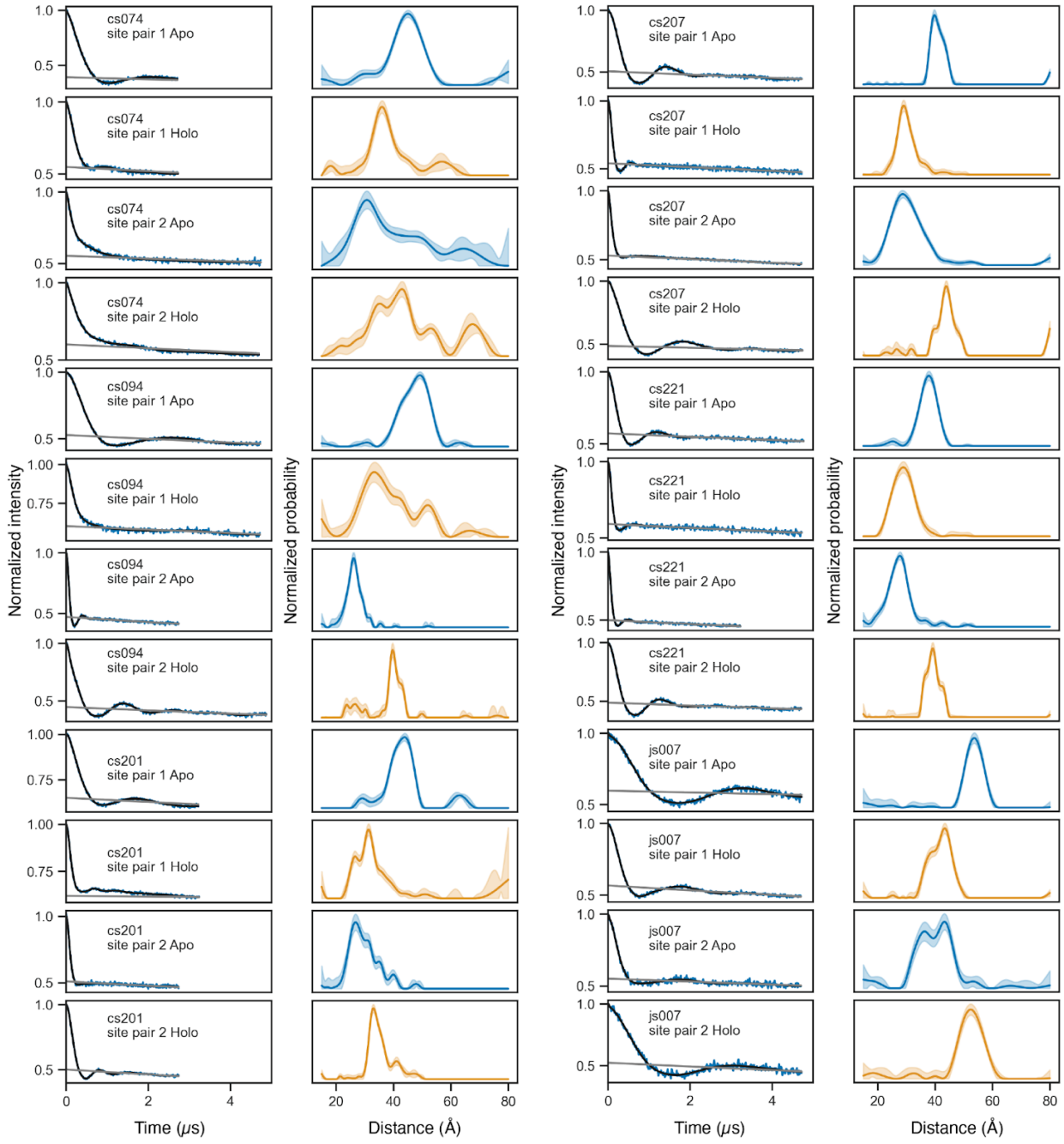

Figure S17: Additional data for DEER experiments.

Raw DEER traces (blue), foreground fits (black), and background fits (gray) are shown on the left. Distance distributions are shown on the right colored by state (apo: blue, holo: orange) with 95% confidence intervals shown as semi-transparent bands.

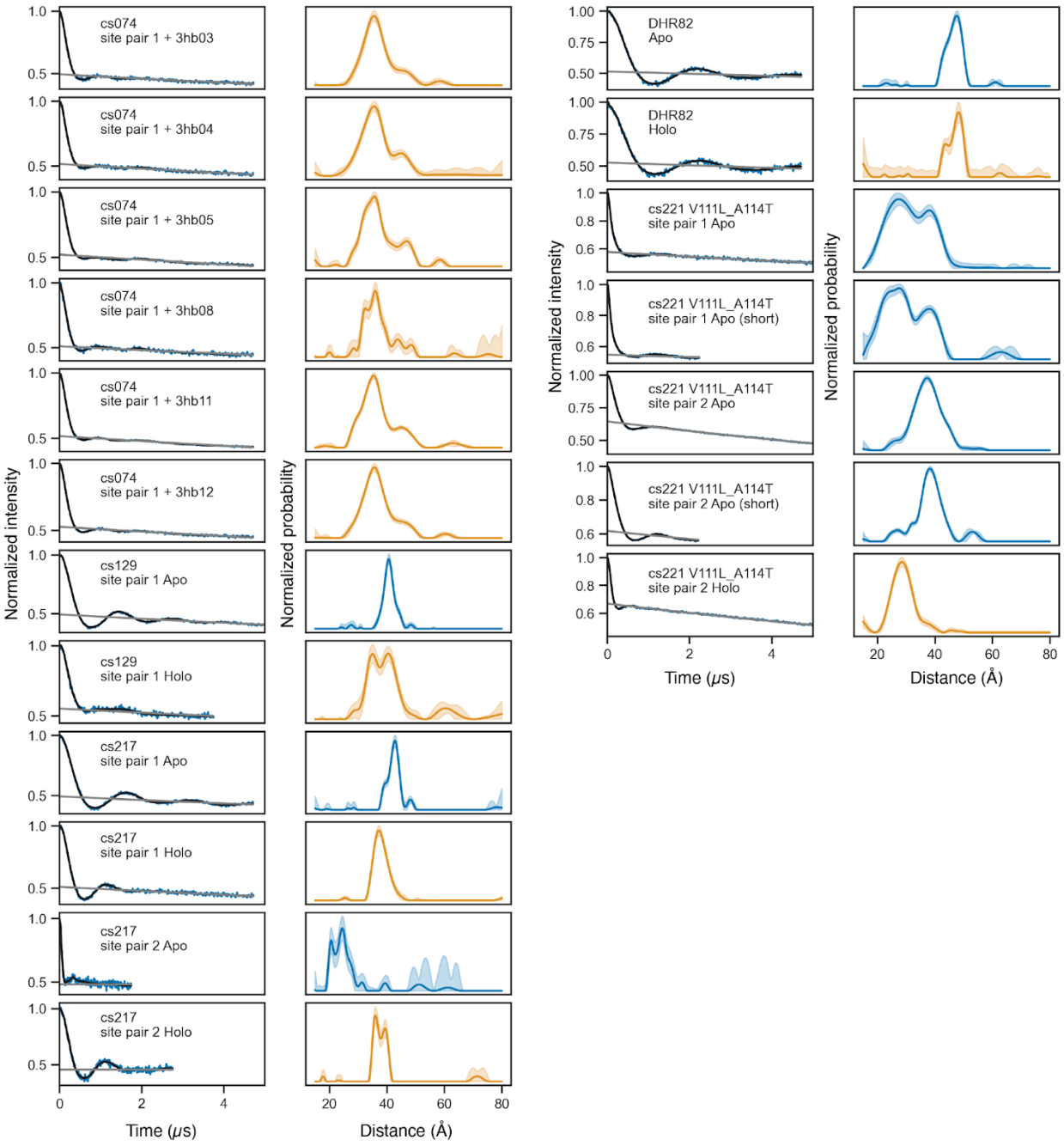

Figure S18: Additional data for DEER experiments.

Raw DEER traces (blue), foreground fits (black), and background fits (gray) are shown on the left. Distance distributions are shown on the right colored by state (apo: blue, holo: orange) with 95% confidence intervals shown as semi-transparent bands.

Supplementary Note 1: Kinetic model for peptide-binding hinges

Our kinetic model comprises the three states X, Y, and YP (state Y bound to the peptide), assuming that a state X bound to the peptide would be sterically unfeasible. The model considers four microscopic rate constants: k_1 and k_{-1} describe the conformational change from X to Y and from Y to X, respectively; k_2 and k_{-2} respectively describe the association and dissociation of state Y and the peptide. We assume that direct transitions from state X to the complex YP or from YP to X do not occur, but always involve Y as intermediate. Given the model of two coupled equilibria as shown in Figure 5B, the fundamental rate laws for the individual states are

$$\frac{d[X]}{dt} = k_{-1}[Y] - k_1[X] \quad (1)$$

$$\frac{d[Y]}{dt} = k_1[X] - k_{-1}[Y] - k_2[Y][P] + k_{-2}[YP] \quad (2)$$

$$\frac{d[YP]}{dt} = k_2[Y][P] - k_{-2}[YP] \quad (3)$$

With [X], [Y], [P] and [YP] being the concentrations for hinge in state X, hinge in state Y, free peptide, and state Y - peptide complex, respectively.

In our FRET system the measured intensity I (acceptor emission upon donor excitation) can be described as

$$I = [X]I_X + ([Y] + [YP])I_Y = [X]I_X + ([H]_{total} - [X])I_Y$$

$$I = [H]_{total}I_Y + [X](I_X - I_Y) \quad (4)$$

With $[H]_{total}$ = [X] + [Y] + [YP] and assuming one intensity I_X for hinges in state X and another intensity I_Y for hinges in state Y or the state Y-peptide complex.
We performed FRET kinetics measurements using a constant concentration of labeled hinge $[H]_{total}$ and varying peptide concentrations $[P]_{total}$ = [P] + [YP] that were at least 10-fold higher than the hinge concentration ($[P]_{total}$ >> $[H]_{total}$). The resulting data can be fit with a single exponential equation

$$I(t) = c_1 + c_2 e^{-k_{app,FRET}\,t}$$

With constants c_1 and c_2 and an apparent rate constant $k_{app,FRET}$. Substituting I from (4) gives

$$[H]_{total} I_Y + [X](I_X - I_Y) = c_1 + c_2 e^{-k_{app,FRET} t}$$

Elimination of constant terms leaves

$$[X](I_X - I_Y) = c_2 e^{-k_{app,FRET} t}$$

$$[X] = \frac{c_2}{(I_X - I_Y)} e^{-k_{app} t}$$

Which leads to

$$\frac{d[X]}{dt} = -k_{app,FRET} \frac{c_2}{(I_X - I_Y)} e^{-k_{app,FRET} t} = -k_{app,FRET} [X]$$

Using the fundamental rate law (1) we get

$$-k_{app,FRET} [X] = k_{-1} [Y] - k_1 [X]$$

Solving for $k_{app,FRET}$ gives

$$k_{app,FRET} = k_1 - \frac{[Y]}{[X]} k_{-1}$$

Which shows that the ratio [Y]/[X] is constant over time, meaning that X and Y are in a pre-equilibrium with exchange rates that are faster than the association rates we observe.

In our FP system, we measure polarization values (pol) that depend on the ratio of bound peptide:

$$pol = \frac{[P]}{[P]_{total}} pol_{free} + \frac{[YP]}{[P]_{total}} pol_{bound} = \frac{[P]_{total} - [YP]}{[P]_{total}} pol_{free} + \frac{[YP]}{[P]_{total}} pol_{bound}$$

$$pol = pol_{free} + (pol_{bound} - pol_{free}) \frac{[YP]}{[P]_{total}}$$

with polarization values pol_{bound} and pol_{free} for bound and unbound peptide, respectively and with $[P]_{total} = [P] + [YP]$. The observed change in polarization upon binding can thus be described as

$$\frac{d[pol]}{dt} = \frac{d[YP]}{dt} \frac{pol_{bound} - pol_{free}}{[P]_{total}}$$

Using $[P]_{total} = [P] + [YP]$ the fundamental rate law (3) can be rewritten as

$$\frac{d[YP]}{dt} = k_2 [Y][P] - k_{-2} [YP] = k_2 [Y] ([P]_{total} - [YP]) - k_{-2} [YP]$$

$$\frac{d[YP]}{dt} = k_2[Y][P]_{total} - (k_2[Y] + k_{-2}) [YP] \quad (5)$$

Given a large excess of hinge over peptide we can assume that

$$[X] + [Y] = [H]_{total} - [YP] \approx [H]_{total}$$

$$[Y] = F_Y [H]_{total} \quad (6)$$

with F_Y defined as the fraction of hinge that is in state Y in equilibrium:

$$F_Y = \frac{[Y]}{[X]+[Y]} = \frac{K_{XY}}{K_{XY}+1}$$

In equilibrium, $[YP] = [YP]_{eq}$ and $\frac{d[YP]}{dt} = 0$ which gives

$$\frac{d[YP]}{dt}(eq) = k_2[Y][P]_{total} - (k_2[Y] + k_{-2}) [YP]_{eq} = 0$$

$$k_2[Y][P]_{total} = (k_2[Y] + k_{-2}) [YP]_{eq} \quad (7)$$

Substituting (7) in (5) gives

$$\frac{d[YP]}{dt} = k_2[Y][P]_{total} - (k_2[Y] + k_{-2}) [YP] = (k_2[Y] + k_{-2}) [YP]_{eq} - (k_2[Y] + k_{-2}) [YP]$$

$$\frac{d[YP]}{dt} = (k_2[Y] + k_{-2}) ([YP]_{eq} - [YP]) \quad (8)$$

The reaction approaches equilibrium with

$$\frac{d([YP]_{eq} - [YP])}{dt} = -\frac{d[YP]}{d[t]} = -(k_2[Y] + k_{-2})([YP]_{eq} - [YP]) \quad (9)$$

Defining the apparent rate constant for association using (6) as

$$k_{app} = k_2[Y] + k_{-2} = k_2 F_Y [H]_{total} + k_{-2} \quad (10)$$

we can rewrite (9) as

$$\frac{d([YP]_{eq} - [YP])}{dt} = k_{app}([YP]_{eq} - [YP])$$

Solving the differential equation gives the displacement from equilibrium $[YP]_{eq}$ - [YP]:

$$([YP]_{eq} - [YP])(t) = c\, e^{k_{app} t}$$

$$[YP](t) = [YP]_{eq} - c\, e^{k_{app} t}$$

Substituting $[YP](0) = [YP]_0$ at $t = 0$ gives $c = [YP]_{eq} - [YP]_0$ and subsequently

$$[YP](t) = [YP]_{eq} - ([YP]_{eq} - [YP]_0)\, e^{k_{app} t}$$

If $[YP]_0 = 0$ as in our kinetic experiments we get

$$[YP](t) = [YP]_{eq} - [YP]_{eq}\, e^{k_{app} t} \quad (11)$$

Single-exponential fits of our FP kinetics experiments show k_{app} to increase linearly with the total hinge concentration, which we can fit using (10):

$$k_{app} = k_2 F_Y [H]_{total} + k_{-2} = k_{on}[H]_{total} + k_{-2}$$

With an observed on rate $k_{on} = k_2 F_Y$

If we assume that variants of a given hinge, such as mutants or stapled versions, have the same microscopic on rate, k_2, we can estimate F_y by comparing observed on rates for variants a and b:

$$\frac{F_{Y,a}}{F_{Y,b}} = \frac{k_{on,a}}{k_{on,b}}$$

Supplementary Note 2: Details on the final hinge design and screening round

Input scaffolds were chosen from a library of 193 previously designed DHRs. Alternative state generation resulted in 1990 backbones covering all 193 input parents. These backbones were run through two one-state sequence design trajectories (Rosetta FastDesign with backbone movement), yielding two poses per input. Designs were leniently filtered on quality of interfaces between the two domains and between hinge domains and peptide, discarding only designs with very poor interfaces (link). After looping, re-design of the loop region, and filtering against loop clashes and loop length (link), only 41.1% (1636) designs survived these steps. Each of these designs was further single-state designed in the resulting state Y with proteinMPNN using 3 different parameter settings for which residues were allowed to be redesigned. The resulting 96 sequences per proteinMPNN parameter set, as well as the parent sequence designed by Rosetta were folded with AF2-IG for a total of 1636*3*97 (476076) sequences. A gate filter was applied that only returned models with RMSD < 1.5, mean pLDDT > 92 and mean PAE interaction < 5, and 30.5% of sequences passed (145491). These designs were further filtered for mean pLDDT > 93 and pTMscore > 0.85, and clustering by state ID (a combination of unique parent scaffold ID, pivot helix index, residue index shift, and docked helix index), and selecting up to 100 designs per state ID cluster. This procedure resulted in 37185 designs from 994 state IDs representing 173 parent DHRs. These designs were next paired with a state X (the corresponding original DHR backbone) by relooping the original parent scaffold with the Blueprint builder as described, and filtered leniently based on Rosetta metrics (link). 85.6% (31823) designs passed this stage.

Next, Rosetta MSD was run, generating 2 outputs per input. Again, lenient Rosetta metric filters were used to discard bad designs (link), resulting in 77.3% (49208) passing. Again, each of these designs was further multistate designed with MPNN-MSD using 3 different parameter settings for which residues were allowed to be redesigned. These designs were clustered again, and up to 50 designs per cluster were selected. A total of 28482 backbones were folded with the effector sequence (state Y) with 1 Rosetta MSD and 96 MPNN-MSD sequences each. A gate filter was applied that only returned models with RMSD < 1.5, mean pLDDT > 92 and mean PAE interaction < 5. 4.9% of the Rosetta sequences (1396 sequences) and 4.3% of the proteinMPNN sequences (118181) passed these filters. However, in the resulting distributions that passed the gate filter, the Rosetta designs had on average lower pLDDT (94.1+/-1.0) and higher mean PAE (4.3+/-0.4) than the proteinMPNN designs (94.5+/-1.1) and (4.1+/-0.5) respectively (see Figure S19 for distributions). The difference in numbers of sequences folded (1 vs. 96) was also reflective of the difference in computational efficiency for the respective sampling methods, with the Rosetta-MSD protocol taking 0.5-3 hours and 6GB RAM per backbone for 2 sequences and the MPNN-MSD protocol taking 0.2-1 hour and 5GB RAM per backbone for 96 sequences. Before folding the hinge backbones without the effector sequence (state X), a final additional filtering step selected designs that predicted with mean pLDDT > 93. This resulted in a total of 107311 designs to be folded. A gate filter was applied that only returned models with RMSD < 1.5 and mean pLDDT > 93. 63.2% (67820) designs passed this gate filter.

Prior to final filtering and ordering, designs were additionally filtered on pTMscore > 0.8, resulting in a final total of 60601 designs. These designs were further filtered after optimizing the peptide sequences to meet length (<=28 aa) and isoelectric point (pI < 5) cutoffs, resulting in 48561 designs. After filtering on many Rosetta metrics (link), a final subset of 9499 designs were considered "orderable." This "orderable" set comprised 305 state IDs, and 53.9% (104/193) input parent scaffolds. From these, the final order of 72 hinges was selected by clustering by state ID, and out of the resulting 305 clusters, selecting 72 different state IDs randomly, and selecting for each state ID a sequence. At this point, sequences were selected preferentially if they included a TRP or TYR residue, for purposes of concentration determination.

While 46 of 72 hinges expressed solubly and were predominantly monomeric, only 14 out of 72 sfGFP-peptide fusions were soluble and showed a significant monomer peak on SEC. Redesigning the non-binding peptide surface of 24 peptides yielded 14 more soluble monomeric sfGFP-peptide fusions. Of the 20 designs for which both components were soluble and monomeric, 9 showed binding as evaluated by SEC mixing experiments, and 3 showed a monodisperse complex peak.

Supplementary Table 1. Crystallographic data collection and refinement

	3hb05 (8FIH)	**3hb12 (8FVT)**	**cs074AB (8FIT)**	**cs207A (8FIN)**	**cs207AB (8FIQ)**
Data Collection					
Space group	P 21 21 2	I 4	P1	P 21 21 21	P 43 21 2
Cell dimensions					
a, *b*, *c* (Å)	66.01, 119.75, 38.80	67.67, 67.67, 82.37	44.11, 45.36, 61.80	22.86, 81.73, 154.94	71.88, 71.88, 127.64
α, *β*, *γ* (°)	90, 90, 90	90, 90, 90	109.27, 94.54, 104.09	90, 90, 90	90, 90, 90
Resolution (Å)	44.35 - 2.2 (2.3 - 2.2)	47.86 - 3.07 (3.86 - 3.07)	57.45 - 2.75 (2.9 - 2.75)	43.66 - 2.3 (2.39 - 2.3)	47.72 - 2.66 (2.8 - 2.66)
R_{merge}	0.030 (1.161)	0.098 (0.415)	0.077 (0.428)	0.1154 (0.5973)	0.1361 (4.34)
R_{pim}	0.030 (0.531)	0.035 (0.145)	0.064 (0.369)	0.03218 (0.1575)	0.03788 (1.226)

	3hb05 (8FIH)	**3hb12 (8FVT)**	**cs074AB (8FIT)**	**cs207A (8FIN)**	**cs207AB (8FIQ)**
$I/\sigma(I)$	9.89 (2.16)	17.39 (5.59)	7.58 (0.57)	15.19 (4.77)	11.84 (0.60)
$CC_{1/2}$	0.999 (0.852)	1 (0.978)	0.986 (0.769)	1 (0.99)	1 (0.661)
Completeness (%)	95.27 (72.82)	99.60 (99.48)	89.39 (90.08)	99.85 (99.93)	97.86 (89.38)
Redundancy	1.4 (1.0)	8.8 (9.1)	2.2 (2.3)	13.7 (14.9)	13.9 (13.3)
Refinement					
Resolution (Å)	44.35 - 2.2 (2.3 - 2.2)	47.86 - 3.07 (3.86 - 3.07)	57.45 - 2.75 (2.9 - 2.75)	43.66 - 2.3 (2.39 - 2.3)	47.72 - 2.66 (2.8 - 2.66)
No. reflections	15519 (1653)	30894 (15864)	10934 (1563)	13722 (1431)	9939 (1246)
R_{work} / R_{free}	0.1908 (0.1892)/ 0.2478 (0.2495)	0.2248 (0.2998)/ 0.2728 (0.3158)	0.2457 (0.2901)/ 0.2984 (0.3345)	0.2403 (0.2360)/ 0.2712 (0.3046)	0.2691 (0.4680)/ 0.3067 (0.4463)
No. atoms					
Protein	2490	1518	3871	2659	1538
Water	81	0	4	24	0
Ligand	10	0	0	0	0
Ramachandran Favored/allowed Outlier (%)	98.70/ 1.30/ 0.00	96.81/ 3.19/ 0.00	97.49/ 1.88/ 0.00	99.42/ 0.58 /0/00	99.48/ 0.52/ 0.00
R.m.s. deviations					
Bond lengths (Å)	0.006	0.004	0.002	0.001	0.002
Bond angles (°)	0.560	0.590	0.440	0.300	0.380

	3hb05 (8FIH)	**3hb12 (8FVT)**	**cs074AB (8FIT)**	**cs207A (8FIN)**	**cs207AB (8FIQ)**
$B_{factors}$ ($Å^2$)					
Protein	41.25	85.94	86.88	48.07	108.65
Water	38.54	n/a	75.76	44.70	n/a
Ligand	64.46	n/a	n/a	n/a	n/a

Supplementary Table 2. DEER experimental metadata and parameters

Sample	λ	Scans	Δt (ns)	SNR	t_0 offset (ns)	τ_2 (µs)	SRT (ms)	α	β	site pair
cs074 D1 Holo	0.45	50	11	50	82.8	3	1.53	0.91	0.51	K36R1 Q211R1
cs074 D1 Apo	0.57	85	11	67	97.2	3	1.53	3.10	0.20	K36R1 Q211R1
cs074 D2 Apo	0.44	31	16	45	73.6	5	2.04	2.50	8.60	E19R1 E179R1
cs074 D2 Holo	0.40	39	16	79	99.2	5	2.04	1.03	9.15	E19R1 E179R1
cs129 D1 Holo	0.45	6	12	50	54	4	2.04	0.62	1.13	E25R1 K176R1
cs129 D1 Apo	0.50	94	22	85	70.4	7	2.04	0.13	2.37	E25R1 K176R1
cs094 D1 Apo	0.47	74	16	72	68.8	5	2.04	1.35	1.03	R30R1 E206R1
cs094 D1 Holo	0.40	83	16	43	80	5	2.04	1.70	2.98	R30R1 E206R1
cs094 D2 Holo	0.55	69	16	72	67.2	5	2.04	0.10	1.36	K12R1 E178R1
cs094 D2 Apo	0.53	49	10	92	68	3	2.04	0.11	8.79	K12R1 E178R1
js007 D1 Apo	0.45	31	16	40	96	5	2.04	1.28	6.63	Q10R1 Q219R1
js007 D1 Holo	0.48	34	16	31	81.6	5	2.04	2.77	1.54	Q10R1 Q219R1
js007 D2 Apo	0.40	40	16	34	78.4	5	2.04	1.51	9.24	D60R1 R190R1
js007 D2 Holo	0.43	43	16	74	97.6	5	2.04	0.45	0.38	D60R1 R190R1
cs207 D2 Holo	0.46	25	8	56	86.4	5	2.04	0.40	1.32	K10R1 D150R1
cs207 D2 Apo	0.50	25	16	62	76.8	5	2.04	0.25	0.82	K10R1 D150R1
cs207 D1 Apo	0.49	85	16	122	84.8	5	2.04	0.24	8.68	D25R1 K130R1
cs207 D1 Holo	0.50	69	16	99	59.2	5	2.04	0.15	0.24	D25R1 K130R1
cs217 D1 Apo	0.52	32	6	35	69.6	2	2.04	0.15	0.18	E47R1 R127R1
cs217 D1 Holo	0.54	17	10	37	47	3	2.04	0.13	6.17	E47R1 R127R1
cs217 D2 Apo	0.51	63	16	91	84.8	5	2.04	0.15	8.61	R27R1 E147R1
cs217 D2 Holo	0.49	74	16	57	72	5	2.04	0.48	8.91	R27R1 E147R1
cs217 D1 Apo	0.52	110	8	35	69.6	3	2.04	0.15	0.18	E47R1 R127R1
cs221 D1 Apo	0.50	69	10	98	82	3.5	2.04	0.02	0.63	E43R1 K131R1
cs221 D1 Holo	0.52	72	16	74	88	5	2.04	0.12	0.24	E43R1 K131R1
cs221 D2 Apo	0.43	47	16	62	76.8	5	2.04	0.86	0.25	R23R1 E150R1
cs221 D2 Holo	0.41	48	16	39	76.8	5	2.04	1.50	9.21	R23R1 E150R1

Sample	λ	Scans	Δt (ns)	SNR	t_0 offset (ns)	τ_2 (µs)	SRT (ms)	α	β	site pair
cs201 D1 Holo	0.50	83	8	154	76	3	2.04	0.14	5.42	R35R1 E207R1
cs201 D1 Apo	0.49	9	8	61	77.6	3	2.04	0.46	9.16	R35R1 E207R1
cs201 D2 Apo	0.35	55	12	91	90	3.5	2.04	0.42	9.23	K24R1 E180R1
cs201 D2 Holo	0.36	90	12	86	76.8	3.5	2.04	0.18	2.93	K24R1 E180R1
DHR82P D1 Apo	0.50	238	16	66	81.6	5	2.04	0.60	8.80	K36R1 A211R1
DHR82P D1 Holo	0.49	92	16	43	86.4	5	2.04	0.37	0.46	K36R1 A211R1
cs074 3hb03 D1 Holo	0.51	78	16	75	73.6	5	2.04	1.17	0.60	K36R1 Q211R1
cs074 3hb04 D1 Holo	0.49	21	16	75	81.6	5	2.04	1.18	1.03	K36R1 Q211R1
cs074 3hb05 D1 Holo	0.50	60	16	106	83.2	5	2.04	0.37	8.76	K36R1 Q211R1
cs074 3hb08 D1 Holo	0.49	21	16	63	83.2	5	2.04	0.19	1.19	K36R1 Q211R1
cs074 3hb11 D1 Holo	0.48	236	16	192	80	5	2.04	0.34	0.00	K36R1 Q211R1
cs074 3hb12 D1 Holo	0.47	73	16	102	84.8	5	2.04	0.77	3.49	K36R1 Q211R1
cs221-mut D1 Apo	0.42	158	18	66	81	7	2.04	1.69	8.68	E43R1 K131R1
cs221-mut D2 Apo	0.36	186	18	151	84.6	7	2.04	0.57	9.16	R23R1 E150R1
cs221-mut D2 Holo	0.41	48	16	43	76.8	5	2.04	1.55	6.76	R23R1 E150R1

λ - Modulation depth
Δt - Pump pulse time step
SRT - Shot repetition time
α - Smoothness regularization parameter
β - Compactness regularization parameter

Methods for Multistate Design of Stimulus-Responsive Protein Hinges

Generating and pairing hinge conformational states

We used curated libraries of DHRs as inputs for generation of hinge conformations. The backbone conformation of a given DHR serves as a template for the first conformational state (state X) of the hinge. The alignment-based alternative state generation protocol was implemented using custom PyRosetta functions. To discard backbone arrangements with significant clashes, we mutate the entire backbone to glycine and score the resulting pose with only the Rosetta “fa_rep” scoreterm(*36*), set to a weight of 1.0, with the “beta_nov16” scorefunction. At this stage, the helical peptide was sometimes extended using a similar alignment/shifting strategy to increase the size of the interface in state Y. We used PyRosetta FastDesign with backbone and jump movement to further improve the backbone and sequence around the tripartite interface

between the first hinge domain, the peptide, and the second hinge domain for scoring purposes. At this stage, we discarded any designs where either of the domains or the peptide didn't form good contacts with the other two chains using interface metrics(*69*). We called the blueprint builder(*77*) in PyRosetta to rebuild the loop region between the hinge domains in state X, based on a secondary structure template of the hinge in state Y.

Two-state sequence design

Initially, we tried many different multi-state design (MSD) algorithms in Rosetta. We first tried an approach where we would iterate between conformational states while performing single-state design (SSD) for each state individually while ramping a custom sequence convergence scoreterm between iterations(*78*). We found this method tended to decorate the surface with hydrophobics in positions that had ambiguous residue-level preferences between the conformational states, so we explicitly penalized excessive surface hydrophobics using constraints that calculated spatial aggregation propensity (SAP)(*79*) on the fly during design. We also used a quasisymmetric multistate design approach in PyRosetta, performing design on both states simultaneously while forcing the packer to consider the chemical context of residue positions linked across the states(*80*). This method seemed to have fewer pathologies in terms of positional sequence selection but scaled poorly in terms of computational performance, so we chose not to use it for large-scale sequence design tasks. Ultimately, we extensively used FastDesign with a version of the annealer originally intended and optimized for multi-conformation, sequence-symmetric design(*81*), since it was the easiest to use and scaled well computationally while being easily tunable to avoid the pathologies of the iterative approach. Once we had sampled sequences and backbones with Rosetta we optionally refined the sequences with proteinMPNN(*41*) multistate design (MPNN-MSD). Using a feature intended for homooligomer symmetry(*32*) we tied corresponding residue positions probabilities together across chains and used MPNN to sample up to 96 sequences per pair of backbones. We then could use AF2 initial guess (AF2-IG)(*47*) to predict the structure of the effector-bound complex (state Y) by threading the MPNN-MSD sequences back onto the backbones using mean predicted Local Distance Difference Test scores (pLDDT), RMSD to reference design model, and mean off-diagonal Predicted Aligned Error matrix (PAE interaction) cutoffs of 93, 1.5, and 5, respectively for AF2. Designs that passed these criteria could be predicted again by AF2(*46*, *82*) to check if they folded to the correct closed position (state X) absent the effector sequence. We observed that sequences designed with MPNN-MSD had much better computational success rates and overall metrics. After two-state design, most hinges were soluble and had significant monomer populations, however many of the effector peptides turned out to be insoluble or formed stable homooligomers. In some cases effector peptides were improved by simple redesign of the peptide surface residues away from the hinge interface, or by truncation of the peptide.

Computational filtering

Concerned with the possibility that hinges designed with this process would randomly oscillate between closed and open conformations in the absence of the effector, we tried to implement additional filters to select only the designs that would have our intended behavior for testing. We chose designs where the hinge sequence scored more favorably in Rosetta(*36*) in the closed conformation relative to the open conformation when the peptide was absent, but scored more favorably in the open conformation with the peptide bound in comparison to the sum of the scores of the closed conformation and the peptide alone. Similarly, we required that the solvent-exposed hydrophobicity, (measured by SAP), would decrease in the closed conformation relative to the open conformation when the peptide was absent, and the bound complex would have less exposed hydrophobics compared to the sum of the exposed hydrophobics of the closed conformation and the peptide alone. We also filtered the bound conformation on interface design metrics, including ddG, cms and SASA.(*69*) This pipeline for designing effector-binding hinges was able to generate very diverse outputs, with large differences in changes in shape and size (Figure S1).

One-sided two-state design for swapped peptide targets

To generate swapped-peptide designs, we started from the state X and state Y backbones of cs074, cs201, cs221, and js007, including peptide backbones. Peptide sequences were replaced, using the sequences of cs074B, cs201B, and cs221B. In cases where the peptide backbone was longer than the new peptide sequence, all combinations of N-terminal and C-terminal truncations of the peptide backbone were tested. In cases where the length of the new peptide sequence exceeded that of the backbone, all possible combinations of N- and C-terminal extensions were tested by adding idealized helical residues. All subsequent design steps locked the peptide sequence. The hinge-peptide interface was designed in PyRosetta FastDesign for one repeat with a fixed backbone followed by two repeats with a flexible backbone, then sequences were improved by design with proteinMPNN using a temperature of 0.2 and model version v_48_020. Structures of proteinMPNN sequences were predicted with AlphaFold2 using the Rosetta design model as an initial guess, and designs that predicted with mean pLDDT < 92, RMSD to reference > 1.5, or mean PAE interaction > 5 were discarded. Poses were combined with state X models from the parent hinges, and residues were linked between states for PyRosetta multistate design with flexible backbone followed by MPNN multistate design with the same settings as the previous proteinMPNN design. An increased success rate was observed when performing proteinMPNN multistate design with a 60%-40% bias toward state Y sequences. State Y structures were predicted with AF2-IG, mean pLDDT, RMSD to reference, and mean PAE interaction cutoffs of 93, 1.5, and 5, respectively. State X structures were also predicted with AlphaFold2 with the same cutoffs, excluding mean PAE interaction. 4 out of 9 possible pairs of parent hinge and peptide sequence produced AlphaFold2-verified models. 20 designs were ordered, expressed, and tested for binding by fluorescence polarization (FP). A common failure mode for these designs was low levels of soluble expression, but 15/19 with sufficient expression for FP displayed detectable binding to

the intended peptide. 9 were selected based on soluble expression levels and on-target affinity for further characterization. Only one design (CSW13, Figure S5) was determined to have no off-target binding to the peptides in the parent set, while 4/9 designs still bound the original peptide with higher affinity than the intended one. The remaining 4 bound the intended peptide with the highest affinity, but also displayed binding to one or more other peptides; an example of this is shown in CSW20 (Figure S5)

Design of 3 helix bundles

Starting from AF2 models of validated hinge-peptide complexes, we sketched rough 3hb backbones in PyMOL by manually positioning two additional helices to buttress the bound helical effector peptide. For each sketch, we extracted the center four residues of the placed helices and used inpainting with RoseTTAFold to generate 1000 3hb backbones scaffolding those fragments onto the effector peptide. During the inpainting process, residues on the effector peptide interfacing with either of the placed helices were allowed to mutate; this and the placements of the four-residue fragments guided inpainting to build valid 3hb backbones that roughly aligned with the sketches. For the best 10% (by RoseTTAFold pLDDT) of backbones generated from each sketch, sequences were optimized using ProteinMPNN. AF2-IG was used to predict the structure of the designed 3hbs with and without their target hinge, selecting only those designs which retained the same structure in both predictions and bound their target hinge with the same interface as the original effector peptide. We experimentally characterized the 1–3 designs per sketch that showed the best PAE interaction in the bound prediction, pLDDT in both predictions, and structural diversity (by eye).

Design of hinge-armed trimers

To fuse hinge cs221 to the asymmetric unit (asu) of a validated C3-symmetric homotrimer(*73*), we manually positioned the two proteins such that they formed a large interface, their termini were near, and the angle of hinge switching was approximately perpendicular to the homotrimer axis of symmetry. We used inpainting with RoseTTAFold to generate 100 loop backbones between the N-terminus of cs221 and the homotrimer asu, allowing residues in the interface between the two proteins to mutate. To improve the visibility of the conformational change in nsEM, we extended the C-terminal end of cs221 by fusing it to LHD101B, a previously validated monomeric protein(*29*). Again, we manually positioned the two proteins such that they formed a large interface and their termini were near, then used inpainting with RoseTTAFold to generate 100 loop backbones between those termini, allowing residues in the interface to mutate. For the best 20% (by RoseTTAFold pLDDT) of backbones generated for each fusion, we optimized sequences of the fusion region using ProteinMPNN. We combined the most confidently predicted (by AF2 pLDDT) LHD101B fusion with each homotrimer asu fusion, modeled each symmetric complex by aligning three copies of each fusion to the original homotrimer, and used AF2-IG to predict the symmetric

structure of the designed fusions. We experimentally characterized the 7 most confidently-predicted designs.

Hinge extension for FRET constructs

Hinges were extended by aligning a copy of the parent DHR to the first repeat of the hinge and another copy of the parent DHR to the last repeat of the hinge. The extended hinge was then obtained by replacing the first and last repeat of the hinge by 2 or more repeats from the parent DHR. For cs221F, the additional repeats were redesigned using proteinMPNN.

Disulfide stapling

A custom PyRosetta script was used to identify candidate positions for disulfides that could lock hinges in one conformation. i-j residue pairs where residue i is in domain 1 of the hinge and residue j is in domain 2 of the hinge were exhaustively evaluated using a 6D hashing protocol(*74*). For each candidate pair, 2 separate pdbs were generated for state X and state Y of the hinge with the identified residues i and j mutated to cysteine. AF2-IG was used to filter candidate pairs, selecting only pairs for which the cysteine side chains in the “target” state showed distances and relative orientations compatible with disulfide formations and for which the “off-target” state showed a large distance between cysteine side chains.

proteinMPNN-based identification of point mutant candidates

ProteinMPNN was used to generate 100 sequences optimized for state X and another 100 sequences optimized for the state Y-peptide complex. For each state, consensus sequences(*75*) were used to identify non-interface positions with distinct residue preferences that were different between both states. For mutations that AF2 predicted to not affect the global structure, individual protein variants carrying these mutations were experimentally tested using the FP peptide binding assay.

Cloning, expression, and protein purification

Genes encoding for proteins and peptides were either purchased as pre-cloned genes from IDT in pet29B expression vectors or purchased as e-blocks from IDT and cloned into custom target vectors using golden gate assembly(*32*). Hinges and 3-helix bundles usually carried a C-terminal SNAC tag(*83*) followed by a 6xHis-tag (Hinge-GSHHWGSTHHHHHH); in some cases, the SNAC tag was omitted (Hinge-GSHHHHHH). Peptides were expressed fused to superfolder green fluorescent protein (sfGFP) in either a sfGFP-(linker)-peptide-(linker)-6xHis construct or sfGFP-GSGSENLYFQS-(linker)-peptide-(linker)-6xHis construct. All proteins were expressed either in LEMO21 or NEB BL21(DE3) E. coli cells by autoinduction using TBII media (Mpbio) supplemented with 50x5052, 20 mM MgSO4 and trace metal mix

and 50 mg/l Kanamycin. Expression cultures were grown at 37°C for 20-24 h or at 37°C for 5-6 h followed by 24 h at 18°C.

After harvesting with centrifugation, cells were lysed at 4°C with sonication in lysis buffer containing (100 mM Tris HCl pH 8, 200 mM NaCl, 50 mM imidazole, 1 mM PMSF, 1 mM DNase, 1 Pierce™ Protease Inhibitor Mini Tablets, EDTA-free per 100 mL) and clarified with ultracentrifugation at 14-20k x g for 20-40 min. The constructs were bound to ~1 mL Ni-NTA resin (Qiagen) and mixed for 10-60 min. The beads were sequentially washed with 15 mL low salt wash buffer (20 mM Tris HCl pH 8, 200 mM NaCl, 50 mM imidazole), 15 mL high salt wash buffer (20 mM Tris HCl pH 8, 1 M NaCl, 50 mM imidazole), and 15 mL low salt wash buffer. Lysates and buffer were flowed over the resin either using gravity or a vacuum manifold. Proteins were eluted in 1.4 mL of elution buffer (20 mM Tris HCl pH 8, 200 mM NaCl, 500 mM imidazole), after a 0.4 mL pre-elution. In constructs with designed disulfides, copper phenanthroline was then added to the elution at a final concentration of 10 mM, and the resulting mixture was incubated overnight to encourage the full formation of the disulfides. In all cases, elutions were further purified by SEC/FPLC on Superdex 75 Increase 10/300 GL or Superdex 200 Increase 10/300 GL columns in TBS (20 mM Tris pH 8, 100 mM NaCl), with 0.5 or 1 mL fractionation between 8 and 20 mL. LC-MS was used to confirm the correct molecular weight of all purified proteins.

Protein purification for crystallography

Constructs were transformed into LEMO21 or NEB BL21(DE3) E. coli and then expressed as 0.5 L cultures in 2L flasks. Proteins were expressed in Studiers M2 autoinduction media with 50 ug/mL kanamycin. Pre-cultures were grown at 37°C for 4 hrs, then 22°C for 14 hr, and cultures were inoculated with 10 mL of preculture. Cells were pelleted at 4,000g for 10 minutes, after which the supernatant was discarded. Pellets were resuspended in 40 mL of lysis buffer (100 mM Tris HCl pH 8, 100 mM NaCl, 400 mM imidazole, 1 mM PMSF, 1 mM DNase). Cell suspensions were lysed by microfluidization on a Microfluidics M-100P at 18,000 psi, and the lysate was clarified at 14,000g for 30 minutes. The His-tagged proteins were bound to 8 mL Ni-NTA resin (Qiagen) during gravity flow and washed with 10 mL lysis buffer and 30 mL high salt wash buffer (25 mM Tris HCl pH 8, 1 M NaCl, 40 mM imidazole), then 10mL SNAC cleavage buffer (100 mM CHES, 100 mM Acetone oxime, 100 mM NaCl, 500 mM GnCl, pH 8.6).(*83*) 40 mL SNAC cleavage buffer and 80 uL 1M NiCl2 were added and columns were closed and shaken on a nutator for 12 hours in order to cleave. After cleavage, the flowthrough was collected and concentrated prior to further purification by SEC/FPLC on a HiLoad 20/600 Superdex 75 pg column in TBS (20 mM Tris pH 8.0, 100 mM NaCl), with 14 mL fractionation between 100 and 290 mL.

Peptide synthesis

Peptides were synthesized in-house on a CEM Liberty Blue microwave synthesizer. All amino acids were purchased from P3 Biosystems. Oxyma Pure was purchased from

CEM, DIC was purchased from Oakwood Chemical, diisopropyl ethylamine (DIEA) and piperidine were purchased from Sigma-Aldrich. Dimethylformamide (DMF) was purchased from Fisher Scientific and treated with an Aldraamine trapping pack prior to use. 5(6)-carboxytetramethylrhodamine carboxylic acid (5(6)-TAMRA) was purchased from Novabiochem. Synthesis was done on a 0.1 mmol scale on CEM Cl-MPA resin. Five equivalents of each amino acid were activated using 0.1 M Oxyma with 2% (v/v) DIEA in DMF, 15.4% (v/v) DIC, and coupled twice on resin for 2 min per coupling with microwave irradiation. For TAMRA labeled peptides, peptides were washed with DMF post-synthesis, then incubated for 3h with 5(6)-TAMRA carboxylic acid (3 eq.), HATU (3 eq.), and DIEA (5 eq.) in DMF, then washed with DMF (3x) followed by DCM (3x) to prepare for global deprotection. Global deprotection was accomplished with a TFA/water/TIPS/2,2'-(ethylenedioxy)diethanethiol (92.5:2.5:2.5:2.5) mixture for 3 hours. This deprotection mixture was concentrated *in vacuo* to 2-3mL, then precipitated in 30 mL of ice-cold ethyl ether, centrifuged, and decanted, then washed twice more with fresh ether and dried under nitrogen to yield crude peptide for high-pressure liquid chromatography (HPLC) purification. The crude peptide was dried and dissolved in a minimal amount of ACN and water to where the entire crude is soluble. This solution was purified on a Zorbax Stablebond C18 (9.4 x 250mm, 5um) column using an Agilent 1260 Infinity HPLC. A linear gradient of water (0.1% TFA) and increasing ACN (0.1% TFA) was used to purify the crude peptides. UV signal was monitored at 214 nm and all peaks were collected. Peak masses were checked using an Agilent G6230B LC-MS and purity was assessed using a C18 column (Higgins Analytical PROTO 300 C18, 10um, 10 x 250mm) on an analytical Agilent 1260 Infinity II HPLC.

SEC binding assay

Individual hinge and sfGFP-fused peptides or 3hb were diluted in 20 mM Tris pH 8, 100 mM NaCl and mixed at approximately 1:1 concentrations. 0.5-1 mL of the resulting samples were injected onto a Superdex 200 Increase 10/300 GL columns and the absorbance at 230 nm was used as a readout for binding. For sfGFP-fused peptides, 473 nm was also used as a readout. Mixtures were at a total concentration of 2.5 μM or higher.

Fluorescence Polarization (FP)

All FP measurements were performed at 25°C in 96-well plates (Corning 3686) using a Synergy Neo2 plate reader and a 530/590 nm filter cube. The buffer for all FP measurements was 20 mM Tris-HCl, 100 mM NaCl, 0.05 % v/v TWEEN20 at pH 8. Titrations were carried out in 96-well format, with 4 replicates per plate and 24 data points per titration (23 steps of two-fold serial dilution of hinges in the presence of TAMRA-labeled peptide at a constant concentration between 0.1 nM and 1 nM) with a final sample volume of 80 μl per well. Titration plates were incubated overnight at room temperature before measuring to ensure complete equilibration. The polarization signal S (as calculated by the Neo2 software) was fitted to the equation

$$S = S_0 + S_1 * f_{AB}$$

$$f_{AB} = \frac{1}{2B_{tot}}\left(A_{tot} + B_{tot} + K_D - \sqrt{(A_{tot} + B_{tot} + K_D)^2 - 4 * A_{tot} * B_{tot}}\right)$$

where f_{AB} is the fraction of peptide that is bound, A_{tot} is the absolute hinge concentration, B_{tot} is the absolute peptide concentration, S_0 is the baseline polarization of free peptide, and $S_1$is the change in polarization upon complex formation.

For FP kinetics experiments a 2x peptide solution and 8 different 2x hinge solutions at different concentrations were prepared separately. 40 µl of each hinge solution were mixed with 40 µl peptide solution using a multichannel pipet and the measurement was started immediately after mixing. Polarization signals S at each concentration were fitted individually to the equation

$$S = S_0 - S_1 * e^{-k_{app}(t_0+t)}$$

where S_0 is the amplitude, S_1 is the polarization at equilibrium, k_{app} is the apparent rate constant, t is the time after start of the measurement, and t_0 is the dead time between mixing and start of the measurement. For each hinge-peptide pair, the apparent rate constants for 8 different concentrations are fitted to the equation

$$k_{app} = k_{off} + k_{on} * A_{tot}$$

where k_{off} and k_{on} are observed off- and on rates and A_{tot} is the absolute hinge concentration.

FRET

AlexaFluor 555 C2 maleimide (donor) and AlexaFluor647 C2 maleimide (acceptor) were purchased from ThermoFisherScientific. Stock solutions at ~5 mM were prepared by dissolving 1 mg of each dye in 200 µl DMSO. Hinge variants containing two cysteines were expressed and purified as described above with the modification that 0.5 mM TCEP was used during lysis, IMAC and SEC, and that the buffer for initial SEC contained 20 mM sodium phosphate (PH 7.0) instead of Tris-HCl. After SEC, 500 µl hinge at a concentration of 50 µM was incubated with 500 µM of a single dy for controls or 250 µM each of two dyes. After 2h incubation at room temperature, samples were purified by SEC using a buffer containing 20 mM Tris-HCl and 100 mM NaCl at pH 8.

The buffer for all FRET measurements was 20 mM Tris-HCl, 100 mM NaCl, 0.05 % v/v TWEEN20 at pH 8. Fluorescence spectra were recorded at room temperature using a

FluoroMax spectrometer in a 1 cm x 1 cm cuvette at a sample volume of 3 ml. FRET titrations and kinetics measurements were performed at 25°C in 96-well plates (Corning 3686) using a Synergy Neo2 plate reader. Excitation wavelength was 520 nm and emission wavelength was 665 nm (except for donor-donor controls for which emission wavelength was 555 nm, see Figure S7B).

Titrations were carried out in 96-well format, with 4 replicates per plate and 24 data points per titration (23 steps of two-fold serial dilution of effector (peptide or 3hb) in the presence of double-labeled hinge at a constant concentration of 2 nM) with a final sample volume of 80 µl per well. Titration plates were incubated overnight at room temperature before measuring to ensure complete equilibration. The fluorescence signal was fitted to the equation

$$S = S_0 + sign * S_1 * f_{AB}$$

$$f_{AB} = \frac{1}{2A_{tot}}\left(A_{tot} + B_{tot} + K_D - \sqrt{(A_{tot} + B_{tot} + K_D)^2 - 4 * A_{tot} * B_{tot}}\right)$$

where f_{AB} is the fraction of hinge that is bound, A_{tot} is the absolute hinge concentration, B_{tot} is the absolute peptide concentration, S_0 is the baseline fluorescence of free hinge, $S_1$is the change in fluorescence upon complex formation, and $sign = -1$ for cs 201F (which shows a decrease in FRET upon binding) and $sign = 1$ for cs074F and cs221F (which show an increase in FRET upon binding.

For FRET kinetics experiments a 2x hinge solution and 8 different 2x effector solutions at different concentrations were prepared separately. 40 µl of each effector solution were mixed with 40 µl hinge solution using a multichannel pipet and the measurement was started immediately after mixing. Fluorescence signals S at each concentration were fitted individually to the equation

$$S = S_0 - sign * S_1 * e^{-k_{app}(t_0+t)}$$

where S_0 is the amplitude, S_1 is the Fluorescence at equilibrium, k_{app} is the apparent rate constant, t is the time after the start of the measurement, and t_0 is the dead time between mixing and start of the measurement, and $sign = -1$ for cs 201F (which shows a decrease in FRET upon binding) and $sign = 1$ for cs074F and cs221F (which show an increase in FRET upon binding. For each hinge-peptide pair, the apparent rate constants for 8 different concentrations are fitted to the equation

$$k_{app} = k_{off} + k_{on} * B_{tot}$$

where k_{off} and k_{on} are observed off- and on rates and B_{tot} is the absolute peptide concentration.

DEER - spin label modeling and site selection

All spin label modeling and distance distribution predictions were performed using chiLife(*71*) with the off-rotamer sampling method(*84*). For each construct, spin label models were made for every site with at least 50 $Å^2$ solvent accessible surface area (SASA) in both conformational states. Pairwise distance distributions were predicted for all modeled spin labels in both states. Site pairs with the largest earth mover's distance (EMD) between the bound and unbound states were manually inspected and site pairs were selected that were predicted to have minimal interference with peptide binding, and conformational change. Two site pairs were chosen for each construct, one predicted to shift the distance distributions to a larger distance upon interaction with substrate and one predicted to shift to a shorter distance.

DEER - sample preparation

Hinge variants carrying two cysteines were purified as described above but with 1 mM TCEP added to the lysis buffer and 0.5 mM TCEP added to an intermediate wash buffer. Directly after elution, 50 µL of 200 mM MTSL solution (in DMSO) was added to the entire 1.3 mL elution. After 1-6 h incubation at room temperature the labeling mixture was sterile filtered and purified by SEC. Successful labeling was confirmed by LC-MS.

Before DEER, 20 µM protein samples were prepared in 20 mM tris, 100 mM NaCl at pH 8.0 in D2O and 20 % d8-glycerol (Cambridge Isotope Laboratories, Inc.) supplemented with 100 µM B-peptide when appropriate. Samples (20 – 40 µL) were transferred to quartz capillaries (Sutter Instruments) with an inner diameter of 1.1 mm and an outer diameter of 1.5 mm, flash frozen with liquid nitrogen and stored at -80 °C.

DEER - measurements

All DEER experiments were performed on an ELEXSYS E580 EPR spectrometer (Bruker) at Q-band (~34 GHz) using an EN5107D2 resonator (Bruker). A cryogen-free cooling system (ColdEdge) was used to maintain a temperature of 50 K. Shaped pulses were generated using a SpinJet arbitrary waveform generator (Bruker). Observer pulses were 60 ns gaussian pulses with a full width at half maximum (FWHM) of 30 ns performed at approximately the center of the field-swept spectrum. Pump pulsers were 150 ns sech/tanh pulses centered 80 MHz above the observer pulses. Sech/tanh pulses were generated using PulseShape (https://gitlab.com/mtessmer/PulseShape) or EasySpin(*85*) with an excitation bandwidth of 80 MHz and a truncation parameter of 10. All sech/tanh pulses were modified to compensate for resonator performance and transmitter nonlinearity. All experiments used 8-step phase cycling and 8-step τ_1 averaging with 16 ns increments from 400 ns to 528 ns. Pump pulse time steps (Δt) and τ_2 times were chosen on a per-sample basis and the values for each sample are

reported in Supplementary Table 2. Additional parameters including t_0 offsets, shot repetition time, the total number of averages, and more are reported in Supplementary Table 2.
DeerLab(*86*) was used to analyze all DEER data to simultaneously fit foreground and background using Tikhonov regularization and compactness regularization(*87*). Akaike information criterion (AIC) and the information complexity criterion (ICC) were used to select regularization parameters for Tikhonov and compactness regularizations respectively. Sample fitting parameters including modulation depth, estimated signal-to-noise, smoothing, and compactness regularization parameters are reported in Supplementary Table 2.

X-Ray crystallography

All crystallization experiments were conducted using the sitting drop vapor diffusion method. Crystallization trials were set up in 200 nL drops using the 96-well plate format at 20 ˚C. Crystallization plates were set up using a Mosquito from SPT Labtech, then imaged using UVEX microscopes and UVEX PS-600 from JAN Scientific. Diffraction quality crystals formed for 3hb05 in 0.2 M Lithium sulfate, 0.1 M Na-Phosphate-citrate pH 4.2, 20% PEG 1000; for 3hb12 1.8 M Ammonium citrate tribasic pH 7.0; for cs074AB in 0.2 M Calcium acetate, 0.1 M Na cacodylate pH 6.5, 40% PEG 300; for cs207A in 0.1 M SPG buffer pH 7, 25% (w/v) PEG 1500; for cs207AB in 0.2 M Magnesium sulfate and 20% (w/v) PEG 3350.

Diffraction data was collected at the Advanced Light Source beamlines 8.2.2/8.2.1. X-ray intensities and data reduction were evaluated and integrated using XDS(*88*) and merged/scaled using Pointless/Aimless in the CCP4 program suite(*89*). Structure determination and refinement starting phases were obtained by molecular replacement using Phaser(*90*) using the designed model for the structures. Following molecular replacement, the models were improved using phenix.autobuild(*91*); efforts were made to reduce model bias by setting rebuild-in-place to false, and using simulated annealing and prime-and-switch phasing. Structures were refined in Phenix(*91*). Model building was performed using COOT(*92*). The final model was evaluated using MolProbity(*93*). Data collection and refinement statistics are recorded in Supplementary Table 1. Data deposition, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank (PDB), http://www.rcsb.org/ with accession code 8FIH (3hb05), 8FVT (3hb12), 8FIT (cs074AB), 8FIN (cs207A) and 8FIQ (cs207AB).

Negative stain electron microscopy

Carbon-coated 400 mesh copper grids (01844-F, TedPella,Inc.) were first glow-discharged using a PELCO easiGlow cleaning System. SEC-purified proteins were diluted to 2 µg/ml with Tris Buffer (100 mM Tris, 40 mM NaCl), and then immediately pipetted onto the glow-discharged grid. The protein solution was allowed to sit on the grid for 30s, before being blotted away with Whatman filter paper. 3 uL of 2% uranyl formate stain was added to the grid and then blotted away after 10s. A second and third

wash of UF stain was added to the grid, allowed to sit for 10s and 30s respectively, before being blotted away. The grid was allowed to air-dry for 5 minutes. Dried grids were then imaged using a FEI Talos L120C TEM (FEI Thermo Scientific, Hillsboro, OR) equipped with a 4K × 4K Gatan OneView camera, at a magnification of 57,000x and pixel size of 2.49 Å. Once a grid-square with satisfactory stain thickness and contrast was identified, EPU software was used to automatically collect 200-400 micrographs across the square. Micrographs were imported into and analyzed using cryoSPARC v4.0.3. 50-100 particles were manually picked and subjected to 2D classification to find coarse 2D averages that could be used as templates for the automated picking of thousands of particles across all micrographs. After automated picking and particle extraction from micrographs, a further round of 2D classification was done to find higher resolution averages of the hinge-bearing cyclic ring proteins in various states and orientations.

Chapter III: Design of Protein-Peptide Interactions

Adapted from Susana Vázquez Torres, Philip J. Y. Leung, Isaac D. Lutz, Preetham Venkatesh, Joseph L. Watson, Fabian Hink, Huu-Hien Huynh, Andy Hsien-Wei Yeh, David Juergens, Nathaniel R. Bennett, Andrew N. Hoofnagle, Eric Huang, Michael J MacCoss, Marc Expòsit, Gyu Rie Lee, Paul M. Levine, Xinting Li, Mila Lamb, Elif Nihal Korkmaz, Jeff Nivala, Lance Stewart, Joseph M. Rogers, David Baker. (2022). De novo design of high-affinity protein binders to bioactive helical peptides. biorxiv

The Peptide Binding Problem

Bioactive Helical Peptides are Difficult Targets of Therapeutic Interest

Peptide hormones, such as parathyroid hormone (PTH), neuropeptide Y (NPY), glucagon (GCG), and secretin (SCT), which adopt alpha-helical structures upon binding their receptors(*94–97*), play key roles in human biology and are well-established biomarkers in clinical care and biomedical research (Fig. 6A). There is considerable interest in their sensitive and specific quantification, which currently relies on antibodies that require substantial resources to generate, can be difficult to produce with high affinity, and often have less-than-desirable stability and reproducibility(*69*). Furthermore, the loop-mediated interaction surfaces of antibodies are not particularly well suited to high specificity binding of extended helical peptides. By contrast, designed proteins can be readily produced with high yield and low cost in *E. coli* and have very high stability, but while there have been considerable advances in *de novo* protein design to generate binders for folded protein targets(*47*, *69*), the design of proteins that bind helical peptides with high affinity and specificity remains an outstanding challenge.

Previous Protein Design Paradigms are Inadequate

Design of peptide-binding proteins is challenging for two reasons. First, existing *de novo* proteins, such as picomolar affinity hyper-stable 50-65 residue minibinders(*69*) have shapes suitable for binding rigid concave regions of folded targets, but not for cradling extended peptides. Second, peptides have less total area, and fewer amino acid residues to interact with and are often partially or entirely unstructured in isolation(*98*); as a result, there can be an entropic cost of structuring the peptide into a specific conformation(*99*), which compromises the favorable free energy of association. Progress has been made in designing peptides that bind to extended beta strand structures(*100*) and polyproline II conformations(*101*) using protein side chains to interact with the peptide backbone, but such interactions cannot be made with the peptide backbone of alpha-helical peptides due to the extensive internal backbone-backbone hydrogen bonds formed by their helical conformations.

Approaches to Designing High-Affinity Peptide Binding Proteins

General Principles for Designing Proteins to Bind Helical Peptides

We set out to develop general methods for designing proteins that bind peptides in helical conformations. To fully leverage recent advances in protein design, we explored both conventional and deep learning-based approaches to backbone sampling and target docking. As a starting point, we reasoned that helical scaffolds with an open groove for a helical peptide could provide a general solution to the helical peptide binding problem: the extended interaction surface between the full length of the helical peptide target and the contacting helices on the designed scaffold could enable the design of high affinity and specificity binding (Fig. 6B).

Parametric Design of Protein Scaffolds with Grooves for Binding Peptides

We began by exploring parametric methods for generating backbones with overall “groove” shapes. Using the Crick parameterization of alpha-helical coiled coils(*28*), we devised a method to sample scaffolds consisting of a three-helix groove supported by two buttressing helices (Fig. 6C, see Supplementary Materials). We assembled a library of these scaffolds sampling a range of supercoiling and helix-helix spacings to accommodate a variety of helical peptide targets (Fig. S19). We then used this library to design binders to PTH, GCG, and NPY, and screened 12 designs for each target using a nanoBiT split luciferase binding assay. Many of the designs bound their targets (3/12, 4/12, and 8/12 to PTH, GCG, and NPY) but with only micromolar affinities (see

Supplementary Materials). These results suggest that groove-shaped scaffolds can be designed to bind helical peptides, but also that design method improvement was necessary to achieve high-affinity binding.

Protein Inpainting with *RFJoint* Improves Binding Affinity

While powerful for generating and sampling a large number of potential scaffolds, the parametric generation approach has the limitation of building only from ideal building blocks in an unnecessarily restricted geometric space, in this case, parametrically ideal alpha helices. Deep learning methods do not have these limitations, and we explored whether RoseTTAFold inpainting (RF*joint*), a model that can jointly design protein sequences and structures(*33*), could be used to improve the modest affinities of our parametrically-designed PTH binders (Figure 7A). We used RF inpainting to extend the binders (non-parametrically) to incorporate additional interactions with the target peptide to take advantage of the full potential binding interface of the peptide. Out of 192 designs tested, 44 showed binding against PTH in initial yeast display screening. Following SEC purification, the best binder was found to bind at 6.1 nM affinity to PTH. Binding was quite specific: very little binding was observed to PTH-related peptide (PTHrp), a related peptide sequence with 34% sequence identity (Figure 7A). Overall, the affinity of the starting PTH binders was improved by approximately three orders of magnitude, and the highest-affinity binder had 19% greater surface area contacting the target peptide. We used the same design strategy to generate higher affinity binders for NPY and GCG. Using weak parametric binders as a starting point, we extended their binding interfaces and generated a ~231 nM affinity binder for GCG and a 3.5 μM binder for NPY after screening 96 designs (Figure S20).

Coarse-Grained Backbone Modeling as an Approach for Peptide Binder Design

As an alternative to *de novo* parametric design of scaffolds that contain grooves, we explored the threading of helical peptides of interest onto already existing designed scaffolds with interfaces that make extensive interactions with helical peptides (Fig. 7B). We started from a library of scaffolds that contained single helices bound by pseudorepetitive helical scaffolds. We then threaded(*102*) sequences of peptides of interest into the bound single helix and filtered to maximize interfacial hydrophobic interactions of the target sequence to the binder scaffold. The binders with the most numerous interfacial hydrophobic residues per target were then redesigned in the presence of the threaded target sequence with ProteinMPNN(*41*) and the complex was

predicted with AF2(*46*), using an initial guess(*47*) and filtered on AF2 and Rosetta metrics. We first attempted this method with SCT, where initial screening using yeast surface display identified 4/66 binders, which were expressed in *E. coli*. Following size exclusion chromatography (SEC) purification of the monomer fraction, all 4 of the designs were found to bind with sub-micromolar affinity using fluorescence polarization (FP), with the highest-affinity design binding with an affinity of 2.7 nM for SCT. Binding specificity was assessed with FP by measuring affinity for GCG, a related hormone to which SCT shares a significant degree of sequence identity (44%) and conformational homology(*94*, *95*). We found that the tightest SCT binder was only 4 fold selective for SCT over GCG, which suggested additional design strategies might be necessary to increase the quality of the binding interface and to achieve high-specificity binding (Fig. 7B). We were able to generalize this approach to gastrointestinal peptide (GIP), as well as GLP1 and GLP2, all hormones crucial for metabolic regulation and gastrointestinal function(*103*). For GIP, we identified 3/12 binders, with the highest affinity binder achieving an on target affinity of 5.5 nM for GIP (Figure 8A). For GLP1, we only screened 6 designs, of which 2 were binders. The tightest GLP1 binder bound with an affinity of 95 nM (Figure 8B). For GLP2, we identified 8/30 binders, with the highest affinity binder being poorly fit by a bimolecular binding isotherm but likely having sub-nanomolar affinity (Figure 8C).

Targeted Backbone Resampling with RF*diffusion*

A long-standing challenge in protein design is to increase the activity of an input native protein or designed protein by exploring the space of plausible closely related conformations for those with predicted higher activity.(*104*) This is difficult for traditional design approaches as extensive full atom calculations are needed for each sample around a starting structure (using molecular dynamics simulations(*105*) or Rosetta full atom relaxation(*66*) methods), and it is not straightforward to optimize for higher binding affinity without detailed modeling of the binder-target sidechain interactions. We reasoned that, in contrast, RF*diffusion* might be able to rapidly generate plausible backbones in the vicinity of a target structure using deep learning(*34*). RF*diffusion* is trained to directly generate a diversity of solutions to specific design challenges starting from random 3D distributions of residues that are progressively denoised. We reasoned that RF*diffusion* could be used for binder optimization, by sampling related conformations around a specific binder structure, guided by the extensive priors of protein structural propensities inherent in RoseTTAfold. During the reverse diffusion (generative) process, RF*diffusion* takes random Gaussian noise as input, and iteratively refines this to a novel protein structure over many (“T”) steps (typically 200). Partly through this denoising process, the evolving structure no longer resembles “pure noise”,

instead resembling a “noised” version of the final structure. We reasoned that ensembles of structure with varying extents of deviation from an input structure could be generated by partially noising to different extents (for example, timestep 70), and then denoising to a similar, but not identical final structure with a more optimal binding configuration (Figure 9A, B). In this manner, by starting from priors of non-zero mean(*106*), we used RF*diffusion* to increase the extent and quality of interactions with the target sequences. We experimented with this approach starting from our parametrically-designed inpainted binders to GCG (with 231 nM affinity) and NPY (with 3.5 μM affinity) (Supplementary Fig. S20). Following partial noising and denoising, and sequence redesign with MPNN, we identified designs that *in silico*, had significantly improved AF2 metrics compared to the starting design. The diversity compared to the starting design could be readily tuned by varying the time point to which the starting design was noised (Figure 9A). Initial screening on yeast display revealed quite high binding success rates, with 25/96 designs binding GCG, and 20/96 binding NPY at 10 nM peptide concentration. The highest affinity designs were expressed in *E. coli*, purified, and their binding affinities were determined using FP. The highest-affinity binders were found to bind at subnanomolar affinities to GCG, and 5.6 nM to NPY (Figure 9C). The designed proteins are quite specific: the GCG binders bound 10 times less tightly to SCT, which was chosen due to its high similarity to GCG. Impressively, the NPY binder did not show any cross-reactivity to peptide YY (PYY), which is a member of the NPY/pancreatic polypeptide family(*107*) and shares a high percentage of sequence similarity (63.5% for the sequences used in the assay).

Discussions on Designing High-Affinity Peptide Binding Proteins

Origins of higher affinity binding

The RF*diffusion* scaffolds bind the peptides with extended helices in a manner not entirely different from our starting groove structures and the other designs described above. What is the origin of their higher affinity? Reasoning that *de novo* building of the designs in the presence of the target, rather than starting from pre-generated scaffolds, could increase the extent of shape matching between binder and target, we computed the contact molecular surface(*69*) for all of our designs in complex with the peptides. The average contact molecular surface for the partially diffused GCG binders and NPY increased by 33% and 29% respectively compared to the starting models, and the Rosetta ddG improved by 29% and 21% (Figure S21A, B). The improvements in the distributions show that the improved binding cannot be explained by the sequence redesign that followed diffusion alone, as in many cases the starting design was nearly or completely out of the score distribution.

Advantages of De Novo Binders over Antibodies and Outlook

Antibodies have served as the industry standard for affinity reagents for many years, but their use is often hampered by variable specificity and stability(*108*, *109*). For binding helical peptides, the computationally designed helical scaffolds described in this paper have a number of structural and biochemical advantages. First, the extensive burial of the full length of an extended helix is difficult to accomplish with antibody loops but very natural with matching extended alpha helices in groove shape scaffolds. Second, designed scaffolds are more amenable to incorporation into biosensors and larger functional assemblies, with tunable control over shape and geometry(*16*). Third, they are more stable, can be produced much less expensively, and could be more easily incorporated into affinity matrices for the enrichment of peptide hormones from human serum. Fourth, peptide binders can achieve high affinity and specificity purely through computational methods, eliminating the need to use animals, which often mount weak responses to highly conserved bioactive molecules. Our results highlight the emergence of powerful new methods for protein design. The inpainting and RF*diffusion* methods were both able to improve on initial designs, and the threading approach circumvented the earlier binder design methods by modeling the target conformation on a per-binder basis. Most impressively, we report binders with specificity and low nanomolar and picomolar affinity to multiple helical peptides. We expect both the *de novo* peptide binder design capability and the ability to resample around initial designs (before or after experimental characterization) to be broadly applicable.

Figures for Design of Protein-Peptide Interactions

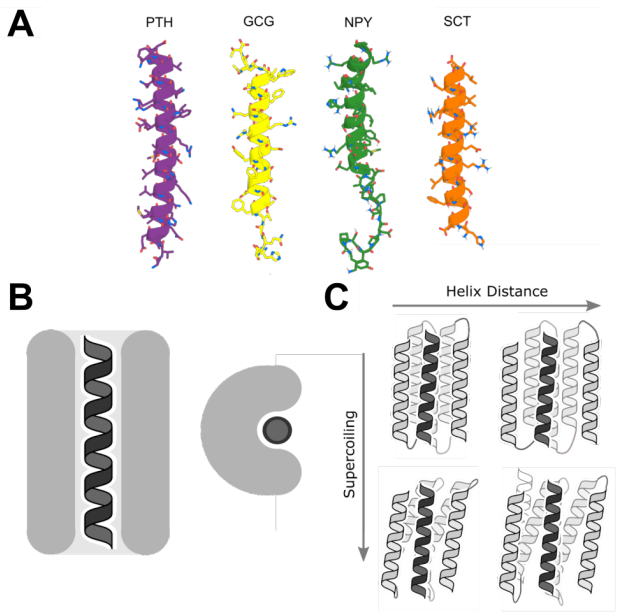

Figure 6. Binding helical peptides in groove scaffolds.

A) Some helical peptide targets: parathyroid hormone (PTH), glucagon (GCG), neuropeptide Y (NPY), and secretin (SCT). **B)** "Open groove" structural solution to the helix binding problem. **C)** Parametric approach to sampling of groove scaffolds varying supercoiling and helix distance to fit different targets.

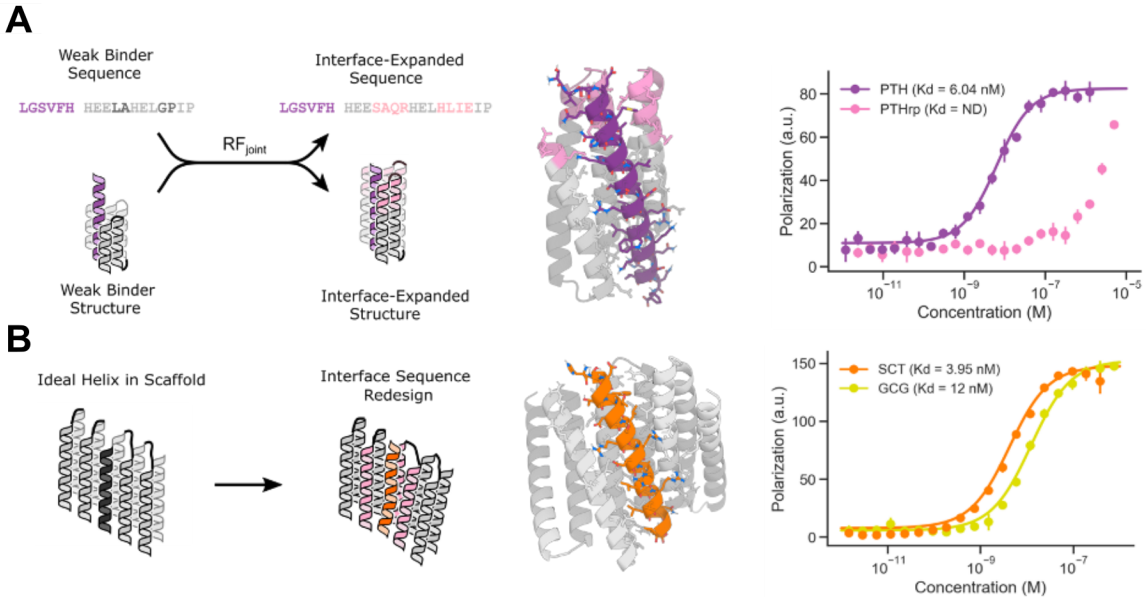

Figure 7. Design strategies for binding helical peptides.

A) Inpainting binder optimization: redesign of parametrically generated binder designs using RF*joint* inpainting to expand the binding interface. Left: schematic illustration of approach. Middle: original parametric scaffold (gray), inpainted design with an extended interface (pink), and PTH target (purple). Right: Fluorescence polarization measurements with TAMRA-labeled targets indicate 6.1 nM binding to PTH and only weak binding to off-target PTH-related peptide (PTHrp). **B)** Threading peptides onto pseudorepetitive protein scaffolds and redesigning. Left: schematic illustration. Right: Design model of SCT based on repeat protein scaffold (grey) and SCT target (orange). Fluorescence polarization measurements with TAMRA-labeled targets indicate 3.95 nM binding to SCT and 12 nM binding to GCG.

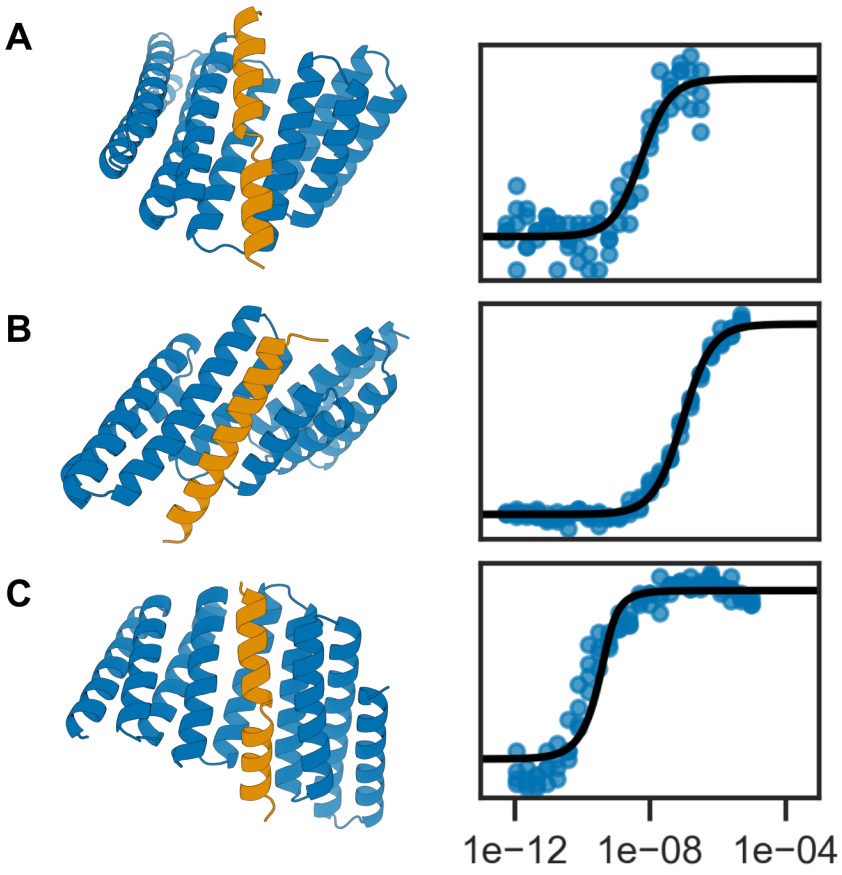

Figure 8. Threading target sequences yields tight binders to many peptide targets

A) Design model of GIP based on repeat protein scaffold (blue) and GIP target (orange). Fluorescence polarization measurements with FAM-labeled target indicates

5.5 nM binding to GIP. **B)** Design model of GLP1 based on repeat protein scaffold (blue) and GLP1 target (orange). Fluorescence polarization measurements with FAM-labeled target indicates 95 nM binding to GLP1. **C)** Design model of GLP2 based on repeat protein scaffold (blue) and GLP2 target (orange). Fluorescence polarization measurements with FAM-labeled target are poorly fit by a bimolecular binding isotherm but may indicate sub-nanomolar binding to GLP2.

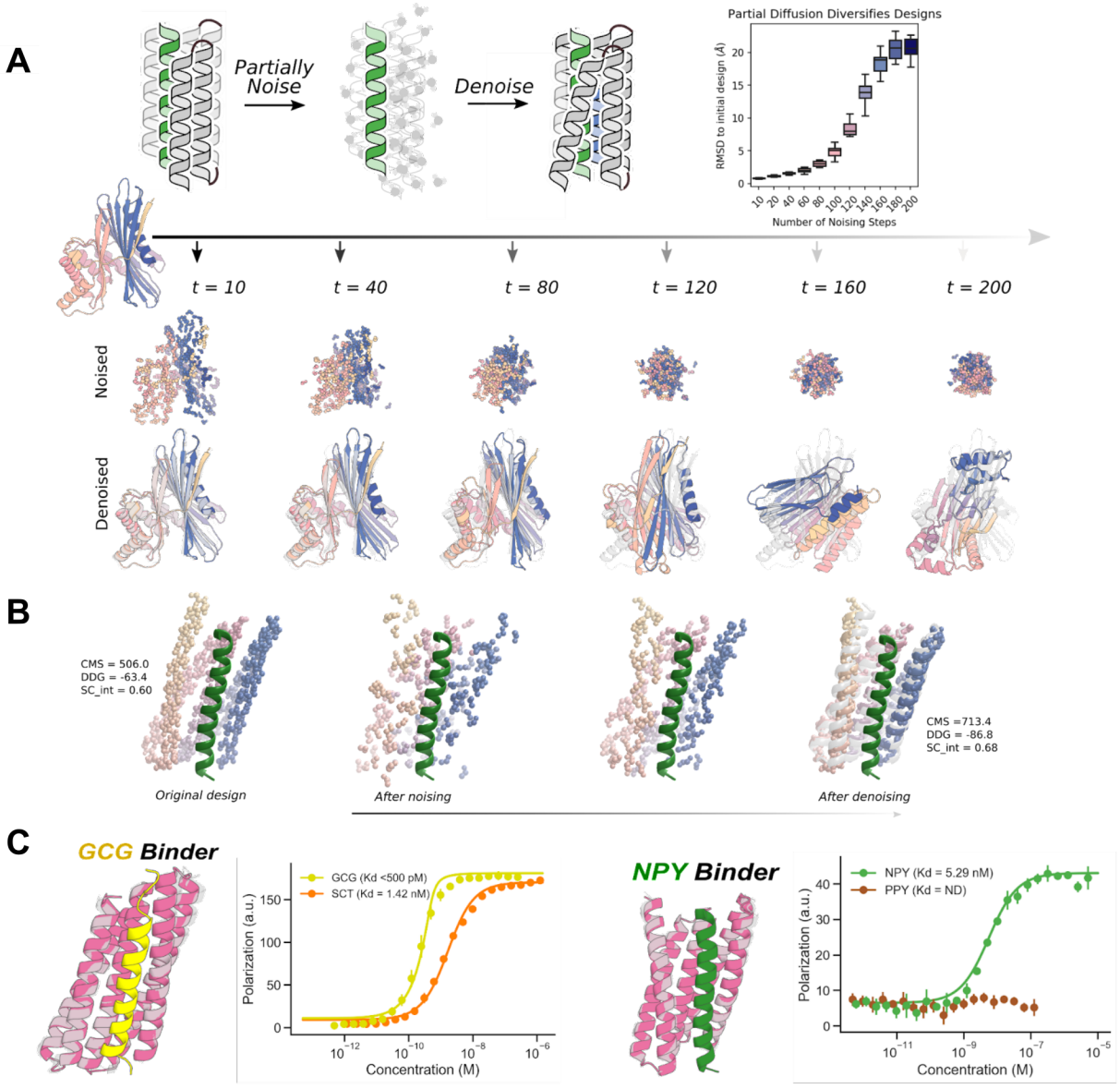

Figure 9. Peptide binder optimization with RF*diffusion*.

A) Top: Schematic showing partial noising and denoising using RF*diffusion*. A starting monomer (left) is partially noised for an increasing number of steps and then denoised resulting in designs (color) increasingly different from the original design (gray). Varying the noising stage from which denoising trajectories are initiated enables control over the extent of introduced structural variation. Bottom left: The distribution of RMSD to initial design vs number of partial noising steps. Bottom right: Starting from initial helix binder designs, we use partial diffusion to design optimized binders with improved shape complementarity. **B)** Partial denoising trajectory starting from an initial NPY binder shown on the left. The final design (color) is shown on the right overlaid over the original design (gray). Contact molecular surface (CMS), Rosetta DDG (DDG) and interface shape complementarity (sc_int) values are reported for the original and optimized binder. **C)** Diffused binders to GCG and NPY. Top left: Design models (gray) and AF2 predictions (pink), of diffused binders to GCG (yellow). Top right: FP measurements with FAM-labeled GCG indicate a sub-nanomolar binding affinity and selectivity over SCT. Bottom left: Design models (gray) and AF2 predictions (pink, metrics in Supplementary Table 1), of diffused binders to NPY (green). Bottom right: FP measurements with FAM-labeled NPY indicate a binding affinity of 5.29 nM and no binding to PYY, demonstrating selectivity.

Authors and Acknowledgements for Design of Protein-Peptide Interactions

Authors and affiliations

Susana Vázquez Torres[†1,2,3], Philip J. Y. Leung[†1,2,4], Isaac D. Lutz[†1,2,5], Preetham Venkatesh[†1,2,3], Joseph L. Watson[1,2], Fabian Hink[6], Huu-Hien Huynh[7], Andy Hsien-Wei Yeh[1,2], David Juergens[1,2,4], Nathaniel R. Bennett[1,2,4], Andrew N. Hoofnagle[7], Eric Huang[8], Michael J MacCoss[8], Marc Expòsit[1,2,4], Gyu Rie Lee[1,2], Paul M. Levine[1,2], Xinting Li[1,2], Mila Lamb[1,2], Elif Nihal Korkmaz[1,2], Jeff Nivala[10,11], Lance Stewart[1,2], Joseph M. Rogers[*6], David Baker[*1,2,9].

1. Department of Biochemistry, University of Washington, Seattle, WA, USA.
2. Institute for Protein Design, University of Washington, Seattle, WA, USA.
3. Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA 98105, USA
4. Graduate Program in Molecular Engineering, University of Washington, Seattle, WA 98105, USA
5. Department of Bioengineering, University of Washington, Seattle, WA, USA.

6. Department of Drug Design and Pharmacology, University of Copenhagen, Jagtvej 160, 2100, Copenhagen, Denmark
7. Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, 98105, USA.
8. Department of Genome Sciences, University of Washington, Seattle, WA 98195
9. Howard Hughes Medical Institute, University of Washington, Seattle, WA 98105, USA
10. School of Computer Science and Engineering, University of Washington, Seattle, WA, USA
11. Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA

†These authors contributed equally to this work
*To whom correspondence should be addressed

Acknowledgements and funding

This work was supported with funds provided by a grant U19 AG065156 from the National Institute for Aging (S.V.T., M.M., E.H., A.H., H.H.H., I.L., D.B.), a gift from Amgen (J.W.), the Audacious Project at the Institute for Protein Design (A.H.-W.Y., D.B.), a gift from Microsoft Gift supporting Computational Protein Structure Prediction and Design at the Institute for Protein Design (D.J., D.B.), the Washington State General Operating Fund supporting the Institute for Protein Design (P.V.), a grant INV-010680 from the Bill and Melinda Gates Foundation Grant (D.J., J.W., D.B.), a NIH NIBIB Pathway to Independence Award (A.H.-W.Y., K99EB031913), a National Science Foundation Training Grant number EF-2021552 (P.L.), NERSC award BER-ERCAP0022018 (P.L.), the Open Philanthropy Project Improving Protein Design Fund (P.L., G.R.L.,D.B.), The Donald and Jo Anne Petersen Endowment for Accelerating Advancements in Alzheimer's Disease Research (N.Ben.), and the Howard Hughes Medical Institute (D.B.). J.M.R. and F.H. were supported by the Novo Nordisk Foundation (NNF19OC0054441 to J.M.R.). H.H.H. is supported by a postdoctoral fellowship provided by the Partnership for Clean Competition. We thank Microsoft and AWS for the generous gifts of cloud computing resources.

Author Contributions

D.B. directed the work. I.L. and S.V.T. designed, screened, and experimentally characterized the parametrically designed groove scaffold peptide binders. P.J.Y.L. and S.V.T. designed, screened and experimentally characterized the threaded peptide

binders. J.L.W. and S.V.T. designed and characterized the inpainted binders. S.V.T, P.V. and P.J.Y.L designed, screened, and experimentally characterized all the different classes of partially diffused peptide binders shown in this manuscript. J.L.W., D.J., and N.R.B. developed the RF*diffusion* algorithm used for peptide binder design. M.E. and G.R.L supported during yeast display binding screening. All authors reviewed and accepted the manuscript.

Supplementary Material for Design of Protein-Peptide Interactions

Figure S19: Parametric groove scaffold library.

45 scaffolds from the library of 18 thousand parametric groove scaffolds, demonstrating a range of supercoiling and helix distances to accommodate a range of helical peptide targets.

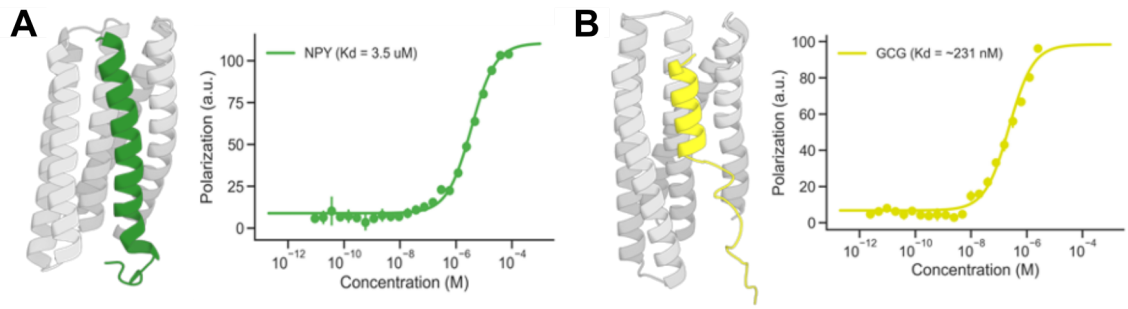

Figure S20. Inpainted peptide binders bound their targets with low affinity.

A) NPY binder. **B)** Glucagon binder. AF2 predictions of the proteins and peptides are shown on the left. FP binding data is shown on the right.

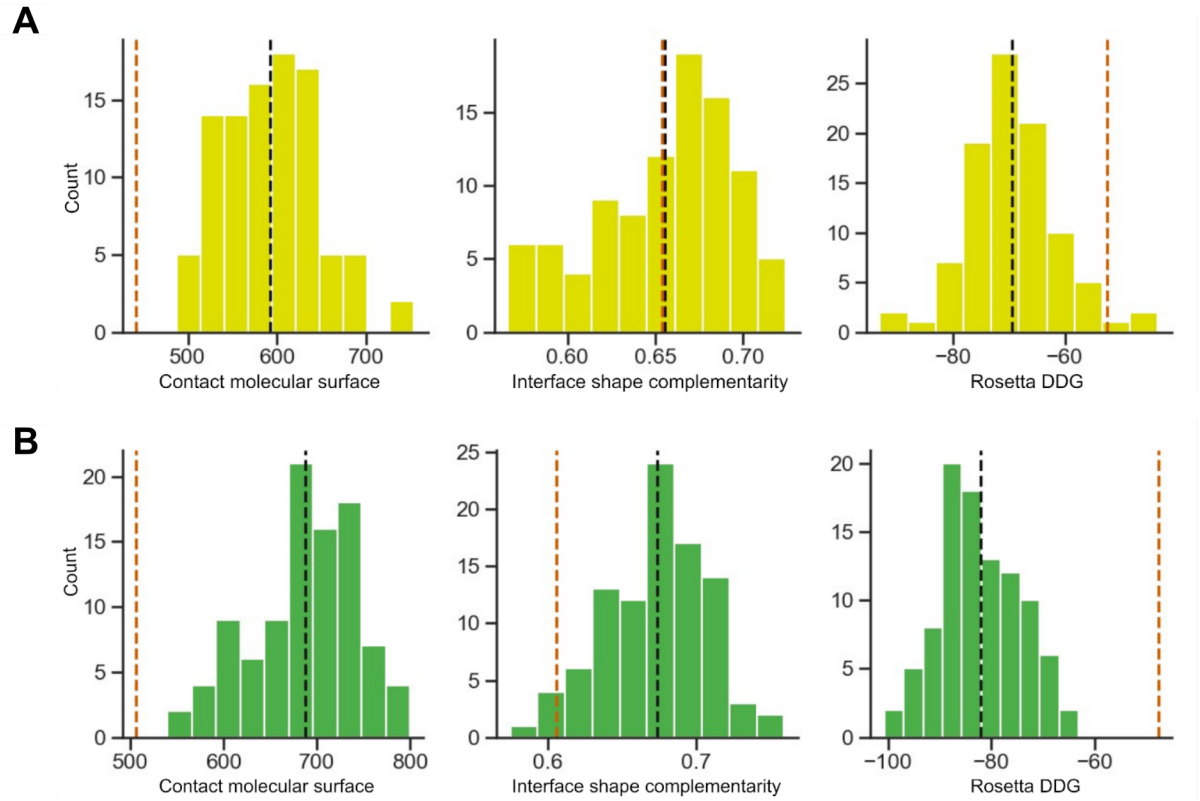

Figure S21. Binding metrics for partially diffused binders.

A) Computational metrics for 96 ordered partially diffused glucagon binders showed significant improvement in contact molecular surface (a measure of interface size and

quality) and Rosetta ddG (a measure of interface predicted energy) over the starting design (vertical red lines). Distribution means are shown in black. **B)** Computational metrics for 96 ordered partially diffused NPY binders showed significant improvement in contact molecular surface, Rosetta ddG, and interface shape complementarity (a measure of interface quality) over the starting design (vertical red lines). Means are shown in black.

Table 3. Fluorophore-labeled peptides used in Fluorescence polarization assays

Peptide name	Sequence	Supplier	Cat #	Fluorophore
PTH-TAMRA	SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNF	In-house	NA	5-TAMRA
PTHrp-FAM	AVSEHQLLHDKGKSIQDLRRRFFLHHLIAEIHTAEIA	Phoenix Pharmaceuticals, Inc.	FG-056-08A	FAM
SCT-FAM	HSDGTFTSELSRLREGARLQRLLQGLV	Phoenix Pharmaceuticals, Inc.	FG-067-03A	FAM
GCG-FAM	HSQGTFTSDYSKYLDSRRAQDFVQWLMNT	Addex Bio	ABBFO2033	FAM
NPY-FAM	SKPDNPGEDAPAEDMARYYSALRHYINLITRQR	Phoenix Pharmace	FG-049-04A	FAM

		uticals, Inc.		
PPY-FAM	IKPEAAGEDASPEELNRYYASLRHYLNLVTRQRY	Phoenix Pharmaceuticals, Inc.	FG-059-02A	FAM
GIP-FAM	YAEGTFISDYSIAMDKIHQQDFVNWLLAQKGKKNDWLHNITQ	Phoenix Pharmaceuticals, Inc.	FG-027-02A	FAM
GLP1-FAM	HAGTFTSDVSSYLEGQAAKEFIAWLVKGRG	Phoenix Pharmaceuticals, Inc.	FG-028-13A	FAM
GLP2-FAM	HADGSFSDEMNTILDNLAARDFINWLIQTKITD	Phoenix Pharmaceuticals, Inc.	FG-028-15A	FAM

Methods for Design of Protein-Peptide Interactions

Parametric design of groove-shaped scaffold library and use for binder design

The parametric groove-shaped scaffold library was sampled using a random sampling approach, where key parameters were selected randomly from distributions. An even distribution of bundle “lengths” was sampled, where each parametric helix was 15-19

residues long. A supercoiling value was randomly selected from a biased distribution favoring more supercoiled scaffolds, given these scaffolds were more likely to fail in the subsequent looping step. An average helix neighbor distance value was randomly selected from a normal distribution informed by native helical bundle geometries. The distance of each helix from its neighbors was independently randomly selected from a much tighter normal distribution centered at the preselected average helix neighbor distance value, to provide some noise within a given scaffold to helix distances and allow for heterogeneous amino acid selections. Values for helix phase and Z displacement were randomly sampled for each helix. The “groove” consisting of 3 helices was first sampled as a helical bundle using the Crick parameterization of alpha-helical coiled coils, around an imaginary central helix where the target was to later be docked. Next, the two buttressing helices were sampled with the same parameterization, but moved radially outward with randomly sampled helix neighbor distances as well as an additional randomly sampled tilt. This process was used to sample a set of 200k arrangements of 5 helices. Next, the Rosetta ConnectChainsMover was used to loop this set into approximately 135k successful scaffold backbones. These backbones were designed and filtered using Rosetta to yield a final library of 18 thousand scaffolds. This library was used to design binders to different helical peptide targets using an adapted version of the miniprotein binder design computational pipeline used by Cao *et al.*[5].

Identification of weak binder hits from parametric designs in pilot experiment

The first helical peptide binder hits were identified in pilot experiments screening for binding using the nanoBiT split luciferase assay (methods). These kinetic binding experiments were performed in cell lysate with no control over protein concentration, so candidate binders were selected qualitatively for showing some increase in luminescence signal over time above background noise, indicating likely binding activity. Additional pilot experiments indicated that this binding activity was all at very weak affinities, likely >100 nM. Therefore, these initial candidates were not further characterized, but rather selected for additional design to yield higher affinity binders.

Identification of weak binders for NPY and GCG using extended parametric designs

We used the RF inpainting approach to extend the binding interfaces of NPY and GCG weak binders hits from parametric design. However, the characterized proteins displayed low affinity binding to their targets, which was not enough for diagnostic applications.

Gene construction of peptide hormone binders

The designed protein sequences were optimized to be both expressed in *S.cerevisiae* and *E. coli*. Linear DNA fragments (eBlocks, Integrated DNA Technologies) encoding design sequences included overhangs suitable for cloning into pETcon3 vector for yeast display(*69*) and Golden Gate cloning into LM627 vector for protein expression(*32*).

Yeast display screening

For the yeast transformation, 50-60 ng of digested pETcon3 and 100 ng of insert (eBlocks, Integrated DNA Technologies) were transformed into *S. cerevisiae* EBY100 strain using the protocol described in ref(*69*). EBY100 cultures were grown in C-Trp-Ura medium supplemented with 2% (w/v) glucose (CTUG). For induction of expression, yeast cells initially grown in CTUG were transferred to SGCAA medium supplemented with 0.2% (w/v) glucose and induced at 30 °C for 16–24 h. Cells were washed with PBSF (PBS with 1% (w/v) BSA) and labeled for 40 minutes with biotinylated peptide targets at room temperature using without-avidity labeling condition.(*69*) After incubation time, cells were washed and resuspended in PBSF for cell sorting (Attune NxT Flow Cytometer, Thermo Fisher Scientific).

NanoBiT screening

Linear gene fragments encoding binder design sequences and target peptide sequences were cloned into *E. coli* expression vectors using Golden Gate assembly; these vectors were pET28b(+) derivatives genetically fusing the smBiT and lgBiT halves of the NanoLuc® Luciferase (Promega) to the binders and peptides respectively. Resulting plasmids were transformed into BL21* (DE3) (Invitrogen) *E. coli* competent cells, then grown in 1mL TBII in 96-deepwell plates at 37C and 600 rpm. After 2 hours, expression was induced with IPTG (0.1 mM) and cells were incubated for an additional 4 hours. Cells were harvested by centrifugation (15 min at 4 kg), then resuspended in 100 uL lysis buffer (10 mM NaP pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mg/mL lysozyme, 10 ug/mL DNAse I, 1 tablet Complete Protease inhibitor / 50 mL). Cells were incubated for 1 hour at room temperature and 600 rpm, then frozen (-80C for 30min) and thawed (37C at 600 rpm for 30min) twice. Lysate was cleared by centrifugation (20 min at 4 kg), and the soluble fraction was then transferred to a 96-well plate for use as stock protein/peptide for conducting the nanoBiT screen. Screens were assembled in

96-well Half Area Black Flat Bottom Polystyrene NBS Microplates (Corning 3686). Binder design smBiT lysate was diluted 12 uL into 1400 uL assay buffer (10 mM NaP pH 7.4, 150 mM NaCl), while target peptide lgBiT lysate was diluted 6 uL into 1400 uL assay buffer. Stock rows in the assay plate were prepared by mixing 40 uL substrate (499.2 uL assay buffer, 20.8 uL Nano-Glo® Luciferase Assay Substrate (Promega)) with 40 uL diluted binder design smBiT lysate, while experimental rows were prepared by adding 50 uL diluted target peptide lgBiT lysate. At read time, 50 uL of the stock row was added to the 50 uL experimental row and mixed quickly and carefully, then luminescence was read immediately for 5 min using a plate reader (Biotek Synergy Neo2).

Peptide synthesis and purification

The PTH-TAMRA peptide was synthesized in-house on a CEM Liberty Blue microwave synthesizer. All L- and D-amino acids were purchased from P3 Biosystems. Oxyma Pure was purchased from CEM, DIC was purchased from Oakwood Chemical, diisopropyl ethylamine (DIEA) and piperidine were purchased from Sigma- Aldrich. Dimethylformamide (DMF) was purchased from Fisher Scientific and treated with an Aldraamine trapping pack prior to use. Synthesis was done on a 0.1 mmol scale on CEM Cl-TCP(Cl) resin. Five equivalents of each amino acid were activated using 0.1 M Oxyma with 2% (v/v) DIEA in DMF, 15.4% (v/v) DIC, and coupled on resin for 4 min with double coupling if needed. This was followed by deprotection using 5 mL of 20% piperidine in DMF for 2 min at 95 °C. Global deprotection was accomplished TFA/Water/TIPS (95:2.5:2.5) for 3 hours. This deprotection mixture was precipitated in 30 mL of ice-cold ethyl ether, centrifuged and decanted, then washed twice more with fresh ether and dried under nitrogen to yield crude peptide for high pressure liquid chromatography (HPLC) purification. The crude peptide was dried and dissolved in a mixture of ACN and water where the entire crude is soluble. This solution was purified on a C18 column in an Agilent HPLC instrument. A linear gradient of increasing ACN with 0.1% TFA was used to purify the samples. UV signal was monitored at 214 nm and all peaks were collected. Peaks were checked using ESI mass spectroscopy for the correct peptide mass. The purified peptide was then lyophilized for further use.

Protein expression and purification in *E. coli* for peptide hormone binders

Protein expression was performed using 50 mL of the Studier autoinduction media supplemented with kanamycin, and grown overnight at 37°C. The cells were harvested by spinning at 4,000 x g for 10 min and then resuspended in lysis buffer (100 mM Tris-HCl, 200 mM NaCl, 50 mM imidazole) supplemented with protease inhibitor tablets (Pierce™ Protease Inhibitor Tablets, EDTA-free). Then, the cells were lysed by sonication in a Qsonica, Q500 with a: 4-pronged horn for 2:30 min ON total, with an

amplitude of 80%. Soluble fractions were clarified by centrifugation at 14,000 x g for 40 minutes, and were subsequently purified by affinity chromatography using bed Ni-NTA resin (Qiagen or Thermo Fisher) on a vacuum manifold. A series of washes using Low-salt buffer (20 mM Tris-HCl, 200 mM NaCl, 50 mM imidazole) and High- salt buffer (20 mM Tris-HCl, 1000 mM NaCl, 50 mM imidazole) were performed prior to elution with Elution buffer (20 mM Tris-HCl, 200 mM NaCl, 500 mM imidazole). After elution, protein samples were filtered and injected into an autosampler-equipped Akta pure system on a Superdex S75 Increase 10/300 GL column at room temperature. The SEC running buffer was 20mM Tris-HCl, 100mM NaCl pH 8. Selected fractions were pooled and concentrated using Spin filters (3 kDa molecular weight cutoff, Amicon, Millipore Sigma) and stored at 4 °C before downstream characterizations. Protein concentrations were determined by absorbance at 280 nm using a NanoDrop spectrophotometer (Thermo Scientific) using their extinction coefficients and molecular weights obtained from their amino acid sequences using the ProtParam tool.

Fluorescence polarization

Fluorescence polarization binding assays were carried out in 96-well plates (Corning 3686), with two-fold serial dilution of designed peptide binders in the presence of 0.5 nM fluorescently labeled peptide targets. Protein and peptide were diluted from their stock concentration into 20mM Tris-HCl pH 8, 100mM NaCl, 0.1% v/v Tween 20, and the protein was titrated in 2-fold serial dilutions onto constant peptide. After incubating the peptide and binder for one hour at room temperature, the fluorescence polarization was measured at the excitation and emission wavelengths of the FAM dye (485/530 nm) or the TAMRA dye (530/590 nm), in a Synergy Neo2 multi-mode plate reader. Titrations were conducted in replicate, and the Kd was fitted with SciPy(*110*). Specifically, curves were fit to N observations of an observed signal, $Signal_i$, at titrated concentrations $[A_{tot}]_i$ according to the following equation:

$$Signal_i = Baseline + Amplitude \frac{AB_{conc}([A_{tot}]_i, [B_{tot}], K_d)}{[B_{tot}]},$$

Where $[B_{tot}]$ is the known total concentration of the binder, $Baseline$ and $Amplitude$ are free parameters, and the concentration of the bound state $[AB]$ is computed as

$$AB_{conc}([A_{tot}]_i, [B_{tot}], K_d) = (([A_{tot}] + [B_{tot}] + K_d) \pm \sqrt{([A_{tot}] + [B_{tot}] + K_d)^2 - 4[A_{tot}][B_{tot}]})/2$$

The unknown parameters (K_D, $Baseline$ and $Amplitude$) were fit using scipy.optimize.curve_fit(*110*), $[B_{tot}]$was additionally fit in the optimization, but only allowed to within 0.5 nM $\pm$ 0.1%.

Peptides used for the assay are shown in Supplementary Table 3.

Chapter IV. Conclusion

These separate works share a common theme: using computational protein design to address challenges where the targets of design may not have a single structural state. In the first work, through the design of stimulus-responsive protein hinges, we show that it is possible to use either energy-based sequence and backbone sampling or probability-based sequence sampling (with a fixed backbone) to achieve the design of proteins with two structural states. We furthermore show that the probability-based approach is more performant and can readily tune the conformational equilibrium between the two states, and we ultimately create a system that can couple biochemical inputs such as protein/peptide binding or redox, *through a conformational change with a tunable transfer function*, to a variety of interesting outputs, including peptide binding or FRET. In the second work, we use a combination of highly efficient coarse-grained backbone modeling and probability-based sequence sampling to achieve high success rates in binding to flexible peptide hormones with low nanomolar affinities. When we incorporate a backbone refinement step using a denoising diffusion on the models for experimental hits made using this and other approaches, we are able to obtain affinities that are sub-nanomolar. Together, these works solve both the design of structured conformational changes and the design of proteins that tightly bind flexible helical peptide hormones. Future work will no doubt combine these capabilities, perhaps resulting in proteins that can conditionally bind or release hormones, or synthetic hormone-responsive systems that hack cellular behavior.

References

1. M. A. Hartman, J. A. Spudich, The myosin superfamily at a glance. *J. Cell Sci.* **125**, 1627–1632 (2012).

2. G. M. Edelman, Antibody structure and molecular immunology. *Science*. **180**, 830–840 (1973).

3. S. G. Wildman, Along the trail from Fraction I protein to Rubisco (ribulose bisphosphate carboxylase-oxygenase). *Photosynth. Res.* **73**, 243–250 (2002).

4. D. Okuno, R. Iino, H. Noji, Rotation and structure of FoF1-ATP synthase. *J. Biochem.* **149**, 655–664 (2011).

5. P.-S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature*. **537**, 320–327 (2016).

6. C. Levinthal, Are there pathways for protein folding? *Journal de Chimie Physique*. **65**, 44–45 (1968).

7. A. R. Dinner, A. Sali, L. J. Smith, C. M. Dobson, M. Karplus, Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339 (2000).

8. J. D. Bernal, Structure of Proteins*. *Nature*. **143**, 663–667 (1939).

9. L. Pauling, Modern Structural Chemistry. *Science*. **123**, 255–258 (1956).

10. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science*. **181**, 223–230 (1973).

11. B. I. Dahiyat, S. L. Mayo, De novo protein design: fully automated sequence selection. *Science*. **278**, 82–87 (1997).

12. B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, Design of a novel globular protein fold with atomic-level accuracy. *Science*. **302**, 1364–1368 (2003).

13. Y.-R. Lin, N. Koga, R. Tatsumi-Koga, G. Liu, A. F. Clouser, G. T. Montelione, D. Baker, Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5478–85 (2015).

14. T. J. Brunette, F. Parmeggiani, P.-S. Huang, G. Bhabha, D. C. Ekiert, S. E. Tsutakawa, G. L. Hura, J. A. Tainer, D. Baker, Exploring the repeat protein universe through computational protein design. *Nature*. **528**, 580–584 (2015).

15. T. J. Brunette, M. J. Bick, J. M. Hansen, C. M. Chow, J. M. Kollman, D. Baker, Modular repeat protein sculpting using rigid helical junctions. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8870–8875 (2020).

16. J. A. Fallas, G. Ueda, W. Sheffler, V. Nguyen, D. E. McNamara, B. Sankaran, J. H. Pereira, F. Parmeggiani, T. J. Brunette, D. Cascio, T. R. Yeates, P. Zwart, D. Baker, Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).

17. H. Shen, J. A. Fallas, E. Lynch, W. Sheffler, B. Parry, N. Jannetty, J. Decarreau, M. Wagenbach, J. J. Vicente, J. Chen, L. Wang, Q. Dowling, G. Oberdorfer, L. Stewart, L. Wordeman, J. De Yoreo, C. Jacobs-Wagner, J. Kollman, D. Baker, De novo design of self-assembling helical protein filaments. *Science*. **362**, 705–709 (2018).

18. Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C. Xu, P.-S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King, D.

Baker, Corrigendum: Design of a hyperstable 60-subunit protein icosahedron. *Nature*. **540**, 150 (2016).

19. Z. Chen, S. E. Boyken, M. Jia, F. Busch, D. Flores-Solis, M. J. Bick, P. Lu, Z. L. VanAernum, A. Sahasrabuddhe, R. A. Langan, S. Bermeo, T. J. Brunette, V. K. Mulligan, L. P. Carter, F. DiMaio, N. G. Sgourakis, V. H. Wysocki, D. Baker, Programmable design of orthogonal protein heterodimers. *Nature*. **565**, 106–111 (2019).

20. S. Bermeo, A. Favor, Y.-T. Chang, A. Norris, S. E. Boyken, Y. Hsia, H. K. Haddox, C. Xu, T. J. Brunette, V. H. Wysocki, G. Bhabha, D. C. Ekiert, D. Baker, De novo design of obligate ABC-type heterotrimeric proteins. *Nat. Struct. Mol. Biol.* **29**, 1266–1276 (2022).

21. Z. Chen, R. D. Kibler, A. Hunt, F. Busch, J. Pearl, M. Jia, Z. L. VanAernum, B. I. M. Wicky, G. Dods, H. Liao, M. S. Wilken, C. Ciarlo, S. Green, H. El-Samad, J. Stamatoyannopoulos, V. H. Wysocki, M. C. Jewett, S. E. Boyken, D. Baker, De novo design of protein logic gates. *Science*. **368**, 78–84 (2020).

22. S. E. Boyken, M. A. Benhaim, F. Busch, M. Jia, M. J. Bick, H. Choi, J. C. Klima, Z. Chen, C. Walkey, A. Mileant, A. Sahasrabuddhe, K. Y. Wei, E. A. Hodge, S. Byron, A. Quijano-Rubio, B. Sankaran, N. P. King, J. Lippincott-Schwartz, V. H. Wysocki, K. K. Lee, D. Baker, De novo design of tunable, pH-driven conformational changes. *Science*. **364**, 658–664 (2019).

23. R. A. Langan, S. E. Boyken, A. H. Ng, J. A. Samson, G. Dods, A. M. Westbrook, T. H. Nguyen, M. J. Lajoie, Z. Chen, S. Berger, V. K. Mulligan, J. E. Dueber, W. R. P. Novak, H. El-Samad, D. Baker, De novo design of bioactive protein switches. *Nature*. **572**, 205–210 (2019).

24. N. B. Woodall, Z. Weinberg, J. Park, F. Busch, R. S. Johnson, M. J. Feldbauer, M. Murphy, M. Ahlrichs, I. Yousif, M. J. MacCoss, V. H. Wysocki, H. El-Samad, D. Baker, De novo design of tyrosine and serine kinase-driven protein switches. *Nat. Struct. Mol. Biol.* **28**, 762–770 (2021).

25. A. Quijano-Rubio, H.-W. Yeh, J. Park, H. Lee, R. A. Langan, S. E. Boyken, M. J. Lajoie, L. Cao, C. M. Chow, M. C. Miranda, J. Wi, H. J. Hong, L. Stewart, B.-H. Oh, D. Baker, De novo design of modular and tunable protein biosensors. *Nature*. **591**, 482–487 (2021).

26. P.-S. Huang, Y.-E. A. Ban, F. Richter, I. Andre, R. Vernon, W. R. Schief, D. Baker, RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One*. **6**, e24109 (2011).

27. P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science*. **346**, 481–485 (2014).

28. G. Grigoryan, W. F. Degrado, Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011).

29. D. D. Sahtoe, F. Praetorius, A. Courbet, Y. Hsia, B. I. M. Wicky, N. I. Edman, L. M. Miller, B. J. R. Timmermans, J. Decarreau, H. M. Morris, A. Kang, A. K. Bera, D. Baker, Reconfigurable asymmetric protein assemblies through implicit negative design. *Science*. **375**, eabj7662 (2022).

30. D. D. Sahtoe, A. Coscia, N. Mustafaoglu, L. M. Miller, D. Olal, I. Vulovic, T.-Y. Yu, I. Goreshnik, Y.-R. Lin, L. Clark, F. Busch, L. Stewart, V. H. Wysocki, D. E. Ingber, J. Abraham, D. Baker, Transferrin receptor targeting by de novo sheet extension. *Proceedings of the National Academy of Sciences*. **118**, e2021569118 (2021).

31. I. Anishchenko, S. J. Pellock, T. M. Chidyausiku, T. A. Ramelot, S. Ovchinnikov, J. Hao, K. Bafna, C. Norn, A. Kang, A. K. Bera, F. DiMaio, L. Carter, C. M. Chow, G. T. Montelione, D. Baker, De novo protein design by deep network hallucination. *Nature*. **600**, 547–552 (2021).

32. B. I. M. Wicky, L. F. Milles, A. Courbet, R. J. Ragotte, J. Dauparas, E. Kinfu, S. Tipps, R. D. Kibler, M. Baek, F. DiMaio, X. Li, L. Carter, A. Kang, H. Nguyen, A. K. Bera, D. Baker, Hallucinating symmetric protein assemblies. *Science*. **378**, 56–61 (2022).

33. J. Wang, S. Lisanza, D. Juergens, D. Tischer, J. L. Watson, K. M. Castro, R. Ragotte, A. Saragovi, L. F. Milles, M. Baek, I. Anishchenko, W. Yang, D. R. Hicks, M. Expòsit, T. Schlichthaerle, J.-H. Chun, J. Dauparas, N. Bennett, B. I. M. Wicky, A. Muenks, F. DiMaio, B. Correia, S. Ovchinnikov, D. Baker, Scaffolding protein functional sites using deep learning. *Science*. **377**, 387–394 (2022).

34. J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models, , doi:10.1101/2022.12.09.519842.

35. J. Ingraham, M. Baranov, Z. Costello, V. Frappier, A. Ismail, S. Tie, W. Wang, V. Xue, F. Obermeyer, A. Beam, G. Grigoryan, Illuminating protein space with a programmable generative model. *bioRxiv* (2022), p. 2022.12.01.518682.

36. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack Jr, R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

37. M. V. Shapovalov, R. L. Dunbrack Jr, A smoothed backbone-dependent rotamer

library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. **19**, 844–858 (2011).

38. H. M. Berman, J. Westbrook, Z. Feng, The protein data bank. *Nucleic acids* (2000) (available at https://academic.oup.com/nar/article-abstract/28/1/235/2384399).

39. C. Norn, B. I. M. Wicky, D. Juergens, S. Liu, D. Kim, D. Tischer, B. Koepnick, I. Anishchenko, Foldit Players, D. Baker, S. Ovchinnikov, Protein sequence design by conformational landscape optimization. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2017228118.

40. A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, J. Ma, R. Fergus, Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2016239118.

41. J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, Robust deep learning–based protein sequence design using ProteinMPNN. *Science*. **378**, 49–56 (2022).

42. S. Ovchinnikov, P.-S. Huang, Structure-based protein design with deep learning. *Curr. Opin. Chem. Biol.* **65**, 136–144 (2021).

43. D. Chivian, T. Robertson, R. Bonneau, D. Baker, Ab initio methods. *Structural bioinformatics*. **44**, 547–557 (2003).

44. K. T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).

45. P. Bradley, K. M. S. Misura, D. Baker, Toward high-resolution de novo structure prediction for small proteins. *Science*. **309**, 1868–1871 (2005).

46. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).

47. N. Bennett, B. Coventry, I. Goreshnik, B. Huang, A. Allen, D. Vafeados, Y. P. Peng, J. Dauparas, M. Baek, L. Stewart, F. DiMaio, S. De Munck, S. N. Savvides, D. Baker, Improving de novo Protein Binder Design with Deep Learning. *bioRxiv* (2022), p. 2022.06.15.495993.

48. J. B. Stiller, R. Otten, D. Häussinger, P. S. Rieder, D. L. Theobald, D. Kern, Structure determination of high-energy states in a dynamic protein ensemble. *Nature*. **603**, 528–535 (2022).

49. S. J. Kerns, R. V. Agafonov, Y.-J. Cho, F. Pontiggia, R. Otten, D. V. Pachov, S. Kutter, L. A. Phung, P. N. Murphy, V. Thai, T. Alber, M. F. Hagan, D. Kern, The energy landscape of adenylate kinase during catalysis. *Nat. Struct. Mol. Biol.* **22**, 124–131 (2015).

50. A. Bisello, M. Chorev, M. Rosenblatt, L. Monticelli, D. F. Mierke, S. L. Ferrari, Selective ligand-induced stabilization of active and desensitized parathyroid hormone type 1 receptor conformations. *J. Biol. Chem.* **277**, 38524–38530 (2002).

51. T. Movassagh, K. H. Bui, H. Sakakibara, K. Oiwa, T. Ishikawa, Nucleotide-induced global conformational changes of flagellar dynein arms revealed by in situ analysis. *Nat. Struct. Mol. Biol.* **17**, 761–767 (2010).

52. W. A. Catterall, G. Wisedchaisri, N. Zheng, The conformational cycle of a prototypical voltage-gated sodium channel. *Nat. Chem. Biol.* **16**, 1314–1320 (2020).

53. E. G. B. Evans, J. L. W. Morgan, F. DiMaio, W. N. Zagotta, S. Stoll, Allosteric conformational change of a cyclic nucleotide-gated ion channel revealed by DEER spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10839–10847 (2020).

54. B. Arragain, G. Effantin, P. Gerlach, J. Reguera, G. Schoehn, S. Cusack, H. Malet, Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes. *Nat. Commun.* **11**, 3590 (2020).

55. A. K. Kim, L. L. Porter, Functional and Regulatory Roles of Fold-Switching Proteins. *Structure*. **29**, 6–14 (2021).

56. J.-H. Ha, S. N. Loh, Protein conformational switches: from nature to design. *Chemistry*. **18**, 7984–7999 (2012).

57. R. Koga, M. Yamamoto, T. Kosugi, N. Kobayashi, T. Sugiki, T. Fujiwara, N. Koga, Robust folding of a de novo designed ideal protein even with most of the core mutated to valine. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 31149–31156 (2020).

58. A. F. Dishman, B. F. Volkman, Design and discovery of metamorphic proteins. *Curr. Opin. Struct. Biol.* **74**, 102380 (2022).

59. N. H. Joh, T. Wang, M. P. Bhate, R. Acharya, Y. Wu, M. Grabe, M. Hong, G. Grigoryan, W. F. De Grado, De novo design of a transmembrane Zn2+-transporting four-helix bundle.

60. 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).

61. V. K. Mulligan, C. S. Kang, M. R. Sawaya, S. Rettie, X. Li, I. Antselovich, T. W. Craven, A. M. Watkins, J. W. Labonte, F. DiMaio, T. O. Yeates, D. Baker, Computational design of mixed chirality peptide macrocycles with internal symmetry. *Protein Sci.* **29**, 2433–2445 (2020).

62. K. Y. Wei, D. Moschidi, M. J. Bick, S. Nerli, A. C. McShan, L. P. Carter, P.-S. Huang, D. A. Fletcher, N. G. Sgourakis, S. E. Boyken, D. Baker, Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7208–7215 (2020).

63. J. Z. Zhang, H.-W. Yeh, A. C. Walls, B. I. M. Wicky, K. R. Sprouse, L. A. VanBlargan, R. Treger, A. Quijano-Rubio, M. N. Pham, J. C. Kraft, I. C. Haydon, W. Yang, M. DeWitt, J. E. Bowen, C. M. Chow, L. Carter, R. Ravichandran, M. H. Wener, L. Stewart, D. Veesler, M. S. Diamond, A. L. Greninger, D. M. Koelle, D. Baker, Thermodynamically coupled biosensors for detecting neutralizing antibodies against SARS-CoV-2 variants. *Nat. Biotechnol.* **40**, 1336–1340 (2022).

64. M. Wang, Q. Guo, K. Zhu, B. Fang, Y. Yang, M. Teng, X. Li, Y. Tao, Interface switch mediates signal transmission in a two-component system. *Proceedings of the National Academy of Sciences*. **117**, 30433–30440 (2020).

65. F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker, F. Players, Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18949–18953 (2011).

66. M. D. Tyka, D. A. Keedy, I. André, F. Dimaio, Y. Song, D. C. Richardson, J. S. Richardson, D. Baker, Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607–618 (2011).

67. N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker, Principles for designing ideal protein structures. *Nature*. **491**, 222–227 (2012).

68. T. M. Lauer, N. J. Agrawal, N. Chennamsetty, K. Egodage, B. Helk, B. L. Trout, Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *J. Pharm. Sci.* **101**, 102–115 (2012).

69. L. Cao, B. Coventry, I. Goreshnik, B. Huang, W. Sheffler, J. S. Park, K. M. Jude, I. Marković, R. U. Kadam, K. H. G. Verschueren, K. Verstraete, S. T. R. Walsh, N. Bennett, A. Phal, A. Yang, L. Kozodoy, M. DeWitt, L. Picton, L. Miller, E.-M. Strauch, N. D. DeBouver, A. Pires, A. K. Bera, S. Halabiya, B. Hammerson, W. Yang, S. Bernard, L. Stewart, I. A. Wilson, H. Ruohola-Baker, J. Schlessinger, S. Lee, S. N. Savvides, K. C. Garcia, D. Baker, Design of protein-binding proteins from the target structure alone. *Nature*. **605**, 551–560 (2022).

70. Berliner, Olga Hankovszky, A Novel Reversible Thiol-Specific Spin Label: Papain Active Site Labeling and Inhibition'. *Anal. Biochem.*

71. M. H. Tessmer, S. Stoll, chiLife: An open-source Python package for in silico spin

labeling and integrative protein modeling. *bioRxiv* (2022), p. 2022.12.23.521725.

72. M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. **373**, 871–876 (2021).

73. J. P. Hallinan, L. A. Doyle, B. W. Shen, M. M. Gewe, B. Takushi, M. A. Kennedy, D. Friend, J. M. Roberts, P. Bradley, B. L. Stoddard, Design of functionalised circular tandem repeat proteins with longer repeat topologies and enhanced subunit contact surfaces. *Commun Biol*. **4**, 1240 (2021).

74. S. Yao, A. Moyer, Y. Zheng, Y. Shen, X. Meng, C. Yuan, Y. Zhao, H. Yao, D. Baker, C. Wu, De novo design and directed folding of disulfide-bridged peptide heterodimers. *Nat. Commun.* **13**, 1539 (2022).

75. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

76. S. V. Torres, P. J. Y. Leung, I. D. Lutz, P. Venkatesh, J. L. Watson, F. Hink, H.-H. Huynh, A. H.-W. Yeh, D. Juergens, N. R. Bennett, A. N. Hoofnagle, E. Huang, M. J. MacCoss, M. Expòsit, G. R. Lee, E. N. Korkmaz, J. Nivala, L. Stewart, J. M. Rodgers, D. Baker, De novo design of high-affinity protein binders to bioactive helical peptides. *bioRxiv* (2022), p. 2022.12.10.519862.

77. L. An, G. R. Lee, De Novo Protein Design Using the Blueprint Builder in Rosetta. *Curr. Protoc. Protein Sci.* **102**, e116 (2020).

78. M. F. Sauer, A. M. Sevy, J. E. Crowe Jr, J. Meiler, Multi-state design of flexible proteins predicts sequences optimal for conformational change. *PLoS Comput. Biol.* **16**, e1007339 (2020).

79. N. Chennamsetty, V. Voynov, V. Kayser, B. Helk, B. L. Trout, Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11937–11942 (2009).

80. S. Warszawski, R. Netzer, D. S. Tawfik, S. J. Fleishman, A “fuzzy”-logic language for encoding multiple physical traits in biomolecules. *J. Mol. Biol.* **426**, 4125–4138 (2014).

81. S. T. Kudlacek, S. Metz, D. Thiono, A. M. Payne, T. T. N. Phan, S. Tian, L. J. Forsberg, J. Maguire, I. Seim, S. Zhang, A. Tripathy, J. Harrison, N. I. Nicely, S. Soman, M. K. McCracken, G. D. Gromowski, R. G. Jarman, L. Premkumar, A. M. de Silva, B. Kuhlman, Designed, highly expressing, thermostable dengue virus 2 envelope protein dimers elicit quaternary epitope antibodies. *Sci. Adv.* **7**, eabg4084

(2021).

82. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat. Methods*. **19**, 679–682 (2022).

83. B. Dang, M. Mravic, H. Hu, N. Schmidt, B. Mensa, W. F. DeGrado, SNAC-tag for sequence-specific chemical protein cleavage. *Nat. Methods*. **16**, 319–322 (2019).

84. M. H. Tessmer, E. R. Canarie, S. Stoll, Comparative evaluation of spin-label modeling methods for protein structural studies. *Biophys. J.* **121**, 3508–3519 (2022).

85. S. Stoll, A. Schweiger, EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **178**, 42–55 (2006).

86. L. Fábregas Ibáñez, G. Jeschke, S. Stoll, DeerLab: a comprehensive software package for analyzing dipolar electron paramagnetic resonance spectroscopy data. *Magn Reson (Gott)*. **1**, 209–224 (2020).

87. L. Fábregas-Ibáñez, G. Jeschke, S. Stoll, Compactness regularization in the analysis of dipolar EPR spectroscopy data. *J. Magn. Reson.* **339**, 107218 (2022).

88. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

89. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).

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

91. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

92. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

93. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall III, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson, MolProbity: More and better reference data for improved all-atom

structure validation. *Protein Sci.* **27**, 293–315 (2018).

94. S. Fukuhara, K. Kobayashi, T. Kusakizako, W. Iida, M. Kato, W. Shihoya, O. Nureki, Structure of the human secretin receptor coupled to an engineered heterotrimeric G protein. *Biochem. Biophys. Res. Commun.* **533**, 861–866 (2020).

95. C. Boesch, A. Bundi, M. Oppliger, K. Wüthrich, 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur. J. Biochem.* **91**, 209–214 (1978).

96. C. Park, J. Kim, S.-B. Ko, Y. K. Choi, H. Jeong, H. Woo, H. Kang, I. Bang, S. A. Kim, T.-Y. Yoon, C. Seok, W. Im, H.-J. Choi, Structural basis of neuropeptide Y signaling through Y1 receptor. *Nat. Commun.* **13**, 853 (2022).

97. N. Shimizu, J. Guo, T. J. Gardella, Parathyroid Hormone (PTH)-(1--14) and-(1--11) analogs conformationally constrained by α-aminoisobutyric acid mediate full agonist responses via the Juxtamembrane region of the PTH-1 receptor. *J. Biol. Chem.* **276**, 49003–49012 (2001).

98. P. E. Wright, H. J. Dyson, Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38 (2009).

99. T. Lazar, A. Tantos, P. Tompa, E. Schad, Intrinsic protein disorder uncouples affinity from binding specificity. *Protein Sci.* **31**, e4455 (2022).

100. F. J. Gisdon, J. P. Kynast, M. Ayyildiz, A. V. Hine, A. Plückthun, B. Höcker, Modular peptide binders - development of a predictive technology as alternative for reagent antibodies. *Biol. Chem.* **403**, 535–543 (2022).

101. K. Wu, H. Bai, Y.-T. Chang, R. Redler, K. E. McNally, W. Sheffler, T. J. Brunette, D. R. Hicks, T. E. Morgan, T. J. Stevens, A. Broerman, I. Goreshnik, M. DeWitt, C. M. Chow, Y. Shen, L. Stewart, E. Derivery, D. A. Silva, G. Bhabha, D. Ekiert, D. Baker, De novo design of modular peptide binding proteins by superhelical matching. *bioRxiv* (2022), p. 2022.11.14.514089.

102. D. T. Jones, W. R. Taylor, J. M. Thornton, A new approach to protein fold recognition. *Nature*. **358**, 86–89 (1992).

103. D. J. Drucker, The role of gut hormones in glucose homeostasis. *J. Clin. Invest.* **117**, 24–32 (2007).

104. D. J. Mandell, T. Kortemme, Backbone flexibility in computational protein design. *Curr. Opin. Biotechnol.* **20**, 420–428 (2009).

105. J. Ludwiczak, A. Jarmula, S. Dunin-Horkawicz, Combining Rosetta with molecular dynamics (MD): A benchmark of the MD-based ensemble protein design. *J. Struct. Biol.* **203**, 54–61 (2018).

106. Y. Liu, L. Chen, H. Liu, De novo protein backbone generation based on diffusion with structured priors and adversarial training. *bioRxiv* (2022), p. 2022.12.17.520847.

107. D. Larhammar, Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul. Pept.* **62**, 1–11 (1996).

108. M. Baker, Reproducibility crisis: Blame it on the antibodies. *Nature*. **521**, 274–276 (2015).

109. A. Bradbury, A. Plückthun, Reproducibility: Standardize antibodies used in research. *Nature*. **518**, 27–29 (2015).

110. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*. **17**, 261–272 (2020).