

De Novo Design of Peptides that Co-assemble into β -sheet Based Nanofibrils

(Short title --- Hierarchical assembly of peptides into nanofibers)

(125-character teaser --- Computational discovery and experimental characterization of co-assembled peptides that form β -sheet-based nanofibrils)

Xingqing Xiao¹, Yiming Wang¹, Dillon T. Seroski², Kong M. Wong³, Renjie Liu², Anant K. Paravastu³, Gregory A. Hudalla², Carol K. Hall^{1*}

1. Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695-7905, United States
2. Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, United States
3. Department of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States

Corresponding authors

E-mail: hall@ncsu.edu (CKH)

Abstract

Peptides' hierarchical co-assembly into nanostructures enables controllable fabrication of multicomponent biomaterials. In this work, we describe a novel computational and experimental approach to design pairs of charge-complementary peptides that selectively co-assemble into β -sheet nanofibers when mixed together, but remain unassembled when isolated separately. The key advance is a *peptide co-assembly design* (PepCAD) algorithm that searches for pairs of co-assembling peptides. Six peptide pairs are identified from a pool of $\sim 10^6$ candidates via the PepCAD algorithm and then subjected to DMD/PRIME20 simulations to examine their co-/self-association kinetics. The five pairs that spontaneously aggregate in kinetic simulations selectively co-assemble in biophysical experiments, with four forming β -sheet nanofibers, and one forming a stable non-fibrillar aggregate. Solid-state NMR, which is applied to characterize the co-assembling pairs, suggests that the *in-silico* peptides exhibit a higher degree of structural order than the previously reported CATCH(+/-) peptides.

Introduction

Certain peptides are known to assemble spontaneously into a variety of nanostructures---nanofibers, nanosheets, nanotubes, nanoparticles, etc., -with applications in a wide variety of fields, including: drug delivery, vaccines, hydrogels, 3-D cell culture, tissue engineering, and protein scaffolds. Great structural variety can be achieved, in principle, via a “bottom-up” strategy in which the peptide amino acid composition, length and sequence pattern are tailored to form a particular structure¹⁻⁷. The big question is, of course, — what are the design rules for programming in a particular self-assembled structure? An even more basic question is which sequences will assemble? Although there are some algorithms that attempt to answer the latter question by correlating the amyloidogenic tendencies of the individual amino acids⁸⁻¹¹, ***there is no efficient computational or experimental approach to discover which sequences form which structures***. Most of the existing β -sheet-forming peptides are derived from naturally-occurring amyloid-forming proteins¹². Others¹³⁻¹⁵ have been designed based on a simple hydrophilic/hydrophobic (HP)_n repeating pattern that is known to form a two-layer fibril with a hydrophobic core¹⁶. The difficulty in *a priori* design of β -sheet-forming peptides comes from the challenge of effectively exploring the enormous amino acid sequence space to discover peptide sequences that form the desired structures. For this reason, systematic investigation of peptide aggregation behaviors in vast sequence space has only been conducted for very short peptides, such as dipeptides¹⁷ and tri-peptides¹⁸, and only for single-component systems. The challenge becomes even more interesting when one considers the possibility of designing two or more peptides that co-assemble to form a single nanostructure.

Recently, peptide co-assembly has emerged as a novel supramolecular design strategy, allowing construction of peptide-based nanofibers with integrated functionalities¹⁹⁻²². Here we focus on selective co-assembly: the formation of stable β -sheet nanofibers by two different peptides, A and B, only when they are both present in solution; otherwise, they remain unassembled in random coil configurations. Researchers have used heuristics to develop pairs of charge-complementary co-assembling peptides by generating highly charged sequence variants of established self-assembling peptides^{23,24}. For instance, the CATCH(4+) (sequence: Ac-QQKFKFKFKQQ-Am) with four positively-charged residues and the CATCH(6-) (sequence: Ac-EQEFEFEFEQE-Am) with six negatively-charged residues²⁴ are both derived from the Q11 peptide (sequence: Ac-QQKFQFKQEQQ-Am)²⁵ and have been shown to coassemble. These peptide pairs have been shown

to co-assemble into bilayer β -sheets that contain the two peptides arranged in a predominantly alternating pattern, although like-charged neighbor mismatches and β -strand polymorphisms have been observed^{26,27}. Although a heuristics-based experimental approach has led to the discovery of several co-assembling pairs, this approach becomes intractable when exploring the vast sequence space for two complementary 11-residue peptides. The addition of computational methodologies capable of designing new selectively co-assembling pairs would greatly accelerate the development of peptide nanostructures, potentially leading to architectures with more precise molecular-level order and organization than has heretofore been possible.

In this work, we describe a computational and experimental protocol, essentially a funnel, which screens large numbers of candidate peptide pairs to identify those that will selectively co-assemble into β -sheet nanofibers with a pre-set structure in experiment, e.g., an automated sequence screening process for 11-mer co-assembling peptide pairs in a pool of more than 300,000 peptides and around 10^6 possible peptide pairs (Fig. 1). The key advance is a Monte Carlo (MC)-type *peptide co-assembly design* (PepCAD) algorithm that is used for *de novo* design of charge-complementary peptide pairs. This method is a logical extension of our previously-developed *peptide binding design* (PepBD) algorithm²⁸⁻³² that we applied to design peptide binders to biomolecular targets with exceptional affinities³³⁻³⁵. Lead compounds from the computational search are subjected to *discontinuous molecular dynamics* (DMD) simulations combined with the knowledge-based PRIME20 force field³⁶⁻⁴⁰ to examine the co-assembly kinetics of the *in-silico* discovered peptide pairs, as well as the self-association of each peptide species when alone. The peptide pairs that can selectively co-assemble in DMD/PRIME20 simulations are then synthesized and their co-assembly versus self-association is examined using transmission electron microscopy (TEM), Fourier-transform infrared spectroscopy (FTIR), and solid-state nuclear magnetic resonance (NMR) spectroscopy. Finally, the structural order and molecular-level compositions are assessed by solid-state NMR and compared to previous co-assembling β -sheet designs. We envision that this new paradigm for *de novo* peptide design will enable rapid development of molecules that assemble into specific supramolecular architectures.

(Figure 1 should be placed here)

Results

Predetermining the molecular architecture of the peptide scaffold. To discover peptide pairs that can co-assemble into a peptide-based nanofiber in solution, we must first decide what fibrillar structure to design — β -sheet nanofiber, cross- α nanofibers, α -helical coiled-coil, etc. Here we choose to design peptides that can assemble into the co-assembled fibril structure formed by a mixture of CATCH(4+) and CATCH(6-) peptides. Discontinuous molecular dynamics (DMD) is combined with the coarse-grained protein force field, PRIME20, to simulate the spontaneous co-assembly of an initially-random system of 24 CATCH(4+) peptides and 24 CATCH(6-) peptides at 330 K and 10 mM into a fibrillar structure. PRIME20 was chosen because it is among the most realistic of the protein coarse-grained models, does not build in any predetermined secondary structure, provides a good representation of amyloid structure in comparison to experiment, and is fast enough to get to the fibrillar stage starting from the random-coil state ([Supplementary Section 1](#)). Our simulation results revealed that the CATCH(4+/6-) peptides preferentially co-assemble into a highly-ordered fibrillar structure with two layers of β -sheets ([Fig. 2a](#)). The structure of this fibril can be characterized in terms of the orientations of the peptides relative to each other. Conformational analysis of the CATCH fibrillar structure as described in [Supplementary Section 2](#) indicates that for a CATCH(6-) peptide in the fibril, its nearest and next-nearest peptides in the same β -sheet are most likely to be a CATCH(4+) that is anti-parallel and a CATCH(6-) that is parallel, respectively; and its nearest peptide on the neighboring β -sheet is most likely to be a CATCH(6-) that is parallel ([Fig. 2b](#)). This preferred organization indicates that the CATCH(4+/6-) peptides prefer to co-aggregate into 2-layers with anti-parallel orientation within each β -sheet and parallel orientation between the sheets ([Fig. 2b](#)). Based on the structural information above, we constructed a 2-layer fibril model using the Discovery Studio 3.5 package and optimized its geometry in the AMBER14 software (as seen in [Supplementary Section 3](#) and [Supplementary Fig. S1](#)). This 2-layer fibril structure is used as an initial conformation in the PepCAD algorithm to discover other potential co-assembling peptides.

(Figure 2 should be placed here)

Sequence evolution of co-assembling peptides. The PepCAD algorithm is a MC-based algorithm for *de novo* design of charge-complementary peptide pairs that can co-assemble into particular supramolecular architectures (as described in [Supplementary Section 4](#) and [Supplementary Fig. S2](#)).

Three different kinds of sequence moves, viz. *intra-chain* residue mutation, *intra-chain* residue exchange, and *inter-chain* residue exchange, are used to perturb the peptide sequences (Fig. 2c), resulting in new trial co-assembling designs, peptides A and B. The peptide backbone scaffold of the 2-layer fibril structure is fixed throughout the design process. To evaluate the merits of these peptides' co-assembly capability, a score function Γ_{score} is introduced in Supplementary Section 5, Equation 1 that takes into account the binding free energy, $\Delta G_{binding}$, between the peptides A and B⁴¹, as well as the intrinsic self-aggregation propensities, P_{agg} , of the individual peptides⁴²⁻⁴⁵. To avoid local searches, we start with three different initial random sequences at two weighting factors λ , viz. $\lambda = 3.0$ and 4.0 , to vary the search pathways, leading to a total of six evolution runs (three for $\lambda = 3.0$ and three for $\lambda = 4.0$). The designed peptides are constrained to have the same sequence pattern "PPHPPHPPHPP" as the CATCH(+/-) peptides, where "H" and "P" refer to hydrophobic and polar amino acids. The peptides are also constrained to have 3 hydrophobic residues, 3 hydrophilic residues, 5 charged residues, and no cysteine, proline, or glycine. Our sequence searches did not start with the known CATCH(4+/6-) peptides because: (1) the peptide co-assembly designs in this work are constrained to contain (5+/5-) residues on the chains, which is different from those for the CATCH(4+/6-) peptides; (2) the sequence searches starting with the CATCH(4+/6-) peptides as initial conditions result in an ineffective exploration of the broad sequence space. The final peptide designs are not sensitive to the initial conditions (sequences) chosen during the Monte Carlo algorithm.

Fig. 2d,e show profiles of score vs. number of evolution steps at $\lambda = 3.0$ and 4.0 , respectively. Since the searches start from random peptide sequences, the scores are high at the initial stage. As the evolution proceeds, the three kinds of sequence moves help the amino acids to rapidly find appropriate sites on the chains, resulting in a sharp drop in the score. Later, the score profile fluctuates considerably due to variations in the identities of the amino acids at the various sites. Favorable sequence moves that decrease the score are always accepted in our procedure, while unfavorable sequence moves that may slightly or significantly increase the score are accepted or rejected according to the MC criterion. By examining the profile of the score over the sequence evolution, we can identify the lowest scores in each profile which correspond to the best peptide sequences for each search. Two pairs of designed co-assembly peptides as well as their corresponding fibril structures are shown in Fig. 2d,e for the case of $\lambda = 3.0$ with the lowest score -25.35 kcal/mol at the 1570th step (Fig. 2d), and for the case of $\lambda = 4.0$ with the lowest score -25.80 kcal/mol at the 5926th step

(Fig. 2e).

The six best-scoring pairs of peptide sequences resulting from the six evolutions, hereafter referred to as Designs 1 through 6, are listed in Table 1 along with their associated scores, binding free energies per peptide and intrinsic self-aggregation propensities per peptide calculated using Supplementary Equation 1. (The latter two quantities are used in the calculation of the peptide's score.) The results from the DMD/PRIME20 kinetics simulations and experiments are also listed and will be discussed later. All of the *in-silico* peptide pairs A and B in Table 1 exhibit negative values for $\Delta\tilde{G}_{binding}$, the first term in the score function, implying that peptides A and B may form fibril-like co-aggregates due to their strong mutual binding affinity. Note that these binding energies are stronger than a typical peptide-biomolecule binding energy because the 2-layer fibril model used for peptide designs is an ideal optimized structure. Each peptide within the fibril model can form at least six backbone hydrogen bonds with its nearest neighboring peptides as well as broadly interact with the other peptides on the neighbor sheet(s), leading to a large value of binding energy. In addition, peptide pairs A and B exhibit weak intrinsic self-aggregation propensities due to their low values of $\tilde{P}_{agg}$, indicating that they are not likely to self-assemble into well-ordered structures when dissolved separately in solution; the lower the value of $\tilde{P}_{agg}$ the weaker its self-aggregation propensity. Without loss of generality, we henceforth assign the label A to the positively charged peptide and the label B to the negatively charged peptide. Interestingly, the design algorithm, not the user-defined sequence specifications, preferentially places positively-charged amino acid, lysine (K), at the N-terminus of peptide A and negatively-charged amino acids, aspartic acid (E) and glutamic acid (D), at the C-terminus of peptide B. Such sequence alignments for peptides A and B likely facilitate the formation of the targeted antiparallel β -sheet structure. The frequent appearances of the amino acids asparagine (N) and threonine (T) on peptide pairs A and B is likely due to their low intrinsic self-aggregation propensities as well as the strong sidechain-sidechain interactions between the carboxamide group (-CO-NH₂) on asparagine (N) and the hydroxyl group (-OH) in threonine (T), which further stabilize the amyloid fibril. These design motifs are not seen in previous co-assembling β -sheet peptide sequences, highlighting the utility of the design algorithm in expanding the design space.

(Table 1 should be placed here)

Computational analysis of the co-/self-assembly properties of *in-silico* peptide pairs. First, we performed explicit-solvent atomistic molecular dynamics (MD) simulations to examine the thermodynamic stability of our designed amyloid fibrils (as described in [Supplementary Section 6](#)). The starting structures of the six 2-layer amyloid fibrils are obtained from the output of the PepCAD algorithm. Simulation results revealed that the six *in-silico* peptide pairs are able to maintain a well-organized 2-layer fibril structure after 100-ns ([Supplementary Fig. S3](#)). We then employed the FoldAmyloid web-server, a bioinformatics method, to estimate the amyloidogenicity (likelihood that the peptides would self-aggregate to form amyloid) of the single peptide species within the six designs⁴⁶. In this method, peptides are predicted to be amyloidogenic if they contain at least 7 consecutive residues that have average self-aggregation scales that are higher than an empirical threshold value of 21.4; otherwise they are predicted to be non-amyloidogenic. The peptide pairs A and B within Designs 1-6 are predicted to be non-amyloidogenic as their average self-aggregation scales are lower than 21.4, implying that these single peptide species exhibit a weak propensity for self-aggregation in solution ([Supplementary Fig. S3](#)).

In addition, the co- and self-association kinetics of these *in-silico* peptide pairs were examined using DMD/PRIME20 simulations. Motivation for this is the possibility that even though the fibril state might be stable according to the atomistic MD simulations, the kinetics might not be fast enough to arrive at an ordered structure. DMD/PRIME20 simulations of a mixture containing 100A and 100B peptides initially in random coil conformations at 10 mM and 330 K were conducted for 5 (or 10) μ s (as described in [Supplementary Section 7](#)). The types of structure formed for all six designs in DMD/PRIME 20 simulations are indicated in [Table 1](#). [Fig. 3](#) shows the aggregation kinetics, reported as the β -sheet content versus simulation time, and final simulation snapshots for Designs (1-6) and CATCH peptides. Our simulation results predict that all the *in-silico* peptide pairs selectively co-assemble into ordered β -sheet fibrillar structures, with the exception of Design 3 ([Fig. 3](#)). The peptide pairs A and B of Designs (1, 2, 4, and 5) co-aggregate rapidly when mixed with each other to form fibril structures with more than 2 β -sheet layers wherein the peptides A and B predominantly adopt an anti-parallel orientation ([Fig. 3](#)). In contrast, Design 6 has the slowest aggregation kinetics ([Fig. 3](#)) but is the only peptide pair that tends to form a 2-layer architecture ([Fig. 3](#)), as the CATCH peptides

do. The co-assembly of the peptides A and B within these designs is initialized by the formation of a small β -sheet fibril nucleus. As the simulation progresses, the nucleus grows and elongates by recruiting random-coil monomeric peptides, and sometimes by associating laterally with other small oligomeric species, to form a single multi-layer beta sheet fibrillar structure (Supplementary Fig. S4). The DMD/PRIME20 simulations of the associated single-component systems reveals that the individual peptides, A or B, in the six designs do not self-associate when alone in solution (Supplementary Fig. S5).

To study the effectiveness of PepCAD, we chose three initial random peptide pairs and three *in-silico* peptide pairs with medium scores from the evolution for co-assembly to test to see if they co-assembled in the DMD/PRIME20 simulations. Simulation results, which are shown in Supplementary Fig. S6 revealed that all three initial peptide pairs do not co-assemble into fibril structures. However, two of the three *in-silico* peptides with medium scores do form fibril-like co-aggregates in simulations, while the other one does not. Further, the individual peptides associated with these initial pairs and pairs with medium scores did not self-assemble in DMD simulations (not shown for brevity). To examine the diversity of sequences, we compared the peptide pairs with medium and best scores (Supplementary Fig. S6 and Table 1), and found that the best-scoring peptides A (and B) exhibit a high similarity with each other but significantly differ from those with medium scores. Sequence evolution in PepCAD achieves the de novo design of peptides that co-assemble into β -sheet based nanofibrils. Future efforts will seek to identify threshold scores that are reliable predictors of co-assembly propensity in simulations and experiments.

(Figure 3 should be placed here)

Experimental analyses of co- and self-assembly of designed peptide pairs. Based on the outcomes of the DMD/PRIME20 simulations, the peptide pairs of Designs 1-6 were synthesized and their selective co-assembly was characterized using transmission electron microscopy (TEM) (Supplementary Section 8) and Fourier-transform infrared spectroscopy (FTIR) (Supplementary Section 9) (Fig. 4). We observed elongated nanofibers with a high degree of lateral association in transmission electron micrographs of equimolar mixtures of Designs 2, 4, and 5 (Fig. 4a). In contrast, the mixture of Design 1 formed shorter nanofibers that were few in number and less laterally

aggregated (Fig. 4a). Unexpectedly, the peptide pair in Design 6 did not form elongated nanofibers, contrary to simulation predictions; instead, Design 6 exclusively formed non-fibrillar aggregates with approximate diameters of 11 ± 1.7 nm (Fig. 4a). Likewise, the peptide pair in Design 3 formed non-fibrillar aggregates, which persisted for 7 days (Supplementary Fig. S8). It is worth noting that these samples were prepared at 1 mM, which is significantly lower than the simulation concentration of 10 mM, yet significantly higher than the minimum co-assembly concentration reported previously for CATCH peptides.

(Figure 4 should be placed here)

Due to the disparity in concentration between simulations and TEM samples, we used FTIR to determine the secondary structure of each peptide in Designs 1-6 alone and in the presence of its complementary partner at 15 mM (Fig. 4b). When alone, the FTIR spectrum of each peptide had local maxima at approximately 1645 and 1675 cm^{-1} (dashed lines). The former indicates that the peptides adopt random coil conformations, and therefore do not undergo considerable self-association. The latter is likely due to residual trifluoroacetic acid remaining from peptide synthesis and purification processes, and likely has no impact on the peptide secondary structure. When paired, the FTIR spectra of Designs 4 and 5 had strong maxima between 1621-1616 cm^{-1} , indicating formation of intermolecular hydrogen bonds consistent with a β -sheet secondary structure. The Design 2 FTIR spectrum had a major peak near 1620 cm^{-1} , but also had significant absorption in the range of 1630-1690 cm^{-1} , suggesting a lesser abundance of β -sheet hydrogen bonds relative to Designs 4 and 5. The Design 3 FTIR spectrum had a strong peak at 1648 cm^{-1} consistent with a random coil and a very weak peak at 1620 cm^{-1} when compared to the spectrum of the Design 2 pair, suggesting an even lesser abundance of β -sheet hydrogen bonds than the other successful designs (Supplementary Fig. S8). Taken with the TEM images, this indicates that DMD correctly identified that Design 3 would not co-assemble into β -sheet fibrils. The spectrum of Design 6 had a shoulder at 1620 cm^{-1} , and a maximum between 1647-1642 cm^{-1} which, taken together, suggested that this peptide pair preferentially adopted random coil conformations. The absence of a strong β -sheet signal in Design 6 FTIR samples suggested that the non-fibrillar oligomers observed in TEM micrographs lacked the considerable backbone hydrogen bonding associated with β -sheet structures (Supplementary Fig. S7).

The FTIR spectrum of Design 1 had a predominant maximum between 1647-1642 cm^{-1} and only a very weak shoulder at 1620 cm^{-1} , indicating that most of the peptides in the mixture adopted random coil conformations. Notably, this suggested that the few nanofibers that are observed in transmission electron micrographs of Design 1 are likely rare relative to those peptides that are unassembled or part of non-fibrillar aggregates.

Informed by the TEM images and FTIR measurements, Designs 1, 2, 4, and 5 were further evaluated for co-assembly behavior by solid-state NMR measurements on co-assembled samples. Designs 3 and 6 were excluded from solid-state NMR analysis due to the lack of nanofibers in the TEM images and a mostly random coil signature in FTIR spectra. Peptide A has a distinct chemical shift peak around 23ppm uniquely attributed to the γ -carbon (C_γ) of the K sidechain. Similarly, peptide B has an identifiable chemical shift peak near 181 ppm uniquely assigned to δ -carbon (C_δ) of E sidechain. In Fig. 5, 1D ^{13}C NMR spectra of Designs 1, 2, 4, and 5 all exhibit peaks at ~ 23 and ~ 181 ppm indicating that peptides A and B are present in appreciable amounts within nanofiber samples. Thus, peptides A and B co-assemble into 2-component nanofibers in all four tested designs. The upfield shift in the measured CO chemical shifts as compared to the value for the same sites in a random coil conformation (Fig. 5, purple shaded region) indicates a β -strand conformation, as was also observed by FTIR and predicted by DMD/PRIME20. Altogether, Designs 1, 2, 4, and 5 from the initial six designs successfully show selective co-assembly into β -sheet-rich nanofibers as originally designed and predicted by simulations.

(Figure 5 should be placed here)

Computationally designed co-assembling peptides show improved structural homogeneity. The ratio (relative abundance) of cationic peptide A to anionic peptide B in the four co-assembled structures was determined using solid-state NMR measurements. The 1D ^{13}C NMR spectra in Fig. 5 was collected in a quantitative manner allowing comparison of chemical shift peak areas. The ratio of Peptide A to Peptide B is reported in Table 2 for Designs 1, 2, 4, and 5 as calculated from the K C_γ and the E C_δ peak areas as detailed in the Supplemental Section 10. Peak linewidths are also shown in Table 2 and are discussed in the following paragraph. The positively charged peptide A is slightly more abundant than the negatively charged peptide B in all tested pairs, consistent with our previous

studies on similar co-assembling β -sheet peptides. Therefore, peptides A and B are likely to arrange in a predominantly alternating $(AB)_n$ pattern although some self-association may occur. Compared to previous designs such as the CATCH(4+/4-) design, the ratio of the two peptide components is closer to unity as shown in Table 2. This improvement in $(AB)_n$ alternation may result from the contribution of the aggregation propensity to the score function which disfavors peptide self-association.

(Table 2 should be placed here)

Measurements of the peak linewidths in 1D ^{13}C NMR spectra of Designs 1, 2, 4, and 5 are compared to previous designs, indicating exceptionally well-ordered nanofibers. Linewidths (full width at half maximum) of the E C_δ and K C_γ chemical shift peaks are reported in Table 2. Broad linewidths can result from the presence of multiple distinct structures or a disordered structure. In contrast, the linewidths observed in nanofibers produced from Designs 1, 2, 4, and 5 are narrow and similar to those observed in protein crystals (0.6 ppm) indicating a very highly ordered structure. Compared to linewidths in the family of CATCH peptides and King-Webb peptides (KW+: Ac-KKFEWEFEKK-Am; KW-: Ac-EEFKWKFKEE-Am) (over 1 ppm)^{26,27}, the linewidths of the computationally identified pairs are almost $2\times$ smaller, suggesting that the computationally designed peptide pairs may be better behaved and produce more structurally-homogeneous nanofibers.

Discussion and conclusion

Here, a computational and experimental protocol is reported to design pairs of charge-complementary peptides that can selectively co-assemble into β -sheet nanofibers when mixed together, but remain unassembled when isolated separately. A *peptide co-assembly design* (PepCAD) algorithm was developed to discover potential selective co-assembling peptides in a fast and efficient manner. The PepCAD algorithm uses a newly-built score function, Γ_{score} , to measure the binding free energy of the co-assembling peptides A and B, as well as their intrinsic self-aggregation propensities. A lower negative value of Γ_{score} during the process of sequence evolution means that the *in-silico* discovered peptides A and B are more likely to form fibril-like co-aggregates, but not fibril-like self-aggregates. As a result, six pairs of charge complimentary co-assembling peptides with the lowest Γ_{score} , viz. Designs 1-6, were identified from a library of $\sim 10^6$ candidate pairs using the

PepCAD algorithm. DMD/PRIME20 simulations were then conducted to examine the co- and self-association kinetics of the six *in-silico* peptide pairs. Designs 1, 2, 4, 5, and 6 formed amyloid-like structures after 5 μ s of simulation time, whereas Design 3 did not co-assemble. Subsequently, the five peptide pairs were synthesized and purified, and their co-assembly *vs.* self-association was examined using TEM, FTIR, and solid-state NMR. Designs 2, 4, and 5 successfully co-assembled into β -sheet nanofibers and did not self-associate; Design 1 formed a combination of β -sheet nanofibers and non-fibrillar aggregates, whereas Design 6 failed to form β -sheet-rich structures. Designs 1, 2, 4, and 5 had solid-state NMR spectra with narrower linewidths and improved ratios of cationic to anionic peptide than the empirically-designed charge-complementary co-assembling peptide pairs, CATCH(+/-) and KW, confirming that the designed peptides exhibit a higher degree of structural order. This improved structural precision, coupled with the observation that none of the designed peptides aggregated when alone, highlights the accuracy of the newly-developed Γ_{score} as a predictor of co- versus self-assembly propensity. Collectively, these observations demonstrate the potential of the PepCAD algorithm for designing co-assembly peptides from an experimentally intractable sequence space. In this design, our first effort at discovering 11-mer co-assembling peptide pairs achieved a respectable success rate of 67%, meaning 4 of the 6 top-scoring peptides co-assembled and did not self-assemble in our experiments. This is encouraging. In the future we plan to further examine/improve the performance of the PepCAD algorithm in peptide co-assembly designs, e.g. by adjusting the lengths of peptides, the combinations of (+/-) charges, and the hydrophobic/hydrophilic sequence patterns. We will also try to determine what the success rate would be if we just consider candidates that have emerged from the design stage or from the design plus kinetic simulation stages.

One lesson learned here is that designing peptide sequences to co-assemble is not as straightforward as one might think. Our early design concept--- to create charge-complementary peptide pairs that selectively co-assemble into amyloid fibrils--- was informed by the thinking that opposite and highly charged peptides should resist self-assembly due to electrostatic repulsion and co-assemble through electrostatic attraction. Computational and experimental observations with co-assembling peptide pairs derived from molecules known to self-assemble demonstrate that simply mixing two peptides with a high degree of opposing (i.e. attractive) charges may speed up the aggregation kinetics, but it does not guarantee exquisite molecular-level co-assembly into β -sheet

nanofibers⁴⁷. A progressive increase in the magnitudes of the opposite charges on the peptide pairs might decrease the binding free energy due to an overwhelming increase in the desolvation penalty^{48, 49}. To capture polarization effects caused by the highly charged residues, we introduced a variable internal dielectric constant model^{50, 51} into the score function of PepCAD to calculate the electrostatic energy and polar solvation energy. By this way, we avoided overestimation of charge-charge interactions in this work. Although the individual peptides generally adopt a β -strand architecture when combined, like-charged neighboring strand imperfections are common and structural polymorphisms are observed^{26, 27}. This occurs even when the CATCH (+/-) sequence pattern, “PPHPHPHPPP”, where “H” and “P” refer to hydrophobic and polar amino acids, is imposed. The PepCAD algorithm adds a much-needed layer of biophysical sophistication to these simple-but-appealing ideas because it accounts for the complexity in sidechain-sidechain interactions, which is impractical through iterative experimentally-driven design processes. For example, the PepCAD algorithm has the ability to bias the fibrillar structure to be parallel/antiparallel within and between sheets. Toward this end, the algorithm preferentially designed peptides with cationic residues at the N-terminus and anionic residues at the C-terminus, in stark contrast to the CATCH(+/-) and KW pairs wherein charged residues are either distributed evenly or in a core/flank arrangement. Furthermore, the PepCAD algorithm can consider a richer diversity of the naturally-occurring amino acids. As a result, the algorithm preferentially designed peptides with 5 charged residues, used a combination of glutamic acid and aspartic acid in the anionic molecule, and placed threonine or asparagine residues at hydrophilic sites. These choices are considerable deviations from the CATCH(+/-) and KW pairs, which included 4, 6, or 7 charged residues, only used glutamic acid, and exclusively placed charged residues or glutamine residues in hydrophilic positions. An additional advantage of the PepCAD algorithm is that it enables us to achieve a “structure-to-sequence” design, viz. an inverse design to identify potential peptide sequences for a desired fibril-like supramolecular architectures. The performance of these types of algorithms has been analyzed by Green, who described a statistical framework for analyzing the performance of hierarchical molecular design methods⁵². In future work aimed at improving PepCAD will use this statistical framework to evaluate the efficiency of peptide design and predict the accuracy of its score function. Our current peptide co-assembly design is based on a fixed peptide backbone scaffold, thereby causing an inevitable bias to sequence evolution. Introducing configurational optimization to relax the peptide scaffold in PepCAD might facilitate

better contacts between residues and promote the stability of fibrils assembled by designed peptides. Hopefully, a new version of PepCAD will enable the efficient design of peptides that assemble into some of the amyloid classes predicted by Sawaya and Eisenberg¹².

The procedures presented here can be thought of as a “funnel” of computational and experimental nominal yes/no tests that allow one to screen a large initial set of candidates to discover pairs of selective co-assembling peptides as illustrated in Fig. 1. The funnel can also be viewed as an inverse design strategy in that the initial set of candidates is not completely random. It has been chosen to have the same length and HP sequence pattern as the CATCH(4+/6-) pairs and to form the two-layer amyloid configuration. (A difference is that each member of the pair must have 5 charged residues). The funnel/inverse design strategy can, in principle, be used to screen a larger (more-random) sequence space, depending on the desired outcome. In Step 1, the funnel is filled with as many candidates as possible that satisfy preconceived notions such as charge complementarity, HP pattern along the chain, etc. The PepCAD algorithm narrows this down by finding pairs whose packing energies and self-aggregation propensities are minimized for a specific structure (e.g. 2 stacked antiparallel beta sheets). In Step 2, DMD/PRIME20 simulations test if the pairs co-assemble but do not self-assemble in a reasonable time frame, 5 μ s. Pairs that fail this test are rejected. In Step 3, the peptides are synthesized, purified and then subjected to biophysical characterization measurements like ThT fluorimetry, FTIR and solid-state NMR. Peptides that fail the early tests in step 3, or are too hard to work with, are rejected. While this funnel protocol worked well, we should point out that peptides that pass step 1 do not always pass step 2, etc. For example, DMD/PRIME20 simulations suggested that Design 6 could co-assemble into a bilayer β -sheet, albeit more slowly than the other designs; yet, biophysical experiments demonstrated that the Design 6 peptides aggregate but do not assemble into β -sheet nanofibers over a month at room temperature (Supplementary Fig. S7). Nevertheless, the protocol is highly promising.

Methods and Materials

Method descriptions on discontinuous molecular dynamics (DMD) simulation and PRIME20 force field are given in Supplementary Section 1. Analysis of the structure of the simulated co-assembled CATCH fibril and construction of peptide scaffold of a 2-layer fibril model are described in Supplementary Sections 2-3. Details regarding the use of *peptide co-assembly design* (PepCAD)

algorithm to *de novo* design co-assembling peptide pairs are described in [Supplementary Section 4](#). The calculations of score function, binding free energy and intrinsic self-aggregation propensity are given in [Supplementary Section 5](#). Atomistic molecular dynamics simulations are performed to examine the thermodynamics stability of the *in-silico* peptide pairs, and the amyloidogenicity of single peptide specie is predicted using the FoldAmyloid web-server, as detailed in [Supplementary Section 6](#). DMD/PRIME20 simulations are conducted to examine the co-/self-association kinetic of the *in-silico* discovered peptide pairs, as detailed in [Supplementary Section 7](#). Nanofiber formation from peptide co-assemblies was observed on a FEI Tecnai Spirit transmission electron microscope as described in [Supplementary Section 8](#). Secondary structure analysis of peptide self- and co-assembly propensity was performed using a Perkin Elmer FTIR spectrophotometer as detailed in [Supplementary Section 9](#). Quantitative 1D ^{13}C spectra were collected for nanofiber samples on an 11.75 T Bruker Avance III spectrometer with a 3.2 mm Bruker MAS probe. NMR sample preparation and pulse sequence parameters are described in more detail in [Supplementary Section 10](#). Custom code in Wolfram Mathematica was used for chemical shift peak analysis with further discussion in [Supplementary Section 10](#).

Acknowledgements

This research was supported by funds from the National Science Foundation grant CBET-1743432. The authors acknowledge the use of instruments at the NMR center at the Georgia Institute of Technology. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation grant number ACI-1548562. X.X. and C.K.H. thank San Diego Supercomputer Center (SDSC) for providing computing time. G.A.H. is a founder, scientific advisory board member, and stockholder of Anchor Biologics, Inc. Patents related to charge-complementary peptides that list G.A.H. and D.T.S. as inventors have been awarded to and applied for by the University of Florida.

Author contributions: C.K.H., A.K.P., and G.A.H. provided the conceptual framework for the study. X.X. developed the PepCAD algorithm and used it to discover the co-assembling peptide pairs. Y.W. conducted the DMD/PRIME20 simulations to examine the co-self association kinetics of the peptide pairs. D.T.S., K.M.W., and R.L. characterized the co-/self-assembly structures of the peptide pairs in biophysical experiments using transmission electron microscopy, Fourier-transform infrared spectroscopy, and solid-state nuclear magnetic

resonance spectroscopy. X.X., Y.W., D.T.S., K.M.W., A.K.P, G.A.H., and C.K.H. wrote the manuscript.

Competing interests: All authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and the Supplementary Materials.

References

- [1] S. Kim, J. H. Kim, J. S. Lee, C. B. Park, Beta-sheet-forming, self-assembled peptide nanomaterials towards optical, energy, and healthcare applications. *Small* **11**, 3623-3640 (2015).
- [2] S. Scanlon, A. Aggeli, Self-assembling peptide nanotubes. *Nano Today* **3**, 22-30 (2008).
- [3] J. D. Hartgerink, J. R. Granja, R. A. Milligan, M. R. Ghadiri, Self-assembling peptide nanotubes. *J. Am. Chem. Soc.* **118**, 43-50 (1996).
- [4] S. Zhang, Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* **15**, 462-468 (1993).
- [5] K. Matsuura, K. Murasato, N. Kimizuka, Artificial peptide-nanospheres self-assembled from three-way junctions of β -sheet-forming peptides. *J. Am. Chem. Soc.* **137**, 10148-10149 (2005).
- [6] S. Vauthey, S. Santoso, H. Gong, N. Watson, S. Zhang, Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles. *Proc. Natl. Acad. Sci. USA* **99**, 5355-5360 (2002).
- [7] J. D. Hartgerink, E. Beniash, S. I. Stupp, Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* **294**, 1684-1688 (2001).
- [8] A. P. Pawar, K. F. DuBay, J. Zurdo, F. Chiti, M. Vendruscolo, C. M. Dobson, Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379-392 (2005).
- [9] G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, M. Vendruscolo, Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425-436 (2008).
- [10] M. Oliveberg, Waltz, an exciting new move in amyloid prediction. *Nat. Methods* **7**, 187-188 (2010).
- [11] I. Walsh, F. Seno, S. C. T. Tosatto, A. Trovato, PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* **42**, W301-W307 (2014).
- [12] M. R. Sawaya, S. Sambashivan, R. Nelson, M. L. Ivanova, S. A. Sievers, M. L. Apostol, M. J.

- Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. Ø., Madsen, C. Riekel, D. Eisenberg, Atomic structures of amyloid cross- β spines reveal varied steric zippers. *Nature* **447**, 453-457 (2007).
- [13] S. Zhang, T. Holmes, C. Lockshin, A. Rich, Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. USA* **90**, 3334-3338 (1993).
- [14] H. Yokoi, T. Kinoshita, S. Zhang, Dynamic reassembly of peptide RADA16 nanofiber scaffold. *Proc. Natl. Acad. Sci. USA* **102**, 8414-8419 (2005).
- [15] S. R. Leonard, A. R. Cormier, X. Pang, M. I. Zimmerman, H. X. Zhou, A. K. Paravastu, Solid-state NMR evidence for β -hairpin structure within MAX8 designer peptide nanofibers. *Biophys. J.* **105**, 222-230 (2013).
- [16] C. J. Wilson, A. S. Bommarius, J. A. Champion, Y. O. Chernoff, D. G. Lynn, A. K. Paravastu, C. Liang, M. C. Hsieh, J. M. Heemstra, Biomolecular Assemblies: Moving from Observation to Predictive Design. *Chem. Rev.* **118**, 11519-11574 (2018).
- [17] C. G. Pappas, R. Shafi, I. R. Sasselli, H. Siccardi, T. Wang, V. Narang, R. Abzalimov, N. Wijerathne, R. V. Ulijn, Dynamic peptide libraries for the discovery of supramolecular nanomaterials. *Nat. Nanotechnol.* **11**, 960-967 (2016).
- [18] P. W. Frederix, G. G. Scott, Y. M. Abul-Haija, D. Kalafatovic, C. G. Pappas, N. Javid, N. T. Hunt, R. V. Ulijn, T. Tuttle, Exploring the sequence space for (tri-) peptide self-assembly to design and discover new hydrogels. *Nat. Chem.* **7**, 30-37 (2015).
- [19] C. A. Hauser, S. Maurer-Stroh, I. C. Martins, Amyloid-based nanosensors and nanodevices. *Chem. Soc. Rev.* **43**, 5326-5345 (2014).
- [20] R. N. Shah, N. A. Shah, M. M. D. R. Lim, C. Hsieh, G. Nuber, S. L. Stupp, Supramolecular design of self-assembling nanofibers for cartilage regeneration. *Proc. Natl. Acad. Sci. USA* **107**, 3293-3298 (2010).
- [21] Y. Liang, P. Guo, S. V. Pingali, S. Pabit, P. Thiagarajan, K. M. Berland, D. G. Lynn, Light harvesting antenna on an amyloid scaffold. *ChemComm.* **48**, 6522-6524 (2008).
- [22] G. A. Hudalla, T. Sun, J. Z. Gasiorowski, H. Han, Y. F. Tian, A. S. Chong, J. H. Collier, Graded assembly of multiple proteins into supramolecular nanomaterials. *Nat. Mater.* **13**, 829-836 (2014).

- [23] P. J. S. King, M. G. Lizio, A. Booth, R. F. Collins, J. E. Gough, A. F. Miller, S. J. Webb, A modular self-assembly approach to functionalized β -sheet peptide hydrogel biomaterials. *Soft Matter* **12**, 1915-1923 (2016).
- [24] D. T. Seroski, A. Restuccia, A. D. Sorrentino, K. R. Knox, S. J. Hagen, G. A. Hudalla, Co-assembly tags based on charge complementarity (CATCH) for installing functional protein ligands into supramolecular biomaterials. *Cell. Mol. Bioeng.* **9**, 335-350 (2016).
- [25] J. H. Collier, P. B. Messersmith, Enzymatic Modification of Self-Assembled Peptide Structures with Tissue Transglutaminase. *Bioconjugate Chem.* **14**, 748-755 (2003).
- [26] Q. Shao, K. M. Wong, D. T. Seroski, Y. Wang, R. Liu, A. K. Paravastu, G. A. Hudalla, C. K. Hall, Anatomy of a selectively coassembled β -sheet peptide nanofiber. *Proc. Natl. Acad. Sci. USA* **9**, 4710-4717 (2020).
- [27] K. M. Wong, Y. Wang, D. T. Seroski, G. E. Larkin, A. K. Mehta, G. A. Hudalla, C. K. Hall, A. K. Paravastu, Molecular complementary and structural heterogeneity within co-assembled peptide β -sheet nanofibers. *Nanoscale* **12**, 4506-4518 (2020).
- [28] X. Xiao, C. K. Hall, P. F. Agris, The design of a peptide sequence to inhibit HIV replication: a search algorithm combining Monte Carlo and self-consistent mean field techniques. *J. Biomol. Struct. Dyn.* **32**, 1523-1536 (2014).
- [29] X. Xiao, P. F. Agris, C. K. Hall, Designing peptide sequences in flexible chain conformations to bind RNA using Monte Carlo, self-consistent mean field and concerted rotation techniques. *J. Chem. Theory Comput.* **11**, 740-752 (2015).
- [30] X. Xiao, P. F. Agris, C. K. Hall, Introducing folding stability into the score function for computational design of RNA-binding peptides boosts the probability of success. *Proteins* **84**, 700-711 (2016).
- [31] X. Xiao, M. E. Hung, J. N. Leonard, C. K. Hall, Adding energy minimization strategy to peptide-design algorithm enables better search for RNA-binding peptides: Redesigned λ N peptide binds boxB RNA. *J. Comput. Chem.* **37**, 2423-2435 (2016).
- [32] X. Xiao, Y. Wang, J. N. Leonard, C. K. Hall, Extended Concerted Rotation Technique Enhances the Sampling Efficiency of Computational Peptide-Design Algorithm. *J. Chem. Theory Comput.* **13**, 5709-5720 (2017).
- [33] J. L. Spears, X. Xiao, C. K. Hall, P. F. Agris, Amino acid signature enables proteins to recognize

modified tRNA. *Biochemistry* **53**, 1125-1133 (2014).

- [34] X. Xiao, Z. Kuang, J. M. Slocik, S. Tadepalli, M. Brothers, S. Kim, P. A. Mirau, C. Butkus, B. L. Farmer, S. Singamaneni, C. K. Hall, R. R. Naik, Advancing peptide-based biorecognition elements for biosensors using *in-silico* evolution. *ACS Sens.* **3**, 1024-1031 (2018).
- [35] X. Xiao, Z. Kuang, B. J. Burke, Y. Chushak, B. L. Farmer, P. A. Mirau, R. R. Naik, C. K. Hall, In silico discovery and validation of neuropeptide-Y-binding peptides for sensors. *J. Phys. Chem. B* **124**, 61-68 (2020).
- [36] M. Cheon, I. Chang, C. K. Hall, Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins* **78**, 2950-2960 (2010).
- [37] M. Cheon, I. Chang, C. K. Hall, Spontaneous Formation of Twisted A β ₁₆₋₂₂ Fibrils in Large-Scale Molecular-Dynamics Simulations. *Biophys. J.* **101**, 2493-2501 (2011).
- [38] D. C. Latshaw, M. Cheon, C. K. Hall, Effects of Macromolecular Crowding on Amyloid Beta (16-22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B* **118**, 13513-13526 (2014).
- [39] Y. Wang, Q. Shao, C. K. Hall, N-terminal Prion Protein Peptides (PrP(120-144)) Form Parallel In-register β -Sheets via Multiple Nucleation-dependent Pathways. *J. Biol. Chem.* **291**, 22093-22105 (2016); correction, Wang et al., **292**, 22655 (2017).
- [40] Y. Wang, S. J. Bunce, S. E. Radford, A. J. Wilson, S. Auer, C. K. Hall, Thermodynamic phase diagram of amyloid- β (16-22) peptide. *Proc. Natl. Acad. Sci. USA* **116**, 2091-2096 (2019).
- [41] H. Gohlke, C. Kiel, D. A. Case, Insights into Protein-Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras-Raf and Ras-RalGDS Complexes. *J. Mol. Biol.* **330**, 891-913 (2003).
- [42] K. F. DuBay, A. P. Pawar, F. Chiti, J. Zurdo, C. M. Dobson, M. Vendruscolo, Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* **341**, 1317-1326 (2004).
- [43] A. P. Pawar, K. F. DuBay, J. Zurdo, F. Chiti, M. Vendruscolo, C. M. Dobson, Prediction of “aggregation-prone” and “aggregation-susceptible” regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379-392 (2005).
- [44] G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, M. Vendruscolo, Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* **380**, 425-436 (2008).

- [45] G. G. Tartaglia, M. Vendruscolo, The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37**, 1395-1401 (2008).
- [46] S. O. Garbuzynskiy, M. Y. Lobanov, O. V. Galzitskaya, FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **26**, 326-332 (2010).
- [47] D. T. Seroski, X. Dong, K. M. Wong, R. Liu, Q. Shao, A. K. Paravastu, C. K. Hall, G. A. Hudalla, Charge guides pathway selection in β -sheet fibrillizing peptide co-assembly. *Communications Chemistry* **3**, 172-11 (2020).
- [48] E. Kangas, B. Tidor. Optimizing electrostatic affinity in ligand-receptor binding: Theory, computation, and ligand properties. *J. Chem. Phys.* **109**, 7522-7545 (1998).
- [49] E. Kangas, B. Tidor. Electrostatic Complementarity at Ligand Binding Sites: Application to Chorismate Mutase. *J. Phys. Chem. B* **105**, 880-888 (2001).
- [50] K. Zhu, M. R. Shirts, R. A. Friesner. Improved methods for side chain and loop predictions via the protein local optimization program: variable dielectric model for implicitly improving the treatment of polarization effects. *J. Chem. Theory Comput.* **3**, 2108-2119 (2007).
- [51] J. Li, R. Abel, K. Zhu, Y. Cao, S. Zhao, R. A. Friesner. The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins* **79**, 2794-2812 (2011).
- [52] D. F. Green. A Statistical Framework for Hierarchical Methods in Molecular Simulation and Design. *J. Chem. Theory Comput.* **6**, 1682-1697 (2010).

Table 1. The sequences of the six *in-silico* discovered peptide pairs, their associated scores (Γ_{score}), binding free energies per peptide ($\Delta\tilde{G}_{binding}$), intrinsic self-aggregation propensities per peptide ($\tilde{P}_{agg}$), the DMD/PRIME20 simulation results and the TEM-observed results. (Unit: kcal/mol)

Designs		Sequences and Sites											Γ_{score}	$\Delta\tilde{G}_{binding}$	$\tilde{P}_{agg}$	DMD/PRI ME20	TEM
		1	2	3	4	5	6	7	8	9	10	11					
1	Peptide A	K	K	K	M	K	V	K	V	N	T	T	-25.40	-25.07	-0.11	multilayer fibril	short nanofiber
	Peptide B	T	N	T	A	D	F	E	F	E	E	D					
2	Peptide A	K	K	K	V	K	V	K	F	T	T	N	-25.35	-24.93	-0.14	multilayer fibril	long nanofiber
	Peptide B	T	N	T	V	D	F	E	Y	E	E	D					
3	Peptide A	K	K	K	W	K	M	K	A	T	N	T	-26.85	-25.87	-0.33	random coils	Not performed
	Peptide B	T	N	T	V	E	V	E	L	D	D	D					
4	Peptide A	K	K	K	V	K	V	K	V	N	T	T	-25.62	-25.16	-0.12	multilayer fibril	long nanofiber
	Peptide B	T	N	T	A	E	F	E	F	E	E	D					
5	Peptide A	K	K	K	V	K	V	K	V	N	T	T	-25.80	-25.62	-0.05	multilayer fibril	aggregated fibrils
	Peptide B	T	N	T	M	D	F	E	Y	E	E	D					
6	Peptide A	K	K	K	V	K	Y	T	F	K	N	T	-25.93	-25.21	-0.18	long two layer fibril	non-fibrillar aggregates
	Peptide B	T	N	T	M	E	V	D	F	D	E	D					

- Designs 1-3 result from setting $\lambda=3.0$, while Designs 4-6 result from setting $\lambda=4.0$.

Table 2. Nanofiber composition and peak linewidth analysis for the computationally identified peptides.

	Ratio of Peptide A to Peptide B	K C _γ Linewidth in ppm (Peptide A)	E C _δ Linewidth in ppm (Peptide B)
CATCH(4+/4-)	2.22	1.098 ± 0.088	0.990 ± 0.065
Design 1	1.73	0.682 ± 0.110	0.430 ± 0.120
Design 2	1.55	0.775 ± 0.064	0.687 ± 0.145
Design 4	1.53	0.522 ± 0.065	0.526 ± 0.088
Design 5	1.78	0.553 ± 0.063	0.553 ± 0.083

Figure Legends

Fig. 1 An overview of our computational and experimental protocol for identifying new pairs of peptides A and B that selectively co-assemble into long-ranged β -sheet nanofibers.

Fig. 2 (a) The DMD/PRIME20 simulation result suggests that the CATCH(4+) and CATCH(6-) peptides preferentially co-assemble into a 2-layer fibril structure that belongs to the 8th class of steric zippers introduced by Sawaya *et al.*¹²) (b) Conformational analysis of the 2-layer amyloid fibril indicates that the CATCH fibril favors having an antiparallel peptide conformation in each β -sheet and that the two neighbor β -sheets align parallel to each other. (c) Three kinds of sequence moves, viz. intra-chain residue mutation, intra-chain residue exchange, and inter-chain residue exchange, are involved in PepCAD to generate new sequences for peptides A and B. Starting from random sequences and setting (d) $\lambda=3.0$ and (e) $\lambda=4.0$, the algorithm searches through large numbers of possibly-co-assembling peptides, A and B. Plots of score *vs* the number of evolution steps are shown on the left. Lower scores imply better peptide designs. The best designs from the two searches are circled in the plots; their corresponding fibril structures are shown on the right.

Fig. 3 DMD/PRIME20 simulations of peptide co-aggregation. Plots of β -sheet content versus simulation time describe the co-aggregation kinetics for mixtures of 100A and 100B peptides in Designs 1-6 and CATCH(+/-). . Snapshots of the final simulation structures of the seven systems are shown as well.

Fig. 4 Experimental characterization of co- and self-assembly of peptide pairs in Designs 1, 2, 4, 5, and 6. (a) Transmission electron micrographs of mixtures of Designs 1, 2, 4, 5, and 6. (b) FTIR spectra of the peptides of Design 1, 2, 4, 5, and 6 alone (dashed lines) and in combination (solid line).

Fig. 5 ¹³C NMR spectra of centrifuged and lyophilized nanofiber samples prepared from equimolar mixtures of Designs 1, 2, 4, 5. The region highlighted in purple represents the range of carbonyl carbon chemical shift values for the two peptides in random coil conformations. The peak highlighted in blue and red correspond to the γ -carbon of the K sidechain and the δ -carbon of the E sidechain.

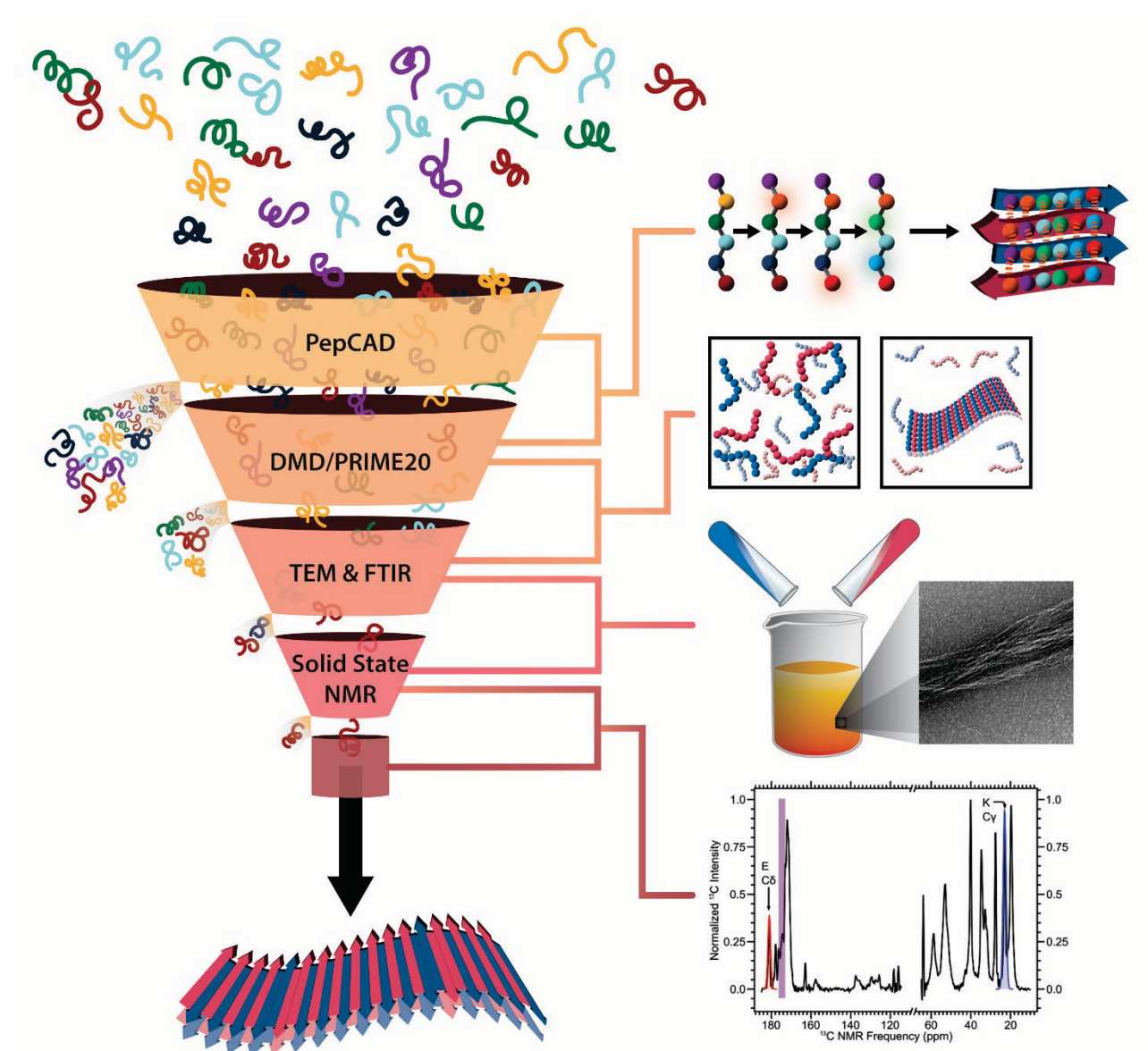


Fig. 1 An overview of our computational and experimental protocol for identifying new pairs of peptides A and B that selectively co-assemble into long-ranged β -sheet nanofibers.

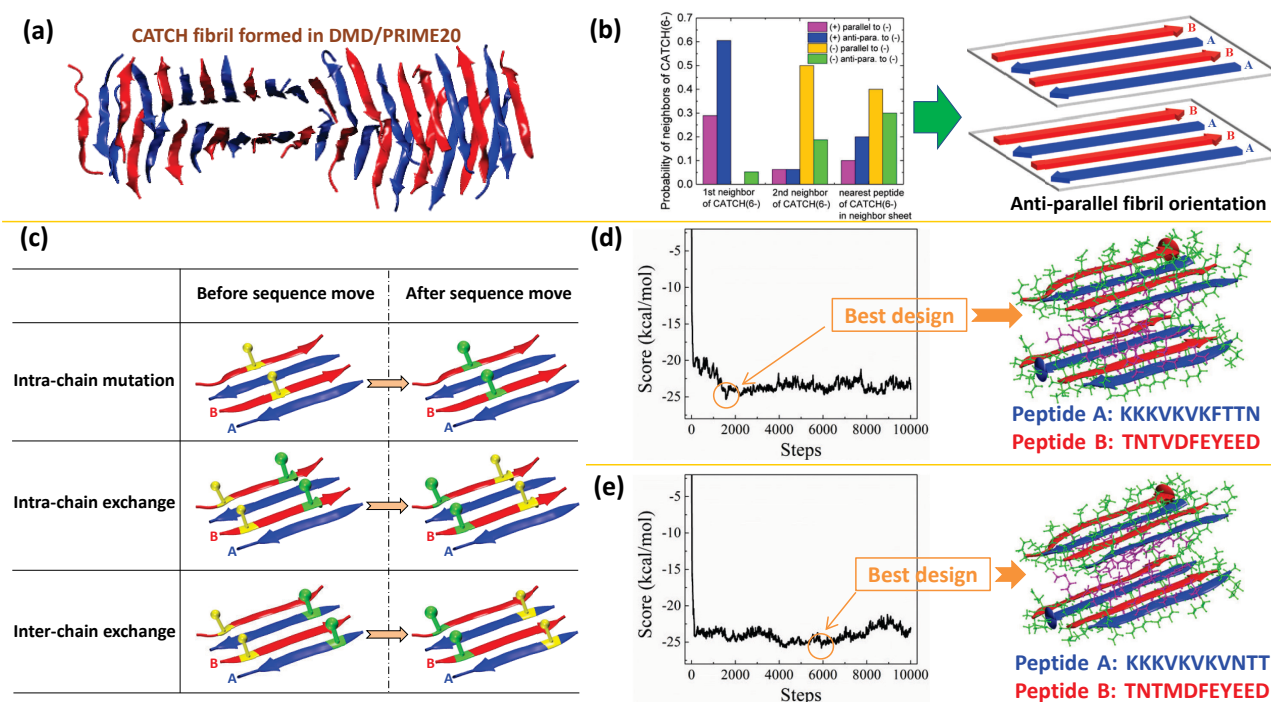


Fig. 2 (a) The DMD/PRIME20 simulation result suggests that the CATCH(4+) and CATCH(6-) peptides preferentially co-assemble into a 2-layer fibril structure that belongs to the 8th class of steric zippers introduced by Sawaya *et al.*¹² (b) Conformational analysis of the 2-layer amyloid fibril indicates that the CATCH fibril favors having an antiparallel peptide conformation in each β -sheet and that the two neighbor β -sheets align parallel to each other. (c) Three kinds of sequence moves, viz. intra-chain residue mutation, intra-chain residue exchange, and inter-chain residue exchange, are involved in PepCAD to generate new sequences for peptides A and B. Starting from random sequences and setting (d) $\lambda=3.0$ and (e) $\lambda=4.0$, the algorithm searches through large numbers of possibly-co-assembling peptides, A and B. Plots of score vs the number of evolution steps are shown on the left. Lower scores imply better peptide designs. The best designs from the two searches are circled in the plots; their corresponding fibril structures are shown on the right.

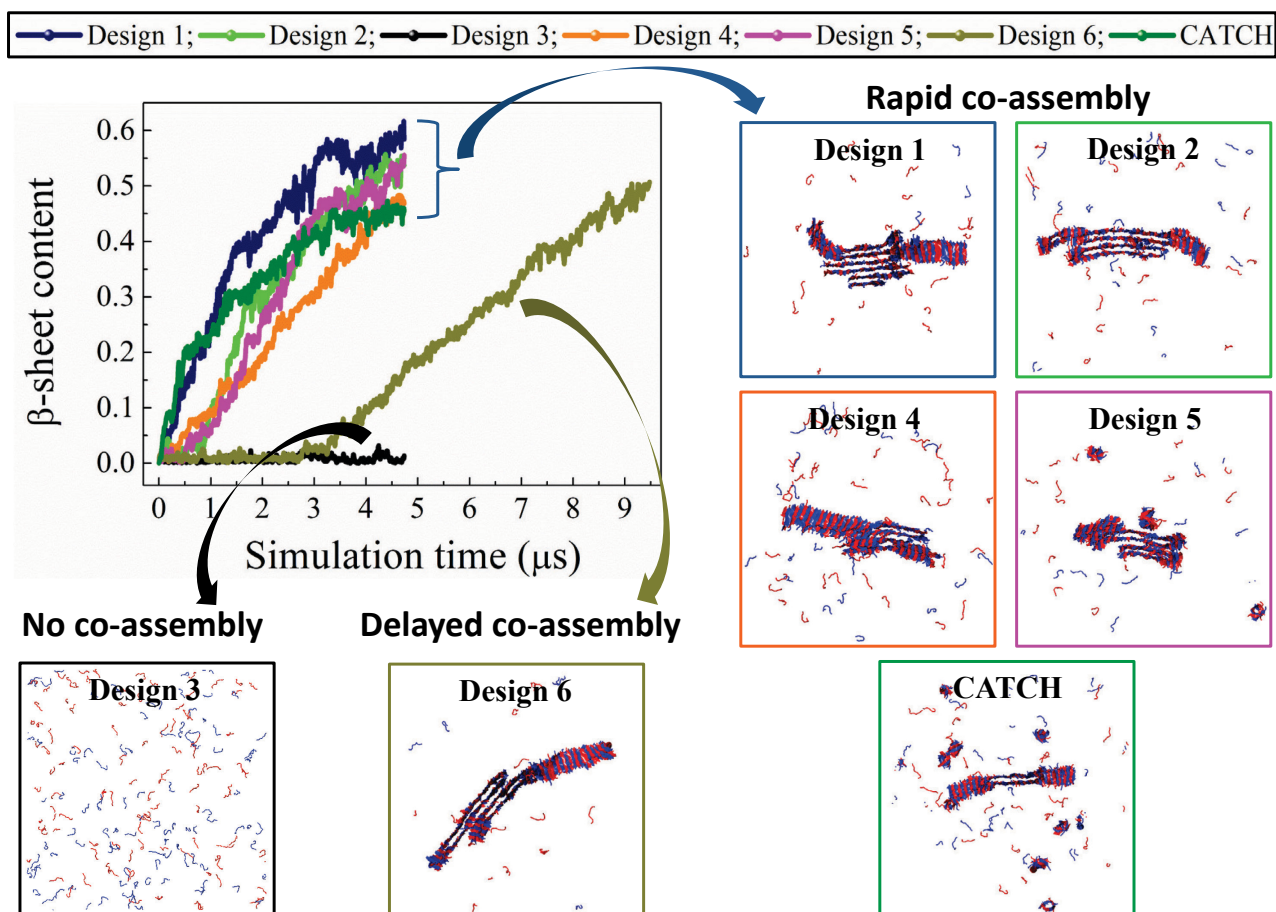


Fig. 3 DMD/PRIME20 simulations of peptide co-aggregation. Plots of β -sheet content versus simulation time describe the co-aggregation kinetics for mixtures of 100A and 100B peptides in Designs 1-6 and CATCH(+/-). Snapshots of the final simulation structures of the seven systems are shown as well.

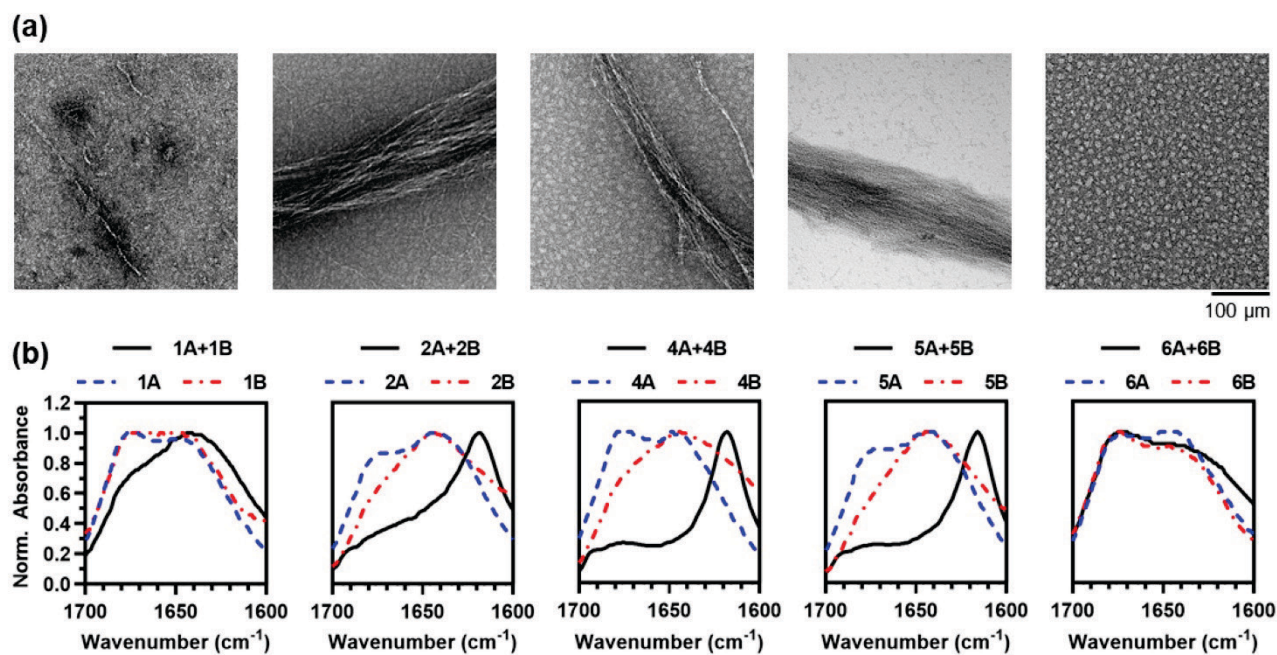


Fig. 4 Experimental characterization of co- and self-assembly of peptide pairs in Designs 1, 2, 4, 5, and 6. (a) Transmission electron micrographs of mixtures of Designs 1, 2, 4, 5, and 6. (b) FTIR spectra of the peptides of Design 1, 2, 4, 5, and 6 alone (dashed lines) and in combination (solid line).

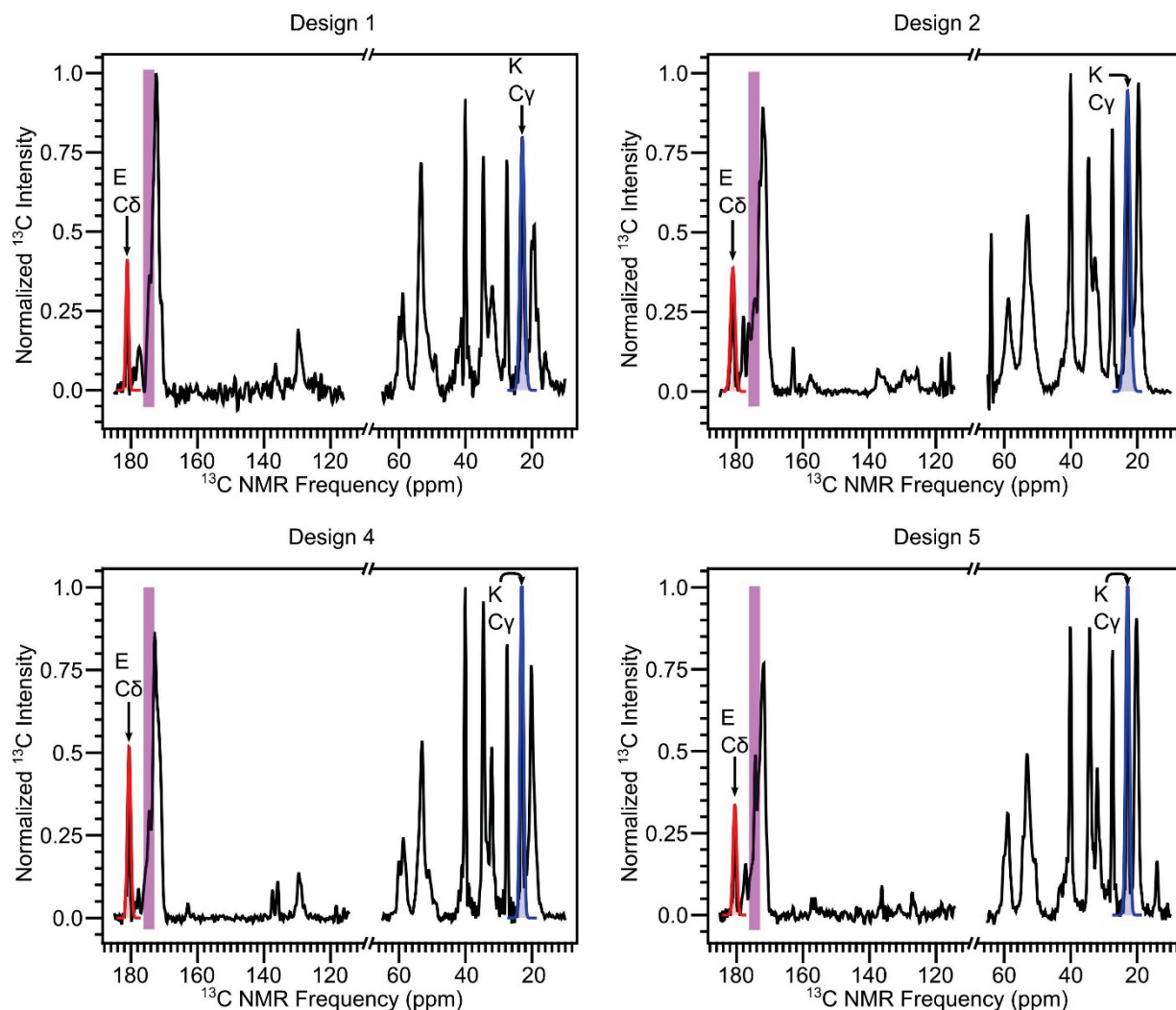


Fig. 5 ^{13}C NMR spectra of centrifuged and lyophilized nanofiber samples prepared from equimolar mixtures of Designs 1, 2, 4, 5. The region highlighted in purple represents the range of carbonyl carbon chemical shift values for the two peptides in random coil conformations. The peak highlighted in blue and red correspond to the γ -carbon of the K sidechain and the δ -carbon of the E sidechain.