



Exploiting conformational dynamics to modulate the function of designed proteins

Enrico Rennella^{a,b,c,1}, Danny D. Sahtoe^{d,e,f}, David Baker^{d,e,f}, and Lewis E. Kay^{a,b,c,g,1}

Edited by G. Marius Clore, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD; received February 23, 2023; accepted March 22, 2023

With the recent success in calculating protein structures from amino acid sequences using artificial intelligence-based algorithms, an important next step is to decipher how dynamics is encoded by the primary protein sequence so as to better predict function. Such dynamics information is critical for protein design, where strategies could then focus not only on sequences that fold into particular structures that perform a given task, but would also include low-lying excited protein states that could influence the function of the designed protein. Herein, we illustrate the importance of dynamics in modulating the function of C34, a designed α/β protein that captures β -strands of target ligands and is a member of a family of proteins designed to sequester β -strands and β hairpins of aggregation-prone molecules that lead to a variety of pathologies. Using a strategy to “see” regions of *apo* C34 that are invisible to NMR spectroscopy as a result of pervasive conformational exchange, as well as a mutagenesis approach whereby C34 molecules are stabilized into a single conformer, we determine the structures of the predominant conformations that are sampled by C34 and show that these attenuate the affinity for cognate peptide. Subsequently, the observed motion is exploited to develop an allosterically regulated peptide binder whose binding affinity can be controlled through the addition of a second molecule. Our study emphasizes the unique role that NMR can play in directing the design process and in the construction of new molecules with more complex functionality.

protein design | conformational dynamics | chemical exchange saturation transfer | NMR spectroscopy | ligand binding

De novo design of customized proteins with new functionalities is one of the most promising and challenging fields in biotechnology (1–3). Essentially the idea is to increase the odds of discovering a protein or family of protein molecules of a particular function by working with the whole range of available protein sequence space, rather than the much smaller range generated by evolution, using the principles of biophysics to guide the rational design process. In addition to informing on new aspects of protein science, the proteins so generated would offer unprecedented opportunities to manipulate biological function, with applications ranging from medicine to the production of protein components essential for industrial processes. Central to the design strategy of novel, functional proteins is the assumption that they fold into the lowest energy states on their respective energy landscapes. Advances in computational strategies, including the ability to use artificial intelligence-based approaches to elucidate accurate three-dimensional folds from the primary amino acid sequence of proteins (4, 5), has greatly increased the potential for the accurate design of molecular function. The molecules so produced are largely generated from single structure-based design strategies seeking an amino acid sequence that folds within a deep energy minimum to assume a protein backbone conformation that is considered optimal for the targeted function of interest. While this approach, based on the generation of presumed static structures has been successful in many cases (6–12), it does not consider the complexity of the conformational landscape of proteins where multiple low energy minima may be visited, allowing for a variety of different functionalities, or at the very least modulating the functionality that was intended to be engineered in the first place. Approaches that produce multi-state proteins, and hence include dynamics in the design strategy (13–17), are a promising avenue for increasing the function of designed molecules, but these are at a very early stage. Like the success of AlphaFold2 (4) and RoseTTAFold (5) that was based on training with a large set of structures, robust design strategies that include low energy states on an energy landscape must await the development of training sets that correlate how an amino acid sequence is able to access a set of conformers rather than only a single one.

Recently Sahtoe, Baker, and coworkers have reported the successful de novo design of a family of α/β protein scaffolds (18) that are able to sequester disordered peptides,

Significance

Designing biomolecules with unique functionalities is an important goal of protein research, with applications ranging from biomedicine to material science. Most design strategies are based on a single target having a unique low energy conformation. However, it is known that proteins can be dynamic, exchanging between multiple conformations of similar free energies. Here we investigate the structural dynamics of a designed protein that binds β -strands, using NMR spectroscopy, establishing that it interconverts between a pair of distinct conformers that compete with binding of the cognate peptide. It is shown that the conformational dynamics can be manipulated to engineer increased functionality in the design and that NMR can be a powerful tool for improving the design process.

Author contributions: E.R., D.D.S., D.B., and L.E.K. designed research; E.R. performed research; E.R. analyzed data; and E.R. and L.E.K. wrote the paper.

Competing interest statement: The authors have patent filings to disclose, D.D.S. and D.B. are inventors on a provisional patent application submitted by the University of Washington for the design, composition, and function of the original WT C34 protein.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: rennella@pound.med.utoronto.ca or kay@pound.med.utoronto.ca.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2303149120/-/DCSupplemental>.

Published April 24, 2023.

β -hairpins, or stretches of sequence within proteins through β -strand pairing interactions. These scaffolds contain a large binding groove engineered between a pair of strands that, in turn, are supported through a network of α -helices. The designs are intended to capture stretches of sequence present in amyloidogenic proteins such as A β 42, microtubule associated protein tau, transthyretin, and serum amyloid A1 and both in vitro and cell-based experiments have established their efficacy. In principle, understanding the structural dynamics of the designed molecules and how their dynamics affect function would be an important step in refinement of the design strategy and for the development of even more potent molecules.

NMR spectroscopy is a powerful technique for obtaining residue-specific information about the dynamic properties of a protein over a wide range of timescales (19–23), providing kinetic, thermodynamic, and structural information on the interconversion between the ground state and a set of low energy conformers that are accessible to it under physiological conditions. Herein we use a series of NMR spin relaxation experiments that are sensitive to conformational exchange processes to map out the energy landscape of one of the designed α/β proteins, so-called C34, to elucidate the structures of the interconverting states, and to establish how the conversion between structures on the landscape affects its engineered function of capturing unfolded β -strands. We establish that unlike predictions of the *apo* C34 structure from AlphaFold2 that suggest that the molecule folds into a stable well-defined conformation, C34 is highly dynamic, with these dynamics attenuating β -peptide binding. Finally, we show how the dynamic properties of C34 can be used to engineer increasing complexity into the system via a few targeted mutations that convert it into an allosterically modulated binder of β -strand peptides.

Results

C34, a Designed β -Strand Binder. The 124-residue C34 is one of a set of designed α/β proteins into which a deep binding groove has been engineered to bind β -strands (18). The predicted structure of C34 indicates an architecture of four α helices and four β strands, with the peptide binding groove positioned between strands β 2– β 3, and stabilized by proximal helices. A preliminary NMR analysis of C34 bound with an optimized octapeptide target (DVRFQVRE) showed the expected topology (18), with the exception that the predicted C-terminal β strand (β 4) was not formed (Fig. 1). However, pervasive conformational exchange in the *apo* state prevented an analysis of the C34 energy landscape at that time, and it was not possible, therefore, to establish the role of dynamics in peptide binding and how it might be exploited to

modulate function. Addressing these questions forms the basis of the work presented here.

“Seeing” the NMR Invisible C-Terminal Region of C34. ^1H - ^{15}N HSQC spectra of the *apo* wild-type (WT) state of C34 (referred to in what follows as *apo* C34 WT) recorded at 25 °C are extensively broadened (18) (SI Appendix, Fig. S1), preventing the complete assignment of resonances from triple-resonance based experiments. Thus, only limited structural information on C34 is available, confirming the presence of four α -helices and a pair of beta strands, β 1 and β 2, (Fig. 2 A, Middle row), that are consistent with expectations based on structural predictions of this protein (Fig. 2 A, Top row). Very few assignments are available for the C-terminal region of C34, which includes a predicted β 3- β 4 hairpin, as this region is severely affected by conformational exchange. The analysis of ^{15}N Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion (24, 25) profiles using a two-site exchange model establishes that the pervasive broadening derives from a unimolecular process (i.e., independent of protein concentration), with a rate of interconversion between states of approximately 400 s^{-1} at 25 °C (Fig. 2B and SI Appendix, Fig. S2). It is not possible to estimate the relative populations of the interconverting states from these data as exchange is in the fast regime on the chemical shift timescale, but it is clear from the quality of spectra recorded on *apo* C34 WT (SI Appendix, Fig. S1) that the populations must be similar (see below).

We have shown previously that addition of an octapeptide that is designed specifically to bind to C34 results in peak narrowing and quenching of the conformational exchange in the WT protein (18) (SI Appendix, Fig. S3). Complete resonance assignments for the bound state were obtained using standard triple-resonance experiments (28), confirming the predicted structure, and validating, therefore, the C34 design strategy (18), except for the region encompassing the predicted fourth strand, β 4, which is flexible in solution and shows little propensity to adopt a beta fold (Fig. 2 A, Bottom row). It is worth mentioning that assignments of the bound state of C34 proved critical for assigning some regions of the *apo* form, even at the high temperature that was used (50 °C for the “*apo*” assignments), where spectral quality was improved, although still rather marginal.

We were interested in elucidating the structural propensities of the C-terminal region of the *apo* form of C34 as a first step toward characterizing the process(es) giving rise to the pervasive conformational exchange in this system. Here we used a strategy that exploits the much higher quality spectra recorded of the bound state to “observe” invisible resonances of the *apo* form using Chemical Exchange Saturation Transfer (CEST) experiments (27, 29–32). In a “traditional” CEST experiment the NMR signals

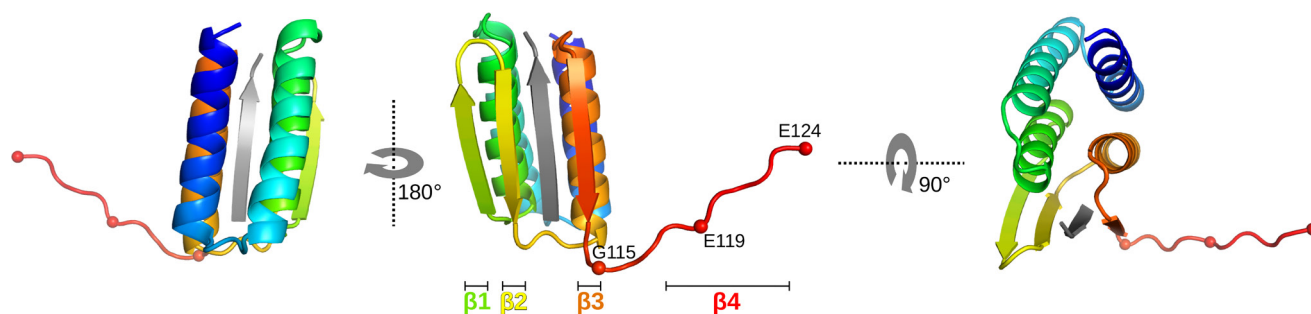


Fig. 1. Cartoon representation of the predicted structure of C34 bound to a cognate octapeptide ligand (gray β -strand) (18). The positions of the putative β -strands, as predicted, are highlighted in the center image; the region corresponding to β 4 is disordered, established from chemical shift assignments of the bound form of C34 (18). Intermolecular ^1H - ^1H NOEs confirm the arrangement of the ligand, bound to a groove between strands β 2 and β 3 (18). Locations of three mutation sites that are used in this study are indicated by spheres. The numbering begins from the SGGGS scar from the TEV cleavage site (i.e., N-terminal S is 1).

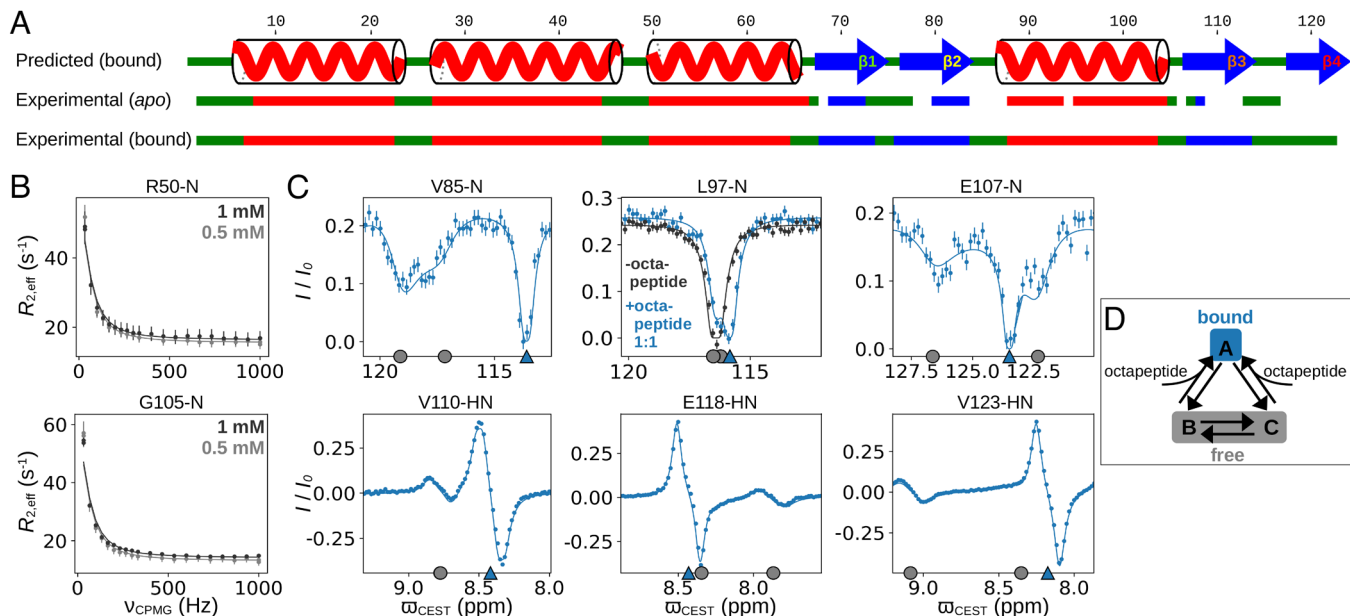


Fig. 2. Overview of the secondary structure and dynamics of C34 WT in the *apo* and bound states. (A) Predicted secondary structure of the bound state from protein design (18) (Top row), and the experimentally determined secondary structures of the *apo* (Center) and bound (Bottom) states based on assigned chemical shifts, using the program SSP (26), where green, red, and blue lines denote random coil, helix, and strand, respectively. In this analysis a residue is considered to be in an α -helix if SSP > 0.5 and in a β -strand if SSP < -0.5. Gaps in the secondary structure profile for *apo* C34 WT are the result of a lack of data due to extensive peak broadening, even though assignment spectra were recorded 50 °C where spectral quality was markedly improved over 25 °C. (B) Experimental data (dots) and fits (continuous lines) for selected ^{15}N CPMG profiles, $R_{2,\text{eff}}$ vs. ν_{CPMG} , (Materials and Methods) from *apo* C34 WT at two protein concentrations of 0.5 mM and 1 mM, 25 °C, 14.1 T. (C) Experimental data (dots) and corresponding fits (continuous lines) of ^{15}N (weak B_1 field of 5 Hz) and ^1H -CEST profiles of a number of residues recorded in the absence (for L97-N only) and in the presence of an octapeptide ligand, 25 °C. I/I_0 is the ratio of peak intensities with (I) and without (I_0) the CEST element (27). The “N” or “HN” in V85-N or V123-HN, for example, indicates that an ^{15}N -CEST or ^1H -CEST profile is displayed. Data are fit to a three-site exchange model, with the gray circles on the x-axis indicating the fitted chemical shift values for spins in the *apo* states, while the blue triangles denote the corresponding values for spins in the bound state. (D) The three-state model used to fit the data.

of rare, invisible states are “observed” via a process in which they are amplified through detection via the highly populated ground state, so long as the fractional populations of the rare states are on the order of 0.5% or higher, and the rare states exchange with the ground conformer with rates less than several hundred/s (27). In the application here, the bound form of C34 is analogous to the ground state (high quality, observable spectra), allowing observation of the resonance positions of the *apo* form via magnetization transfer. Fig. 2C and SI Appendix, Fig. S4 show a number of ^{15}N - and ^1H -CEST profiles (circles) recorded on a sample of 1:1 (450 μM) C34 WT–octapeptide, 25 °C, comprising an approximately equal ratio of peptide-bound and *apo* molecules. Notably, the ^{15}N CEST curves for V85 and E107 show three dips, with the major dips at the resonance positions of the corresponding peaks in the bound state (“A” in Fig. 2D, highlighted by blue triangles in Fig. 2C), with the additional pairs of dips indicating that there are at least two *apo* conformations (“B” and “C”; denoted with gray circles in Fig. 2C) in exchange with the bound form. Not all of the CEST profiles show three well-defined dips, for example the ^{15}N CEST curve for L97 where only a pair of dips is observed. As is the case in standard 1D NMR spectra, the number of peaks (dips) in CEST profiles is a function of the shift differences between corresponding nuclei in major and minor states, $\Delta\omega_{\text{AB}} (= \omega_{\text{B}} - \omega_{\text{A}}$, ppm) and $\Delta\omega_{\text{AC}}$ in the present application. Fits of all of the CEST profiles to a three-site exchange model (Fig. 2D) establish that $k_{\text{ex,AB}} (= k_{\text{AB}} + k_{\text{BA}})$ and $k_{\text{ex,AC}}$ are a few s^{-1} (SI Appendix, Fig. S4C), while the exchange between the two free states, $k_{\text{ex,BC}}$ is $\sim 420 \text{ s}^{-1}$. The value of $k_{\text{ex,BC}}$ was obtained by a joint fit of the CEST and CPMG data, and was in close agreement with the rate obtained by fits of the CPMG datasets of *apo* C34 exclusively (see above). Although the populations of states B and C are not very well defined, the best fit suggests that they are similar, (B:C = 1:1.76,

SI Appendix, Fig. S4D), explaining the excessive broadening of some of the *apo* state peaks.

It is worth emphasizing that the CEST experiments recorded with octapeptide allow the observation of resonances from the *apo* states (B and C) that could not be directly detected in the absence of the octapeptide. For example, of the residues illustrated in Fig. 2C only L97 could be observed in conventional ^1H - ^{15}N HSQC spectra of *apo* C34 (^{15}N CEST profile of this residue recorded on a sample without the octapeptide is shown in black). For this residue and others that could be observed in spectra recorded without ligand it must be that $\Delta\omega_{\text{BC}} \sim 0$, since otherwise extensive broadening would occur and the corresponding peak would not be observed in the *apo* C34 ^1H - ^{15}N HSQC dataset. Thus, only single dips were present in CEST profiles of C34 without added ligand for the peaks present in spectra (see also G105 of SI Appendix, Fig. S4A). As expected, these dips overlapped with those from states B/C in corresponding CEST profiles recorded with octapeptide (Fig. 2C, L97 and SI Appendix, Fig. S4A, G105).

Characterizing the Underlying Structural Dynamics in the *apo* State. Analysis of the CEST data for the 1:1 C34 WT–octapeptide mixture clearly establishes that the C terminus of the protein interconverts between at least two distinct conformations in the *apo* form. The goal, then, is to determine what these states might be. However, the relatively fast interconversion between them ($\sim 400 \text{ s}^{-1}$) complicates assignment of a given “*apo*” dip unequivocally to either state B or state C, and thus it is difficult to obtain the correct set of chemical shifts for each of the conformers. Further, similar fits are obtained for any three-state model (i.e., the triangular model shown in Fig. 2D or the linear model $\text{A} \rightleftharpoons \text{B} \rightleftharpoons \text{C}$, SI Appendix, Table S1). Nevertheless, very strong insight into the structural rearrangements that accompany the interconversion

between the two *apo* states can still be obtained, using the fitted chemical shift differences from analysis of the CEST data and the strategy described below. From the three-state fits of the 1:1 C34 WT:octapeptide CEST profiles, yielding chemical shift differences $\Delta\varpi_{AB}$ and $\Delta\varpi_{AC}$, a per-residue $\Delta\varpi_{\max\text{WT}} = \max(\Delta\varpi_{AB}, \Delta\varpi_{AC})$ value is calculated, where “max” refers to the maximum of the two shift differences without considering their signs [i.e., $\max(-7 \text{ ppm}, 2 \text{ ppm}) = -7 \text{ ppm}$]. The value $\varpi_{\max\text{BC}} = \Delta\varpi_{\max\text{WT}} + \varpi_A$ is then calculated, where $\varpi_{\max\text{BC}} \in \{\varpi_B, \varpi_C\}$; $\varpi_{\max\text{BC}}$ is thus the chemical shift of the nucleus in the free state that is maximally displaced from the corresponding shift in the octapeptide bound conformer. Fig. 3*A* plots ($^1\text{H}^N$, ^{15}N , $^{13}\text{C}^\alpha$, ^{13}CO) $\Delta\varpi_{\max\text{WT}}$ values of the WT *apo* state (first four rows) obtained directly from fits of CEST datasets. In this figure we include only $\Delta\varpi_{\max\text{WT}}$ values above a critical threshold for each nucleus (see figure legend), but it is important to emphasize that $\Delta\varpi_{\max\text{WT}}$ values are obtained for almost all of the residues in C34 (SI Appendix, Fig. S5). The largest $|\Delta\varpi_{\max\text{WT}}|$ values are observed for the β regions in the protein, as expected since dissociation of the octapeptide from the binding cleft located between β_2 and β_3 (Fig. 1) would produce large changes in the chemical shifts of nuclei in these two strands. Interestingly, negative $^{13}\text{C}^\alpha$ and ^{13}CO $\Delta\varpi_{\max\text{WT}}$ values and positive $^1\text{H}^N$ $\Delta\varpi_{\max\text{WT}}$ were measured for the β_4 region (Fig. 3*A* top four rows) suggesting that this stretch, which is unstructured in the presence of the octapeptide (Fig. 2*A*), assumes a β -structure in at least one of the two free states (Fig. 3*A* and *B*, rows corresponding to WT, and SI Appendix, Fig. S6; recall that $\Delta\varpi = \varpi_{\text{apo}} - \varpi_{\text{peptide-bound}}$). Fig. 3*B* shows “secondary structure propensities” of WT *apo* C34 calculated from $\varpi_{\max\text{BC}}$ values, supplemented by chemical shifts obtained by assignments from triple resonance datasets where available, as determined by the program SSP (26). As SSP values were calculated from $\varpi_{\max\text{BC}}$, (i.e., either ϖ_B or ϖ_C , depending on which shift is maximally distant from ϖ_A) it is important to note that the secondary structure determined is for a “hybrid” *apo* state that is maximally distinct from the peptide bound conformation.

The utility of $\Delta\varpi_{\max\text{WT}}$ values for providing structural insights into states B and C becomes evident when comparing these shift differences to those measured for mutants of C34 that were designed in an attempt to stabilize each of the two alternative *apo* states. As the β_4 region appears to be structured in at least one of the two free states (Fig. 3*A* and *B*) and unstructured in the bound conformation (Fig. 2*A*), we hypothesized that a mutant prepared by deletion of this region (referred to as C34 $\Delta\beta_4$, residues 1 to 115) would still bind the octapeptide, but that one of the *apo* conformations might be eliminated. Additional mutants were prepared in an attempt to generate a peptide-bound mimic in which β_4 would assume the role of the octapeptide. A sequence alignment of strand β_4 with the octapeptide shows that there is some similarity between the two ($\beta_4 = {}^{118}\text{EERYRVE}^{124}$ vs. the octapeptide = DVRFQVRE), suggesting that in the *apo* form of C34 the β_4 strand could fold into the binding cleft to form a second conformation of the protein. With this in mind, we generated two mutants that resembled the octapeptide more closely, one containing the point mutation E119V (new $\beta_4 = {}^{118}\text{EVRVYRVE}^{124}$, referred to as C34 E119V) and a second containing an insertion at position 124, whereby E124 is replaced by an RE sequence, referred to as C34 E124RE (new $\beta_4 = {}^{118}\text{EERYRVRE}^{125}$).

The NMR spectra of all three mutants were of high quality, with narrow resonances, even in the absence of the octapeptide (SI Appendix, Fig. S1), indicating that the underlying exchange process had been quenched. Complete resonance assignments of all mutants could be achieved via analysis of standard

triple-resonance experiments (28), from which the secondary structures of the *apo* proteins were obtained (Fig. 3*B* and SI Appendix, Fig. S6). These were similar for both C34 E119V and C34 E124RE, with β_4 formed in the *apo* states of both mutants. It was also possible to assign the octapeptide bound states of C34 $\Delta\beta_4$ and C34 E124RE, showing that, as with peptide bound C34 WT, all α helices and β strands were formed, with the exception of the region corresponding to the predicted β_4 , that was unstructured (18) (SI Appendix, Fig. S7). Notably, the chemical shifts of peptide-bound C34 $\Delta\beta_4$ and peptide-bound C34 E124RE were very similar to those of the peptide-bound C34 WT (SI Appendix, Fig. S7), as expected, since the region corresponding to β_4 in the *apo* state is unstructured in the ligated state (Fig. 2*A*). As will be discussed in the subsequent section the octapeptide binds only very weakly to the E119V mutant of C34 so that it is not possible to assign the peptide-bound state.

What is the relation between C34 $\Delta\beta_4$, C34 E119V, C34 E124RE and the exchanging states in C34 WT? To address this, we calculated $\Delta\varpi = \varpi_{\text{apo}} - \varpi_{\text{peptide-bound}}$ values for both C34 $\Delta\beta_4$ ($\Delta\varpi_{\Delta\beta_4}$) and C34 E124RE ($\Delta\varpi_{\text{E124RE}}$) from the measured chemical shifts in the octapeptide bound and free states and compared these with $\Delta\varpi_{\max\text{WT}}$ (Fig. 3*A* and *C*, Left). Notably, $\Delta\varpi_{\max\text{MUT}} = \max\{\Delta\varpi_{\text{E124RE}}, \Delta\varpi_{\Delta\beta_4}\}$ values are very well correlated with $\Delta\varpi_{\max\text{WT}}$ over the complete sequence (Fig. 3*A* and *C*, Left; note that frequency differences in Hz ($\Delta\nu$) as opposed to ppm ($\Delta\varpi$) are plotted in Figure 3*c*). This strong correlation is due to the fact that in the interval including residues S1 - G84 the agreement between $\Delta\varpi_{\Delta\beta_4}$ and $\Delta\varpi_{\max\text{WT}}$ values is very good, significantly better than for $\Delta\varpi_{\text{E124RE}}$ so that $\Delta\varpi_{\max\text{WT}} \sim \Delta\varpi_{\max\text{MUT}} \sim \Delta\varpi_{\Delta\beta_4}$ in this region. In contrast, the agreement between $\Delta\varpi_{\text{E124RE}}$ and $\Delta\varpi_{\max\text{WT}}$ for residues from the C-terminal half of β_3 to the end of C34 (residues V110-E124), that includes β_4 , is high, so that $\Delta\varpi_{\max\text{WT}} \sim \Delta\varpi_{\max\text{MUT}} \sim \Delta\varpi_{\text{E124RE}}$ in this region (Fig. 3*C*, Right). Given that the conformations of the peptide bound forms of C34 $\Delta\beta_4$ and C34 E124RE are very similar, both to each other, and to peptide bound C34 WT (SI Appendix, Fig. S7), it follows that the exchange model of Fig. 2*D*, where one of the states (B or C) is closely aligned with C34 $\Delta\beta_4$ and the other with C34 E124RE, is fully consistent with the data (Fig. 4*A*).

Having established that the *apo* forms of C34 $\Delta\beta_4$ and C34 E124RE are excellent models for the exchanging states comprising *apo* C34 WT, we next determined their topologies by analyzing the 3D ^{15}N -edited NOESY spectra of these mutants along with the NOESY spectrum of C34 E119V. The structure of C34 $\Delta\beta_4$ shows a closed peptide binding cleft due to the formation of a three-stranded β sheet comprising β_1 , β_2 , and β_3 , with strands β_2 and β_3 paired in a parallel fashion, and strands β_1 and β_2 oriented in an antiparallel manner with respect to each other. In contrast, in both C34 E119V and C34 E124RE, β_4 mimics the octapeptide, by folding into the binding cleft as an antiparallel β -strand between β_2 and β_3 . The strand orientations and NOE contacts between them are shown in Fig. 4*B* for both C34 $\Delta\beta_4$ and C34 E124RE, along with strips from NOESY spectra and NOE contacts for C34 $\Delta\beta_4$, C34 E124RE, and C34 E119V in SI Appendix, Figs. S8–S10.

Fig. 4*C* presents a model of the conformational dynamics of C34 WT that is consistent with the mutagenesis, chemical shift, and NOE data described above. A pair of *apo* states with closed binding pockets, formed either by the arrangement of parallel β_2 - β_3 strands (FC1 in Fig. 4*C* and analogous to the structure of C34 $\Delta\beta_4$), or by the folding of β_4 into the peptide binding pocket (FC2, and analogous to the structures of C34 E124RE and C34 E119V), and corresponding to states B and C of Fig. 2*D*, interconvert rapidly. A third *apo* state with an open binding pocket

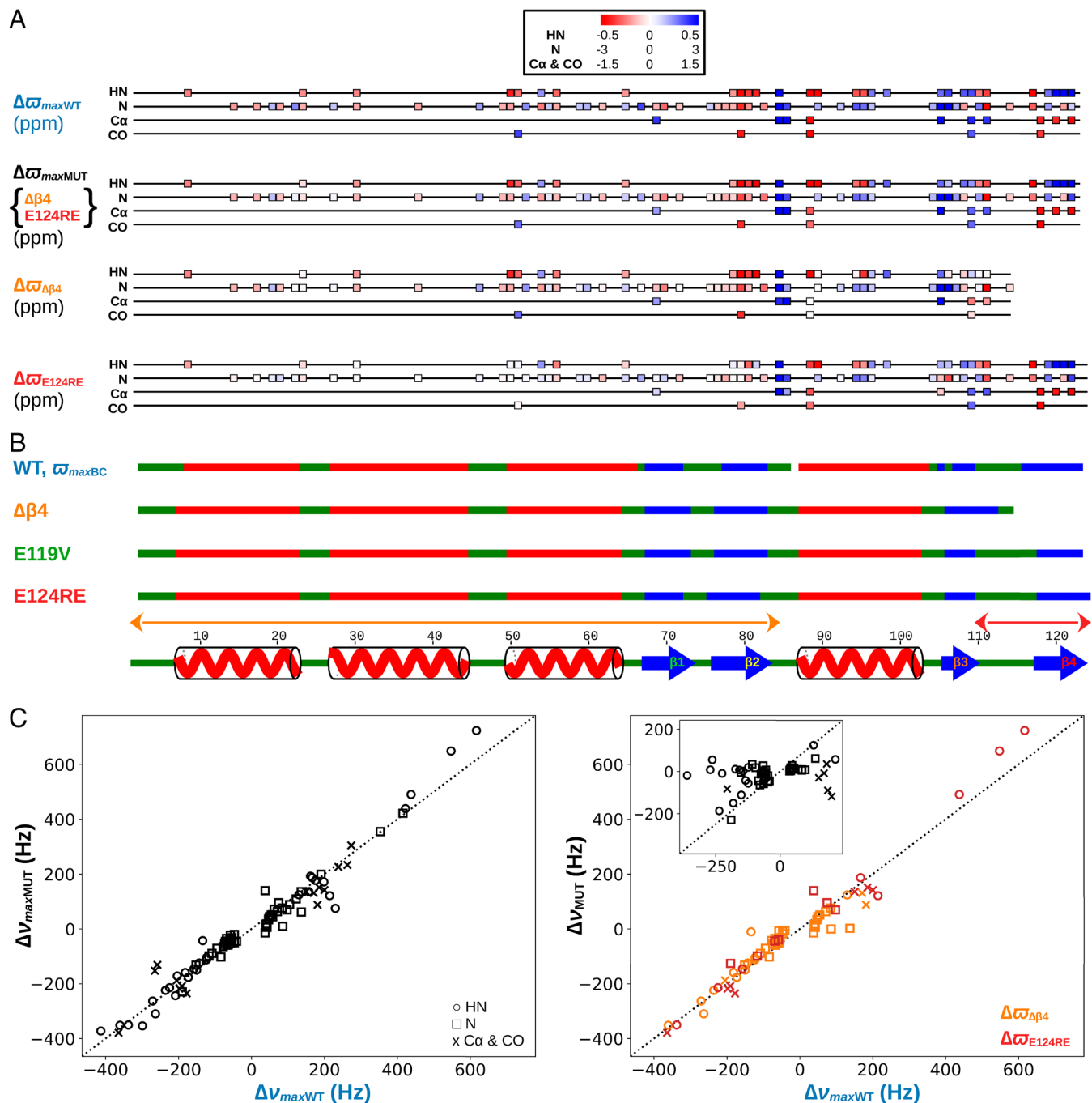


Fig. 3. Mutations in strand $\beta 4$ allow the separate observation of each of the interconverting *apo* states. (A) Differences in chemical shifts between *apo* and bound states of C34 WT, $\Delta\omega_{\max WT} = \max(\Delta\omega_{\Delta\beta 4}, \Delta\omega_{\Delta C})$, based on analysis of ^1H , ^{15}N , $^{13}\text{C}\alpha$ and ^{13}CO CEST profiles (top set of four rows), as described in the text. $\Delta\omega_{\max MUT} = \max\{\Delta\omega_{\text{E124RE}}, \Delta\omega_{\Delta\beta 4}\}$ values (second set of 4 rows) calculated from $\Delta\omega = \omega_{\text{apo}} - \omega_{\text{peptide-bound}}$ for C34 $\Delta\beta 4$ ($\Delta\omega_{\Delta\beta 4}$), and C34 E124RE ($\Delta\omega_{\text{E124RE}}$) (bottom two series of rows). The chemical shift differences for C34 $\Delta\beta 4$ and C34 E124RE were obtained directly from analysis of triple resonance datasets. Only a subset of nuclei is shown, for which $|\Delta\omega_{\max WT}|$ values are larger than 0.2 (^1H), 0.6 (^{15}N), 1.2 ($^{13}\text{C}\alpha$), and 1.0 (^{13}CO) ppm to ensure that major and minor dips in CEST profiles are not extensively overlapped, and thus increase the accuracy of extracted chemical shift differences (27) from which $\Delta\omega_{\max WT}$ is calculated. A much more extensive set of $\Delta\omega$ values is available, however, (SI Appendix, Fig. S5), from which the secondary structure of C34 WT can be calculated; secondary structures of mutant variants of C34 are obtained directly from measured chemical shifts, as high-quality spectra can be recorded. The $\Delta\omega$ values are color coded as indicated in the *Inset* at the top of the figure (ppm). (B) Experimentally determined secondary structure propensities (26) for *apo* states of C34 WT, C34 $\Delta\beta 4$, C34 E119V, and C34 E124RE, along with a cartoon representation of the experimentally derived secondary structure of the *apo* state of E124RE from NMR chemical shifts, where green, red, and blue lines denote random coil, helix, and strand, respectively. A residue is considered to be in an α -helix if $\text{SSP} > 0.5$ and β -strand if $\text{SSP} < -0.5$. The orange and red lines above the cartoon extend between S1-G84 and V110-E124, respectively, where $\Delta\omega_{\max MUT} \sim \Delta\omega_{\Delta\beta 4}$ (orange) and $\Delta\omega_{\max MUT} \sim \Delta\omega_{\text{E124RE}}$ (red). (C) Linear correlation plots of $\Delta v_{\max WT}$ and $\Delta v_{\max MUT}$ (Left), and $\Delta v_{\max WT}$ and $\Delta v_{\max MUT} \in \{\Delta v_{\text{E124RE}}, \Delta v_{\Delta\beta 4}\}$ (Right). The Δv frequencies are shown as their values in Hz at a magnetic field of 14.1 T (for example, $\Delta v = 600 \cdot \Delta\omega$ for ^1H , where $\Delta\omega$ is in ppm). The *Inset* shows that poor correlations are obtained between Δv_{E124RE} and $\Delta v_{\max WT}$ over the range from S1-G84 and between $\Delta v_{\Delta\beta 4}$ and $\Delta v_{\max WT}$ for the C-terminal region of the sequence.

(FO) must also exist to allow the binding of the octapeptide to C34 WT, but we have not been able to observe this conformer in any of our experiments, likely reflecting its low population. Our attempts to stabilize FO by mutagenesis, whereby a valine residue

at the terminus of $\beta 3$ (V106) was replaced by a bulkier leucine residue in the $\Delta\beta 4$ background so as to create steric clashes between position 106 and the backbone atoms of R50 and A51 of $\alpha 3$, thus preventing insertion of $\beta 3$ into the peptide binding

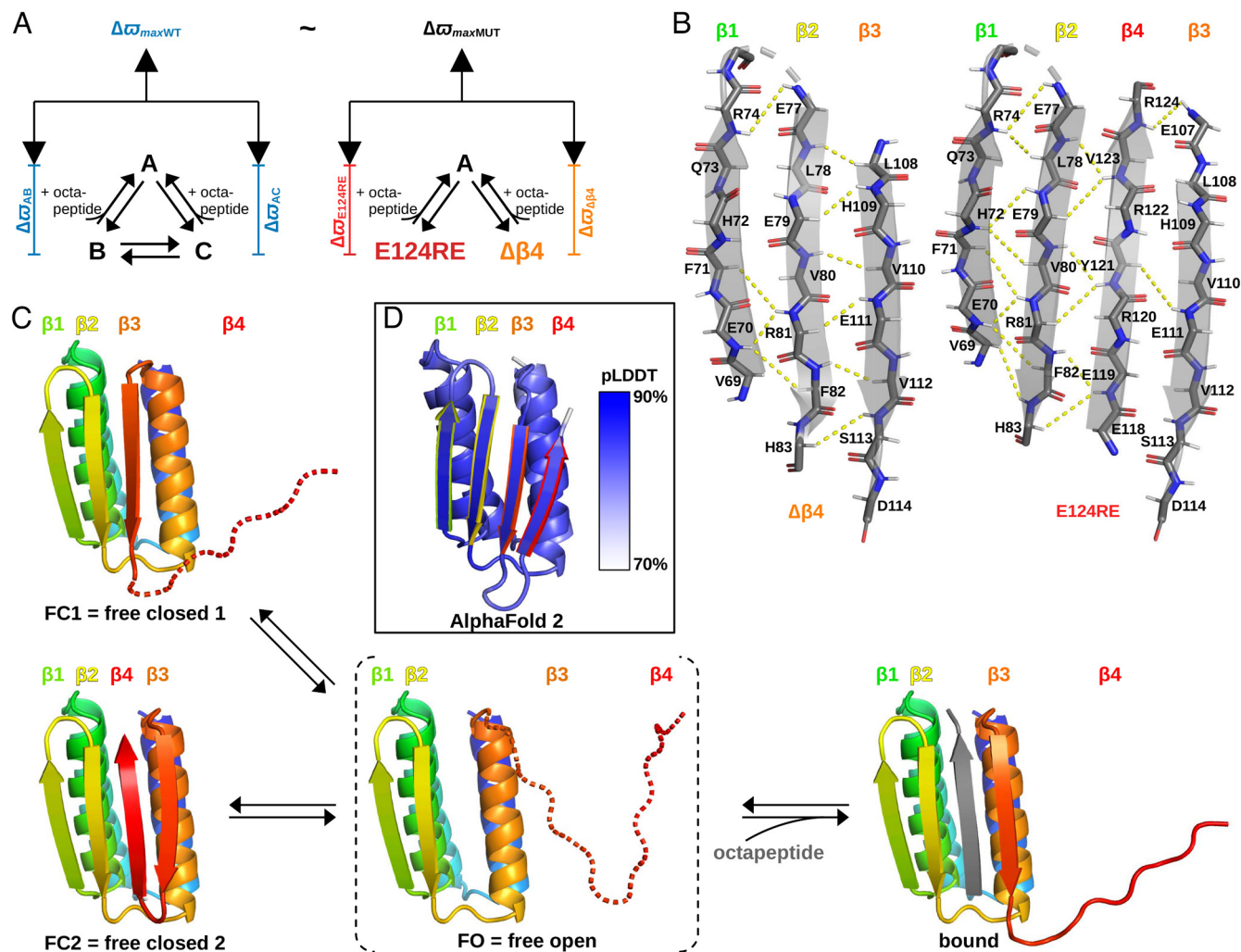


Fig. 4. Peptide binding model for C34 WT. (A) Comparison of $\Delta\omega_{maxWT}$ and $\Delta\omega_{maxMUT}$ values allows assignment of WT *apo* states (B and C of Fig. 2D) to structures resembling C34 E124RE and C34 $\Delta\beta 4$, as described in the text. (B) Schematic representation of β strands of C34 $\Delta\beta 4$ and C34 E124RE showing $^1H^N$ - $^1H^N$ and $^1H^N$ - $^1H^N$ NOE connectivities with dashed lines. (C) Structural models for the two closed states of WT *apo* C34, FC1 and FC2. These conformers interconvert rapidly (~ 400 s $^{-1}$), leading to pervasive broadening of *apo* C34 spectra. Binding of ligand (gray β -strand) proceeds via formation of an open C34 conformation (shown in dashed parenthesis, as we were not able to observe it experimentally). The dashed C-terminal region for FC1 indicates that it is not possible to establish whether $\beta 4$ is formed in this state from our data, and structures of regions corresponding to $\beta 3$ and $\beta 4$ in FC2 are not known for the FO state. The structure of the bound state is the one predicted by Rosetta (18), modified by our experimental analysis showing that $\beta 4$ is not formed, and verified by analysis of NOEs. (D) AlphaFold2 (4) model of WT *apo* C34, along with per-residue confidence scores (pLDDT – predicted Local Distance Difference Test). The scores range from 0 to 100% (100% = highest confidence) and are color coded on the structure in shades of blue as indicated on the bar to the right.

cleft (i.e., preventing formation of FC1), failed, as did increasing the temperature to 50 °C.

An AlphaFold2-based structure of WT *apo* C34 is shown in Fig. 4D, along with the confidence of the model color coded on the structure. Notably, the locations of all four helices and β -strands are established with a high degree of confidence and the structure closely resembles that of FC1. There is no indication, based on the model, of alternate conformations such as FC2, and hence the pervasive dynamics present in this system is not captured in the artificial intelligence-based structure, highlighting the importance of experimental data in exploring the energy landscapes of dynamical systems.

C34 Conformational Heterogeneity Modulates Ligand Binding.

Fig. 5A highlights spectral regions from $^1H^N$ - ^{15}N HSQC datasets recorded at 25 °C, either without (single black contours) or with added octapeptide at a 1:10 C34–octapeptide ratio (45 μ M C34: 450 μ M peptide), with full spectra shown in SI Appendix, Fig. S3. While the effect of octapeptide addition is evident on the spectra of C34 WT, C34 $\Delta\beta 4$ and C34 E124RE (see also SI Appendix,

Fig. S11), it has little effect on the spectrum of C34 E119V. From inspection of peak intensities, it is clear that addition of 10-fold excess peptide results in a fully bound C34 $\Delta\beta 4$ conformation, a mostly bound C34 WT state, and approximately equal amounts of bound and unbound C34 E124RE, while there is very little binding to C34 E119V. An in-depth analysis based on quantifying peak volume ratios of corresponding peaks reporting on bound and unbound states of C34 (Fig. 5B and C and Materials and Methods) yields well defined dissociation constants K_D (SI Appendix, Fig. S11) that increase as C34 $\Delta\beta 4$ (highest affinity, ~ 55 μ M) < C34 WT (~ 200 μ M) < C34 E124RE (~ 400 μ M) < C34 E119V ($\sim 8,000$ μ M) (Fig. 5D and SI Appendix, Tables S2–S6).

A simple calculation (SI Appendix) leads to a predicted K_D value of approximately 90 μ M for C34 $\Delta\beta 4$, similar to the experimentally determined value of 60 μ M. Note that the FC2 state of Fig. 4B is not populated in the case of C34 $\Delta\beta 4$, accounting for a great part of the observed affinity increase for the peptide (i.e., strand $\beta 4$ binding can no longer compete with octapeptide binding). In contrast, the C34 E119V and C34 E124RE mutations stabilize the FC2 state, leading to marginally weaker octapeptide

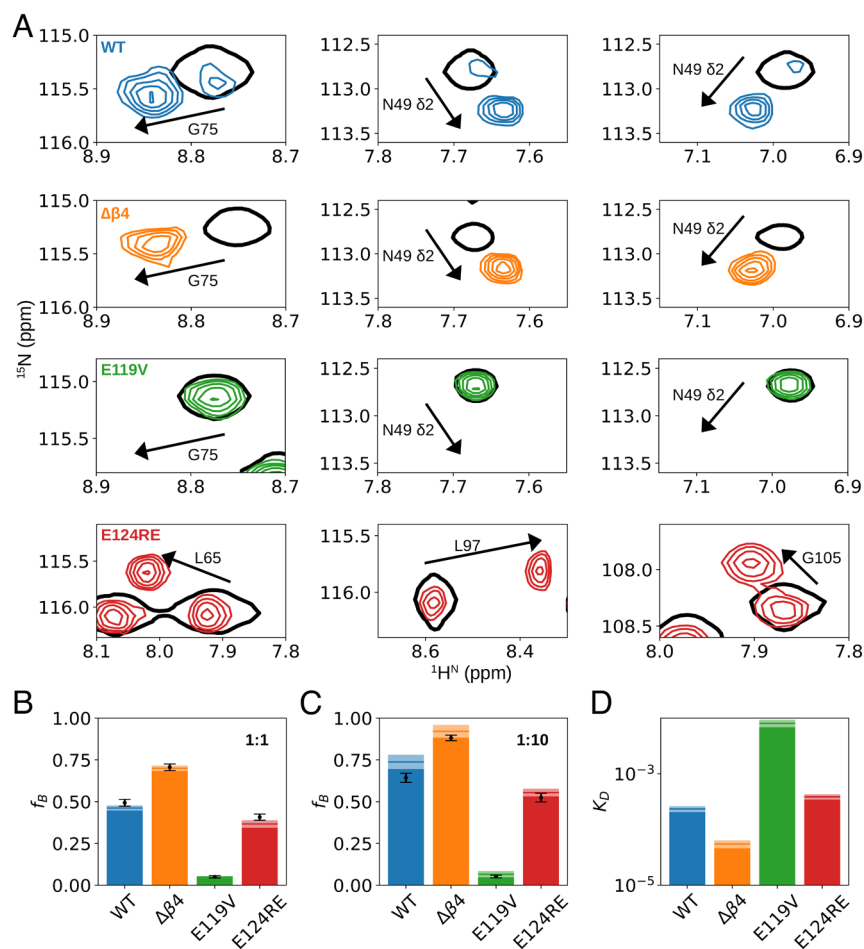


Fig. 5. Cognate ligand binding to C34 variants. (A) ^1H - ^{15}N HSQC spectral regions from datasets of C34 WT (blue), C34 $\Delta\beta 4$ (orange), C34 E119V (green) and C34 E124RE (red) recorded at 25 °C, either without (single black contours) or with added octapeptide at a 1:10 C34–octapeptide ratio. A separate region for the E124RE spectrum has been shown for comparison (SI Appendix, Fig. S12). (B) Experimental (bars) and best fitted (black dots) values of peptide bound C34 protein fractions at [C34] = 450 μM , [octapeptide] = 450 μM . (C) Bound fractions at [C34] = 45 μM , [octapeptide] = 450 μM . (D) Fitted values of K_D . The light shaded areas at the top of the bars in B–D indicate the error (± 1 SD). Experimental and fitted values of the bound fractions are also shown in SI Appendix, Tables S3–S6.

binding for C34 E124RE ($K_D \sim 400$ μM) or dramatically impaired binding for C34 E119V (~ 10 mM).

Regulating Ligand Binding by Allostery. We have exploited the C34 conformational dynamics (Fig. 4) to design a new protein variant whereby the affinity of the octapeptide–C34 interaction is allosterically modulated through the binding of a second ligand. This was achieved by taking advantage of the interaction between the SH3 domain of yeast actin binding protein (Abp1p) and its cognate peptide-binding partner from the yeast actin patch kinase, Ark1p [referred to as Ark1p peptide (33–35); $K_D \sim 1$ μM (33, 36)]. The Ark1p peptide was introduced into the loop between $\beta 3$ and $\beta 4$ of the C34 E119V mutant, along with glycine–serine linkers on either side of the addition (Fig. 6), generating a C34 E119V $\beta 3,4$ -Ark1p construct (Fig. 6A). We have chosen the E119V mutant background as the E to V substitution stabilizes the closed C34 state, and, of the mutants considered (see above), it is the strongest at inhibiting octapeptide binding. By designing the length of the loop such that the Ark1p peptide folds back on itself when $\beta 4$ (in red in Fig. 6A) binds in the groove between $\beta 2$ and $\beta 3$ (FC2 conformation of Fig. 4C), the Ark1p peptide would be inaccessible to the added Abp1p SH3 domain unless $\beta 4$ is released. The hope, then, is that formation of the C34–octapeptide complex would become controlled through binding of the SH3 domain. We have assigned the spectrum of *apo* C34 E119V $\beta 3,4$ -Ark1p and calculated its secondary structure, which parallels that of *apo* C34

E119V in the absence of the Ark1p peptide (Fig. 6B). Additionally, ^1H - ^1H NOE contacts establish the β -sheet topology of *apo* C34 E119V $\beta 3,4$ -Ark1p (Fig. 6C and SI Appendix, Fig. S13), showing that the FC2 structure is the one populated in the *apo* state (in the absence of SH3), and confirming the design strategy.

Fig. 6D highlights regions of ^1H - ^{15}N HSQC spectra of C34 E119V $\beta 3,4$ -Ark1p recorded either without (single black contours) or with peptide in a 1:10 C34 E119V $\beta 3,4$ -Ark1p:octapeptide ratio (multiple contours), and without (Fig. 6D, Top) or with (Fig. 6D, Bottom) added Abp1p SH3 domain. As expected, addition of the SH3 domain increases the affinity toward octapeptide, as the “*apo*” peaks disappear and are replaced by correlations reporting on the bound state. A more quantitative analysis shows that for a 1:1.3 C34 E119V $\beta 3,4$ -Ark1p:Abp1p SH3 domain ratio, the K_D for dissociation of the octapeptide is ~ 100 μM , a decrease of approximately fivefold relative to the case where the Abp1p SH3 domain is absent (Fig. 6E–G and SI Appendix, Tables S7 and S8).

Discussion

Structural biology is often focused on the generation of a single structure by which to explain mechanism or, in the case of design, the construction of a sequence that can fold into a single conformation that is efficient for carrying out a particular function. However, protein function is based on sampling of multiple

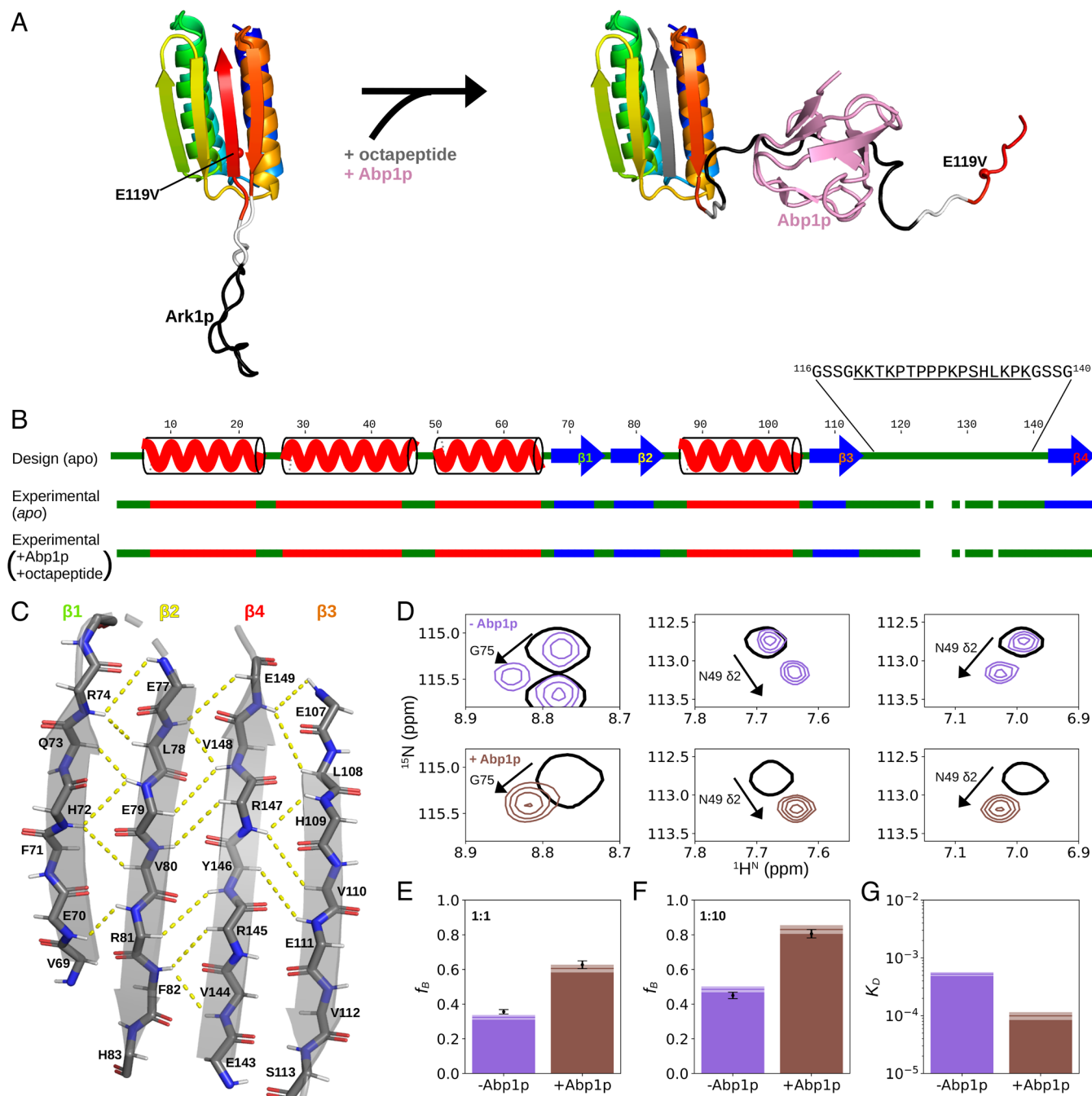


Fig. 6. Regulating peptide binding through allosteric modulation. (A) Schematic illustrating the design strategy whereby addition of a ligand (Abp1p SH3 domain) to a C34 E119V β3,4-Ark1p construct containing an Ark1p peptide (black) regulates binding of the cognate octapeptide ligand. The position of the E119V mutation is shown by a sphere. The Ark1p peptide is engineered into the C34 construct with GSSG spacers (white), as indicated. (B) The predicted design secondary structure of apo C34 E119V β3,4-Ark1p (top row), along with experimentally derived secondary structures for the apo and Abp1p SH3+octapeptide bound C34 states. The sequences of the added Ark1p peptide (underlined) and linkers are shown. (C) Schematic of β-sheet architecture of apo C34 E119V β3,4-Ark1p showing $^1\text{H}^\alpha\text{-}^1\text{H}^\text{N}$ and $^1\text{H}^\text{N}\text{-}^1\text{H}^\text{N}$ NOE connectivities (dashed lines). (D) Regions of $^1\text{H}^\text{N}\text{-}^{15}\text{N}$ HSQC spectra recorded either without octapeptide (single black contours) or with a 1:10 C34 E119V β3,4-Ark1p:octapeptide ratio (multiple contours) and without (Top) or with (Bottom) added Abp1p SH3 domain. Spectra were acquired with 45 μM C34 E119V β3,4-Ark1p and 0 μM (Top) or 60 μM Abp1p SH3 domain (Bottom). (E) Experimental (bars) and best fitted (black dots) values of C34 bound fractions at [C34 E119V β3,4-Ark1p] = 450 μM, [octapeptide] = 450 μM, without or with [Abp1p] = 600 μM. (F) Bound fractions at [C34 E119V β3,4-Ark1p] = 45 μM, [octapeptide] = 450 μM, without or with [Abp1p] = 60 μM. (G) Fitted K_D values. The light shaded areas at the top of the bars in E–G indicate the errors (± 1 SD) in the fits. Experimental and fitted values of bound fractions are also shown in [SI Appendix, Tables S7 and S8](#).

low-lying positions on an energy landscape, rather than on a single conformation that represents the lowest energy state of the molecule (37–40). There is thus a strong need to establish the functionally relevant motions in a system of interest and, beyond static structures, to generate a sufficient library of functional dynamic trajectories, so as to efficiently incorporate the effects of dynamics into a design strategy.

The recently designed α/β scaffold proteins that are optimized for binding of β -strands and β -hairpins (18) present an interesting case in point. Here we have examined one such protein, C34, that binds a cognate β -strand partner in a deep groove between a pair of β -strands. While the structure of the bound form of C34 follows expectations based on the design strategy, with the exception that a predicted strand at the C terminus of C34 is not formed, the

apo form of the protein is highly dynamic, in particular, in regions proximal to or including the β -strands that are important functionally (i.e., for peptide binding). These dynamics are in contrast with the predictions from deep-learning protein structure prediction algorithms that predict an ordered structure for *apo* C34 with high confidence (Fig. 4D). The pervasive broadening of these dynamic regions complicates a quantitative study of *apo* C34, that would be required to improve the design in an iterative manner, since many of the expected peaks are not observed in spectra. Here we developed a strategy based on CEST whereby the invisible spectral regions are “brought into focus” through the transfer of magnetization to the NMR-visible peptide-bound conformer, using a system with a 1:1 octapeptide–C34 ratio. Thus, it becomes possible to read out the chemical shifts of nuclei in the *apo* conformation through the bound state. The analysis of chemical shift differences between bound and free conformations for *apo* C34 WT and for a number of mutants that stabilize C34 into different conformational states establishes that *apo* C34 exchanges between a pair of closed conformers, with β 4 positioned in the binding groove in one of the conformers, effectively mimicking the β -strand C34 binding partner in the bound form of the protein. The conformational dynamics thus attenuate the binding of the ligand to C34. They also present an opportunity to engineer allostery into the system, whereby binding of a second target (Abp1p SH3 domain) to an engineered site (Ark1p peptide) in the molecule modulates the binding of cognate peptide.

Our study reveals an important role for low-lying conformational states in controlling function and the need, therefore, to account for multiple conformations in protein design strategies. In this context, minimum energy molecular dynamics trajectories on energy landscapes could provide critical insights into the conformers that must be considered. Experimentally such information could come from studies of single particles via electron cryo-microscopy or solutions of proteins using NMR spectroscopy. In the latter approach, such studies can be performed at physiological temperatures that would not attenuate potentially significant motions, and with sensitivity to the presence of relatively small percentages (~1%) of conformers that can, nonetheless, assume functional roles. The development of databases of low-energy protein conformations for a given system will provide important direction for protein design. In this regard, NMR spectroscopy will assume a significant role, both in elucidating conformers of significance, as illustrated here, and in testing computational approaches that will be developed to assist this process.

Materials and Methods

Protein Production. All C34 constructs are encoded in a pET29b vector and have a N-terminal His₆ tag followed by a TEV cleavage site. All proteins were expressed in *Escherichia coli* BL21(DE3) cells using M9 minimal media supplemented with ¹⁵NH₄Cl and [U-¹³C] D-glucose as the sole nitrogen and carbon sources, respectively. Cells were induced at OD ~ 0.7 by the addition of 0.25 mM IPTG and then incubated at room temperature for 16 h with constant shaking at 220 to 240 rpm. Protein purification was done at room temperature using 50 mM Tris buffers at pH 8. Cells were lysed by sonication on ice in the presence of 0.5 M NaCl and a trace of DNase I. Clarified cell extracts were loaded onto a NiNTA column, purified by a wash step with 30 mM imidazole, and then the His₆ labeled protein was eluted using 0.5 M imidazole. The His₆ site was then cleaved by TEV protease during a dialysis step to remove the imidazole, in the presence of 0.1 M NaCl and 1 mM EDTA. A second NiNTA step was done but keeping the flow-through this time, before size exclusion chromatography using a prepacked 16/600 Superdex 75 column in the presence of 0.1 M NaCl. Precipitation of the C34 proteins was observed upon storage in the fridge at 4 °C but solubility is quickly restored by incubation at 50 °C.

Octapeptide Synthesis. A C34 target peptide (DVRFOVRE) was produced as previously described (18) via solid-phase synthesis using Fmoc-protected amino acids and purified by RP-HPLC.

NMR Experiments. All NMR experiments were recorded using one of the following spectrometers: Bruker AVANCE III HD 14.1 T, Varian INOVA 14.1 T, or Bruker AVANCE III HD 18.8 T, all equipped with cryogenically cooled, triple-resonance probes. Measurements were performed at 25 °C, except for those related to resonance assignments of *apo* C34 WT for which a temperature of 50 °C was used in an attempt to minimize the negative effects of pervasive chemical exchange on spectral quality. Samples varied in protein concentration from 0.045 mM to 1 mM, dissolved in a buffer comprised of 50 mM Bis-Tris pH 6.4, 50 mM NaCl, 1.5 mM Na₂N₃, 1 mM EDTA, 2% D₂O. All datasets were processed with the *NMRPipe* software package (41).

Assignment of NMR Resonances and Calculation of SSP. Three-dimensional CBCA(CO)NH (42), HNCACB (43), HNCO (44), HN(CA)CO (45), and HNN (46) spectra were recorded with uniformly [¹³C,¹⁵N]-labeled samples (~0.2 to 1 mM in protein, depending on the variant of C34 assigned) and analyzed using the program CARRA (<http://cara.nmr.ch/doku.php>). ¹³C α , ¹³C β , ¹³CO, ¹⁵N, and ¹H^N resonances were referenced to the DSS ¹H signal using the indirect referencing method (47) and then analyzed with SSP software (26) (<http://abragam.med.utoronto.ca/~JFKlab/SSP>).

NOESY Experiments. Three-dimensional gradient-enhanced (48) [¹H,¹⁵N,¹H] NOESY-HSQC (28) experiments were recorded with a mixing time of 150 ms. The ¹H α resonances required for the identification of ¹H α -¹H^N NOE contacts were assigned using the HBHA(CO)NH experiment (28, 49).

CPMG Experiments. ¹⁵N CPMG experiments (16 h for samples at either 0.5 mM or 1 mM) were recorded using a pulse scheme in which ¹⁵N magnetization is maintained in-phase during the CPMG interval (24) using a 16.7 kHz ¹H continuous-wave decoupling field, with 0013 phase-cycling of the refocusing pulses comprising the CPMG train (25, 50), and a constant-time relaxation interval, *T*, of 30 ms (51). The variation of the effective transverse ¹⁵N relaxation rate, *R*_{2,eff}, as a function of the frequency of application of chemical shift refocusing pulses *v*_{CPMG}, with *v*_{CPMG} given by 1/(2 δ) where δ is the delay between successive refocusing pulses, is plotted in Fig. 2B and SI Appendix, Fig. S2. A total of 22 values of *v*_{CPMG} were sampled between 33 and 1,000 Hz, with two repeats for error estimation (52). An additional reference plane without the constant-time relaxation interval was also recorded to obtain the reference intensity values (*I*₀) required for the calculation of *R*_{2,eff} = 1/*T* · ln(*I*₀/*I*).

CEST Experiments. ¹H^N, ¹⁵N, ¹³C α , and ¹³CO CEST experiments were recorded on a sample comprised of a 1:1 C34 WT–octapeptide ratio (concentration of 450 μ M of each), with approximately 50% of C34 in the bound form (Fig. 5B). Additionally, ¹⁵N CEST datasets were acquired for *apo* C34 WT. The ¹H^N CEST used cosine-modulated waveforms (53, 54) to selectively excite regions of the spectrum, along with suppression of NOE effects during the CEST period by selection of two-spin longitudinal order at the termination of the CEST element (30). Two ¹H^N experiments were acquired. In the first, the effective CEST field was 5 Hz, with the excitation field comprised of five cosine terms [*N* = 5 corresponding to five frequencies; see equation 1 of Yuwen et al. (53)] and the radio frequency sampled over a width of *SW* = 500 Hz (51 planes; 44 h). In the second experiment, a field of 10 Hz was used, with *N* = 3 and *SW* = 800 Hz (81 planes; 75 h). The duration of the CEST element *T*_{EX} was 0.5 s in both experiments. ¹⁵N experiments (27) were recorded using a weak effective *B*₁ field generated by DANTE pulses (55), with a pair of experiments acquired corresponding to effective *B*₁ fields of 5 Hz (10 Hz), *SW* = 500 Hz (600 Hz) and *T*_{EX} = 1 s (0.5 s). The number of planes acquired per data set was 51 (*B*₁ = 5 Hz) and 61 (*B*₁ = 10 Hz) for total experimental times of ~35 h/dataset. For the ¹³C α (110 planes, 140 h) and ¹³CO (104 planes; 150 h) CEST experiments (31, 32) a continuous weak *B*₁ field of 30 Hz was applied during the CEST period (*T*_{EX} = 0.5 s) and full ¹³C α or ¹³CO spectral ranges were sampled. For each CEST experiment a reference plane was acquired without the relaxation element (27). A non-TROSY enhanced sensitivity quadrature detection scheme (48) was used for all CEST experiments.

Analysis of CPMG and CEST Data. CPMG and CEST data were fitted by numerical integration of the Bloch-McConnell equation (56) via minimization of a χ^2 target function using the program ChemEx (57) (<https://github.com/gbouvnignis/>).

ChemEx). Data for C34 WT in the absence (CEST and CPMG) and in the presence (CEST) of octapeptide were analyzed together using the three-state model of Fig. 2D, by constraining $[A] = 0$ for the data recorded on the *apo* sample.

Analysis of C34–octapeptide Titration Data. Affinities of peptide for the variants of C34 have been calculated based on ratios of *apo* and bound signal intensities as quantified from ^1H - ^{15}N HSQC spectra. However, because of millisecond-timescale dynamics in *apo* C34 the ratio of peak intensities may not accurately represent the fraction of peptide-free and bound forms of C34. For this reason, we have only selected those peaks for analysis where the ratio measured in a standard enhanced sensitivity HSQC spectrum (48) was replicated to within error in ^1H and ^{15}N CPMG (24, 58) planes with relaxation intervals of 10 and 30 ms, respectively, and with only a single chemical shift refocusing pulse applied at the center of the delays. Not surprisingly, bound/free intensity ratios are generally larger when calculated from intensities in the CPMG planes because of additional exchange broadening of resonances originating from *apo* C34 during the CPMG relaxation periods. A number of peaks with consistent ratios in all experiments could be selected, however (shown in *SI Appendix, Tables S3–S8*).

Molecular Modeling. All structural models and representations were constructed with Pymol (The PyMOL Molecular Graphics System, Version 2.5.0 Schrödinger, LLC.) using the original C34 model (18) and the published Abp1p SH3 domain - Ark1p peptide (59) structure. Backbone dihedral angles were manually changed:

1) to depict the random coil nature of the region corresponding to $\beta 4$ in Fig. 1, or to $\beta 3 + \beta 4$ of FO in Fig. 4 or 2) to fold $\beta 4$ inside the β -strand of FC2 in Figs. 4 and 6A (left hand side).

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Alex Sever, University of Toronto, for the gift of the Abp1p SH3 domain and Mila Lamb, Paul Levine, and Xinting Li, University of Washington, for the octapeptide used in this work. This work was supported by grants from the Canadian Institutes of Health Research (FDN-503573, L.E.K.), the Natural Sciences and Engineering Research Council of Canada (FDN-455908, L.E.K.), a gift from Gates Ventures (D.D.S. and D.B.), the Audacious Project at the Institute for Protein Design (D.B.), a gift from Amgen (D.B.), and a grant from DARPA supporting the Harnessing Enzymatic Activity for Lifesaving Remedies Program (HR001120S0052 contract HR0011-21-2-0012, D.B.).

Author affiliations: ^aDepartment of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; ^bDepartment of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada; ^cDepartment of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada; ^dDepartment of Biochemistry, University of Washington, Seattle, WA 98195; ^eInstitute for Protein Design, University of Washington, Seattle, WA 98195; ^fHHMI, University of Washington, Seattle, WA 98195; and ^gProgram in Molecular Medicine, The Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada

1. P.-S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
2. D. N. Woolfson, A brief history of de novo protein design: Minimal, rational, and computational. *J. Mol. Biol.* **433**, 167160 (2021).
3. M. Defresne, S. Barbe, T. Schiex, Protein design with deep learning. *Int. J. Mol. Sci.* **22**, 11741 (2021).
4. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
5. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
6. D. Röthlisberger *et al.*, Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
7. J. B. Siegel *et al.*, Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309–313 (2010).
8. N. H. Joh *et al.*, De novo design of a transmembrane Zn²⁺-transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
9. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837–844 (2016).
10. D.-A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
11. N. Anand *et al.*, Protein sequence design with a learned potential. *Nat. Commun.* **13**, 746 (2022).
12. J. Dauparas *et al.*, Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
13. X. I. Ambroggio, B. Kuhlman, Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154–1161 (2006).
14. K. M. Hart, C. M. W. Ho, S. Dutta, M. L. Gross, G. R. Bowman, Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **7**, 12965 (2016).
15. 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).
16. K.-Y.M. Chen, D. Keri, P. Barth, Computational design of G Protein-Coupled Receptor allosteric signal transductions. *Nat. Chem. Biol.* **16**, 77–86 (2020).
17. J. Vucinic, D. Simoncini, M. Ruffini, S. Barbe, T. Schiex, Positive multistate protein design. *Bioinformatics* **36**, 122–130 (2020).
18. D. D. Sahtoe *et al.*, Design of amyloidogenic peptide traps. *bioRxiv* [Preprint] (2023). <https://doi.org/10.1101/2023.01.13.523785> (Accessed 17 March 2023).
19. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton, "Chapter 5–Relaxation and dynamic processes" in *Protein NMR Spectroscopy*, J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton, Eds. (Academic Press, ed. 2, 2007), pp. 333–404.
20. H. Arthanari, K. Takeuchi, A. Dubey, G. Wagner, Emerging solution NMR methods to illuminate the structural and dynamic properties of proteins. *Curr. Opin. Struct. Biol.* **58**, 294–304 (2019).
21. A. Bax, G. M. Clore, Protein NMR: Boundless opportunities. *J. Magn. Reson.* **306**, 187–191 (2019).
22. Y. Tokunaga, T. Viennet, H. Arthanari, K. Takeuchi, Spotlight on the ballet of proteins: The structural dynamic properties of proteins illuminated by solution NMR. *Int. J. Mol. Sci.* **21**, 1829 (2020).
23. T. R. Alderson, L. E. Kay, NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. *Cell* **184**, 577–595 (2021).
24. D. F. Hansen, P. Vallurupalli, L. E. Kay, An improved ^{15}N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J. Phys. Chem. B* **112**, 5898–5904 (2008).
25. B. Jiang, B. Yu, X. Zhang, M. Liu, D. Yang, A ^{15}N CPMG relaxation dispersion experiment more resistant to resonance offset and pulse imperfection. *J. Magn. Reson.* **257**, 1–7 (2015).
26. J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: Implications for fibrillation. *Protein Sci.* **15**, 2795–2804 (2006).
27. P. Vallurupalli, G. Bouvignies, L. E. Kay, Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).
28. M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
29. T. Yuwen, A. Sekhar, L. E. Kay, Separating dipolar and chemical exchange magnetization transfer processes in 1H-CEST. *Angew. Chem. Int. Ed.* **56**, 6122–6125 (2017).
30. T. Yuwen, L. E. Kay, A new class of CEST experiment based on selecting different magnetization components at the start and end of the CEST relaxation element: An application to 1H CEST. *J. Biomol. NMR* **70**, 93–102 (2018).
31. G. Bouvignies, P. Vallurupalli, L. E. Kay, Visualizing side chains of invisible protein conformers by solution NMR. *J. Mol. Biol.* **426**, 763–774 (2014).
32. P. Vallurupalli, L. E. Kay, Probing slow chemical exchange at carbonyl sites in proteins by chemical exchange saturation transfer NMR spectroscopy. *Angew. Chem. Int. Ed.* **52**, 4156–4159 (2013).
33. J. Haynes *et al.*, The biologically relevant targets and binding affinity requirements for the function of the yeast actin-binding protein 1 Src-homology 3 domain vary with genetic context. *Genetics* **176**, 193–208 (2007).
34. D. G. Drubin, J. Mulholland, Z. Zhu, D. Botstein, Homology of a yeast actin-binding protein to signal transduction proteins and myosin-I. *Nature* **343**, 288–290 (1990).
35. P. Vallurupalli, D. F. Hansen, L. E. Kay, Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11766–11771 (2008).
36. P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch, L. E. Kay, Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18473–18477 (2007).
37. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
38. K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
39. L.-Q. Yang *et al.*, Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* **32**, 372–393 (2014).
40. M. D. Miller, G. N. Phillips, Moving beyond static snapshots: Protein dynamics and the Protein Data Bank. *J. Biol. Chem.* **296**, 100749 (2021).
41. F. Delaglio *et al.*, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
42. S. Grzesiek, A. Bax, Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293 (1992).
43. M. Wittekind, L. Mueller, HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson. B* **101**, 201–205 (1993).
44. L. E. Kay, M. Ikura, R. Tschudin, A. Bax, Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **196**, 499–514 (1990).
45. R. T. Clubb, V. Thanabal, G. Wagner, A constant-time three-dimensional triple-resonance pulse scheme to correlate intrareidue ^1H , ^{15}N , and ^{13}C chemical shifts in ^{15}N - ^{13}C -labeled proteins. *J. Magn. Reson.* **196**, 213–217 (1992).
46. S. C. Panchal, N. S. Bhavesh, R. V. Hosur, Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ^{15}N sequential correlations in (^{13}C , ^{15}N) labeled proteins: Application to unfolded proteins. *J. Biomol. NMR* **20**, 135–147 (2001).
47. D. S. Wishart *et al.*, ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140 (1995).
48. L. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).
49. S. Grzesiek, A. Bax, Amino acid type determination in the sequential assignment procedure of uniformly $^{13}\text{C}/^{15}\text{N}$ -enriched proteins. *J. Biomol. NMR* **3**, 185–204 (1993).
50. G. N. B. Yip, E. R. P. Zuiderweg, A phase cycle scheme that significantly suppresses offset-dependent artifacts in the R2-CPMG ^{15}N relaxation experiment. *J. Magn. Reson.* **171**, 25–36 (2004).
51. F. A. Mulder, N. R. Skrynnikov, B. Hon, F. W. Dahlquist, L. E. Kay, Measurement of slow (μs – ms) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* **123**, 967–975 (2001).

52. D. M. Korzhnev *et al.*, Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590 (2004).
53. T. Yuwen, G. Bouvignies, L. E. Kay, Exploring methods to expedite the recording of CEST datasets using selective pulse excitation. *J. Magn. Reson.* **292**, 1–7 (2018).
54. M. Leninger, W. M. Marsiglia, A. Jerschow, N. J. Traaseth, Multiple frequency saturation pulses reduce CEST acquisition time for quantifying conformational exchange in biomolecules. *J. Biomol. NMR* **71**, 19–30 (2018).
55. T. Yuwen, L. E. Kay, G. Bouvignies, Dramatic decrease in CEST measurement times using multi-site excitation. *ChemPhysChem* **19**, 1707–1710 (2018).
56. H. M. McConnell, Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28**, 430–431 (1958).
57. G. Bouvignies, L. E. Kay, Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. *J. Phys. Chem. B* **116**, 14311–14317 (2012).
58. T. Yuwen, L. E. Kay, Revisiting 1HN CPMG relaxation dispersion experiments: A simple modification can eliminate large artifacts. *J. Biomol. NMR* **73**, 641–650 (2019).
59. E. J. Stollar *et al.*, Structural, functional, and bioinformatic studies demonstrate the crucial role of an extended peptide binding site for the SH3 domain of yeast Abp1p*. *J. Biol. Chem.* **284**, 26918–26927 (2009).