

De Novo Design of a β -Helix Tau Protein Scaffold: An Oligomer-Selective Vaccine Immunogen Candidate for Alzheimer's Disease

Adekunle Aina, Shawn C. C. Hsueh, Ebrima Gibbs, Xubiao Peng, Neil R. Cashman, and Steven S. Plotkin*



Cite This: *ACS Chem. Neurosci.* 2023, 14, 2603–2617



Read Online

ACCESS |

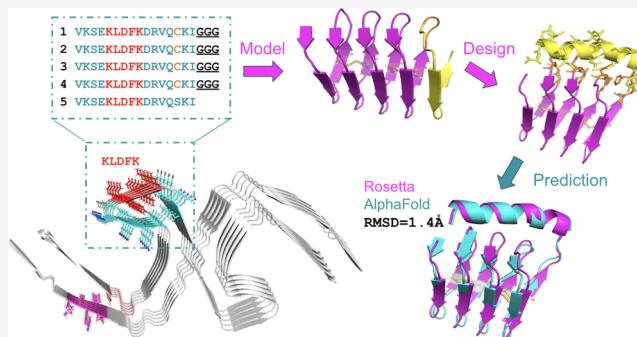
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Tau pathology is associated with many neurodegenerative disorders, including Alzheimer's disease (AD), where the spatio-temporal pattern of tau neurofibrillary tangles strongly correlates with disease progression, which motivates therapeutics selective for misfolded tau. Here, we introduce a new avidity-enhanced, multi-epitope approach for protein-misfolding immunogen design, which is predicted to mimic the conformational state of an exposed epitope in toxic tau oligomers. A predicted oligomer-selective tau epitope $^{343}\text{KLDFK}^{347}$ was scaffolded by designing a β -helix structure that incorporated multiple instances of the 16-residue tau fragment $^{339}\text{VKSEKLDFKDRVQSKI}^{354}$. Large-scale conformational ensemble analyses involving Jensen-Shannon Divergence and the embedding depth \mathcal{D} showed that the multi-epitope scaffolding approach, employed in designing the β -helix scaffold, was predicted to better discriminate toxic tau oligomers than other “monovalent” strategies utilizing a single instance of an epitope for vaccine immunogen design. Using Rosetta, 10,000 sequences were designed and screened for the linker portions of the β -helix scaffold, along with a C-terminal stabilizing α -helix that interacts with the linkers, to optimize the folded structure and stability of the scaffold. Structures were ranked by energy, and the lowest 1% (82 unique sequences) were verified using AlphaFold. Several selection criteria involving AlphaFold are implemented to obtain a lead-designed sequence. The structure was further predicted to have free energetic stability by using Hamiltonian replica exchange molecular dynamics (MD) simulations. The synthesized β -helix scaffold showed direct binding in surface plasmon resonance (SPR) experiments to several antibodies that were raised to the structured epitope using a designed cyclic peptide. Moreover, the strength of binding of these antibodies to in vitro tau oligomers correlated with the strength of binding to the β -helix construct, suggesting that the construct presents an oligomer-like conformation and may thus constitute an effective oligomer-selective immunogen.

KEYWORDS: protein design, epitope scaffolding, tau protein, molecular dynamics, protein misfolding, cyclic peptides, ensemble comparison, Alzheimer's disease



INTRODUCTION

Tau is a microtubule-associated protein mainly found in the axonal part of neurons.¹ The assembly of α and β tubulins results in the formation of microtubules. Tau binding to tubulin promotes the assembly of microtubules and stabilization of the microtubule network, which is required for axonal transport and neuronal health.^{2–4} A hallmark of pathology in Alzheimer's disease as well as other tauopathies is the formation of tau oligomers and fibrils,^{5,6} which in turn destabilizes microtubules and thus intracellular trafficking.⁷ Destabilization of microtubules leads to neuronal and synaptic loss^{8,9} and promotes neuroinflammation.^{10,11} Neurodegenerative disorders involving tau pathology, collectively known as tauopathies, include frontotemporal dementia (FTD) or Pick's disease (PiD), chronic traumatic encephalopathy (CTE), corticobasal degeneration (CBD), progressive supranuclear

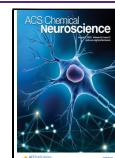
palsy (PSP), and Alzheimer's disease (AD).¹² The presence of tau lesions strongly correlates with the progression of these diseases.^{10,13–15} Alternative splicing gives rise to alternate isoforms having variable numbers of microtubule-binding domains.¹² During pathology, a given isoform may aggregate on a different energy landscape¹⁶ to various fibril morphologies that correspond to different disease phenotypes.^{17–19}

While the pathogenesis is not yet fully understood, the aberrant presence of extracellular amyloid- β (A β) plaques and

Received: January 6, 2023

Accepted: June 22, 2023

Published: July 17, 2023



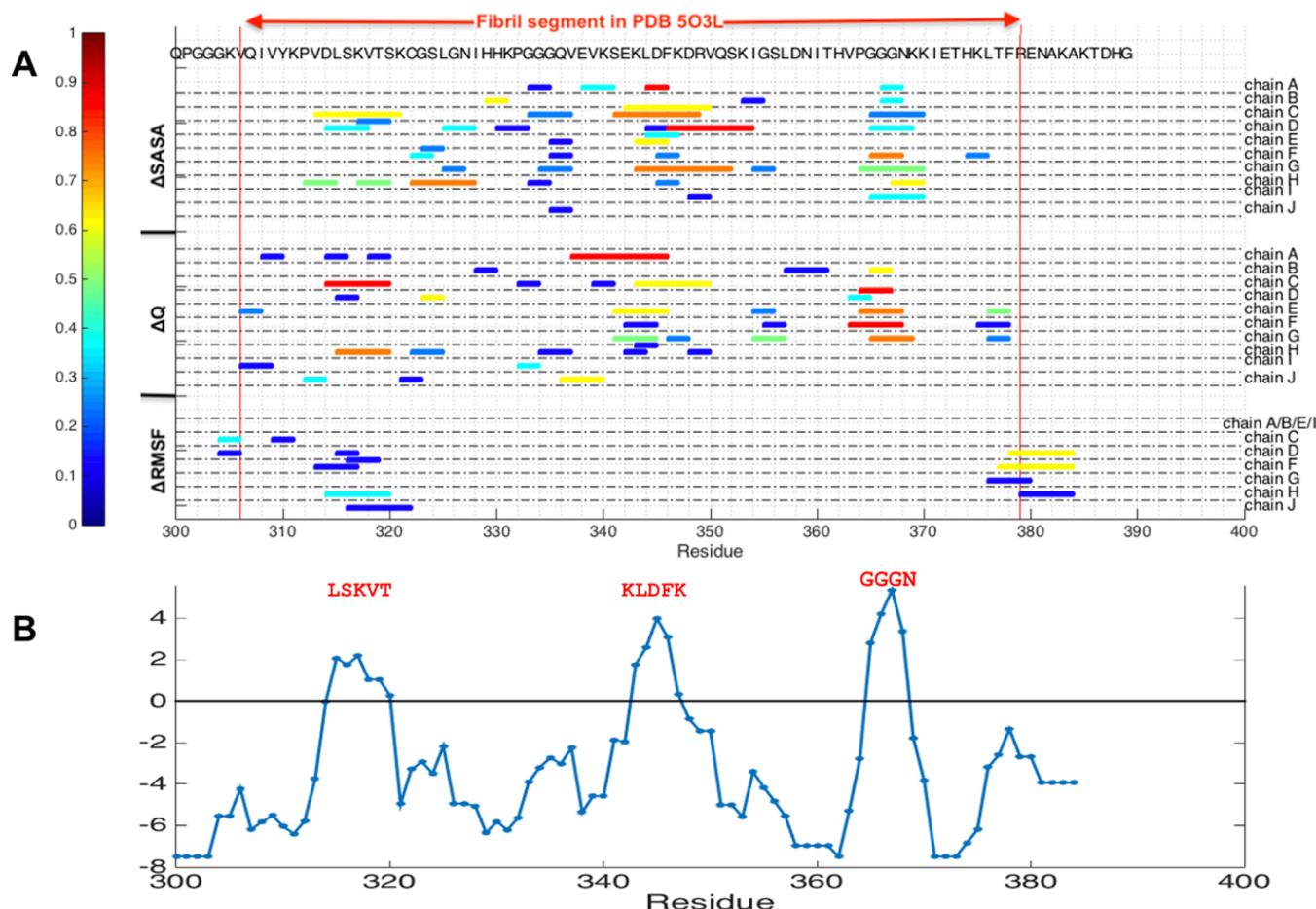


Figure 1. Prediction of tau epitopes. (A) Per residue normalized increase in solvent accessible surface area (ΔSASA), loss of native contacts (ΔQ), and local increase in root mean squared fluctuations (ΔRMSF).^{28,33} (B) Per residue weighted sum of ΔSASA , ΔQ , and ΔRMSF over all chains with the boundary chains in the 5×2 -chain, 2-prototubular structure A, B, I, and J having half the weight of other chains. The zero point was set to a numerical value (7.5) that identified epitopes that are within 5 amino acid long. The sequences shown in red font are the predicted epitopes for tau.

intracellular tau-containing neurofibrillary tangles in AD brains make both $\text{A}\beta$ and tau proteins the two main therapeutic targets.⁶ Safely targeting therapeutically relevant epitopes on $\text{A}\beta$ or tau is nevertheless a herculean task.²⁰ Immunogens (or vaccines) are highly effective therapeutic agents employed in the public health sector to trounce diseases. After two decades of research and development, vaccines for treating AD have been unsuccessful.^{21–24} However, the recent approval of the first immunotherapy against AD by the United States Food and Drug Administration (FDA), although amidst some controversy, has reactivated interest in immunotherapeutic strategies for combating AD. AADvac1, the first immunotherapy targeting tau, was advanced to clinical trials in 2013.²⁵ Over the following decade, more than a dozen immunogens targeting tau have been developed and advanced to clinical trials.¹⁰ A strategy employed in some administrations of peptide or protein immunogens in neurodegenerative disease is the presentation of either a whole protein or a peptide containing an epitope of interest. Nevertheless, it is the aggregated, oligomeric forms of the protein that have been shown to be toxic.⁶ This motivates a strategy wherein key epitopes of the protein are scaffolded in such a way as to present multiple copies of the epitope in a polyvalent immunogen.

The problem of designing an effective vaccine to treat a protein-misfolding disease such as AD is two-fold, involving

both epitope identification and selective targeting to pathogenic species. Epitopes must be identified that are disease-specific and induce strong and long-lasting immunity,²⁶ and the presentation of the immunogenic, disease-specific epitopes must selectively target pathogenic species while sparing healthy proteins. The epitope in the healthy protein often has an identical and unmodified primary amino acid sequence as the toxic species. Vaccines induce an ensemble of immunogen-specific antibodies, which recognize various epitopes on the immunogen. T cells recognize MHC-presented linear peptide sequences, while B cells recognize both linear and conformational epitopes that are exposed on antigenic surfaces.^{24,27} In the present application, an effective vaccine must induce antibodies that selectively bind to misfolded conformations in toxic forms of the protein over healthy “native” conformations.²⁸ Like its predecessors tested in clinical trials, there remain concerns regarding the efficacy of the recently approved $\text{A}\beta$ -targeting antibody aducanumab for the treatment of AD.^{29–32} It is therefore important to continue to innovate new approaches to combat AD.

In this paper, we rationally design an immunogen that captures the conformational ensemble of misfolding-specific epitopes in tau protein, characterized by regions most readily disordered and solvent exposed in a stressed, partially disordered fibril. We achieved this by designing a scaffold that is primarily a β -helix that incorporates multiple instances

of the targeted epitope, which is stabilized by a single α -helix. Conformational analyses show that this scaffold presents the misfolding-specific epitopes in conformations distinct from those in either the intrinsically disordered monomer, or the amyloid fibril, more effectively than the approach of scaffolding a single epitope. The designed immunogen is predicted to be suitable for eliciting therapeutic antibodies that are selective for toxic tau oligomers in Alzheimer's disease.

RESULTS AND DISCUSSION

Prediction of Tau Epitopes. We predicted misfolding-specific epitopes in tau protein using the procedure described in the *Prediction of Epitopes* section, which has previously been used to predict misfolding-specific epitopes in α -synuclein,^{28,34} superoxide dismutase (SOD1),³³ and $A\beta$.^{35,36} An equally weighted sum of three measures of local disorder from the fibril structure (PDB 5O3L; see Figure S1) including loss of native contacts (ΔQ), increased root mean squared fluctuations (ΔRMSF), and increased solvent accessible surface area (ΔSASA) were used to predict epitopes (Figures S2–S4). Figure 1A shows the per residue normalized values for ΔSASA , ΔQ , and ΔRMSF . As a necessary condition for a candidate epitope, a contiguous string of residues must satisfy a consistent decrease in Q , an increase in SASA, and a value of ΔRMSF greater than the average across all residues in 8 of 10 independent simulations for all metrics. Figure 1B shows the per residue weighted sum of ΔSASA , ΔQ , and ΔRMSF over all chains with the boundary chains A, B, I, and J in the fibril structure having half the weight of other chains. Mathematically, we plotted the function Ω (eq 1) vs the residue index.

$$\begin{aligned} \Omega = & \frac{1}{2}(R_A + R_B + R_I + R_J) + R_C + R_D + R_E + R_F \\ & + R_G + R_H \end{aligned} \quad (1)$$

where

$$\begin{aligned} R = & (\Delta \text{SASA}/\Delta \text{SASA}_{\max}) + (\Delta Q/\Delta Q_{\max}) \\ & + (\Delta \text{RMSF}/\Delta \text{RMSF}_{\max}) \end{aligned} \quad (2)$$

The zero point of function Ω was set at 7.5 to identify epitopes that are about 5-residue long. Three motifs above the zero point, ³¹⁵LSKVT³¹⁹, ³⁴³KLDFK³⁴⁷, and ³⁶⁵GGGN³⁶⁸, were thus predicted as misfolding-specific epitopes for tau.

Robustness of Epitopes. The disorder-prone epitope, KLDFK, was predicted solely by analyzing the unfolding trajectories of a single tau fibril structure (PDB 5O3L). While a more robust measure would consider collective coordinates prediction for other tau fibril structures in the Protein Data Bank (PDB), such analysis would currently require a significant amount of computational resources. On the other hand, the disorder-prone epitope, KLDFK, is found to be at least moderately exposed in all of the five tau fibrils we examined, including 5O3L,¹⁷ 5O3T,¹⁷ 7P66,¹⁹ 7P68,¹⁹ and 6TJO³⁷ (Figure S7). We note that these fibril structures correspond to different tauopathies (Alzheimer's, Globular Glial Tauopathy, and Corticobasal degeneration) and need not expose the same epitopes. Examination of the SASA profiles in Figure S7 shows that the 5-residue segment containing ³⁴³KLDFK³⁴⁷ is partly solvent exposed, having a SASA that exceeds more than 70% of the other windows in all fibril structures analyzed, yet the epitope is generally flanked by regions having even higher solvent exposure. We may thus hypothesize that this epitope

would experience a significant change in solvent exposure upon stress to the fibril. The pairwise local distance test (lldt)³⁸ shows that 5O3L and 5O3T as well as 7P66 and 7P68 are mutually similar, resulting in 3 distinct structure classes (Figure S7G).

Immunogenicity of Epitopes. The immunogenicity of the epitopes predicted from Collective Coordinates³³ was examined using the Epitopia³⁹ server. Figure S10 plots the immunogenicity for each residue on a scale between 1 (lowest) and 5 (highest). By this metric, ³⁴³KLDFK³⁴⁷ had below-average immunogenicity (2.5 vs 3.3 average for 5 aa segments). Nevertheless, antibodies have been raised to cyclic peptides containing KLDFK, as discussed further in Results: *Binding Strength of Antibodies to the β -Helix Immunogen Correlates with Their Binding to In Vitro Oligomers*. Interestingly, ³¹⁵LSKVT³¹⁹, another misfolding-prone epitope predicted from the collective coordinates algorithm, did have high predicted immunogenicity.

In summary, a protofibril region is used as a model to predict oligomer-selective epitopes. These regions of peptide sequence are conformationally distinct from either those in an isolated monomer (because they are presented in a stressed fibril conformation) or those in the fibril itself (because they were selected as being disordered from the fibril when stressed). This recipe has been previously successful in identifying oligomer-selective antibodies in $A\beta$ ^{35,36} and α -synuclein.^{28,34} The problem then turns to how to properly scaffold the predicted epitope in order to present it in a conformation similar to that in the stressed fibril. For $A\beta$ and α -synuclein, this has been successfully achieved by cyclic peptide "glycindel" scaffolds.²⁸ We now turn to this analysis for the above-predicted epitopes.

Conformational Selectivity of Single-Epitope Cyclic Peptide Scaffolds. Our goal at this stage is to design scaffolds that are selective for toxic tau oligomers. In other words, we aim to design scaffolds that present epitopes in similar conformations to those present on toxic tau oligomers that are different from conformations presented by healthy tau monomers, or fibril, as we wish to target paracrine species that have the potential to spread pathology by prion-like propagation.^{6,40}

To this end, we computationally designed "glycindel" cyclic peptides (see Methods: *Epitope Scaffolding Using Cyclic Peptides*) to scaffold two of the three predicted tau epitopes, including ³¹⁵LSKVT³¹⁹ and ³⁴³KLDFK³⁴⁷ epitopes. We did not construct scaffolds for the ³⁶⁵GGGN³⁶⁸ predicted tau epitope because it consists of mostly glycine (GLY) residues, which are less immunogenic than other side chains. Each constructed cyclic peptide scaffold consists of a single instance of either the ³¹⁵LSKVT³¹⁹ or ³⁴³KLDFK³⁴⁷ epitope flanked between one to four variable numbers of GLY on the N- and/or the C-terminus. Hence, we constructed a total of 4×4 or 16 cyclic peptide scaffolds for each of the ³¹⁵LSKVT³¹⁹ and ³⁴³KLDFK³⁴⁷ epitopes.

To determine the conformational selectivity of the constructed cyclic peptide scaffolds in silico, we sampled the conformational ensemble of each epitope in four different contexts: The fibril, oligomer, monomer, and each of $16 \times 2 = 32$ cyclic peptide scaffolds, by performing all-atom MD simulations in explicit solvent (see Methods: *Sampling Conformational Ensembles*). Each cyclic peptide epitope ensemble was then compared to the same epitope ensemble

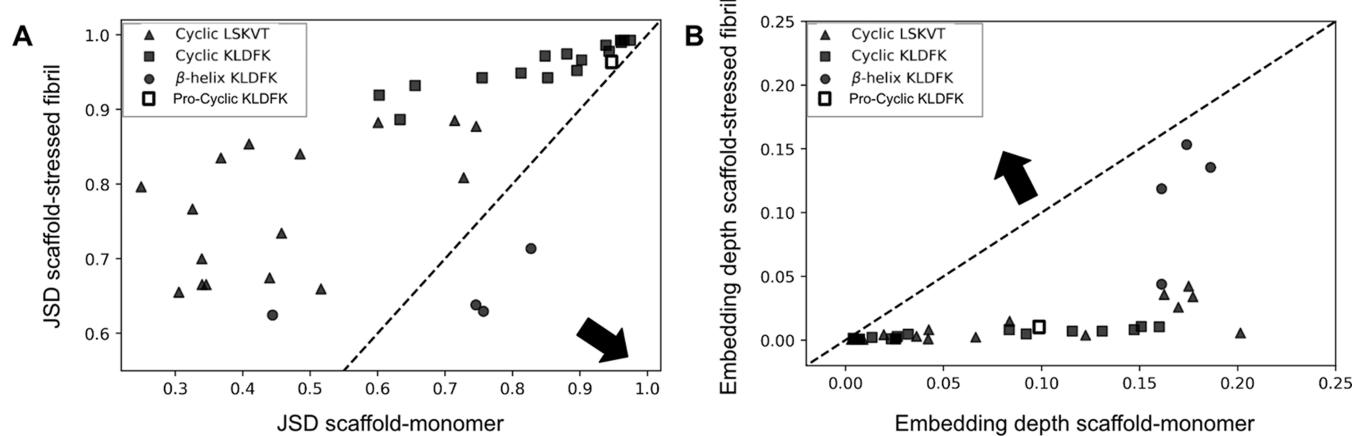


Figure 2. Conformational ensemble similarity. (A) Similarity between two conformational ensembles was quantified using JSD (see Methods: Comparing Conformational Ensembles), which is normalized to lie between 0 (for identical ensembles) to 1 (for entirely different ensembles). Shown is the similarity between the epitope ensemble in the context of a scaffold and the oligomer, vs the similarity between the epitope ensemble in the context of the scaffold and the monomer, for cyclic peptide “glycindel” scaffolds of the LSKVT epitope (triangles), cyclic peptide glycindel scaffolds of the KLDFK epitope (squares), CRISPro-designed cyclic CPPPPKLDFKGPGG scaffold (open squares), and the β-helix scaffold of the four KLDFK epitope repeats (circles). The arrow indicates the direction of the favorable conformational profile, and the dashed line is the line $y = x$. (B) Same as for panel (A), but for the embedding depth \mathcal{D} . Now, data points toward the upper left above the dashed $y = x$ line are favorable; interestingly, none of the points fall in this regime, but the β-helix scaffold epitopes do cluster differently and have the largest embedding depth with the oligomer model. Cyclic peptide ensembles have variable embedding depth with the monomer ensemble, but they show little embedding in the oligomer model ensemble.

in the context of the fibril, oligomer, and monomer. We used Jensen–Shannon Divergence (JSD) to quantify the degree of similarity between two conformational ensembles (see Methods: Comparing Conformational Ensembles). We also apply a metric we have recently introduced to measure the degree to which one ensemble is subsumed by another ensemble, the embedding depth \mathcal{D} .²⁸

Glycindel scaffolds were generically distinct from the fibril, with JSD values between scaffold and fibril ranging from 0.96 to 1.00. We thus focused on comparisons with the monomer and stressed fibril oligomer model ensembles.

We seek epitope scaffolds with both high similarity (low JSD) to the oligomer ensemble and low similarity (high JSD) to the monomer ensemble. Similarly, we seek epitope scaffolds with both high embedding depth (high \mathcal{D}) in the oligomer ensemble and (low \mathcal{D}) in the monomer ensemble. Figure 2A shows the conformational ensemble similarity (JSD) to the oligomer vs JSD to the monomer for each of the 32 cyclic peptide scaffolds (16 scaffolds each for tau³¹⁵LSKVT³¹⁹ and³⁴³KLDFK³⁴⁷ epitopes).

Several observations are apparent in Figure 2A. In general, ³¹⁵LSKVT³¹⁹ scaffolds sample epitope ensembles that are more similar to both the monomer and the oligomer ensembles than ³⁴³KLDFK³⁴⁷ scaffolds. There is a positive correlation between the similarity between scaffold and oligomer and the similarity between scaffold and monomer, with Pearson correlation coefficient $r = 0.71$ for all the cyclic peptide scaffolds, a moderate correlation for the ³¹⁵LSKVT³¹⁹ epitope scaffolds ($r = 0.56$), and strong correlation for the ³⁴³KLDFK³⁴⁷ epitope scaffolds ($r = 0.92$). This correlation shows that as the cyclic peptide scaffolds become less similar to the monomer (increasing JSD), which is desirable, they also become less similar to the oligomer (increasing JSD), which is undesirable. It is worth noting that the JSD between monomer and oligomer is 0.54, more similar than any of the scaffolds to the oligomer and more similar than any of the KLDFK scaffolds to the monomer. The more similar the monomer and oligomer

ensembles are, the stronger the correlation we expect in the data of Figure 2A.

The desired scaffold should be more similar to the oligomer than the monomer, which would put the data point for a scaffold below the $y = x$ dashed line in Figure 2. The result here shows an unfavorable result that the typical conformations explored by glycindel cyclic peptide scaffolds of ³¹⁵LSKVT³¹⁹ and ³⁴³KLDFK³⁴⁷ epitopes are predicted to be more similar to those of an isolated monomer of tau than an oligomer model using a stressed fibril. We note that this result is in contrast to glycindel scaffolding of epitopes for α-synuclein, which were able to predict scaffolds having greater similarity to oligomers.²⁸

Figure 2B shows the embedding depth \mathcal{D} of each of the 32 cyclic peptide ensembles in the oligomer ensemble vs \mathcal{D} in the monomer ensemble. Here, the glycindels show variable embedding depth in the monomer ensemble but show very little embedding in the oligomer model ensemble (points are spread horizontally along the bottom of the plot, below the line $y = x$).

Design of Avidity-Enhanced Multi-Epitope β-Helix Scaffolds. From the above result, all of the 32 cyclic peptide scaffolds are predicted to explore monomer-like conformations more often than oligomer-like conformations. Though this does not rule out the possibility that antibodies raised against such a peptide may be selective to the oligomer-like conformations occasionally explored by the ensemble, it suggests that the probability for such a scenario is low. We therefore designed a single-chain β-helical scaffold consisting of multiple instances of the ³⁴³KLDFK³⁴⁷ epitope. We chose the ³⁴³KLDFK³⁴⁷ epitope because of its location in the tau fibril structure (Figure 6A). This epitope is in the 4th microtubule-binding domain of tau; it does not overlap with the epitopes of any antibodies currently in clinical trials for AD.^{6,41} The closest epitope is that of Zagotenemab, which has a discontiguous epitope consisting of residues 7–9/312–322.⁴² The epitope ³⁴³KLDFK³⁴⁷ is a motif that is part of a larger contiguous

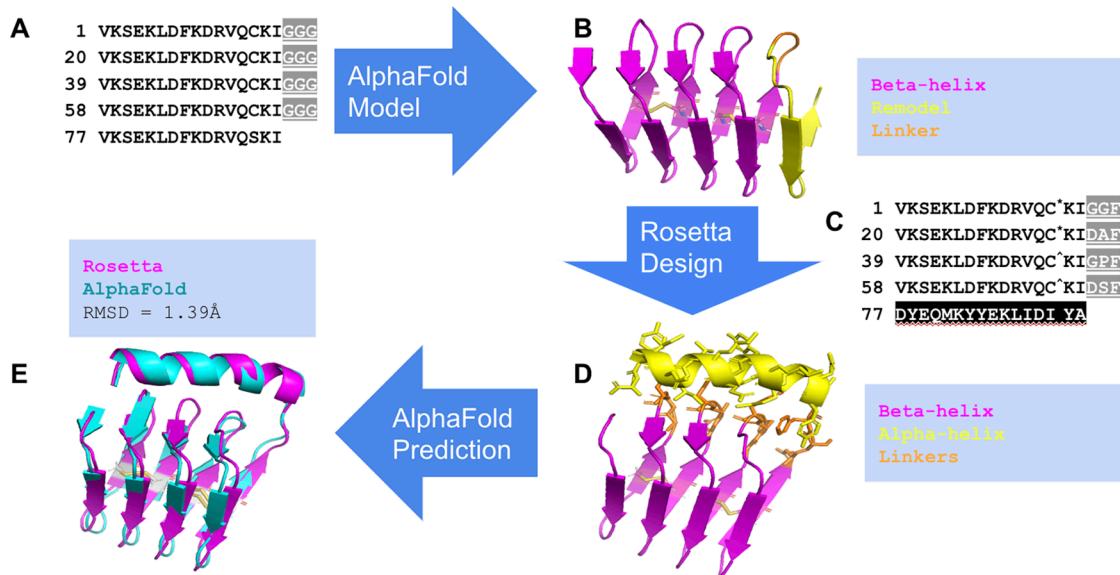


Figure 3. β -Helix scaffold design strategy. (A) Five copies of tau-derived 16-residue $^{339}\text{VKSEKLDFKDRVQSKI}^{354}$ repeating sequence are linked serially by glycine tripeptide linkers (GGG); S352C substitutions are made in the first four repeats. These were used to model (B) a β -helix structure, with disulfide bridges between residues C14 and C33 and residues C52 and C71. (C) Redesign of the last repeat (black box) and linker residues (gray boxes) to improve the stability of (D) the β -helix scaffold by creating interactions between the designed residues. Last repeat is shown in yellow and linkers are shown in orange, both with licorice side chains. (E) Designed β -helix structure (magenta) and AlphaFold predicted structure (cyan) have C_α root mean squared deviation (RMSD) of 1.39 Å. All images were rendered with PyMOL molecular visualization system (<https://pymol.org>).

sequence in the tau monomer, $^{339}\text{VKSEKLDFKDRVQSKI}^{354}$ (tau339-354), that has close spatial proximity between its N- and C-termini within the tau fibril structure PDB 5O3L. Tau structures in the PDB are polymorphic, often depending on the disease phenotype. We have chosen a tau structure associated with AD; however, other tau structures in the PDB also satisfy the condition that the N- and C-termini of V339 and I354 are in close spatial proximity (e.g., PDBs 6NWP and 6NWQ involved in chronic traumatic encephalopathy). The strategy used here should potentially be applicable to this disease as well. β -helical structures have been observed in nature in other contexts, for example, antifreeze proteins.^{43,44}

To design β -helix peptide scaffolds, we tethered 5 copies of tau339-354 sequence with glycine (GLY) tripeptides (GGG) (see Figure 6). We then made a S352C (serine 352 to cysteine (CYS)) mutation in the first four copies of the sequence, with the intention of designing two disulfide bridges that would improve the stability of the scaffold (Figure 3A). Indeed, AlphaFold⁴⁵ predicts a β -helix structure (Figure 3B) with the S352C substitutions but predicts a disordered peptide without the S352C substitutions (see Methods: Structure Prediction with AlphaFold). All 5 AlphaFold models predicted a β -helix structure for the five-repeat sequence (with S352C substitutions) with high average per residue confidence scores of 89, 92, 89, 88, and 85% for the 1st through 5th AlphaFold models, respectively (see Figure S5C).

We had also examined a 4-repeat β -helix scaffold (Figure S6A). Although the AlphaFold confidence (PLDDT) scores for this construct were high (Figure S6B), the free energy surface as a function of Q , obtained by Hamiltonian replica exchange molecular dynamics (HREMD) simulations, did not predict a stable native-like structure (Figure S6C, see also Results: Designed β -Helix Scaffold is Structurally and Energetically Stable). This prompted additional design as described further below.

The formation of β -helix structure always put cysteine C352 in successive repeats of tau339-354 within 4.8 Å of each other. Two disulfide bonds, between cysteines in the 1st and 2nd repeats and between cysteines in the 3rd and 4th repeats, were predicted as intended by only the third and fifth AlphaFold models; the other three models predicted only one of the two disulfide bonds.

We thus used the 3rd AlphaFold model with the two designed disulfide bonds and higher confidence score (89%) as the initial β -helix structural model for a redesign. To further improve the stability of the β -helix structural model, we remodeled the amino acids starting from the last GGG linker and including the linker and the last tau339-354 repeat (see Methods: β -Helix Structure Modeling). The last tau339-354 repeat was remodeled to an α -helix structure in a conformation that was docking and interacting with all the linkers (Figure 3D). For both linker and α -helix, we allowed for the possibility of any of the 20 naturally occurring amino acids to occur in all remodeled positions. Finally, all four linkers and α -helix residues were designed to optimize the sequence and structure of the β -helix scaffold (Figure 3D) (see Methods: β -Helix Sequence Design and Structure Optimization).

Ten thousand (10,000) decoys of the β -helix scaffold were generated during the design phase using Rosetta.^{46,47} Figure S5A plots the Rosetta energy per residue of the designed decoys against the C_α root mean squared deviation (RMSD) from the starting Rosetta-remodeled structure, with energy = 1.98 Rosetta energy units (REU). The RMSD for all designed decoys is below 1 Å except for one outlying decoy with an RMSD of 1.13 Å. All the designed decoys have energies below -3 REU per residue, which is significantly lower than the energy of the starting structure.

AlphaFold Predicts the Designed β -Helix Structure. To validate the feasibility that a designed sequence folds into the designed β -helix structure, AlphaFold was used to predict

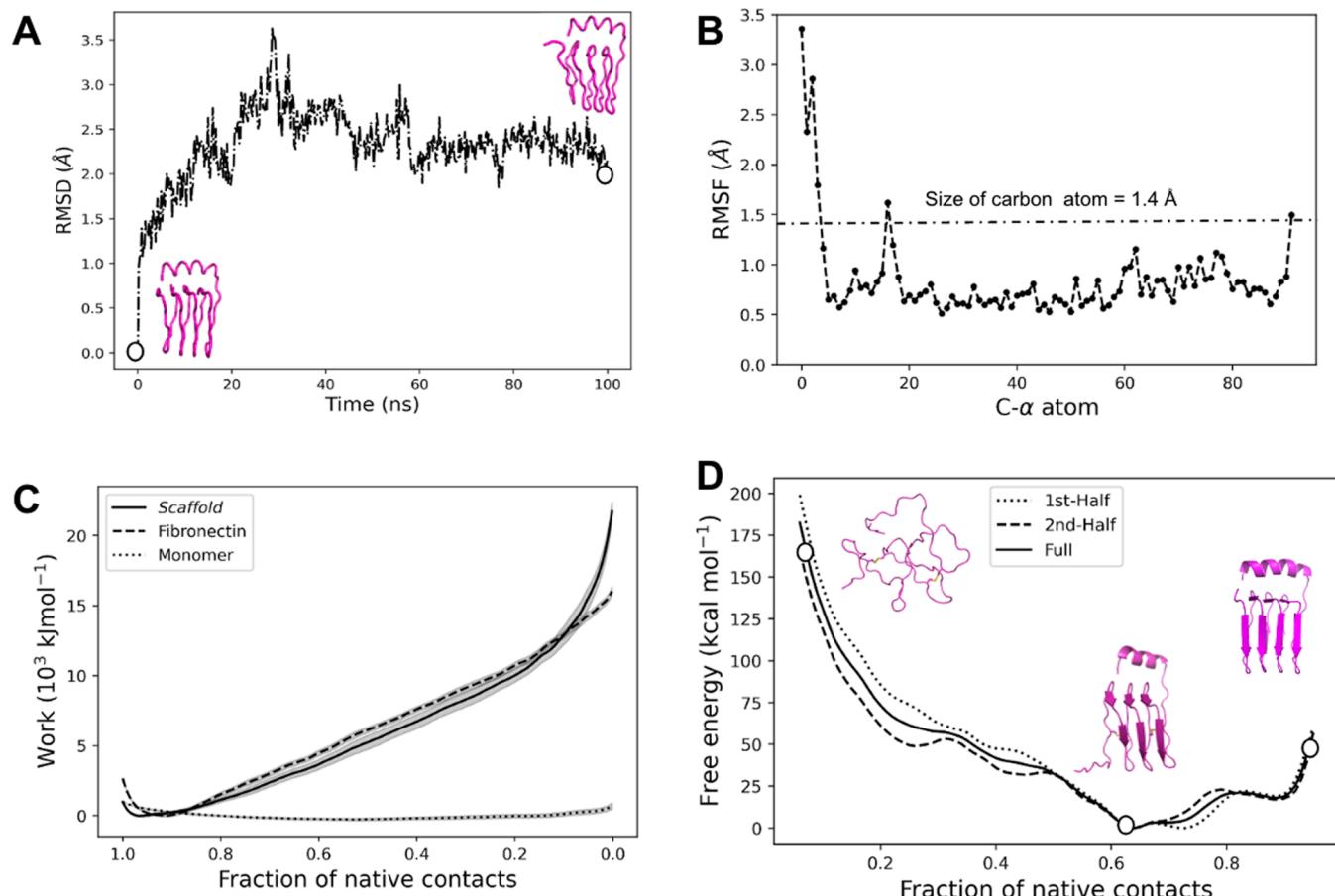


Figure 4. Computational characterization of the stability of β -helix scaffold. The (A) global and (B) local structural stability in the free energy basin of the Rosetta-designed structure. Inset images in panel (A) show snapshots of the initial and final structures. The secondary structure evolution over the 100 ns simulation time is shown in Figure S11. (C) Mechanical stability measured by the average cumulative work done during forced unfolding simulations (solid line), compared to fibronectin,^{50,51} a stable extracellular matrix protein with a similar length (dashed line), and an isolated (intrinsically disordered) tau monomer (dotted line), which serves as a negative control. (D) Thermodynamic stability determined from the relative free energy surface along the fraction of native contacts Q (eq 6) order parameter. Snapshots from the first half (dotted), second half (dashed), and full 80 ns trajectories (solid) in each umbrella are shown (see text). Inset images show representative snapshots of the structures at the free energy minimum, where $Q = 0.64$ and where $Q = 1$ and $Q = 0.1$.

the structure from the designed sequences alone (see Methods: **Structure Prediction with AlphaFold**). A distribution of energy was constructed for the 10,000 decoys, and the lowest 1% of energy defined a cutoff that included 82 unique sequences. The structures of these 82 unique sequences were then predicted with AlphaFold. AlphaFold outputs 5 models per sequence, with each model predicting one structure. Table S1 shows the average per residue confidence (PLDDT) score for the median ranking model of the 5, and the C_α RMSD from the Rosetta-designed β -helix structure for the three top-ranking models for 28 selected sequences. The 28 selected sequences are those from the 82 lowest energy sequences that satisfy two criteria: (1) they have an average per residue confidence score for the median ranking model $\geq 50\%$, and (2) each of the three top-ranking models has a C_α RMSD from the remodeled β -helix structure $\leq 5 \text{ \AA}$.

To characterize the structural and energetic stability of a representative designed sequence in silico, several selection criteria were imposed. We selected a sequence that (1) has a low energy; it is among the 1% lowest energy decoys, corresponding to an energy cutoff of -3.41 REU per residue, (2) has a high structure prediction score; a median confidence score (across 5 models) ≥ 70 , which is the threshold for high

confidence structure prediction for AlphaFold.⁴⁵ That is, at least three AlphaFold models must have high confidence scores (see Figure S5B) and (3) has the designed structure across all the top three models; it is predicted to have (i) the designed β -helix structure and (ii) correctly forms the two designed disulfide bonds. Table S1 shows that model 0 is the only sequence satisfying these stringent selection criteria. The Rosetta-designed structure and AlphaFold predicted structure for model 0 have C_α RMSD of 1.39 \AA (Figure 3E).

Designed β -Helix Scaffold Is Structurally and Energetically Stable. One of the most important pharmacological properties for a peptide such as the designed tau β -helix scaffold intended for use as an immunogen is stability. To this end, the stability of the selected β -helix scaffold (i.e., model 0) was characterized in silico. All simulations were performed in explicit solvent (0.15 mM NaCl) at 300 K and 1 bar using GROMACS⁴⁸ with the CHARMM36m force field⁴⁹ and periodic boundary conditions.

Metastability. To check the structural stability of the scaffold in the free energy basin of the designed structure, a 100 ns all-atom MD simulation starting from the designed structure was performed and the last 50 ns was used for analysis. Heavy-atom root mean square deviation (RMSD) and

C_{α} root mean square fluctuation (RMSF) were used to quantify the global and local stability, respectively. The RMSD vs time plot in Figure 4A shows that the scaffold is globally structurally stable within the simulation time. The RMSD from the starting structure of the sampled conformations remains below ≈ 3 Å. Apart from five N-terminus residues with RMSF between 1.5 and 3.5 Å (Figure 4B), the per residue local fluctuations remain below the size of a carbon atom (1.4 Å) for all residues, suggesting a high structural stability for the scaffold in the free energy basin of the designed structure.

Mechanical Stability. To quantify mechanical stability, 10 independent forced unfolding simulations were performed (see Methods: Mechanical Unfolding Simulations). The cumulative average work done by the applied forces to unfold the β -helix scaffold structure as a function of the native contact order parameter Q (eq 6) was taken as a measure of mechanical stability. The work performed in a given trajectory generally averages out the stochastic effects of forces and is a smooth function of the order parameter.^{52,53} Here, the work performed was averaged over 10 independent simulations to further average out stochastic effects during the unfolding process. For comparison, the analogous calculation was performed to assess the mechanical stability of isolated fibronectin type III domain of human tenascin (PDB ID: 1TEN)^{50,51} and tau monomer (chain A from PDB 5O3L),¹⁷ which serve as positive and negative controls respectively. The force field and simulation model (all-atom in TIP3P solvent) used for all proteins are the same, so no additional normalization across time, energy, or force is required for accurate comparison.⁵⁴ Figure 4C shows that the mechanical stability of the scaffold is comparable to that of fibronectin, a stable extracellular matrix protein with a similar length (92 vs 90 amino acids).

Thermodynamic Stability. We also determined the thermodynamic stability of the β -helix scaffold by performing a Hamiltonian replica exchange molecular dynamics (HREMD) simulation to determine the free energy surface as a function of Q (see Methods: Hamiltonian Replica Exchange Molecular Dynamics). The HREMD simulation consists of 100 replicas with harmonic restraints centered at positions equally spaced between the fraction of native contacts $Q = 0$ to $Q = 1$. Each replica was simulated for 250 ns for a total of 25 μ s. A total of 8000 conformations were sampled at 10 ps intervals from the last 80 ns of each replica to determine the relative free energy of unfolding along Q for the scaffold. Figure 4D shows a largely downhill free energy surface towards higher degrees of nativeness, until the free energy minimum of the scaffold at $Q = 0.64$. There is a weak metastable minimum near $Q \approx 0.9$, but entropic forces bias the global minimum towards a lower degree of native contacts. Nevertheless, the global minimum ensemble has a large amount of native structural similarity: The average RMSD of this ensemble is 1.25 Å, and if we ignore the largely disordered N-terminal 13 residues, the RMSD of the remainder is 0.45 Å. Figure 4D also shows representative snapshots of the conformations sampled at $Q = 0.64$, $Q = 0.1$, and $Q = 1$. The $Q = 0.64$ snapshot shows that most of the lost contacts are the result of the flexible N-terminus (cf. Figure 4B). Overall, this result shows that the designed scaffold is thermodynamically stable in a largely structured, β -helical conformation with a stabilizing α -helix. This result is in contrast to similar analyses applied to a 4-repeat β -helix construct (without a stabilizing α -helix), which does not exhibit a stable native-like structure (Figure S6C). We note that as an immunogen proxy for an

oligomer, which generally is conformationally labile and thus elusive to structural determination,⁶ some degree of the partial disorder may actually be a favorable feature of the designed scaffold.

Proteolytic Stability. Cleavage analysis by the Procleave⁵⁵ server indicates that the designed β -helix structure was predicted to have cellular stability comparable or better than ubiquitin (Figure S8). Specifically, the highest-scoring cleavage site for the β -helix protein scored higher than ubiquitin for only 8 out of the 27 proteases examined.

Aggregation Propensity. The aggregation propensity of the designed β -helix scaffold structure, as determined by the AGGRESCAN3D server,⁵⁶ was found to be comparable to ubiquitin, and less than that of transthyretin, an amyloid-prone protein.⁵⁷ Figure S9 shows plots of the aggregation score vs residue index, along with the mean aggregation score and the fraction of residues that have an aggregation score larger than zero. The overall aggregation propensity of the construct is observed to be intermediate to ubiquitin and transthyretin and closer to ubiquitin. This may not be surprising since the region of the tau sequence selected for scaffolding was the first to disorder upon the simulated application of stress to the fibril. Consistent with this, the region containing this epitope is the least aggregation-prone according to AGGRESCAN3D (Figure S9E).

Designed β -Helix Scaffold Is Conformationally Selective for a Stressed Protomeric Model of Tau Oligomers. To determine the conformational selectivity of the designed β -helix scaffold for a stressed fibril model of tau oligomers (cf. Figure 2A), the conformation of the predicted ³⁴³KLDFK³⁴⁷ epitope was sampled in the context of monomer, stressed protomeric oligomer, and β -helix scaffold (see Methods: Sampling Conformational Ensembles). Four different KLDFK ensembles, corresponding to the four KLDFK instances in the β -helix scaffold sequence, were separately analyzed for the β -helix scaffold. Each KLDFK scaffold ensemble was then compared with the monomer and the oligomer ensembles, respectively, using the Jensen–Shannon divergence (JSD) similarity measure (see Methods: Comparing Conformational Ensembles) and the embedding depth D .

Here, we compare the conformational selectivity of a multi-epitope β -helix scaffold to single-epitope cyclic peptide scaffolds for the tau ³⁴³KLDFK³⁴⁷ epitope. Figure 2A shows that the β -helix scaffold is predicted to be more selective for tau oligomers than are cyclic peptide scaffolds. For the same conformational similarity to the oligomer, three of the four β -helix ensembles clearly show lower similarity (higher JSD) to the monomer than KLDFK cyclic peptide scaffolds. Similarly, for the same conformational dissimilarity to the monomer, three β -helix ensembles show higher similarity (lower JSD) to the oligomer than all of the KLDFK cyclic peptide scaffolds. The N-terminal KLDFK motif has an ensemble that is more similar to the monomer than that of the other three epitope repeats. This is expected since the N-terminus of the β -helix scaffold is relatively flexible, assuming the accuracy of the force field used in the simulations (CHARMM36m⁴⁹) (Figure 4D).

Figure 2B shows the corresponding embedding depths D for the 4 repeats of the epitope in the β -helix scaffold. The β -helix scaffolded epitopes also embed deeper in the monomer ensemble than the oligomer ensemble; however, the 3 most structured β -helix epitopes cluster together and have significantly higher embedding within the oligomer ensemble than do any of the glycindels. The N-terminal weakly

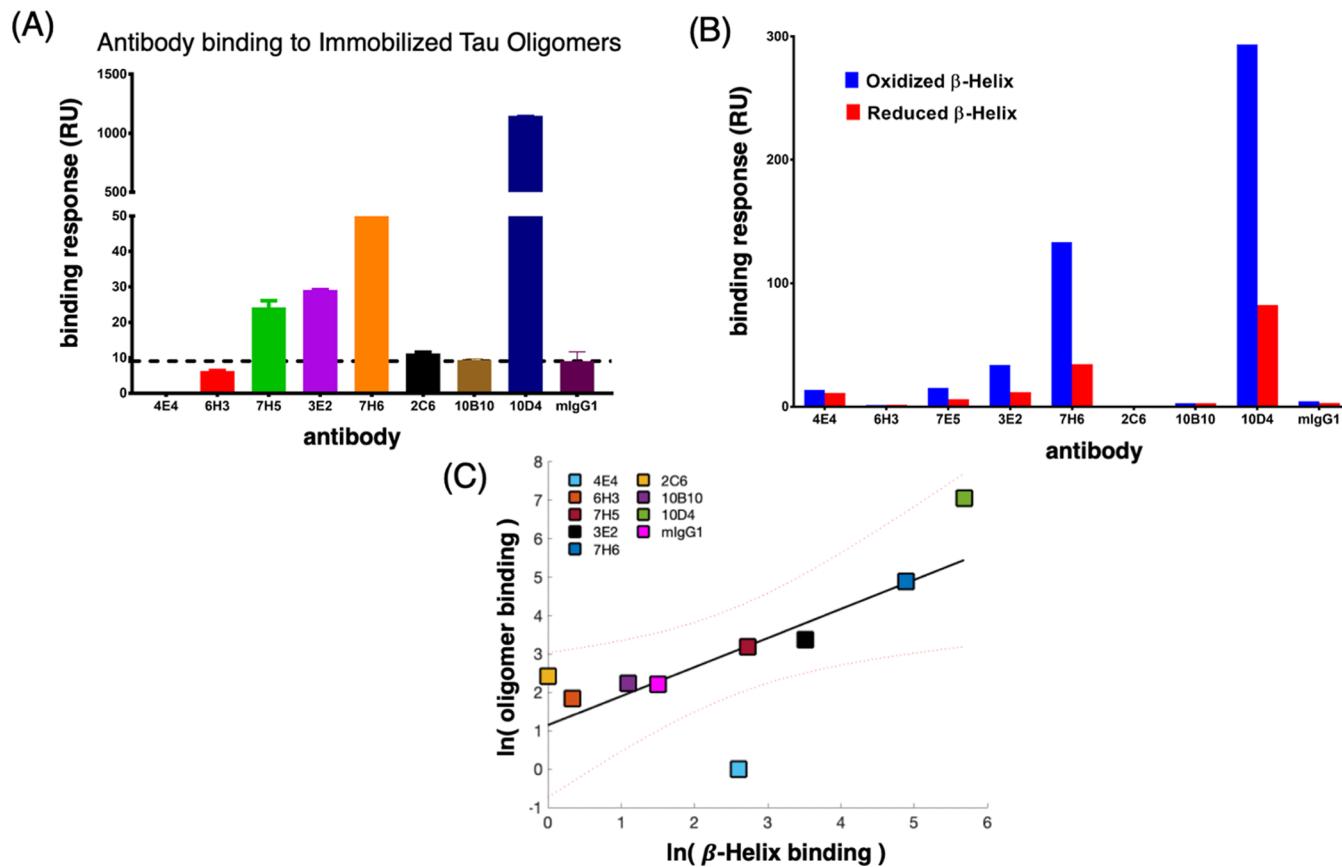


Figure 5. Experimental binding of antibodies to β -helix construct and tau oligomers. The (A) antibodies raised to tau cyclic peptide show variable binding response (BRU) to in vitro tau oligomers. (B) Disulfide reduction of the β -helix construct to generate a more unfolded form resulted in decreased antibody binding, indicating that the epitope's conformation in the β -helix favors antibody binding. (C) Strength of antibody binding to oligomers correlates significantly with the strength of binding to the β -helix construct ($r = 0.74, p = 0.02$), suggesting that the β -helix construct may present an oligomer-like conformational epitope and be a useful oligomer-selective immunogen.

structured epitope in the β -helix scaffold also behaves more monomer-like, falling within the distribution of the glycindel scaffolds.

Previously, Gibbs et al. have constructed single-epitope cyclic peptide glycindel scaffolds to successfully generate A β oligomer-selective antibodies³⁶ and α -synuclein oligomer-selective antibodies,³⁴ following a similar procedure to the one described in the *Epitope Scaffolding Using Cyclic Peptides* section. Therefore, this analysis cannot rule out the ability of a single-epitope cyclic peptide scaffold to raise oligomer-selective antibodies, and in fact, glycindel scaffolds²⁸ of collective coordinate-predicted epitopes³³ have been used to raise antibodies showing greater binding response to tau oligomers than tau monomers in SPR assays.⁵⁸ Specifically, antibodies raised from glycindel scaffolds of the epitope³⁴³KLDF³⁴⁶ showed greater selectivity to oligomers/monomers than a pan-tau control antibody (in SPR binding response units, $RU_{\text{olig}}/RU_{\text{mon}} \approx 2$ on average and up to a ratio of ≈ 4 for some antibodies, while the pan-tau antibody had $RU_{\text{olig}}/RU_{\text{mon}} \approx 1.4$). We note that this is not a huge improvement in selectivity, and the β -helical scaffold may result in even further enhanced selectivity. The selectivity of glycindels did show scaffold dependence (see Methods: *Epitope Scaffolding Using Cyclic Peptides*), e.g., antibodies raised by (3,2)KLDF were more selective to oligomers than those raised by either (3,1)KLDF or (4,1)KLDF. As well, co-incubation with antibodies raised against (3,2)KLDF showed the largest

reduction of the 3 scaffolds in the seeding of aggregation by PFFs in Tau RD P301S FRET Biosensor HEK cells.⁵⁸ Taken together, this evidence suggests that conformational presentation of the epitope is important in determining the selectivity and effectiveness of the resulting antibodies. In light of the predictive data in Figure 2, this stresses the potential fruitfulness of pursuing more accurate scaffolding strategies such as the one we have proposed here.

Binding Strength of Antibodies to the β -Helix Immunogen Correlates with Their Binding to In Vitro Oligomers. Several monoclonal antibodies have been obtained previously from mice immunized with the Tau cyclic peptide sequence CPPPKLDFKPGGG.⁵⁸ These include monoclonal antibodies 4E4, 3E2, 10B10, 6H3, 7H6, 10D4, 7E5, and 2C6, which were investigated here. An isotype negative control (mIgG1) was also used for comparison.

Although these antibodies were not raised using the β -helix construct, they were raised to a cyclized conformational epitope that overlaps with the sequence of the epitope in the β -helix construct (KLDFK), and some of these antibodies showed binding to in vitro tau oligomers (Figure 5A). The above cyclic peptide sequence was determined from the CRISPro computational method, as described in the *Methods* section. The ensemble of this specific sequence was not explicitly investigated, and although Figure 2 implies that is likely distinct from the ensembles of the β -helix construct, the

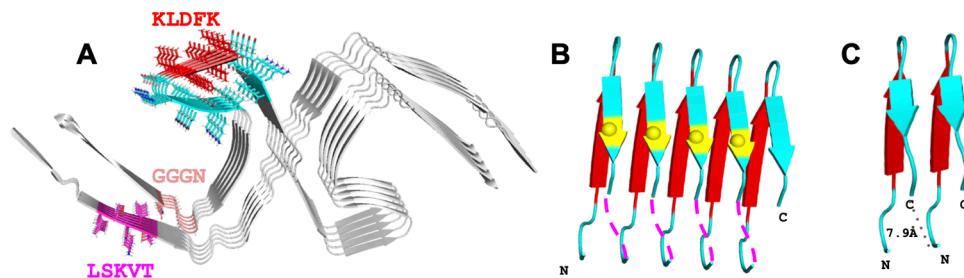


Figure 6. Modeling tau β -helix structure. (A) Tau fibril from Alzheimer's disease brain (PDB 5O3L) showing the side chains on the left protofilament in the figure for the predicted epitopes: $^{315}\text{LSKVT}^{319}$ (magenta), $^{343}\text{KLDFK}^{347}$ (red), and $^{365}\text{GGGN}^{368}$ (light orange) and the tau339-354 fragment $^{339}\text{VKSEKLDKFDRVQSKI}^{354}$ (non-epitope portion in cyan) that subsumes the $^{343}\text{KLDFK}^{347}$ epitope. (B) Segments of five chains of the tau339-354 fragment are connected together with interchain linkers to form a β -helix structure. Magenta dash lines indicate how glycine tripeptide linkers connect the C-terminus of one chain to the N-terminus of the next chain, and S352 is shown in yellow spheres. (C) 7.9 Å distance between the C-terminus of one chain and the N-terminus of the next chain in the tau fibril (black dotted line). All images were rendered with a PyMOL molecular visualization system (<https://pymol.org>).

ensembles may still partially overlap, suggesting some of these antibodies may be cross-reactive.

Experiments to examine the conformational specificity of these antibodies to the β -helix construct, vs an unfolded monomeric form, were performed by comparing the binding of the above set of antibodies to both oxidized and reduced β -helix constructs. The reduced construct was used as a proxy for a more unstable, unfolded form of the construct. Antibodies raised to the epitope in the cyclic peptide ensemble were observed to bind stronger to oxidized (well-folded) β -helix protein, than to the reduced (more poorly folded) protein (Figure 5B), for all antibodies tested. This result suggests that the targeted epitope is at least partially structured, and loss of structure reduces binding.

Figure 5C shows that the strength of antibody binding to oligomers correlated significantly with the strength of binding to the β -helix construct ($r = 0.74$, $p = 0.02$). This suggests that the β -helix construct presents an oligomer-like conformation and, thus, that the β -helix construct may be a useful oligomer-selective immunogen.

METHODS

Prediction of Epitopes. To predict tau epitopes, two conformational ensembles were generated, one for tau fibril and the other for a partially unfolded fibril, which serves as a model for tau oligomers. This oligomer-selective epitope prediction method posits that a partially unfolded protofibril ensemble is enriched in oligomer-selective conformational epitopes³³ and has been supported by previous experimental evidence.^{33,36}

The fibril ensemble was generated by performing a 30.1 ns molecular dynamics (MD) simulation, starting from a tau fibril structure (PDB 5O3L) extracted from an AD patient brain.¹⁷ To generate the partially unfolded fibril ensemble, ten independent 150 ns bias MD simulations were performed in order to average over the stochastic process of unfolding. During the first 50 ns of the bias simulations, the number of native contacts was linearly decreased to 60% of its native value. A native contact was defined as being formed when heavy atoms belonging to residues with a sequence separation ≥ 3 are within 4.8 Å over 95% of the time in a 100 ns equilibrium fibril simulation. This determines the total number N of native contacts. After 50 ns, the bias of 60% native contacts was fixed for the remaining 100 ns of simulation.

The bias potential $V(Q, t)$ is implemented to globally unfold the tau fibril structure along the fraction of native contacts Q reaction coordinate, which takes the following form

$$V(Q, t) = \frac{1}{2}k(Q - Q_0(t))^2 \quad (3)$$

where

$$Q = \frac{\sum_{ij}^N Q_{ij}(r_{ij})}{\sum_{ij}^N \langle Q_{ij}(r_{ij}) \rangle_{\text{native}}} \quad (4)$$

In eq 3, the sum in the numerator runs over the N native contacts that may be formed in a given conformation, while the sum in the denominator runs over the N native contacts in a member of the native conformational ensemble and is then averaged over conformations sampled during the 100 ns equilibrium simulation of the native fibril structure. Q_{ij} in eq 4 is defined by a contact function for each atom pair i, j with contact separated by distance r_{ij} using the following formula

$$Q_{ij}(r_{ij}) = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \quad (5)$$

In eq (5), we take $r_0 = 4.8$ Å, $n = 6$, and $m = 12$, to obtain a sigmoidal function that rapidly goes to one as r_{ij} falls below r_0 and rapidly goes to zero as r_{ij} becomes larger than r_0 . All simulations were performed in explicit solvent (TIP3P water model and 150 mM NaCl) at 300 K and 1 bar using GROMACS⁴⁸ and the CHARMM36m force field⁴⁹ with periodic boundary conditions.

Using the fibril ensemble as a reference, epitopes were predicted based on the consensus of three metrics that quantified local disorder, including loss of native contacts (ΔQ), increased root mean squared fluctuations (ΔRMSF), and increased solvent accessible surface area (ΔSASA). For a more detailed description of this approach for predicting epitopes, see refs 28 and 33.

Epitope Scaffolding Using Cyclic Peptides. As described in the Results: Prediction of Tau Epitopes section, 3 regions emerge from the collective coordinates prediction as candidate misfolding-specific epitopes: $^{315}\text{LSKVT}^{319}$, $^{343}\text{KLDFK}^{347}$, and $^{365}\text{GGGN}^{368}$ (see Figure 6A). The $^{365}\text{GGGN}^{368}$ epitope was not pursued further due to its lack of sequence complexity and expected lack of immunogenicity. To scaffold a single instance of the predicted $^{315}\text{LSKVT}^{319}$ or $^{343}\text{KLDFK}^{347}$ epitope, we computationally constructed “glycindel” cyclic peptides following the procedure described in ref 28. Briefly, each cyclic peptide scaffold consists of the epitope flanked on the N- and C-termini by n and m glycines, respectively. We refer to these scaffolds as cyclo(C-G_n-(EPITOPE)-G_m) or (n, m)EPITOPE, where C is a cysteine residue for conjugating a scaffold to carriers such as Keyhole limpet hemocyanin (KLH) in order to increase immunogenicity. Cyclic peptide construction is implemented computationally by mutating the corresponding flanking residues to glycine or cysteine in the native fibril structure using SCWRL4.⁵⁰ The topology of the cyclic peptide is obtained by head-to-tail backbone linkage using a locally written Python script,²⁸ which is then energy minimized using a steepest descent algorithm in GROMACS.⁴⁸

β -Helix Structure Modeling. The initial β -helix structure was modeled using a repeating unit of 16 amino acid sequence spanning the tau fragment $^{339}\text{VKSEKLDFKDRVQSKI}^{354}$ (tau339-354),^{43,44} which subsumes the $^{343}\text{KLDFK}^{347}$ predicted epitope (see Figure 6A). Five tau339-354 repeating sequences were connected serially with glycine tripeptide (GGG) linkers to form a single chain, β -helix peptide with a total of 92 residues (Figure 6B).

The distance between the C_α carbon of the C-terminal residue of tau339-354 on one chain and the C_α carbon of the N-terminal residue of tau339-354 on the next consecutive chain in the tau fibril structure (PDB 5O3L) was calculated to be $\approx 7.9 \text{ \AA}$ (Figure 6C). We thus hypothesized that an ideal linker length would be about 3 amino acids and investigated the distribution of distances in a non-redundant database of 26,900 structures.⁶⁰ Figure 7 shows the distribution of

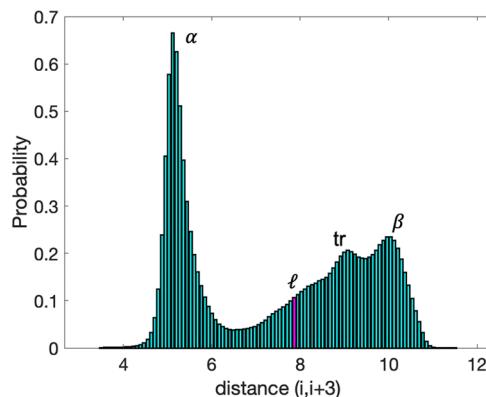


Figure 7. Distribution of $(i, i + 3)$ distances in a non-redundant database of protein structures. The 7.9 \AA distance between consecutive chains to be linked in our designed structure (ℓ in the plot) is well-within the distribution of distances between C_α atoms separated by three amino acids ($i, i + 3$). The distribution shows several features that are labeled in the figure: α (α -helical secondary structure), β (β -helical secondary structure), and tr (transitional elements between secondary structure and loop or turn).

distances between C_α atoms of amino acids separated by $(i, i + 3)$. We find several expected features, including peaks corresponding to α -helical and β -sheet secondary structures, along with an additional peak at $\approx 9 \text{ \AA}$ separation corresponding to transitions from (α, β) secondary structures to loops or turns.⁶¹ These sequences were often initiated or terminated by glycines or prolines. The 7.9 \AA distance of our linkers (magenta bar labeled ℓ in Figure 7) is in the middle of this distribution, supporting the hypothesis that a 3-amino linker was compatible in length and would not frustrate our designed structure.

In order to further stabilize the designed scaffold, the serine at amino acid position 352 was mutated to cysteine for all but the last tau339-354 repeat, amounting to four S352C mutations in the structure (Figure 1B). The S352C substitution introduces two disulfide bonds in the scaffold, which are introduced to improve the structural stability of the β -helix structure.

The structure of the designed sequence was predicted using AlphaFold⁴⁵ (see Methods: **Structure Prediction with AlphaFold** for further details), a deep learning algorithm for accurate protein structure prediction. Five models were generated by AlphaFold, and the model with the highest average confidence score that also formed the designed disulfide bonds between residues 14 and 33 and between residues 52 and 71 was selected as the initial model for the β -helix structure.

To further improve the stability of the initial model, the structure of the C-terminal repeat sequence was remodeled to form an α -helix structure using the RosettaRemodel algorithm⁶² while imposing the constraint that the distance between the N- and the C- termini C_α atoms of the whole β -helix structure is $\approx 3.5 \text{ \AA}$ (see Figure 3D). This remodeling step creates an interaction interface between the C-

terminal α -helix and the GGG linkers. Furthermore, the remodeling allows for head-to-tail cyclization of the scaffold if needed since the termini are now in spatial proximity.

β -Helix Sequence Design and Structure Optimization.

Starting from the designed β -helix structure with C-terminal α -helix discussed above, the Rosetta protein design software^{46,47} was used to further optimize the sequence and overall structure of the modeled tau β -helix. PyRosetta⁶³ (Linux release 311 64 bit) was used for all Rosetta designs. The FastDesign protocol was used to pack all residues and to apply sequence design to (only) the four linkers and the C-terminal α -helix residues, using the REF15 scoring function.⁶⁴ All 20 naturally occurring amino acid residues were allowed in the designed residue positions. A MoveMapFactory was set up to allow the minimization of all bond lengths, bond angles, and torsion angles. The design protocol was repeated 10,000 times to generate 10,000 decoys of the β -helix structure. The Python code for structure remodeling, sequence design, and structure optimization using Rosetta can be found on github (see Methods: **Code Availability**).

Structure Prediction with AlphaFold. DeepMind's AlphaFold (version 2), a deep learning structure prediction algorithm that showed the best performance at CASP14 by a significant margin^{45,65} (<https://github.com/deepmind/alphafold>), was installed locally and trained, searching through the full database current as of Jan 01, 2020, and the model obtained was used to predict the structures of selected Rosetta-designed β -helix scaffold decoys from their primary sequence alone.

Mechanical Unfolding Simulations. To determine the mechanical stability of the tau β -helix scaffold, unfolding simulations were performed. Ten independent forced unfolding simulations were performed along the reaction coordinate Q (the fraction of native contacts) over the course of 100 ns each, using PLUMED,⁶⁶ OpenMM⁶⁷ simulation engine, and the CHARMM36m force field.⁴⁹ A native contact was defined to be between heavy atoms within 0.45 nm , belonging to residues α and β with a sequence separation $|\alpha - \beta| > 3$. The reaction coordinate Q of an arbitrary conformation X is defined as follows⁶⁸

$$Q(X) = \frac{1}{N} \sum_{(i,j)} \frac{1}{1 + \exp[\beta(r_{ij}(X) - \lambda r_{ij}^0)]} \quad (6)$$

where the sum runs over N pairs of native contacts (i, j) , $r_{ij}(X)$ is the distance between atoms i and j in conformation X , and r_{ij}^0 is the native distance between atoms i and j . The parameters β and λ were taken to be 50 nm^{-1} and 1.5 , respectively. A time-dependent harmonic potential $V(Q, t)$ (eq 3) was applied to move the center of the bias linearly from $Q_0 = 1$ to $Q_0 = 0$ over the simulation time. The spring constant k in eq 3 was taken to be 10^7 kJ mol^{-1} .

Hamiltonian Replica Exchange Molecular Dynamics. Hamiltonian replica exchange molecular dynamics (HREMD) simulations^{69,70} were performed to determine the free energy surface (i.e., the potential of mean force (PMF)) along the fraction of native contacts Q (eq 6). The setup consists of 100 replicas with time-independent harmonic restraints centered at positions equally spaced between $Q = 1$ and $Q = 0$. The strength of the harmonic restraint k was taken to be 10^5 kJ mol^{-1} for all replicas for the production run (80 ns for each replica). This value of harmonic restraint corresponds to an RMS value of $\Delta Q = 0.009$ (this was about twice the $\sqrt{k_B T/k} = 0.005$ value expected from equipartition theorem). Before the production run, the protein must be unfolded and equilibrated in each harmonic potential restraint. The unfolding run is performed in 10 ns for each replica, and the system in each of the 100 harmonic potentials is equilibrated for 160 ns in the equilibration run.

The initial configurations before unfolding in each of the replicas were obtained by seeding all 100 replicas with the Rosetta-designed native structure and applying the time-dependent harmonic restraints with corresponding center linearly decreased from $Q = 1$ to $Q = Q_{0i}$ during the unfolding run. Here, Q_{0i} , $0 < i < 99$, is the center of the harmonic restraint used for each replica during the equilibration and production runs. Q_{0i} is different for each of the replicas, increasing linearly from 0.01 to 1 (see Figure 8). Because the interactions

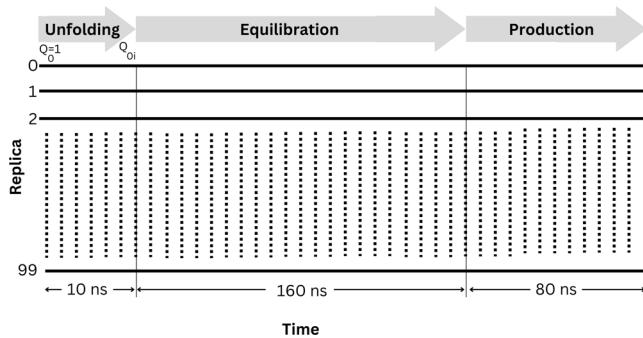


Figure 8. Setup used for Hamiltonian replica exchange molecular dynamics (HREMD) simulations. The HREMD setup consists of 100 replicas (horizontal lines), each simulated in an unfolding run to different degrees of unfolding Q_{0i} for 10 ns, followed by an equilibration run for 160 ns with each replica in its own harmonic potential centered at Q_{0i} , followed by a production run in the same harmonic potential for 80 ns.

stabilizing the protein are stronger at higher nativeness, in practice, we linearly decrease the strength of the harmonic restraints during the unfolding run as a function of time, from a spring constant of $k = 10^7$ to $k = 10^5$ kJ mol $^{-1}$. The spring constant then remains fixed at $k = 10^5$ kJ mol $^{-1}$ during the equilibration and production runs. Exchanges between neighboring replicas were attempted every 1 ps (500 steps). The simulation, including the initial and the production runs, ran for 250 ns per replica for a total of 25 μ s for all replicas. Conformations were sampled at 10 ps intervals from each replica in the production run and used to determine the unfolding free energy profile along Q by solving the multistate Bennett acceptance ratio (MBAR) equations.^{51,72}

Sampling Conformational Ensembles. All-atom molecular dynamics (MD) simulations were performed in explicit solvent (TIP3P water model, 150 mM NaCl) at a temperature of 300 K and pressure of 1 bar using GROMACS⁴⁸ and CHARMM36m force field⁴⁹ to generate ensembles for tau fibril, tau oligomer, tau monomer, cyclic peptide scaffolds, and β -helix scaffold.

Fibril Ensemble. The fibril ensemble was generated by performing 30.1 ns MD simulation starting from a tau fibril structure (PDB 5O3L). Conformations were sampled at equal time intervals of 1 ns for a total of $30.1 \times 10 \times 10 = 3010$ single-chain conformations (there are 10 chains in the system) in the fibril ensemble.

Oligomer Ensemble. The oligomer was modeled as a partially unfolded fibril. Simulations were performed as described in the **Prediction of Epitopes** section. Conformations were sampled at equal time intervals of 1 ns from the last ≈ 42 ns of 10 independent biased simulations of 10 chains, for a total of $42 \times 10 \times 10 = 4200$ single-chain conformations in the oligomer ensemble.

Monomer Ensemble. The free energy landscape of an intrinsically disordered protein (IDP) such as tau is rugged and weakly funneled. Therefore, it is very challenging to generate an equilibrium ensemble for an IDP through direct MD simulation. The strategy that we have used previously in refs 28 and 73 was therefore employed. In brief, two steps are involved. First, a highly diverse conformational ensemble with 10,000 configurations was generated by using the pivot algorithm,^{74,75} along with crankshaft moves⁷⁶ that randomize torsion angles between two randomly selected fixed backbone atoms. Second, each feasible conformation generated was energy minimized, and then, a 3 ns MD simulation was performed. The last conformation from each simulation was collected to obtain a total of 7166 conformations in the monomer ensemble.

Cyclic Peptide Ensembles. Each cyclic peptide scaffold was simulated for 300 ns using conventional MD. Conformations were then sampled at equal time intervals of 100 ps from the last 250 ns of the simulation for a total of 2500 conformations in each cyclic peptide scaffold ensemble.

β -Helix Scaffold Ensembles. For the β -helix peptide ensembles, we performed 100-replica HREMD simulations for a total of 25 μ s (see Methods: **Hamiltonian Replica Exchange Molecular Dynamics**). A number N_i of conformations determined from eq 7 were sampled randomly from the i th replica so that replicas were Boltzmann-weighted

$$N_i = 1250 \times e^{-\Delta F_i} \quad (7)$$

Here, ΔF_i is the relative free energy (in $k_B T$) of the i th replica (see Figure 4D). A total of 2393 conformations were sampled for each of the four instances of KLDFK motif in the β -helix scaffold.

Comparing Conformational Ensembles. Conformational ensembles are compared using Jensen–Shannon divergence (JSD)^{77,78} and embedding depth (\mathcal{D})²⁸ measures. JSD was implemented using the ENCORE software.⁷⁹ JSD is a symmetrized version of the Kullback–Leibler divergence,⁸⁰ D_{KL} , which is a measure of the difference between two probability distributions P and Q , defined by

$$JSD(P, Q) = \frac{1}{2}D_{KL}(P, M) + \frac{1}{2}D_{KL}(Q, M) \quad (8)$$

where

$$M = \frac{P + Q}{2} \quad (9)$$

and

$$D_{KL}(P, M) = \int_{-\infty}^{\infty} P(x) \log \left(\frac{P(x)}{M(x)} \right) dx \quad (10)$$

The probability distribution representing an ensemble is obtained by creating a root mean squared deviation (RMSD) matrix of all structures in all ensembles to be compared. The dimension of this high-dimensional RMSD matrix is then reduced using the stochastic proximity embedding (SPE) method.⁸¹ Since the value of the JSD depends on the dimension of the SPE-reduced space, weighted average JSD values from three-dimensional (3D) to 11D are used, with weights taken as the inverse of the SPE residuals in each dimension (this weights higher dimensions more strongly). The maximum value of JSD in eq 8 is $\log 2$. JSD was then normalized to lie between 0 (for identical ensembles) to 1 (for entirely different ensembles) by dividing by $\log 2$.

The embedding depth $\mathcal{D}_{Q|P}$ or \mathcal{D}_{Q-in-P} , when used to compare two conformational ensembles represented by probability distributions P and Q , quantifies the extent to which the ensemble Q is subsumed by ensemble P . For a point $x = x_o$ in a distribution $P(x)$, the embedding depth $\mathcal{D}_{\delta(x-x_o)|P}$ (eq 11) is defined as the fraction of the distribution that has a lower probability than point $x = x_o$.²⁸

$$\mathcal{D}_{\delta(x-x_o)|P} = \int_{-\infty}^{\infty} dx P(x) \Theta(P(x_o) - P(x)) \quad (11)$$

In eq 11, $\Theta(P)$ is the Heaviside step function, which returns 1 if $P > 0$, otherwise 0.

The embedding depth of one distribution $Q(x)$ within another $P(x)$ can be found by integrating eq 11 over the distribution $Q(x)$

$$\mathcal{D}_{Q|P} = \int_{-\infty}^{\infty} dx Q(x) \int_{-\infty}^{\infty} dx' P(x') \Theta(P(x) - P(x')) \quad (12)$$

The probability distribution representing an ensemble for embedding depth calculation is also obtained by creating a RMSD matrix of all structures in all ensembles to be compared. The dimension of this high-dimensional RMSD matrix is however reduced using the multidimensional scaling (MDS) method.^{82–84} Since the value of \mathcal{D} also depends on the dimension of the MDS-reduced space, a weighted average of \mathcal{D} values from 3D to 11D is used, with weights taken as the inverse of the MDS residuals in each dimension (similar to JSD above, this applies more weight to higher dimensions). The embedding depth between two distributions is in general non-reciprocal, in that $\mathcal{D}_{Q|P} \neq \mathcal{D}_{P|Q}$. The mode of a distribution has the

maximum value of $\mathcal{D} = 1$ while two entirely different ensembles have the minimum value of $\mathcal{D} = 0$. The embedding depth of a distribution within itself is $\mathcal{D}_{\text{PIP}} = 1/2$.²⁸

Epitope Robustness Analysis. The SASA (averaged across chains) is plotted as a function of residue position, for each of the five tau fibril structures in Figure S7. A rolling average window of 5 amino acids was applied. The pairwise local distance test (lLdt)³⁸ was used to compare PDB fibril structures to determine their similarity (Figure S7G).

Protein Cleavage Analysis. Cleavage analysis was performed by the Procleave server,³⁵ with default parameters on all proteases. The proteolytic propensity of the designed β -helix scaffold structure was compared to ubiquitin (PDB 1UBQ⁸⁵) as a negative control. Plots were made of the top 10 cleavage site scores for 27 specific proteases (Figure S8).

Aggregation Propensity Analysis. Aggregation propensity analysis was performed using the AGGRESCAN3D server³⁶ with default parameters. The designed β -helix scaffold structure was compared to ubiquitin (PDB 1UBQ⁸⁵) as a negative control and transthyretin (PDB 1TFP⁸⁶) as a positive control for a known aggregation-prone protein.⁵⁷ Plots of the aggregation score vs. residue index, along with the mean aggregation score and the fraction of residues that have an aggregation score larger than zero, are shown in Figure S9.

Immunogenicity Analysis. A predicted immunogenicity score was obtained from the Epitopia server.³⁹ Figure S10 plots the immunogenicity for each residue on a scale from 1 to 5, for the tau fibril structure (PDB 5O3L). The number plotted in Figure S10 is averaged over the 10 chains in the fibril structure.

Design Method for the Cyclic CPPPKLDFKGPGG Scaffold.

The centroid structures of the biased fibril ensemble, as well as the monomer ensemble of the KLDFK segment, were both identified. Both centroid structures were scaffolded using a (4, 4) glycindel scaffold, consisting of four glycines on both the C- and N-terminus. To identify Gly → Pro mutations that would reinforce conformational similarity to the centroid structure of the biased fibril while avoiding similarity to the centroid structure of the monomer ensemble, the CRISPro software⁸⁷ was employed. This analysis led to the mutations at five specific proline mutation sites, modifying the sequence from CGGGGKLDFKGGGG to CPPPKLDFKGPGG. A cyclic peptide with this latter sequence was used in immunizations to obtain the monoclonal antibodies described in the Methods: **Surface Plasmon Resonance (SPR) Experiments** section.

Peptide Synthesis. Peptide synthesis was performed by LifeTein, LLC. (Hillsborough, New Jersey) following standard manufacturing procedures. Mass spectrometry QC indicated the correct molecular weight to within a protonation state and was consistent with disulfide bond formation (Theoretical value 11336.99, Observed on MS at 11336.05). HPLC indicated a purity value at 85.04% by peak area. Designed constructs were conjugated to keyhole limpet hemocyanin (KLH).

Surface Plasmon Resonance (SPR) Experiments. Several monoclonal antibodies have been obtained previously from mice immunized with the Tau cyclic peptide sequence CPPPKLDFKGPGG. These include antibodies 4E4, 3E2, 10B10, 6H3, 7H6, 10D4, 7E5, and 2C6 (see Figure 5). An isotype control (mIgG1) was also used for comparison.

Stable tau oligomers (consisting primarily of trimers and pentamers) were generated from recombinant full-length human tau monomers (wild-type tau—441 amino acids, 2N4R, >97% purity) and were obtained from SynAging SAS (Vandœuvre-lès-Nancy, France).

Tau β -helix construct reduction was performed by incubating 0.8 mg mL⁻¹ (90 μ M) sample with 10 mM DTT (90× molar excess) and 2 M EDTA for 30 min at room temperature.

Approximately 400 RUs of oxidized and reduced β -helix construct were immobilized for SPR binding studies. Antibodies were diluted to 150 μ g mL⁻¹ and injected over immobilized β -helix constructs or oligomers.

Code Availability. A Python script for remodeling and design of the tau beta-helix protein scaffold is available at <https://github.com/PlotkinLab/Beta-Helix-Design>.

CONCLUSIONS

In this paper, we have presented a novel approach for scaffolding epitopes in order to target toxic forms of proteins implicated in protein-misfolding diseases. This approach entails the design of scaffolds that connect several chains in an aggregated structure by designed linkers to incorporate multiple repeats of a target epitope. This construction is intended to guide the conformational ensemble explored by the epitope toward those conformations present in aggregated conformations, which we hypothesize would present the epitope in disease-associated forms of the underlying protein. The prediction of what epitopes to scaffold in this manner would depend on the application; in our context, the epitopes were predicted to be those on the accessible surface of an oligomer.

The method was applied to tau, a microtuble-binding protein whose loss of function and tangle formation in neurons is associated with many neurodegenerative diseases, including Alzheimer's disease and other tauopathies. First, oligomer-selective tau epitopes were predicted by stressing the fibril to induce a partially disordered fibril, which serves as a predictive model for epitopes that are likely to be exposed on the surface of oligomers.

Four copies of the predicted tau ³⁴³KLDFK³⁴⁷ epitope were incorporated in the final design of a β -helix scaffold. The structure of the designed scaffold was validated with AlphaFold, and it is gratifying to see that the Rosetta-designed structure and AlphaFold-predicted structure are nearly identical, with a C_{α} RMS deviation of 1.39 Å. In silico characterization of the designed scaffold shows that it is predicted to be thermodynamically stable, which is an important property of a molecule intended for use as an immunogen. Furthermore, we showed that for this epitope in tau protein, the multi-epitope scaffolding approach is predicted to be better in discriminating models of abnormal forms of tau from isolated tau monomer, than single-epitope scaffolding approaches, such as in cyclic peptide "glycindel" scaffolds. The multi-epitope scaffolding approach introduced here can straightforwardly be applied to proteins implicated in other protein aggregation-related diseases, such as tau protein in chronic traumatic encephalopathy (CTE), α -synuclein in Parkinson's disease, SOD1 in amyotrophic lateral sclerosis (ALS), TDP43 in frontotemporal dementia (FTD), and Huntingtin in Huntington's disease.

The experimental characterization included here showed that immunological antibodies raised to a cyclic peptide of the KLDFK epitope could bind to the β -helical construct, that their binding is weakened by the loss of structure of this construct, and interestingly, that their strength of binding to the β -helical construct is correlated with their strength of binding to *in vitro* oligomers of tau. This supports the notion that the β -helix construct may present an oligomer-selective conformational epitope and may thus be useful as a potential immunogen. It is currently unknown how effective such an immunogen would be in the human immune system to raise an oligomer-selective antibody response.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acschemneuro.3c00007>.

Snapshots of tau fibrils (Figure S1); “Fireplots” of the change in SASA upon stressing the fibril (Figure S2); “Fireplots” of the loss of native contacts upon stressing the fibril (Figure S3); “Fireplots” of the change in RMSF upon stressing the fibril (Figure S4); Rosetta energy and AlphaFold confidence for designed β -helix structures (Figure S5); instability of 4-repeat β -helix scaffold (Figure S6); commonly high SASA of KLDFK motif among tau fibrils (Figure S7); protease analysis of β -helix scaffold (Figure S8); aggregation propensity of β -helix scaffold (Figure S9); immunogenicity propensity of β -helix scaffold (Figure S10); secondary structure evolution of β -helix scaffold (Figure S11); models of tau β -helix scaffold (Table S1) ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

Steven S. Plotkin — Department of Physics and Astronomy, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada; Genome Science and Technology Program, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada;
 orcid.org/0000-0001-8998-877X; Email: steve@phas.ubc.ca

Authors

Adekunle Aina — Department of Physics and Astronomy, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada;  orcid.org/0000-0002-8215-7452

Shawn C. C. Hsueh — Department of Physics and Astronomy, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada;  orcid.org/0000-0001-5348-7877

Ebrima Gibbs — Djavad Mowafaghian Centre for Brain Health, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

Xubiao Peng — Department of Physics and Astronomy, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada; Present Address: Center for Quantum Technology Research, School of Physics, Beijing Institute of Technology, Beijing 100081, China;  orcid.org/0000-0002-6494-9560

Neil R. Cashman — Djavad Mowafaghian Centre for Brain Health, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acschemneuro.3c00007>

Author Contributions

A.A. and S.S.P. contributed to conceptualization; A.A. and S.S.P. contributed to methodology; A.A. contributed to coding— β -helix design; A.A., S.C.C.H., X.P., and S.S.P. contributed to software; A.A., S.C.C.H., X.P., and S.S.P. contributed to data analysis; A.A., S.C.C.H., X.P., and S.S.P. contributed to visualization; E.G., N.R.C., and S.S.P. contributed to experiment—design and method; E.G., N.R.C., and S.S.P. contributed to experiment—data analysis; A.A. contributed to writing—original draft; S.S.P., A.A., and

S.C.C.H. contributed to writing—revision, review, and editing; S.S.P. and N.R.C. contributed to supervision; S.S.P., A.A., and N.R.C. contributed to funding acquisition.

Funding

This research was supported by the Canadian Institutes of Health Research Transitional Operating Grant 2682, and by Alberta Innovates Research Team Program Grant PTM13007, Compute Canada Resources for Research Groups RRG 3071, and UBC ARC Sockeye Advanced Research Computing (<https://doi.org/10.14288/SOCKEYE>, 2019). A.A. has received funding from a UBC Four-Year Fellowship and a MITACS Accelerate Scholarship.

Notes

The authors declare the following competing financial interest(s): S.S.P. was the Chief Physics Officer of ProMIS Neurosciences until October 2020 and continues to consult for ProMIS Neurosciences. N.R.C. is the Chief Scientific Officer of ProMIS Neurosciences. A.A. and S.S.P. have filed US provisional patent application 63/435,074 on discoveries described in this manuscript. S.S.P. and N.R.C. are inventors on patent applications PCT/CA2016/051306 (Publication US 2018/0330045, describing methods and systems for predicting misfolded protein epitopes by collective coordinate biasing) and PCT/CA2020/050722 (Publication US 2022/0218805, describing immunogens, antibodies, and methods of their making as well as their use); applicant being University of British Columbia, and licensee ProMIS Neurosciences. The early stages of the work presented here were financially supported in part by ProMIS Neurosciences. A.A., S.C.C.H., E.G., X.P., N.R.C., and S.S.P. have received consultation compensation from ProMIS Neurosciences.

ACKNOWLEDGMENTS

The authors thank Santanu Sasidharan for his help in rendering the experimental data.

REFERENCES

- (1) Dujardin, S.; Bégard, S.; Caillierez, R.; Lachaud, C.; Carrier, S.; Lieger, S.; Gonzalez, J. A.; Deramecourt, V.; Déglon, N.; Maurage, C.-A.; Frosch, M. P.; Hyman, B. T.; Colin, M.; Buée, L. Different tau species lead to heterogeneous tau pathology propagation and misfolding. *Acta Neuropathol. Commun.* **2018**, *6*, 1–12.
- (2) Iqbal, K.; Liu, F.; Gong, C.-X.; Grundke-Iqbali, I. Tau in Alzheimer Disease and Related Tauopathies. *Curr. Alzheimer Research* **2010**, *7*, 656–664.
- (3) Weingarten, M. D.; Lockwood, A. H.; Hwo, S. Y.; Kirschner, M. W. A protein factor essential for microtubule assembly. *Proc. Natl. Acad. Sci.* **1975**, *72*, 1858–1862.
- (4) Klein, C.; Krämer, E.-M.; Cardine, A.-M.; Schraven, B.; Brandt, R.; Trotter, J. Process Outgrowth of Oligodendrocytes Is Promoted by Interaction of Fyn Kinase with the Cytoskeletal Protein Tau. *J. Neurosci.* **2002**, *22*, 698–707.
- (5) Wischik, C. M.; Novak, M.; Thøgersen, H. C.; Edwards, P. C.; Runswick, M. J.; Jakes, R.; Walker, J. E.; Milstein, C.; Roth, M.; Klug, A. Isolation of a fragment of tau derived from the core of the paired helical filament of Alzheimer disease. *Proc. Natl. Acad. Sci.* **1988**, *85*, 4506–4510.
- (6) Plotkin, S. S.; Cashman, N. R. Passive immunotherapies targeting $\text{A}\beta$ and tau in Alzheimer’s disease. *Neurobiol. Dis.* **2020**, *144*, 1–26.
- (7) Zilka, N.; Filipcik, P.; Koson, P.; Fialova, L.; Skrabana, R.; Zilkova, M.; Rolkova, G.; Kontsekova, E.; Novak, M. Truncated tau from sporadic Alzheimer’s disease suffices to drive neurofibrillary degeneration in vivo. *FEBS Lett.* **2006**, *580*, 3582–3588.

- (8) Gómez-Isla, T.; Hollister, R.; West, H.; Mui, S.; Growdon, J. H.; Petersen, R. C.; Parisi, J. E.; Hyman, B. T. Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease. *Ann. Neurol.* **1997**, *41*, 17–24.
- (9) Tai, H.-C.; Serrano-Pozo, A.; Hashimoto, T.; Frosch, M. P.; Spires-Jones, T. L.; Hyman, B. T. The Synaptic Accumulation of Hyperphosphorylated Tau Oligomers in Alzheimer Disease Is Associated With Dysfunction of the Ubiquitin-Proteasome System. *Am. J. Pathol.* **2012**, *181*, 1426–1435.
- (10) Novak, P.; Kontsekova, E.; Zilka, N.; Novak, M. Ten Years of Tau-Targeted Immunotherapy: The Path Walked and the Roads Ahead. *Front. Neurosci.* **2018**, *12*, No. 798.
- (11) Zilka, N.; Kazmerova, Z.; Jadhav, S.; Neradil, P.; Madari, A.; Obetkova, D.; Bugos, O.; Novak, M. Who fans the flames of Alzheimer's disease brains? Misfolded tau on the crossroad of neurodegenerative and inflammatory pathways. *J. Neuroinflammation* **2012**, *9*, 1–9.
- (12) Arendt, T.; Stieler, J. T.; Holzer, M. Tau and tauopathies. *Brain Res. Bull.* **2016**, *126*, 238–292.
- (13) Nelson, P. T.; Alafuzoff, I.; Bigio, E. H.; et al. Correlation of Alzheimer Disease Neuropathologic Changes With Cognitive Status: A Review of the Literature. *J. Neuropathol. Exp. Neurol.* **2012**, *71*, 362–381.
- (14) Murray, M. E.; Lowe, V. J.; Graff-Radford, N. R.; Liesinger, A. M.; Cannon, A.; Przybelski, S. A.; Rawal, B.; Parisi, J. E.; Petersen, R. C.; Kantarci, K.; Ross, O. A.; Duara, R.; Knopman, D. S.; Jack, C. R.; Dickson, D. W. Clinicopathologic and ¹¹C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. *Brain* **2015**, *138*, 1370–1381.
- (15) Josephs, K. A.; Petersen, R. C.; Knopman, D. S.; Boeve, B. F.; Whitwell, J. L.; Duffy, J. R.; Parisi, J. E.; Dickson, D. W. Clinicopathologic analysis of frontotemporal and corticobasal degenerations and PSP. *Neurology* **2006**, *66*, 41–48.
- (16) Chen, X.; Chen, M.; Schafer, N. P.; Wolynes, P. G. Exploring the interplay between fibrillization and amorphous aggregation channels on the energy landscapes of tau repeat isoforms. *Proc. Natl. Acad. Sci.* **2020**, *117*, 4125–4130.
- (17) Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017**, *547*, 185–190.
- (18) Falcon, B.; Zhang, W.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Ghetti, B.; Scheres, S. H. W.; Goedert, M. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **2018**, *561*, 137–140.
- (19) Shi, Y.; Zhang, W.; Yang, Y.; et al. Structure-based classification of tauopathies. *Nature* **2021**, *598*, 359–363.
- (20) Hoskin, J. L.; Sabbagh, M. N.; Al-Hasan, Y.; Decourt, B. Tau immunotherapies for Alzheimer's disease. *Expert Opin. Invest. Drugs* **2019**, *28*, 545–554.
- (21) Malonis, R. J.; Lai, J. R.; Vergnolle, O. Peptide-Based Vaccines: Current Progress and Future Challenges. *Chem. Rev.* **2020**, *120*, 3210–3229.
- (22) Mantile, F.; Prisco, A. Vaccination against β -Amyloid as a Strategy for the Prevention of Alzheimer's Disease. *Biology* **2020**, *9*, 1–21.
- (23) Kwon, S.; Iba, M.; Kim, C.; Masliah, E. Immunotherapies for Aging-Related Neurodegenerative Diseases—Emerging Perspectives and New Targets. *Neurotherapeutics* **2020**, *17*, 935–954.
- (24) Apostolopoulos, V.; Bojarska, J.; Chai, T. T.; et al. A Global Review on Short Peptides: Frontiers and Perspectives. *Molecules* **2021**, *26*, No. 430.
- (25) Kontsekova, E.; Zilka, N.; Kovacech, B.; Novak, P.; Novak, M. First-in-man tau vaccine targeting structural determinants essential for pathological tau–tau interaction reduces tau oligomerisation and neurofibrillary degeneration in an Alzheimer's disease model. *Alzheimer's Res. Ther.* **2014**, *6*, 1–12.
- (26) Skwarczynski, M.; Toth, I. Peptide-based synthetic vaccines. *Chem. Sci.* **2016**, *7*, 842–854.
- (27) Sanchez-Trincado, J. L.; Gomez-Perosanz, M.; Reche, P. A. Fundamentals and Methods for T- and B-Cell Epitope Prediction. *J. Immunol. Res.* **2017**, *2017*, 1–14.
- (28) Hsueh, S. C. C.; Aina, A.; Roman, A. Y.; Cashman, N. R.; Peng, X.; Plotkin, S. S. Optimizing Epitope Conformational Ensembles Using α -Synuclein Cyclic Peptide "Glycindel" Scaffolds: A Customized Immunogen Method for Generating Oligomer-Selective Antibodies for Parkinson's Disease. *ACS Chem. Neurosci.* **2022**, *13*, 2261–2280.
- (29) Sevigny, J.; et al. The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature* **2016**, *537*, 50–56.
- (30) Tolar, M.; Abushakra, S.; Hey, J. A.; Porsteinsson, A.; Sabbagh, M. Aducanumab, gantenerumab, BAN2401, and ALZ-801—the first wave of amyloid-targeting drugs for Alzheimer's disease with potential for near term approval. *Alzheimer's Res. Ther.* **2020**, *12*, 1–10.
- (31) Knopman, D. S.; Jones, D. T.; Greicius, M. D. Failure to demonstrate efficacy of aducanumab: An analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019. *Alzheimer's Dementia* **2021**, *17*, 696–701.
- (32) Mullard, A. Controversial Alzheimer's drug approval could affect other diseases. *Nature* **2021**, *595*, 162–163.
- (33) Peng, X.; Cashman, N. R.; Plotkin, S. S. Prediction of Misfolding-Specific Epitopes in SOD1 Using Collective Coordinates. *J. Phys. Chem. B* **2018**, *122*, 11662–11676.
- (34) Gibbs, E.; Zhao, B.; Roman, A.; Plotkin, S. S.; Peng, X.; Hsueh, S. C. C.; Aina, A.; Wang, J.; Shyu, C.; Yip, C. K.; Nam, S.-E.; Kaplan, J. M.; Cashman, N. R. Rational Generation of Monoclonal Antibodies Selective for Pathogenic Forms of Alpha-Synuclein. *Biomedicines* **2022**, *10*, No. 2168.
- (35) Silverman, J. M.; Gibbs, E.; Peng, X.; et al. A Rational Structured Epitope Defines a Distinct Subclass of Toxic Amyloid-beta Oligomers. *ACS Chem. Neurosci.* **2018**, *9*, 1591–1606.
- (36) Gibbs, E.; Silverman, J. M.; Zhao, B.; Peng, X.; Wang, J.; Wellington, C. L.; Mackenzie, I. R.; Plotkin, S. S.; Kaplan, J. M.; Cashman, N. R. A Rationally Designed Humanized Antibody Selective for Amyloid Beta Oligomers in Alzheimer's Disease. *Sci. Rep.* **2019**, *9*, No. 9870.
- (37) Zhang, W.; Tarutani, A.; Newell, K. L.; Murzin, A. G.; Matsubara, T.; Falcon, B.; Vidal, R.; Garringer, H. J.; Shi, Y.; Ikeuchi, T.; Murayama, S.; Ghetti, B.; Hasegawa, M.; Goedert, M.; Scheres, S. H. W. Novel tau filament fold in corticobasal degeneration. *Nature* **2020**, *580*, 283–287.
- (38) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722–2728.
- (39) Rubinstein, N. D.; Mayrose, I.; Martz, E.; Pupko, T. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinf.* **2009**, *10*, 287.
- (40) McAlary, L.; Plotkin, S. S.; Yerbury, J. J.; Cashman, N. R. Prion-Like Propagation of Protein Misfolding and Aggregation in Amyotrophic Lateral Sclerosis. *Front. Mol. Neurosci.* **2019**, *12*, No. 262.
- (41) Ji, C.; Sigurdsson, E. M. Current Status of Clinical Trials on Tau Immunotherapies. *Drugs* **2021**, *81*, 1135–1152.
- (42) Jicha, G. A.; Bowser, R.; Kazam, I. G.; Davies, P. Alz-50 and MC-1, a new monoclonal antibody raised to paired helical filaments, recognize conformational epitopes on recombinant tau. *J. Neurosci. Res.* **1997**, *48*, 128–132.
- (43) Liou, Y.-C.; Tocilj, A.; Davies, P. L.; Jia, Z. Mimicry of ice structure by surface hydroxyls and water of a β -helix antifreeze protein. *Nature* **2000**, *406*, 322–324.
- (44) Leinala, E. K.; Davies, P. L.; Jia, Z. Crystal Structure of β -Helical Antifreeze Protein Points to a General Ice Binding Model. *Structure* **2002**, *10*, 619–627.
- (45) Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (46) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302*, 1364–1368.

- (47) Leaver-Fay, A. et al. *Methods in Enzymology*; Academic Press, 2011; Chapter 19, Vol. 487, pp 545–574.
- (48) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (49) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (50) Clarke, J.; Hamill, S. J.; Johnson, C. M. Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **1997**, *270*, 771–778.
- (51) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins 1 Edited by P. E. Wright. *J. Mol. Biol.* **1998**, *277*, 985–994.
- (52) Das, A.; Plotkin, S. S. SOD1 exhibits allosteric frustration to facilitate metal binding affinity. *Proc. Natl. Acad. Sci.* **2013**, *110*, 3871–3876.
- (53) Seifi, B.; Aina, A.; Wallin, S. Structural fluctuations and mechanical stabilities of the metamorphic protein RfaH. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 289–300.
- (54) Habibi, M.; Rottler, J.; Plotkin, S. S. As Simple As Possible, but Not Simpler: Exploring the Fidelity of Coarse-Grained Protein Models for Simulated Force Spectroscopy. *PLOS Comput. Biol.* **2016**, *12*, No. e1005211.
- (55) Li, F.; Leier, A.; Liu, Q.; Wang, Y.; Xiang, D.; Akutsu, T.; Webb, G. I.; Smith, A. I.; Marquez-Lago, T.; Li, J.; Song, J. Procleave: Predicting Protease-specific Substrate Cleavage Sites by Combining Sequence and Structural Information. *Genomics, Proteomics Bioinf.* **2020**, *18*, 52–64.
- (56) Zambrano, R.; Jamroz, M.; Szczasiuk, A.; Pujols, J.; Kmiecik, S.; Ventura, S. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* **2015**, *43*, W306–W313.
- (57) Johnson, S. M.; Wiseman, R. L.; Sekijima, Y.; Green, N. S.; Adamski-Werner, S. L.; Kelly, J. W. Native State Kinetic Stabilization as a Strategy To Ameliorate Protein Misfolding Diseases: A Focus on the Transthyretin Amyloidoses. *Acc. Chem. Res.* **2005**, *38*, 911–921.
- (58) Plotkin, S. S.; Cashman, N. R.; Kaplan, J. Conformation-Specific Epitopes in Tau, Antibodies Thereto and Methods Related Thereof. United States provisional application serial number 62/853,121 (2020).
- (59) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 778–795.
- (60) He, Z.; Zhang, C.; Xu, Y.; Zeng, S.; Zhang, J.; Xu, D. MUFOOLD-DB: a processed protein structure database for protein structure prediction and analysis. *BMC Genomics* **2014**, *15*, No. S2.
- (61) Imai, K.; Mitaku, S. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics* **2005**, *1*, 55–65.
- (62) Huang, P.-S.; Ban, Y.-E. A.; Richter, F.; Andre, I.; Vernon, R.; Schief, W. R.; Baker, D. RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS One* **2011**, *6*, No. e24109.
- (63) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010**, *26*, 689–691.
- (64) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048.
- (65) Pereira, J.; Simpkin, A. J.; Hartmann, M. D.; Rigden, D. J.; Keegan, R. M.; Lupas, A. N. High-accuracy protein structure prediction in CASP14. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 1687–1699.
- (66) The PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670–673.
- (67) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **2017**, *13*, No. e1005659.
- (68) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci.* **2013**, *110*, 17874–17879.
- (69) Jang, S.; Shin, S.; Pak, Y. Replica-Exchange Method Using the Generalized Effective Potential. *Phys. Rev. Lett.* **2003**, *91*, No. 058305.
- (70) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2*, 217–228.
- (71) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, No. 124105.
- (72) Ding, X.; Vilseck, J. Z.; Brooks, C. L. Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. *J. Chem. Theory Comput.* **2019**, *15*, 799–802.
- (73) Das, A.; Sin, B. K.; Mohazab, A. R.; Plotkin, S. S. Unfolded protein ensembles, folding trajectories, and refolding rate prediction. *J. Chem. Phys.* **2013**, *139*, No. 09B625_1.
- (74) Madras, N.; Sokal, A. D. The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *J. Stat. Phys.* **1988**, *50*, 109–186.
- (75) Hadizadeh, S.; Linhananta, A.; Plotkin, S. S. Improved measures for the shape of a disordered polymer to test a mean-field theory of collapse. *Macromolecules* **2011**, *44*, 6182–6197.
- (76) Favrin, G.; Irbäck, A.; Sjunnesson, F. Monte Carlo update for chain molecules: biased Gaussian steps in torsional space. *J. Chem. Phys.* **2001**, *114*, 8154–8158.
- (77) Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151.
- (78) Lindorff-Larsen, K.; Ferkinghoff-Borg, J. Similarity Measures for Protein Ensembles. *PLoS One* **2009**, *4*, No. e4203.
- (79) Tiberti, M.; Papaleo, E.; Bengtsen, T.; Boomsma, W.; Lindorff-Larsen, K. ENCORE: Software for Quantitative Ensemble Comparison. *PLOS Comput. Biol.* **2015**, *11*, No. e1004415.
- (80) Kullback, S.; Leibler, R. A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
- (81) Agrafiotis, D. K. Stochastic proximity embedding. *J. Comput. Chem.* **2003**, *24*, 1215–1221.
- (82) Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **1964**, *29*, 1–27.
- (83) Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **1964**, *29*, 115–129.
- (84) Borg, I.; Groenen, P. Modern Multidimensional Scaling: Theory and Applications. *J. Educ. Meas.* **2003**, *40*, 277–280.
- (85) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (86) Sunde, M.; Richardson, S. J.; Chang, L.; Pettersson, T. M.; Schreiber, G.; Blake, C. C. F. The Crystal Structure of Transthyretin from Chicken. *Eur. J. Biochem.* **1996**, *236*, 491–499.
- (87) Rawi, R.; Shen, C.-H.; Kwong, P. D.; Chuang, G.-Y. CRISPro: An Automated Pipeline for Protein Conformation Stabilization by Proline. *J. Chem. Inf. Model.* **2018**, *58*, 2189–2192.