

ARTICLE



<https://doi.org/10.1038/s41467-020-20689-w>

OPEN

# Structural analysis of cross $\alpha$ -helical nanotubes provides insight into the designability of filamentous peptide nanomaterials

Fengbin Wang<sup>1</sup>, Ordy Gnewou<sup>2</sup>, Charles Modlin<sup>2</sup>, Leticia C. Beltran<sup>1</sup>, Chunfu Xu<sup>2</sup>, Zhangli Su<sup>1</sup>, Puneet Juneja<sup>3</sup>, Gevorg Grigoryan<sup>4,5</sup>, Edward H. Egelman<sup>1</sup> & Vincent P. Conticello<sup>2,3</sup>✉

The exquisite structure-function correlations observed in filamentous protein assemblies provide a paradigm for the design of synthetic peptide-based nanomaterials. However, the plasticity of quaternary structure in sequence-space and the lability of helical symmetry present significant challenges to the de novo design and structural analysis of such filaments. Here, we describe a rational approach to design self-assembling peptide nanotubes based on controlling lateral interactions between protofilaments having an unusual cross- $\alpha$  supramolecular architecture. Near-atomic resolution cryo-EM structural analysis of seven designed nanotubes provides insight into the designability of interfaces within these synthetic peptide assemblies and identifies a non-native structural interaction based on a pair of arginine residues. This arginine clasp motif can robustly mediate cohesive interactions between protofilaments within the cross- $\alpha$  nanotubes. The structure of the resultant assemblies can be controlled through the sequence and length of the peptide subunits, which generates synthetic peptide filaments of similar dimensions to flagella and pili.

<sup>1</sup>Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA. <sup>2</sup>Department of Chemistry, Emory University, Atlanta, GA 30322, USA. <sup>3</sup>The Robert P. Apkarian Integrated Electron Microscopy Core (IEMC), Emory University, Atlanta, GA 30322, USA. <sup>4</sup>Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA. <sup>5</sup>Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA. ✉email: [vcontic@emory.edu](mailto:vcontic@emory.edu)

Peptide-based filamentous assemblies have been employed with great success over the past several decades for the fabrication of structurally ordered materials within the nano-scale size regime for applications in biomedicine and nanotechnology<sup>1–3</sup>. The molecular design of these materials has relied thus far on relatively simple sequence-structure correlations derived from structural informatics analysis of native protein folds. The mode of assembly involves self-association of structural subunits (i.e., protomers) into long non-covalent polymers that display helical symmetry. While *de novo* design has been employed frequently to identify suitable candidate peptides for construction of synthetic helical assemblies, the number of structures that have been characterized at near-atomic resolution is relatively limited<sup>4–13</sup>. As recent high-resolution structural analyses have demonstrated for both native and synthetic protein filaments, atomic models for helical assemblies based on low-resolution data are incomplete and often in error<sup>14–17</sup>. In the infrequent cases in which high resolution analysis has been performed on designed peptide filaments, the observed structures can differ significantly from the models that were employed as the basis for the design. These results suggest that our current knowledge of the design principles that underlie the formation of helical peptide and protein assemblies remains limited in scope, even for the simple structural motifs that have been employed as substrates in these studies. In addition, it raises the question of whether quaternary structure, i.e., interactions at the interfaces between protomers, is robust in sequence space, that is, designable<sup>18</sup>, and can be employed for reliable and predictable design of synthetic peptide filaments<sup>12,19–26</sup>.

Given the difficulties involved in correctly predicting quaternary structure, the *de novo* design of synthetic peptide assemblies requires validation of the predicted model through structural determination at near-atomic level resolution. This structural information is essential for understanding the inter-subunit contacts that stabilize the interfaces within filamentous peptide assemblies, especially as these interactions are critically important for functional properties such as mechanical response and bioactivity<sup>27–29</sup>. Structural polymorphism is frequently observed for designed helical peptide assemblies, as well as for biologically derived protein filaments that have been assembled *in vitro*<sup>19–23</sup>. Diverse populations of structurally distinct filaments can arise from the same peptide or protein sequence. Selection for monomorphic variants of defined and predictable structure represents a significant challenge to *de novo* design of peptide-based materials. Slight variations in preparative conditions can result in the formation of distinctly different structural variants in a manner that has yet to be understood at the level of atomic interactions<sup>24–26</sup>. Therefore, structural information that results from near-atomic resolution analyses of helical assemblies, yielding reliable atomic models, is essential for interrogating the limits of our knowledge of peptide design as well as for reverse engineering of structure to generate assemblies reliably and predictably.

We recently described the structural analysis of helical filaments derived from two related synthetic coiled-coil peptides, Form I and Form II (Fig. 1 and Supplementary Fig. 1)<sup>12</sup>. Atomic models were generated using Rosetta modeling of the peptides into the density map derived from electron cryomicroscopy (cryo-EM) with direct electron detection<sup>30</sup>. Although cryo-EM has come to dominate structural biology within the past several years, and has emerged as the main technique for determining the atomic structure of macromolecular complexes<sup>31–33</sup>, applications of this method in chemistry and materials science have been relatively sparse<sup>10–12,34,35</sup>. However, the resolution revolution in cryo-EM has enabled direct structural determination of designed helical assemblies at near-atomic resolution that were previously

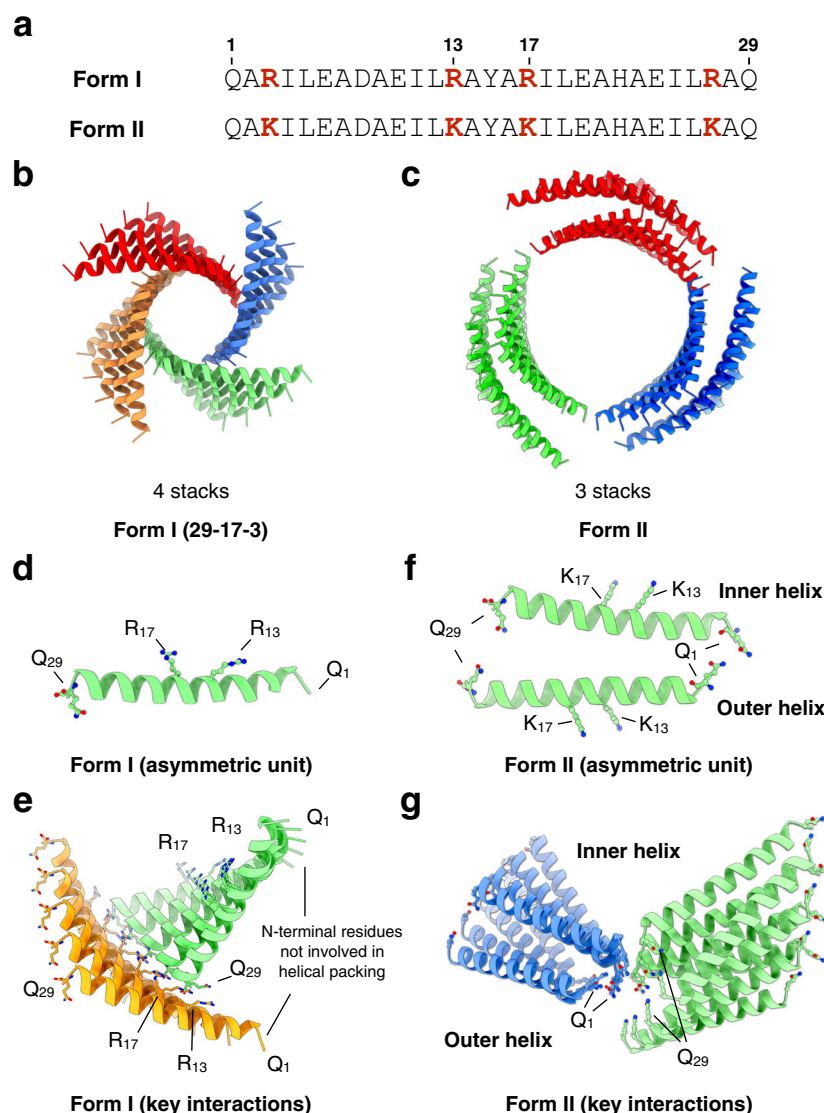
inaccessible. While the sequences of the Form I and Form II peptides differed solely in the substitution of arginine residues in the former sequence for lysine residues in the latter sequence (Fig. 1a), the corresponding atomic models for the filaments displayed significant differences in helical symmetry and the nature of the cohesive interactions between protomers. These differences were not anticipated in that the corresponding sequence substitutions occurred at positions that were not expected to participate directly in interfacial interactions (Supplementary Fig. 1)<sup>12,30</sup>. Moreover, site-directed mutagenesis of one or two residues within the respective sequences of Form I and Form II resulted in interconversion between the two alternative helical structures (*vide infra*). While the structural transition could be rationalized on the basis of the differences between the atomic models of the respective filaments, neither structure was predicted *a priori*, nor could the structures be considered as robust in sequence space, due to the significant impact of limited mutagenesis. Indeed, the dramatic change in quaternary structure due to the semi-conservative mutagenesis of one or two amino acids is suggestive of nearly chaotic behavior.

Here we report seven high-resolution cryo-EM structures of filamentous peptide nanotubes based on a cross- $\alpha$  supramolecular architecture that provide insight into the designability of helical peptide filaments. In addition, emergent structural behavior is observed among the corresponding helical filaments. A structural interaction is identified based on an appropriately placed pair of arginine residues (RxxxR) within the corresponding peptide sequences. This local interaction motif, designated as an arginine clasp, is robust within the context of these cross- $\alpha$  assemblies and can guide the coalescence of helical protofilaments into structurally defined cross- $\alpha$  nanotubes. Despite this observation, the arginine clasp motif does not appear to have been evolutionarily sampled as a mechanism to mediate similar helix-helix interactions in naturally occurring protein structures and, therefore, cannot be considered as natively designable.

## Results

**Structural differences between the Form I and Form II filaments.** The respective assemblies of Form I and Form II epitomized an unusual packing motif, recently designated as a cross- $\alpha$  structure<sup>36</sup> (alternatively, an  $\alpha$ -sheet structure<sup>37</sup>). In contrast to most structurally characterized filaments based on synthetic  $\alpha$ -helical coiled-coils<sup>38–46</sup>, the chain axes of the helical protomers are oriented in a plane nearly perpendicular to the long axis of the assembly. The initial reconstruction of Form II resulted in an atomic model of relatively low (circa 7-Å) resolution, in contrast to the higher (3.6-Å) resolution of the corresponding model for Form I (PDB ID: 3J89, [<https://www.rcsb.org/structure/3j89>]). In order to better compare the two models, cryo-EM data were collected on a fresh preparation of Form II filaments (Fig. 1c). Helical reconstruction afforded an atomic model at 4.2-Å resolution with a different helical symmetry (i.e., a rise of 1.93 Å and a rotation of 124.4°) than that proposed for the original model<sup>12</sup>. The higher resolution model of Form II is consistent with a parallel orientation of the inner-outer helix pair, which was suggested but could not be unambiguously determined in the previous analysis<sup>12</sup>.

A structural comparison of the two assemblies indicated that helical protomers were arranged in four (Form I) or three (Form II) cross-helical stacks (i.e., protofilaments) that self-associated laterally to form single-walled and double-walled nanotubes, respectively (Fig. 1). The axial interactions along the respective stacking interfaces were conserved between the different filaments. These interfaces were stabilized through a series of heterotypic knobs-into-holes (KIH) interactions between

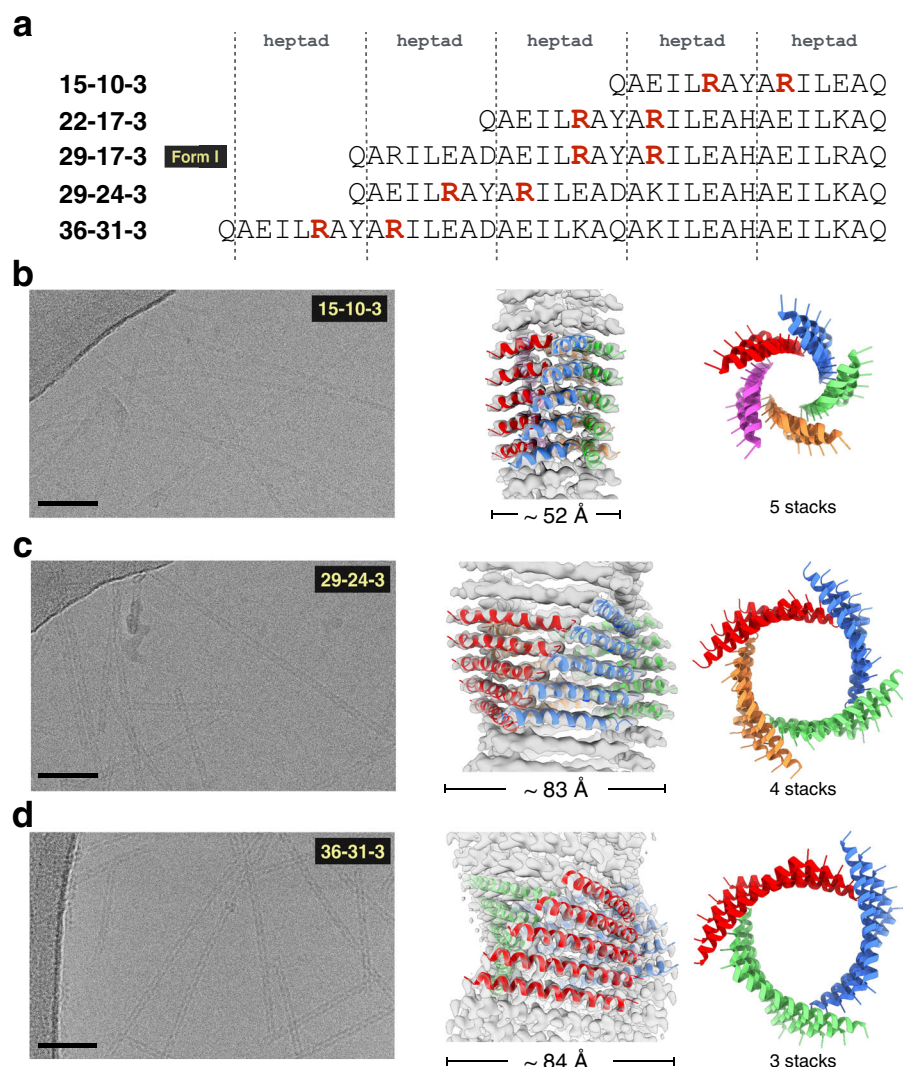


**Fig. 1 Structures and critical interactions in the Form peptide filaments.** **a** The 29-residue sequences of Form I and Form II peptides. **b, c** Cryo-EM structure of Form I (**b**) and Form II (**c**) filaments. **d** The building block (asymmetric unit) of Form I peptide with key residues shown in sticks. **e** Essential interactions between adjacent helix stacks in Form I filaments that maintain the helical packing. **f** The building block (asymmetric unit) of the re-examined Form II peptide with key residues shown in sticks. **g** Essential interactions between adjacent helix stacks in Form II filaments that maintain the helical packing.

hydrophobic residues at the a/d and c/f faces of axially adjacent helices within a protofilament (Supplementary Fig. 1)<sup>12,37</sup>. The main distinguishing feature between the Form I and Form II assemblies was the mode of lateral association between the respective cross  $\alpha$ -helical stacks. The Form I structure displayed a unique interaction involving a pair of arginine residues, R13 and R17, that formed a network of hydrogen bonds and electrostatic interactions with the C-terminal residues of helices in an adjacent protofilament (Fig. 1e). Mutagenesis of either of these residues to lysine resulted in conversion of the assembly to a Form II-like structure<sup>12</sup>. Conversely, a K13R, K17R double mutant of the Form II sequence resulted in conversion of the resultant assembly to a Form I-like structure, which suggested that the presence of both arginine residues was necessary for formation of the Form I structure. High-resolution cryo-EM structural analysis led to identification of the critical interactions that were responsible for the structural differences between the Form I and Form II filaments, as well as the ability to directly interrogate the

importance of these interactions through site-directed mutagenesis.

These results suggested that the RxxxR structural unit, in which the two arginine residues, R13 and R17, are arranged as  $i, i + 4$  within the Form I sequence, defined a side-to-end helical interaction motif within the assembly<sup>47</sup>. While interactions involving individual RxxxR motifs may be relatively weak, many such interactions occurred at the lateral interfaces between protofilaments along the contour length of the Form I structure. Consequently, the influence of these localized interactions on the stability of the helical assembly was magnified significantly. Numerous instances of the stabilizing effect of structural interactions between appropriately substituted  $i, i + 4$  residues have been described within the sequences of helical peptides<sup>48–57</sup>. In addition, specific interaction motifs such as the left-handed leucine zipper, i.e., LxxxLxx<sup>58–60</sup>, and the right-handed glycine zipper<sup>61–63</sup>, i.e., GxxxG, have been observed to promote and stabilize lateral (side-to-side) interactions in coiled-coils and



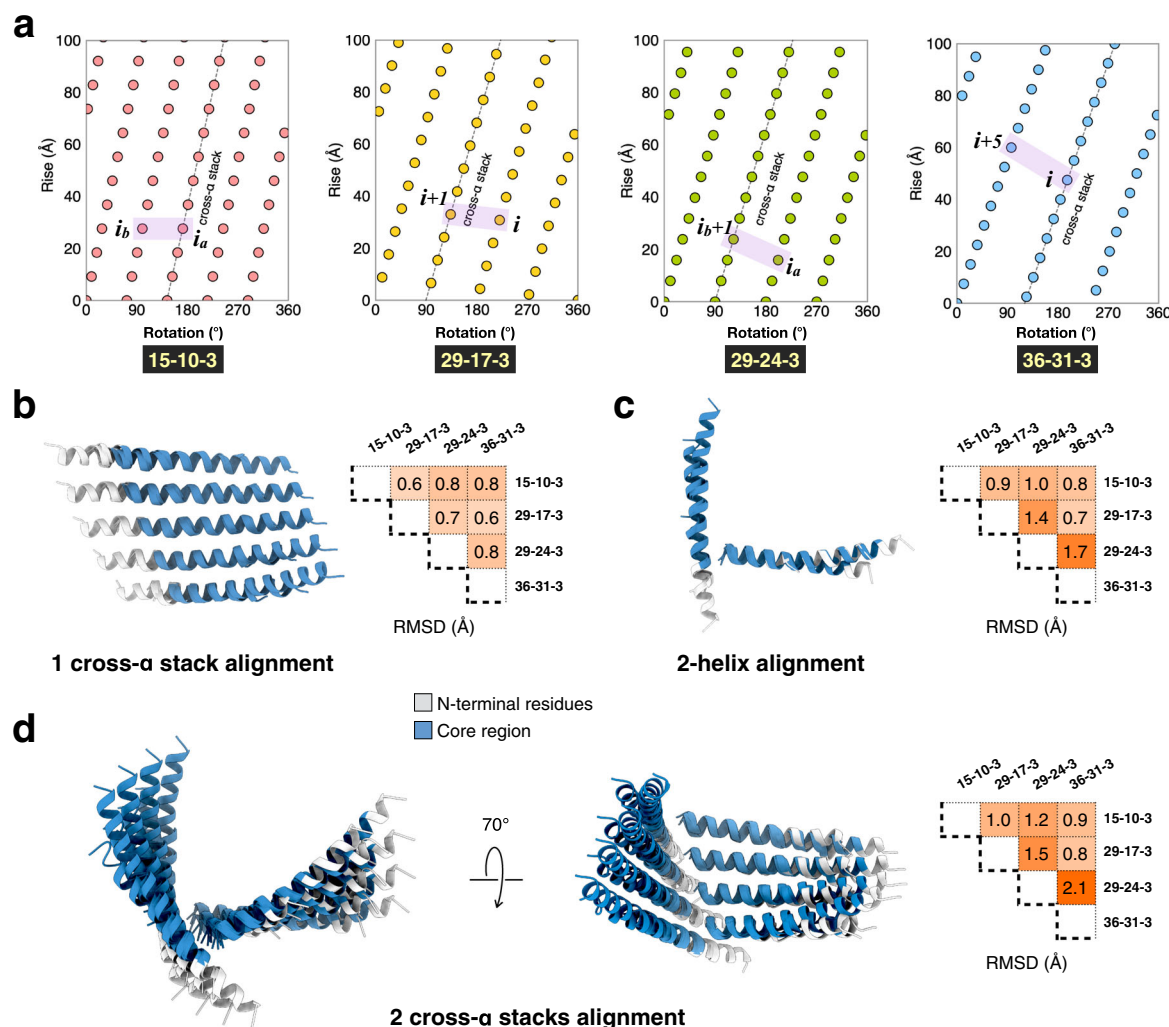
**Fig. 2 Cryo-EM of arginine clasp peptide filaments based on Form I.** **a** Sequences of the Form I peptide and four designed peptides derived from Form I. **b–d** Cryo-EM of peptides 15-10-3, 29-24-3, and 36-31-3. Representative raw micrographs are shown on the left (scale bar 500 Å) out of a total number of 84, 279, and 314 images recorded, respectively. The cryo-EM reconstructions are shown in the middle, with models fitted into the maps. The top views of the atomic models are shown on the right.

transmembrane helical bundles, respectively<sup>64</sup>. The crossing angle,  $\Omega$ , between laterally interacting pairs of helices in the Form I structure was circa  $94^\circ$ . In contrast, the leucine and glycine zipper motifs define side-to-side association between helices and display more acute crossing angles ( $\Omega = -20^\circ$  and  $+40^\circ$ , respectively)<sup>58,63,65</sup>. PISA<sup>66</sup> analysis indicated that the lateral interface between interacting helices in the Form I protofilaments buried approximately  $275 \text{ \AA}^2/\text{peptide}$ . For comparison, the coiled-coil interaction along the axial direction of the Form I protofilaments defined an interface of approximately  $700 \text{ \AA}^2$  of surface area/peptide with a helix crossing angle of circa  $-11.5^\circ$ . We designated this RxxxR interaction motif as an arginine clasp, since it mediated a distinctive local interaction between appropriately oriented protomers on structurally adjacent protofilaments.

**Dependence of filament structure on peptide length.** In order to test the utility of the arginine clasp in mediating interactions within a defined structural context, we designed a series of peptides derived from the Form I sequence (Fig. 2, Supplementary Figs. 2 and 3). In the latter structure, the N-terminal heptad

sequence, upstream of the RxxxR motif, was unstructured and did not contribute to the helix-helix interactions. We reasoned that the major contribution to the structural integrity of the helical filament was the region in Form I sequence space from the RxxxR motif to the C-terminus of the peptide. The corresponding peptide segment contained the critical structural elements responsible for the lateral and axial interactions. Therefore, the designed series of Form I-like peptides encompassed minimal peptide sequences in which the structurally determinative distance between the RxxxR motif and the C-terminus was systematically varied over the range from 2 to 5 heptad repeats (Fig. 2). In each case, the most N-terminal heptad contained the RxxxR motif and, in order to prevent competition, lysine residues occupied all of the other basic amino acid sites at solvent-contacting b- and e-positions within the respective heptad sequences. Lysines contribute to stabilizing electrostatic interactions with glutamic acid residues that occur between peptides within a cross- $\alpha$  stack, but minimally to the cohesive interactions between stacks<sup>12</sup>. The nomenclature of the peptides was changed to one that reflected the structural elements that were considered to be critical to the design. All peptides were designated as (x-y-z), in which x refers





**Fig. 3 Structural conservation of the Form I-based arginine clasp filaments.** **a** Helical nets for the peptide 15-10-3, 29-17-3 (Form I), 29-24-3, and 36-31-3 filaments. The helical nets show the unrolled surface lattice viewed from the outside of the filament. One of the right-handed n-start helices-associated with a cross- $\alpha$  protofilament within the respective assembly-is indicated with a straight line in the corresponding helical net diagram. Adjacent helices shown in **c** with the conserved interactions are highlighted in a purple box. **b** Structural alignment of a single cross- $\alpha$  stack containing five helices (left) and their C $\alpha$  RMSD (right). **c** Structural alignment of two adjacent helices (left) and their C $\alpha$  RMSD (right). **d** Structural alignment of two cross- $\alpha$  stacks (left) and their C $\alpha$  RMSD (right).

to peptide length,  $y$  refers to the number of residues from the first arginine to the C-terminus, and  $z$  to the number of residues between the two arginines (Fig. 2). The original Form I sequence corresponded to peptide 29-17-3 under this system of nomenclature.

The designed peptide sequences were based on the assumption that the helical symmetry of the respective filaments and, consequently, the axial and lateral interfaces of the Form I filament, would be conserved in the resultant assemblies. In this scenario, the RxxxR motifs within a given protofilament would interact selectively with the C-terminal residues within adjacent helices oriented  $(i+1)$  along a left-handed 1-start helix oriented similarly to that of the Form I filament (Figs. 1b and 3a). If a given peptide self-associated into a Form I-like structure, the diameter of the resultant nanotubes would depend on the length of the peptide segment from the first arginine to the C-terminus (i.e., the  $y$ -value). The Form I assembly displayed a nearly square cross-section, in which the nanotube diameter was determined by the seventeen amino acid length of the interacting peptide segments within a laterally adjacent pair of protofilaments within the assembly (Fig. 1). If the Form I-like structure were retained,

then the width of the assembly should be predictable from the position of the arginine clasp motif and the length of the peptide, as these parameters determine the cross-sectional dimensions of the corresponding assemblies. If the lengths of the interacting peptide segments were shorter than for Form I, a narrower nanotube would result and, if longer, the corresponding nanotube should display a wider diameter.

All four peptides in the (x-y-3) series adopted an  $\alpha$ -helical conformation and formed thin filaments in aqueous buffers (Supplementary Figs. 4 and 5). Three filaments, derived from peptides 15-10-3, 29-24-3, and 36-31-3, were selected for further analysis using cryo-EM imaging with direct electron detection. The filament derived from 22-17-3 was not analyzed further, as it was expected that the structure would be similar to the previously described Form I/29-17-3, since it corresponded to a truncation of the unstructured N-terminal heptad. The iterative helical real space reconstruction (IHRSR) algorithm<sup>67–70</sup> was employed to generate 3D reconstructions at final resolutions (Supplementary Table 1) of 4.2-Å, 4.1-Å, and 4.0-Å for the 15-10-3, the 29-24-3, and the 36-31-3 filaments, respectively (Fig. 2b). Surprisingly, the number of protofilaments (i.e., cross- $\alpha$  helical stacks) within the

nanotubes was found to have an inverse relationship to peptide length. Five protofilaments interacted to form the 15-10-3 filament, four interacted in the 29-24-3 filament, and three interacted in the 36-31-3 filament. The Form I/29-17-3 structure fits the observed trend, in that the 17 amino acid length of the structurally determinative segment lay between the 10 amino acid and 24 amino acid segment length of 15-10-3 and 29-24-3, respectively.

High-resolution structural data enabled a detailed comparison between the arrangement of peptides within the respective atomic models (Fig. 3). In each case, the  $\alpha$ -helical protomers within a protofilament stacked along a right-handed *n*-start helix (Fig. 3a), in which the value of *n* corresponded to the number of protofilaments within a nanotube. For example, each 5-start helix of 15-10-3 passed through every fifth protomer and therefore corresponded to the axial stacking direction in a protofilament (Figs. 2b and 3a). Similarly, the 4-start helices of 29-17-3 and 29-24-3 and the 3-start helices of 36-31-3, passing through every fourth subunit and every third subunit, respectively, were associated with the axial stacking direction of the protofilaments within the corresponding assemblies. The arrangement of protofilaments within the respective assemblies can be clearly discerned in the corresponding helical net diagrams (Fig. 3a). The right-handed curvature of the protofilament derived from a two-residue progressive displacement between successive helical protomers along the rise of the *n*-start helices. This displacement was a consequence of a heterotypic coiled-coil interaction along the stacking interface between the two offset hydrophobic faces (a/d and c/f) of the protomers, as predicted by Walshaw and Woolfson for type III coiled-coil structures (Supplementary Fig. 1)<sup>37</sup>. Structural overlays of helices within a protofilament (Fig. 3b) showed reasonably good alignment (root mean square standard deviation (RMSD) < 1 Å), which presumably resulted from the constraints of KIH packing at the coiled-coil interfaces despite differences that were observed at the lateral packing interfaces (Fig. 3c, d).

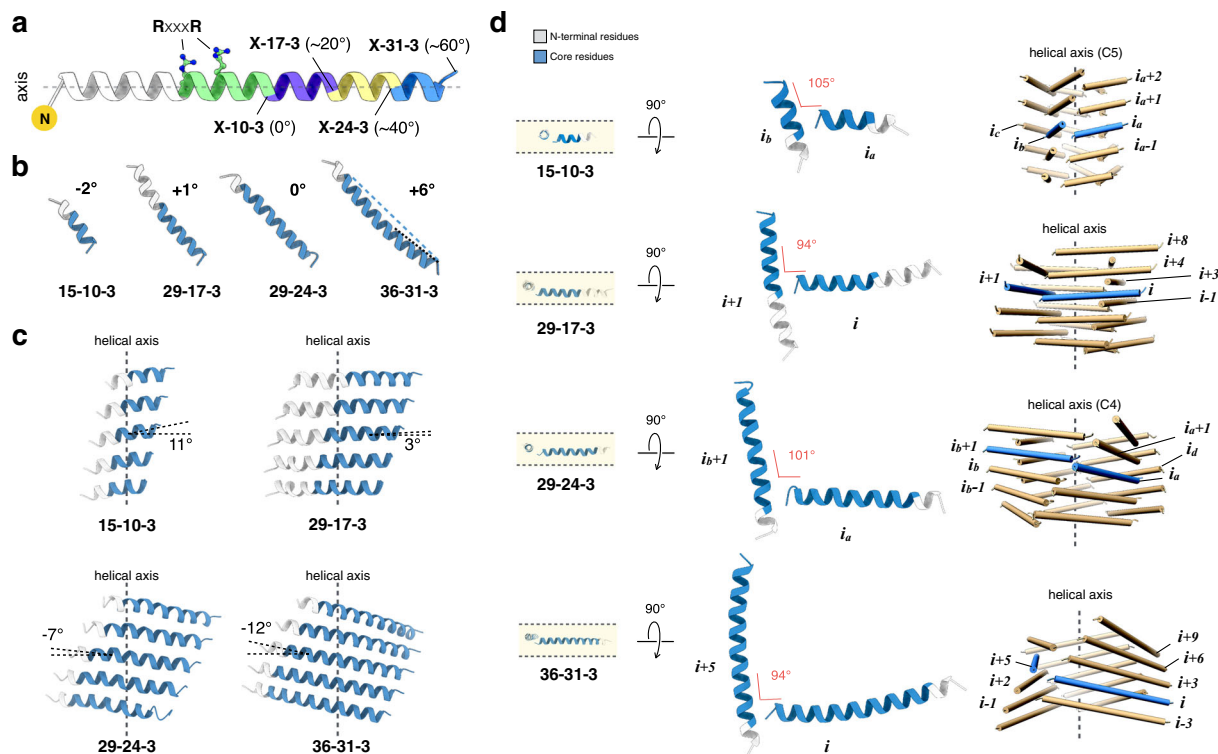
Structural comparison of the filaments suggested that three mechanisms could be distinguished through which the packing of peptides adjusted itself in order to accommodate the increase in peptide length (Fig. 4). First, the tilt angle of the core peptide with respect to the plane normal to the filament axis became progressively more negative with increasing peptide length (Fig. 4c). Despite the differences in cross-sectional geometry for the respective filaments, the helix crossing angles were remarkably similar and lay within the range from 95° to 105° (Fig. 4d). We postulate that the arginine clasp interaction required that the crossing angle was within this range in order to effectively interact with the C-terminus of an adjacent helix. Adjustment of the tilt angle of helical protomers provided a mechanism to satisfy this geometric constraint. Second, for the longer peptide sequences, the lateral interaction between protofilaments could not be maintained between nearest neighbor helices (Figs. 3a and 4d). For the 29-24-3 peptide, the interaction occurred between the C-terminus of peptide (*i<sub>a</sub>*) and the RxxxR motif of peptide (*i<sub>b</sub>* + 1), in contrast to the C5-symmetric structure of 15-10-3 that involved helices *i<sub>a</sub>* and *i<sub>b</sub>*. Similarly, the filament structure of 36-31-3 revealed that the C-terminus of peptide (*i*) interacted with the axially offset RxxxR motif of peptide (*i* + 5) within the 1-start (C1) symmetry of the assembly. The length and flexibility of the arginine side-chains enabled the conservation of the RxxxR lateral interaction between protofilaments despite differences in peptide length. Finally, while the local helix crossing angle associated with the arginine clasp interaction was maintained at 94°, the distal segments of the 36-31-3 helices bend at a more acute crossing angle of 80° in order to maintain the overall helical symmetry of the filament. These observations suggested that peptide length

determined the helical symmetry of the respective cross- $\alpha$  nanotubes through selection of number and relative orientation of protofilaments that could maintain the structurally critical cohesive interactions at the axial and lateral interfaces (Fig. 2).

**Effect of arginine position.** The two arginine residues in the RxxxR clasp motif were positioned on the same side of the helix in an *i, i* + 4 orientation. This placement enabled lateral association between protofilaments in the structures of the cross- $\alpha$  nanotubes described above. However, a structural arrangement in which two arginines are placed in an *i, i* + 3 orientation, that is, as an RxxR motif, could potentially promote a similar interaction. The two arginines in the RxxR motif remain on the same face of the helical protomer, but would be arranged in a different orientation (angular displacement = -60° and axial displacement = 4.5 Å) in comparison to the RxxxR motif (angular displacement = +40° and axial displacement = 6.0 Å). Previous research demonstrated that appropriately selected residues in an *i, i* + 3 orientation could mediate stabilizing intra- and inter-helix interactions<sup>48,57</sup>.

The Form I sequence was modified to test the effect of arginine placement on the structure of the corresponding peptide assemblies (Fig. 5). Arginines were placed in an *i, i* + 3 (RxxR) arrangement within a heptad repeat located proximal to the N-terminus. The sequence at the axial interface was conserved in order to maintain the cross- $\alpha$  stacking interactions within a protofilament. The resultant peptide was designated as 29-20-2 using the same (x-y-z) nomenclature described above, with the value of *z* equal to 2 for the RxxR motif (Supplementary Figs. 6 and 7). The 29-20-2 peptide adopted an  $\alpha$ -helical conformation in solution and formed high aspect-ratio tubular filaments (Supplementary Figs. 8 and 9). However, the observed widths of the filaments were significantly larger, circa 12 nm, than the corresponding values for peptides in the (x-y-3) series. A 3.8-Å resolution helical reconstruction from the corresponding cryo-EM images afforded an atomic model for the 29-20-2 filament (Supplementary Table 2). Unexpectedly, the filament formed a double-walled cross- $\alpha$  nanotube that was similar in structure to that of the Form II filament and a structurally related mutant (Fig. 5). Remarkably, the arginine residues in RxxR motif were not involved in the lateral interactions between cross- $\alpha$  helical protofilaments. Instead, cohesive interactions were mediated through hydrogen-bonding between the N- and C-terminal glutamine residues from helices on laterally adjacent protofilaments, as in the Form II-related structures (Fig. 1g).

Each nanotube was derived from coalescence of three protofilaments derived from the dimeric protomer. However, these were arranged in different symmetries; C3 for 29-20-2 and C1 for Form II. In either case, the hydrophobic stacking interactions were oriented along the direction of the three 3-start helices of the respective nanotubes and coincided with the cross- $\alpha$  protofilaments (Fig. 5g). Among all three structures, the helices within the dimeric asymmetric unit could be superimposed with only minimal differences (Fig. 5e). Structural overlays of laterally interacting protofilaments indicated a conservation of the interfaces between the three filament structures (Fig. 5f). The main difference between the structures was manifested in terms of the lateral interactions between the termini of the protofilaments. These interactions occurred along the direction of the left-handed 12-start and 11-start helices for 29-20-2 and Form II, respectively, as illustrated in the the helical net diagrams of the corresponding assemblies (Fig. 5g). This lability of helical symmetry is common among filamentous peptide assemblies and has been observed even for structures in which the interfacial interactions between protomers are largely conserved. These



**Fig. 4 Three degrees of freedom to maintain the conserved arginine clasp contacts.** **a** The building blocks of four types of peptide filaments. The N-terminal residues not involved in filament packing are colored white. The RxxxR motif is highlighted. The core region of X-10-3 peptide is colored in green. The core region of X-17-3, X-24-3, and X-31-3 have one (magenta), two (yellow) or three (blue) additional heptads, respectively. **b** The helix tilt angle of the core region in four types of peptide filaments. **c** The cross- $\alpha$  stack within the respective protofilaments, indicating tilt angle from the plane normal to the helical axis. **d** Two adjacent helices that make the key contacts in these peptide filaments. The adjacent helices are oriented to be in the same plane (left), and the angle between the two helices are indicated (middle). The relative positions of these two adjacent helices (blue) in the filaments are shown (right).

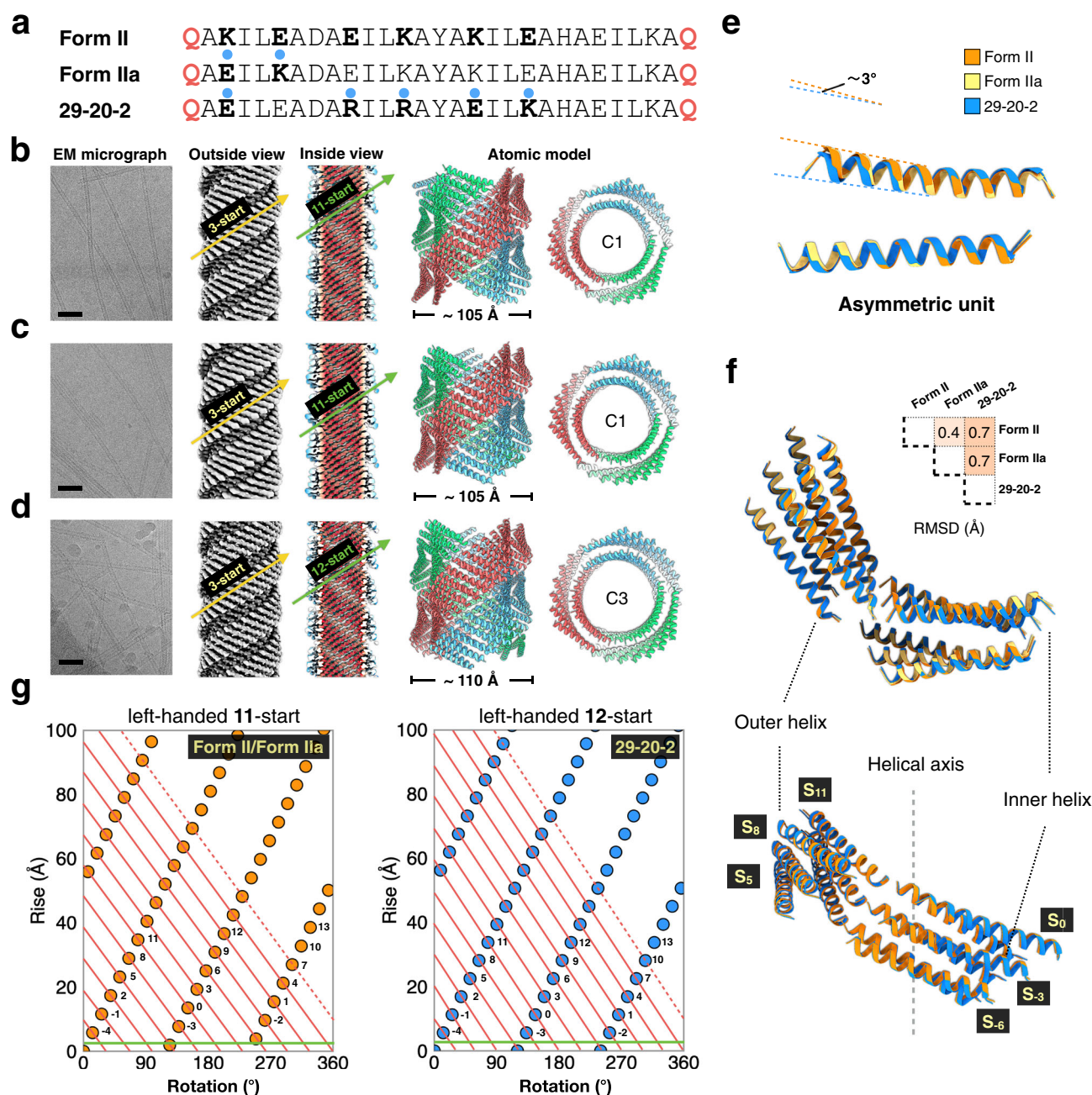
differences can arise solely as a result of minor changes in interfacial packing orientation between protomers in the filament<sup>10,17,23,71–73</sup>.

The structural analysis of the 29-20-2 filament provided evidence that the formation of the arginine clasp interaction required not only the presence of the two arginines, but also the correct orientation, RxxxR, of the corresponding residues within the sequences of the respective helical protomers. The relatively minor substitution of an RxxR motif for an RxxxR motifs drove the lateral association down an alternative pathway that resulted in formation of a double-walled nanotube in which lateral interactions occurred between terminal glutamine residues. The latter interaction must be relatively weak as it can only sustain lateral association for a peptide having this single length. In contrast, the RxxxR motif could mediate lateral interactions over a wider range of peptide lengths from 2 to 5 heptad repeats. However, the axial interactions along the contour length of a protofilament for the Form II-like assemblies appeared quite robust and comparable in energy and buried surface area to that of Form I protofilaments.

**Native designability of the helical interfaces.** The conserved structural context of the RxxxR motif in the Form I-like assemblies raised the question of its native designability, i.e., whether it occurred as a structural feature within natural proteins in which the helices were oriented similarly to those participating in the lateral interactions of Form I-like filaments. To address this question, we defined a local lateral interface motif from each of the filament structures solved here, and identified sub-structures

in the PDB that displayed similar backbone geometries. The results of this analysis are reported for the 36-31-3 filament (Fig. 6), although similar results were obtained for the other Form I-like assemblies. The motif comprised a seven-residue helical fragment containing the RxxxR motif and a five-residue fragment from the C-terminus of a laterally interacting helix (Fig. 6a). Despite this compact size, the closest matching geometries from the PDB for this lateral interaction exhibited root mean squared deviations (RMSDs) over backbone atoms from the query between 0.6–0.8 Å. Furthermore, even these were rare geometries, as the RMSD of subsequent closest matches rose quickly (Fig. 6c). To demonstrate the rarity of this geometry, we compared these statistics with those of a control motif of equal complexity, corresponding to a seven-residue helix fragment, as described above, but a different five-residue helix fragment from an axially adjacent partner within the same protofilament (Fig. 6a). In all cases, this reference motif gave much more robust statistics, with many low-RMSD matches and sequence preferences consistent with the amino-acid choices in the assembly. The latter results suggested that the stacking interactions within a protofilament were locally consistent with typical helix-helix interactions, which implies that the axial helix stacking interactions are well represented in naturally occurring proteins and, therefore, are natively designable. On the other hand, the sequence statistics of the closest hits to the lateral interaction motif did not reveal a strong RxxxR signal. The only strongly conserved position corresponded to the second Arg, but the sequence logo analysis indicated that its overwhelming preference was for Gly (Fig. 6d). Upon closer inspection, we found the nature of this preference to be entirely



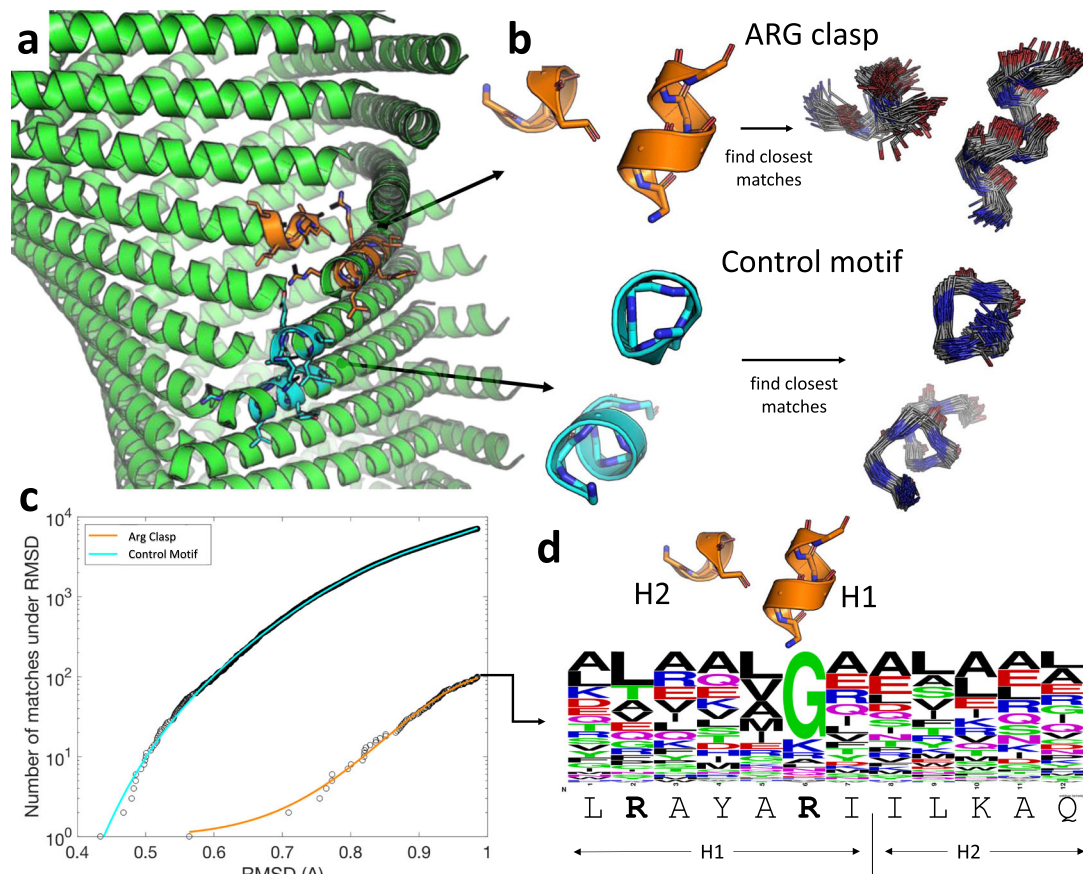


**Fig. 5 Structural comparison of the Form II-like peptides.** **a** Sequences of Form II, Form IIa and 29-20-2 peptides. Residues changed from original the Form II sequences are highlighted with blue dots. **b–d** Cryo-EM of peptides Form II, Form IIa, and 29-20-2. Representative raw micrographs are shown on the left (scale bar 500 Å) out of a total number of 171, 233, and 347 images recorded, respectively. The cryo-EM reconstructions are shown in the middle, one view from outside highlighting right-handed 3-start helices, and another view from the inside of the nanotube lumen highlighting the left handed 11-start or 12-start helices. The side and top views of the atomic models are shown on the right. **e** Structural alignment of the asymmetric unit (helix-dimer) of the three peptide filaments. **f** Structural alignment of six asymmetric units of the three peptide filaments. The alignment RMSD is shown. **g** Helical nets for the Form II and 29-20-2 filaments. The helical nets show the unrolled surface lattice viewed from the outside of the filament. The right-handed 3-start helices can be clearly discerned for each assembly, although not explicitly highlighted. The positions of the left-handed 11- and 12-start helices are identified and are defined by the red lines crossing the green horizontal line. Subunits labeled  $n$  and  $n + 11$  are connected with red lines.

unrelated to the arginine clasp. Rather, some of the closest matches arose in the context of an internal helix-helix interaction between two closely-approaching helices, necessitating a Gly at the position above (Supplementary Fig. 10). The results suggested that the arginine clasp motif, which involves a helix terminus interacting with the side of another helix at a nearly 100° angle, was not common within the structures of naturally occurring proteins and, consequently, not natively designable.

This outcome led us to speculate whether structural motifs other than RxxxR could mediate the lateral interaction between helical stacks and promote nanotube formation. The 36-31-3 filament was employed as the starting point for a computational optimization using the design engine dTERMen<sup>74,75</sup>. A solubility constraint was applied in which the occupancy of the solvent-contacting b/e/g residues within the heptad repeats was constrained to polar residues. Computational optimization

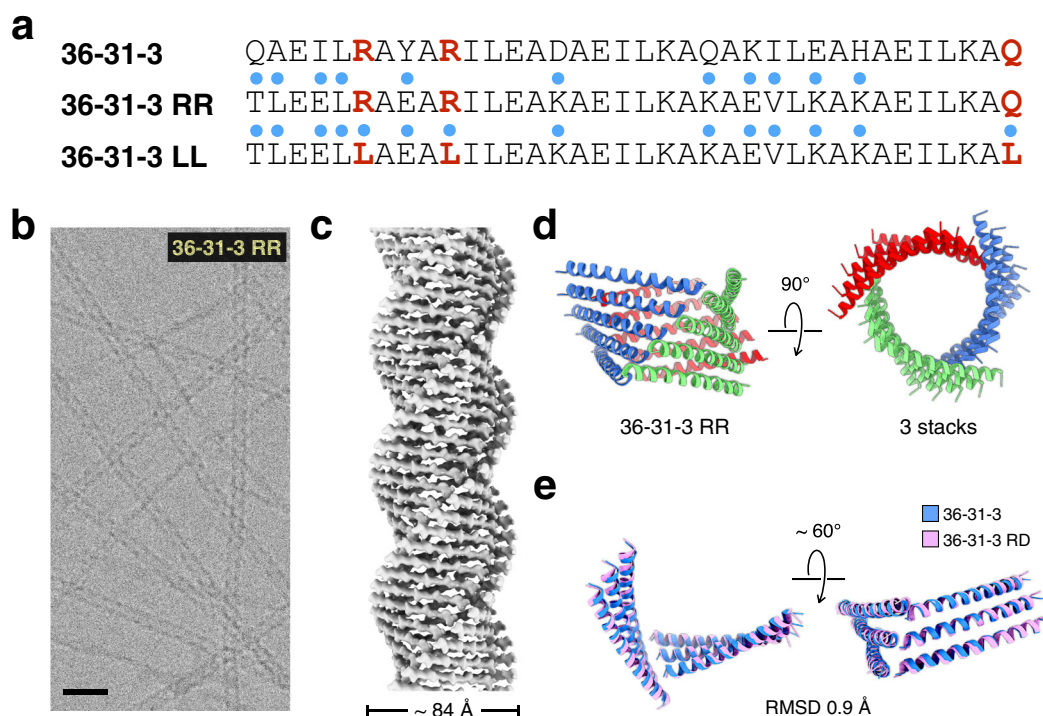




**Fig. 6 Structural analysis of arginine clasp designability.** **a** The 36-31-3 assembly showing the definition of the motif used to interrogate for the presence of the arginine clasp in native structures (orange), as well as a control motif of the same size and complexity (cyan). **b** Structural alignments of resulting 100 closest matches for both motifs. The control motif, corresponding to the helix stacking interaction within a protofilament, clearly corresponds to a well-populated structural attractor, while the arginine clasp motif brings in much more diverse hits. **c** RMSD divergence plots, showing the number of matches below a given RMSD cutoff for both motifs. Matches to the Arg clasp motif are rare as the RMSD required to get up to a certain number of matches grows much more rapidly than for the control. **d** Sequence logo of the closest non-redundant sequence matches to the Arg clasp fragment, showing the query sequence underneath. The native statistics do not bear out the Arg clasp sequence motif, and the one conserved position corresponds to a Gly rather than an Arg. This preference is due to close helix-helix interactions (see Supplementary Fig. 10) and represents a different structural circumstance than the close-to-right angle docking of one helix into another, as seen in our assemblies.

returned a sequence in which the RxxxR motif was replaced with an LxxxL motif and the C-terminal glutamine was replaced with a leucine residue (Fig. 7a). In the corresponding model, the network of hydrogen-bonding interactions at the lateral interface was replaced with hydrophobic interactions. In a control experiment, a second computational optimization was performed in which positions 6, 7, 10, and 36 were constrained to those in the original peptide sequence. The arginine residues at positions 6 and 10 defined the RxxxR motif, while residues at positions 7 and 36 (Ala and Gln, respectively) appeared to be important for organizing this motif. A comparison between the two sequences, 36-31-3\_LL and 36-31-3\_RR, could provide additional insight into importance of structural preservation of the arginine clasp interaction. Even in the absence of explicit sequence constraints, the resultant computational models retained most of the original amino acid residues at the a/d and c/f positions that mediated the heterotypic hydrophobic interaction at the axial stacking interfaces of the protofilament (Fig. 7a). Coupled with the structural similarity of axial stacking interactions between the Form I-like and Form II-like assemblies, these results provided further support that the helix-helix interactions along the axial interface were natively designable (Fig. 6c).

The computationally designed peptides, 36-31-3\_LL and 36-31-3\_RR, displayed an  $\alpha$ -helical conformation and formed high aspect-ratio filaments of similar width to those of 36-31-3 (Supplementary Figs. 2–5). However, filaments derived from the 36-31-3\_LL peptide unraveled into individual protofilaments over time (Supplementary Fig. 5e). In contrast, 36-31-3\_RR filaments persisted over weeks in solution (Fig. 7b and Supplementary Fig. 5f). The differences in stability between the two computationally designed filaments suggested that the cohesive interactions were much weaker in the absence of the arginine clasp interaction. The most probable explanation resided in the replacement of the hydrogen-bonding network of the arginine clasp interaction within 36-31-3\_RR with the network of hydrophobic interactions within 36-31-3\_LL. In order to confirm that the arginine clasp was conserved, filaments of 36-31-3\_RR were analyzed by cryo-EM. A 4.5-Å resolution reconstruction of the 36-31-3\_RR filaments displayed a nearly identical structure to that of the original 36-31-3 filaments, with very slight differences in the rise (2.49 versus 2.51 Å/subunit) and the rotation (124.5 versus 124.0 degrees/subunit) for the respective 1-start helices of the 3D models. Accordingly, repeating the statistical analysis of this assembly's RxxxR motif produced very similar results as for



**Fig. 7 Cryo-EM of the computationally re-designed (36-31-3) peptide nanotube.** **a** Sequences of the computationally redesigned peptides 36-31-3\_RR and 36-31-3\_LL in comparison to the parent sequence. Residues changed from original the 36-31-3 sequences are highlighted with blue dots on top. **b** Representative cryo-EM image of 36-31-3\_RR peptide filaments out of a total number of 166 images recorded. Scale bar is 20 nm. **c** Cryo-EM reconstruction of the 36-31-3\_RR filaments. **d** Side view and top view of the atomic model. **e** Alignment of two adjacent cross- $\alpha$  stacks between 36-31-3 (blue) and 36-31-3\_RR (magenta).

the original 36-31-3 filament, indicating the native rarity of the geometry.

## Discussion

Tayeb-Fligelman et al., reported the initial example of a cross- $\alpha$  fibril structure from crystallographic analysis of PSMa3, a short cytolytic peptide secreted from virulent strains of *S. aureus*<sup>36</sup>. The helices were observed to pack in a stacked bilayer array, in which individual peptides were oriented perpendicular to the long axis of the crystallographically defined filament. In contrast to the cross- $\alpha$  nanotubes described here, translational symmetry restricted coiling of the cross- $\alpha$  filaments. Moreover, the composition of the hydrophobic interface was rich in phenylalanine residues and did not display the KIH packing characteristically associated with coiled-coils. Subsequently, Zhang et al.<sup>7</sup>, reported a series of peptides based on designed coiled-coil sequences that assembled into filaments. Crystallographic analysis demonstrated that the peptides formed cross- $\alpha$  super-helical arrays through stacking of parallel dimers in an alternating antiparallel orientation. The filaments displayed limited interfaces over which KIH packing was observed between protomers, in contrast to the cross- $\alpha$  protofilaments in the nanotubes that we have described. Supercoiling of filaments was only observed under conditions in which the cross- $\alpha$  orientation was not maintained in the extended structure. In addition, our cross- $\alpha$  nanotubes differed from classical coiled-coils, in that the helical packing in a protofilament displayed an open or extended mode of association that is reminiscent of solenoidal tandem repeat proteins<sup>10,11,76,77</sup>. Most coiled-coils form closed assemblies of defined oligomerization state, which are often, but not exclusively, based on cyclic symmetry in which the helical protomers are aligned nearly parallel to the super-helical axis<sup>78,79</sup>. Surprisingly, the closest structural analogs of these cross- $\alpha$  nanotubes were observed among helical

supramolecular assemblies derived from oligoureia foldamers having similar polar patterning based on pentad repeat sequences<sup>80</sup>. Crystallographic characterization of the latter assemblies revealed that the oligoureia nanotubes were composed of protofilaments corresponding to stacked helical subunits that propagated along right-handed 2-start or 6-start helices.

Cross- $\beta$  assemblies<sup>81</sup> represent a natural point of comparison to the cross- $\alpha$  nanotubes. The respective protofilaments differ in that cross- $\beta$  architecture is primarily stabilized through main-chain hydrogen-bonding interactions along the contour length of the assembly. In contrast, the cohesive interactions between protomers within the cross- $\alpha$  protofilaments depended exclusively on side chain interactions. Surprisingly, few near-atomic resolution structures of synthetic cross- $\beta$  nanotubes derived from designed  $\beta$ -strand sequences have been reported, presumably due to the propensity for structural polymorphism and concomitant difficulty in isolation of monomorphic variants<sup>24–26</sup>. However, the formation of cross- $\beta$  nanotubes of uniform diameter has been reported for conformationally constrained cyclic peptide sequences<sup>8,82</sup>. More recently, cryo-EM helical reconstruction of Orb2 filaments from *D. melanogaster* brain revealed a minimal diameter nanotube composed of three identical cross- $\beta$  protofilaments in a C3-symmetric arrangement<sup>83</sup>. Individual cross- $\beta$  protofilaments were based on  $\beta$ -hairpin protomers interacting through a combination of main-chain and side-chain hydrogen-bonding interactions. Similar to the cross- $\alpha$  nanotubes, side-chain hydrogen bond formation provided the cohesive interactions between protofilaments that stabilized the cross- $\beta$  assembly. These results suggested that directional non-covalent interactions, e.g., the formation of complementary hydrogen-bonded networks or metal ion coordination<sup>84,85</sup>, could provide a mechanism for rational design of defined higher-order structure in synthetic peptide and protein assemblies.

The de novo design of peptide and protein assemblies requires not only the control of interfacial interactions, but also the folding energetics of the individual subunits<sup>10,18</sup>. In this study, the protomer sequences were based on the highly designable coiled-coil motif, which has been demonstrated to display robust folding energetics and structural stability in sequence space. The axial stacking interactions within a protofilament were highly conserved between assemblies differing in both the number of protofilaments and nature of the lateral interactions between protofilaments. The structure of the axial interfaces within a protofilament were consistent with helix-helix interactions in canonical coiled-coils, even though the open-ended supramolecular architecture of the helical assemblies was quite distinct from the closed cyclic oligomers that are typically associated with such helical bundles. In contrast, the lateral interactions between protofilaments have not been commonly observed within the structures of native proteins having similar helix-helix orientations. Notably, we corroborated that the lateral interface between Form I-like protofilaments was based on the unusual interaction of an RxxxR motif with the C-terminal residues of a helix in a structurally adjacent protofilament. Our investigation of the designability of this arginine clasp motif suggested that it does not appear to have been evolutionarily sampled as a mechanism to mediate helix-helix interactions in native protein structures and cannot be considered as natively designable. Nevertheless, the RxxxR motif appeared robust as a design element within a structurally related family of cross  $\alpha$ -helical assemblies and was preferred to other more designable interaction motifs, such as the common LxxxL motif, within an otherwise identical peptide sequence context. We demonstrated that within a designed series of coiled-coil peptides that the RxxxR motif could be employed to control nanotube structure through peptide length. The absence of the RxxxR motif drove the assembly down an alternative pathway, which suggested that the arginine clasp interaction was not robust outside of this relatively limited sequence context<sup>12</sup>.

From this perspective, we propose that the RxxxR clasp may be representative of a more general class of local interaction motifs that can potentially mediate quaternary interactions within a specific structural context in peptide and protein assemblies<sup>18</sup>. The influence of these localized interaction motifs may be magnified due to the presence of multiple copies in a structurally similar context along the contour length of the helical assembly, which enables selection for specific modes of self-association of protofilaments into higher order structures. Similar local interaction motifs, e.g., Phe-Phe diads<sup>86–88</sup> and glutamine/asparagine ladders<sup>83,89</sup>, among others, have been identified that mediate higher-order interactions in cross- $\beta$  assemblies. The relatively low information content of such short sequence motifs hinders a systematic sequence-based analysis of their role in higher order assembly processes. The challenge will be to effectively identify such motifs in order to exploit them for de novo design of helical assemblies. This process requires the application of high resolution structural analysis in order to identify the presence of such structurally determinative interactions. Cryo-EM is the primary method for structural characterization of helical filaments. More broadly applied structural investigations of naturally occurring and synthetic peptide and protein filaments represent the most promising approach to identify local structural interactions that contribute to specificity and stability within these assemblies.

## Methods

**Peptide synthesis and purification.** Chemical reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) or Anaspec, Inc. (Fremont, CA) unless otherwise specified. Peptides 15-10-3, 22-17-3, 29-24-3, 36-31-3, 36-31-3\_RR, 36-31-3\_LL and 29-20-2 were obtained from SynPeptide Co., LTD. (Shanghai, China). All purchased chemical reagents and peptides ordered were

used without further purification. The Form II and Form IIa peptides were prepared via microwave-assisted solid phase peptide synthesis on a CEM Liberty Blue Automated Microwave Peptide Synthesizer as the N-acetyl, C-amide capped derivatives. A PAL-PEG-PS resin from Applied Biosystems (Foster City, CA) was used for synthesis of both peptides. Standard Fmoc protection chemistry was utilized in conjunction with coupling cycles consisting of HBTU/OXIMA-mediated activation protocols and base-induced deprotection (20% piperidine in N, N-dimethylformamide with 0.1 M hydroxybenzotriazole) of the Fmoc group. After synthesis, the DMF/resin mixture was filtered and rinsed with acetone then air-dried. The peptides were cleaved from the resin by incubation at room temperature for 3 h in a cocktail consisting of 92.5% trifluoroacetic acid (TFA), 2.5% distilled water, 2.5% triisopropylsilane, and 2.5% 2,2'-(ethylenedioxy)-diethanethiol. Cleavage was followed by filtration and subsequent precipitation in diethyl ether. The precipitate was allowed to desiccate overnight. The crude peptides were resolubilized in 3 mL of a 50:50 mixture of acetonitrile and water (0.1% TFA additive) and purified using a Shimadzu LC-20AP reversed-phase high-pressure liquid chromatography (HPLC) equipped with a C18 column. The peptides were eluted with a linear gradient of water-acetonitrile with 0.1% TFA. Peptide mass was confirmed using ESI mass spectrometry. Purified HPLC fractions were lyophilized, sealed, and stored at  $-30^{\circ}\text{C}$ .

**Peptide assembly.** Stock solutions of all peptides (3 mg/mL) were prepared by solubilizing purified, lyophilized peptide (0.75 mg) in 250  $\mu\text{L}$  of acetate buffer (10 mM, pH 4.0). The solution was titrated with dilute sodium hydroxide solution to adjust the final pH value to 4.0. The solutions were allowed to assemble from 24 h to 2 weeks depending on the sequence. Peptides 36-31-6\_RR and 36-31-6\_LL were thermally annealed using the following thermal cycle protocol: (1) rapid heating to  $90^{\circ}\text{C}$  for 30 min and (2) cooling to  $25^{\circ}\text{C}$  at a rate of  $0.2^{\circ}\text{C}/\text{minute}$ .

**Circular dichroism spectropolarimetry.** CD measurements were performed on a Jasco J-1500 CD spectropolarimeter using 0.10 mm thick quartz plates (Hellma Analytics). Three CD spectra were collected and averaged for each peptide sample. Each spectrum was acquired at a scan rate of 100 nm/min from 190–260 nm with a bandwidth of 2 nm and a data pitch of 0.2 nm.

**Negative stain TEM analysis.** TEM Grids were prepared using dilute solutions of peptide (3 mg/mL) in aqueous buffer (10 mM acetate, pH 4.0). Samples were prepared by depositing a peptide solution (4  $\mu\text{L}$ ) onto a 200-mesh carbon-coated copper grid from Electron Microscopy Services (Hatfield, PA). After 90 s of incubation on the grid, the excess liquid was wicked away, leaving a thin film of sample. An aliquot (4  $\mu\text{L}$ ) of negative stain solution (1% uranyl acetate) was deposited onto the thin film. After 1 min of staining, the remaining moisture was wicked away, and the grid was dried overnight in a desiccator. Electron micrographs were captured on a Hitachi HT-7700 transmission electron microscopy, equipped with a tungsten filament and AMT CCD camera, operating at an accelerating voltage of 80 kV.

**Cryo-EM image analysis.** The peptide sample (ca. 2–4  $\mu\text{L}$ ) was applied to glow-discharged Quantifoil (1.2/1.3 or 2/2) or lacey carbon grids, and then plunge frozen using a Vitrobot Mark IV (FEI). Cryo-EM data sets on peptides 15-10-3, 29-24-3, 29-20-2 and 36-31-3\_RR were collected on a 200 keV Talos Arctica with a K2 camera (Emory University) at  $1.04 \text{ \AA}/\text{pixel}$  and a total dose of ca.  $55 \text{ e}/\text{\AA}^2$ . The cryo-EM images for the dataset of peptide 36-31-3 were collected on a 300 keV Titan Krios with a K2 camera (National Cryo-EM Facility at NCI) at  $1.06 \text{ \AA}/\text{pixel}$  and a total dose of ca.  $50 \text{ e}/\text{\AA}^2$ . Cryo-EM datasets on peptides Form II and Form IIa were collected on a 300 keV Titan Krios with a K3 camera (University of Virginia) at  $1.08 \text{ \AA}/\text{pixel}$  and a total dose of ca.  $51 \text{ e}/\text{\AA}^2$ . Cryo-EM movies were recorded in counting mode using EPU v2.4 (Thermo Fisher). The micrographs were first motion corrected and dose weighted by MotionCorr v2<sup>90</sup>, and then a CTF correction was applied by multiplying the images with the theoretical CTF using CTFFIND3. Filament images corresponding to ca. 20 electrons/ $\text{\AA}^2$  were extracted using e2helixboxer (EMAN2)<sup>91</sup>. For each peptide filament, a list of possible helical symmetries was calculated from the averaged power spectrum of peptide particles. To determine the correct helical symmetry from the list, first if the full dataset has more than 30,000 particles, a subset containing 30,000 particles was generated. And then the initial volume was generated from those 30,000 particles with random assigned azimuthal angles. After that, possible helical symmetries were tested in Spider v2.10 using IHRSR<sup>67,92</sup> by trial and error until recognizable amino acid side chain densities could be observed in the correct symmetry (see Supplementary Figs. 11 and 12 for representative examples of this process). Specifically, the IHRSR reconstructions were divided into 3 steps, where 4x binned, 2x binned, un-binned particles were used to accelerate the progress. Helical symmetries were tightly locked in the first two steps, and then slightly relaxed in the third step. All helical parameters, including rise, rotation and point-group symmetry were imposed at the end of each IHRSR cycle. After determining the correct symmetry, final reconstructions using the all particles were run in both RELION<sup>93</sup> v3.0 and SPIDER<sup>92</sup> v2.10. The resulting volumes from both approaches were sharpened with the same negative B-factor, and then selected by eye based on the quality of the side chain densities. The statistics are listed in Supplementary Tables 1 and 2.



**Model building.** As the cryo-EM maps of all peptide samples reach to 4.5-Å resolution or higher, right-handed  $\alpha$ -helices were used to determine the helical hand of the EM volumes. It was clear from the map that each peptide molecule forms a single  $\alpha$ -helix. Therefore, an  $\alpha$ -helix model was generated in UCSF Chimera<sup>94</sup> v1.14 from the corresponding peptide sequence and then docked in the respective EM map. This single  $\alpha$ -helix model was adjusted manually in Coot<sup>95</sup> v0.8 to best fit into the map. Finally, this adjusted single  $\alpha$ -helix model was used to generate a model filament using the determined helical symmetry, which was then refined against the full cryo-EM map using real space refinement in PHENIX<sup>96</sup> v1.16. Final geometries of the atomic models were validated with the MolProbity<sup>96</sup> implementation in PHENIX. The refinement statistics are shown in Supplementary Tables 1 and 2. Cryo-EM maps and atomic coordinates have been deposited with the Electron Microscopy Data Bank and Protein Data Bank with accession codes given in Supplementary Tables 1 and 2. Model versus map FSC calculations were employed to estimate the resolution of the reconstructions and are reported in Supplementary Fig. 13.

**Motif analysis.** A search database was created by filtering the PDB, as of 04/19/2020, for the following: X-ray structures with resolution of 2.6 Å or better, containing at least 60% protein residues, with no more than 5000 residues and no more than 26 chains in the biological unit. This resulted in a total of 146,052 biological-unit entries. All-to-all sequence clustering of chains within these biological units was then performed at 50% sequence identity using the USEARCH tool v10<sup>97</sup>. Using this information, a minimally non-redundant subset of biological units was selected by eliminating a biological unit A if another one in the set contained a super set of its chains (i.e., had chains belonging to the same clusters as chains of A and additional chains). Resolution was used as the tie-breaker when two biounits had equivalent sets of chains. In the end, 23,915 biounits survived with a total of 176,442 chains and 43,449,344 residues. This set of structures was used to search for the Arg clasp structural motif (or the control motif) using the MASTER program<sup>98,99</sup>. All non-redundant matches (i.e., less than 50% sequence identity in the matching region—the matching fragments themselves + /- 15 residues on each side) with backbone RMSD below 1.0 Å were sought and used for analysis.

**Computational design.** dTERMen<sup>74,75</sup> was used to perform design. dTERMen is a general-purpose method for computational protein design that is based on the concept of structural degeneracy of proteins. Central to this approach is the concept of a TERM (short for tertiary motif)—a fragment of protein structure—consisting of one or more disjoint structural segments, that is reused across unrelated proteins. This reuse, which we described in several studies<sup>100–102</sup>, allows one to directly mine the PDB for sequence-structure relationships. Specifically, repeated TERM instances allow one to identify sequence features necessary for stabilizing the corresponding structural motif. Thus, dTERMen works by automatically breaking the desired target structure *T* down into its constituent TERMS, deducing sequence-structure relationships for each using rapid database searches (via the underlying method MASTER<sup>98</sup>) and combines these to build a model for what sequences would most likely stabilize target structure *T*.

A python-based implementation of dTERMen was employed in this study, which makes extensive use of the structure search engine MASTER to identify TERM matches and extract their sequence statistics. The structural database for underlying MASTER searches used in this study was prepared on 01/22/2019 following the procedure above. As the design template, a 15-chain portion of the 36-31-3 assembly was used (shown in color in Fig. 7c), with one chain (chain A) fully surrounded by all symmetry mates with which it could make any appreciable interactions. The symmetry-based design feature in dTERMen was used (–image flag), with chain A as the central unit and all others being images. The energy table resulting from running dTERMen described the sequence landscape compatible with folding to the 36-31-3 geometry, based on structural statistics of constituent motifs. Given this table, integer linear programming (ILP) was used to identify the best-scoring sequence under a solubility/fold-specificity constraint, requiring solvent-facing  $\alpha$  positions (i.e., 8, 15, 22, and 29) to be polar (i.e., one of Asp, Glu, Gly, His, Lys, Asn, Gln, Arg, Ser, Thr, Tyr). The resulting sequence, TLEELLAELILKAKAEILKAKAEVLKAKAEILKAL, lacked the Arg clasp motif, which was not surprising in light of our analysis of this motif's natural designability. To test the importance of this motif for assembly formation, we generated another design by re-running the optimization with the additional constraint that preserved the Arg clasp motif (i.e., left identities of positions 6, 7, 10, and 36 at Arg, Ala, Arg, and Gln, respectively). This resulted in the sequence: TLEELRAEILKAKAEILKAKAEVLKAKAEILKAKA, which was called 36-31-3\_RR, while the original design was called 36-31-3\_LL.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The reconstruction maps were deposited in the Electron Microscopy Data Bank with accession numbers of EMD-21812, EMD-21813, EMD-21814, EMD-21815, EMD-21816, EMD-21817, and EMD-21818. The corresponding filament models were deposited in the Protein Data Bank with accession numbers of 6WKX, 6WKY, 6WL0, 6WL1, 6WL7,

6WL8, and 6WL9. Additional data that support the findings of this study are available from the corresponding author upon request.

## Code availability

The codes for MASTER and dTERMen are freely available for noncommercial purposes from the Grigoryan Lab at Dartmouth College through the following links; <https://grigoryanlab.org/master> and <https://grigoryanlab.org/dtermen>, respectively.

Received: 8 July 2020; Accepted: 2 December 2020;

Published online: 18 January 2021

## References

- Acar, H. et al. Self-assembling peptide-based building blocks in medical applications. *Adv. Drug Deliv. Rev.* **110–111**, 65–79 (2017).
- Ekiz, M. S., Cinar, G., Khalily, M. A. & Guler, M. O. Self-assembled peptide nanostructures for functional materials. *Nanotechnology* **27**, 402002 (2016).
- Ramakers, B. E., van Hest, J. C. & Lowik, D. W. Molecular tools for the construction of peptide-based materials. *Chem. Soc. Rev.* **43**, 2743–2756 (2014).
- Cormier, A. R., Pang, X., Zimmerman, M. I., Zhou, H. X. & Paravastu, A. K. Molecular structure of RADA16-I designer self-assembling peptide nanofibers. *ACS Nano* **7**, 7562–7572 (2013).
- Nagy-Smith, K., Moore, E., Schneider, J. & Tycko, R. Molecular structure of monomeric peptide fibrils within a kinetically trapped hydrogel network. *Proc. Natl Acad. Sci. USA* **112**, 9816–9821 (2015).
- Lee, M. et al. Zinc-binding structure of a catalytic amyloid from solid-state NMR. *Proc. Natl Acad. Sci. USA* **114**, 6191–6196 (2017).
- Zhang, S. Q. et al. Designed peptides that assemble into cross-alpha amyloid-like structures. *Nat. Chem. Biol.* **14**, 870–875 (2018).
- Chen, K. H., Corro, K. A., Le, S. P. & Nowick, J. S. X-ray crystallographic structure of a giant double-walled peptide nanotube formed by a macrocyclic beta-sheet containing abeta16-22. *J. Am. Chem. Soc.* **139**, 8102–8105 (2017).
- Kajander, T., Cortajarena, A. L., Mochrie, S. & Regan, L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr. D. Biol. Crystallogr.* **63**, 800–811 (2007).
- Shen, H. et al. De novo design of self-assembling helical protein filaments. *Science* **362**, 705–709 (2018).
- Hughes, S. A. et al. Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs. *Proc. Natl Acad. Sci. USA* **116**, 14456–14464 (2019).
- Egelman, E. H. et al. Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure* **23**, 280–289 (2015).
- Nagy-Smith, K. et al. Molecular, local, and network-level basis for the enhanced stiffness of hydrogel networks formed from coassembled racemic peptides: predictions from Pauling and Corey. *ACS Cent. Sci.* **3**, 586–597 (2017).
- Fujii, T. et al. Structure of a type III secretion needle at 7-Å resolution provides insights into its assembly and signaling mechanisms. *Proc. Natl Acad. Sci. USA* **109**, 4461–4466 (2012).
- Loquet, A. et al. Atomic model of the type III secretion system needle. *Nature* **486**, 276–279 (2012).
- Wang, Y. A., Yu, X., Silverman, P. M., Harris, R. L. & Egelman, E. H. The Structure of F-Pili. *J. Mol. Biol.* **385**, 22–29 (2009).
- Costa, T. R. D. et al. Structure of the bacterial sex F pilus reveals an assembly of a stoichiometric protein-phospholipid complex. *Cell* **166**, 1436–1444 (2016). e1410.
- Zhang, J., Zheng, F. & Grigoryan, G. Design and designability of protein-based assemblies. *Curr. Opin. Struct. Biol.* **27**, 79–86 (2014).
- Adamcik, J. & Mezzenga, R. Amyloid polymorphism in the protein folding and aggregation energy landscape. *Angew. Chem. Int. Ed. Engl.* **57**, 8370–8382 (2018).
- Galkin, V. E., Orlova, A., Schroder, G. F. & Egelman, E. H. Structural polymorphism in F-actin. *Nat. Struct. Mol. Biol.* **17**, 1318–1323 (2010).
- Galkin, V. E. et al. The structure of the CS1 pilus of enterotoxigenic *Escherichia coli* reveals structural polymorphism. *J. Bacteriol.* **195**, 1360–1370 (2013).
- Egelman, E. H. A tale of two polymers: new insights into helical filaments. *Nat. Rev. Mol. Cell Biol.* **4**, 621–630 (2003).
- Lu, A. et al. Plasticity in PYD assembly revealed by cryo-EM structure of the PYD filament of AIM2. *Cell Discov.* **1**, 15013 (2015).
- Guenther, E. L. et al. Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat. Struct. Mol. Biol.* **25**, 311–319 (2018).

25. Close, W. et al. Physical basis of amyloid fibril polymorphism. *Nat. Commun.* **9**, 699 (2018).
26. Zhang, W. et al. Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *Elife* **8**, (2019).
27. Wu, Y. et al. A supramolecular vaccine platform based on alpha-helical peptide nanofibers. *ACS Biomater. Sci. Eng.* **3**, 3128–3132 (2017).
28. Wu, D. et al. Polymers with controlled assembly and rigidity made with click-functional peptide bundles. *Nature* **574**, 658–662 (2019).
29. Fries, C. N. et al. Controlled lengthwise assembly of helical peptide nanofibers to modulate CD8+ T-cell responses. *Adv. Mater.* **32**, 2003310 (2020).
30. DiMaio, F. et al. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015).
31. Ramakrishnan, V. The ribosome emerges from a black box. *Cell* **159**, 979–984 (2014).
32. Subramaniam, S., Earl, L. A., Falconieri, V., Milne, J. L. S. & Egelman, E. H. Resolution advances in cryo-EM enable application to drug discovery. *Curr. Opin. Struct. Biol.* **41**, 194–202 (2016).
33. Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443–1444 (2014).
34. Chen, Y. X. et al. Structural determination of a filamentous chaperone to fabricate electronically conductive metalloprotein nanowires. *ACS Nano* **14**, 6559–6569 (2020).
35. Feng, Z. et al. Artificial intracellular filaments. *Cell Rep. Phys. Sci.* **1**, 100085 (2020).
36. Tayeb-Fligelman, E. et al. The cytotoxic *Staphylococcus aureus* PSMalpha3 reveals a cross-alpha amyloid-like fibril. *Science* **355**, 831–833 (2017).
37. Walshaw, J. & Woolfson, D. N. Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. *Protein Sci.* **10**, 668–673 (2001).
38. Pandya, M. J. et al. Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry* **39**, 8728–8734 (2000).
39. Papapostolou, D. et al. Engineering nanoscale order into a designed protein fiber. *Proc. Natl Acad. Sci. USA* **104**, 10853–10858 (2007).
40. Sharp, T. H. et al. Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of de novo design. *Proc. Natl Acad. Sci. USA* **109**, 13266–13271 (2012).
41. Burgess, N. C. et al. Modular design of self-assembling peptide-based nanotubes. *J. Am. Chem. Soc.* **137**, 10554–10562 (2015).
42. Potekhin, S. A. et al. De novo design of fibrils made of short alpha-helical coiled coil peptides. *Chem. Biol.* **8**, 1025–1032 (2001).
43. Melnik, T. N. et al. Shift of fibril-forming ability of the designed alpha-helical coiled-coil peptides into the physiological pH region. *Protein Eng.* **16**, 1125–1130 (2003).
44. Kajava, A. V., Potekhin, S. A., Corradin, G. & Leapman, R. D. Organization of designed nanofibrils assembled from alpha-helical peptides as determined by electron microscopy. *J. Pept. Sci.* **10**, 291–297 (2004).
45. Zimenkov, Y., Conticello, V. P., Guo, L. & Thiyagarajan, P. Rational design of a nanoscale helical scaffold derived from self-assembly of a dimeric coiled coil motif. *Tetrahedron* **60**, 7237–7246 (2004).
46. Xu, C. et al. Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *J. Am. Chem. Soc.* **135**, 15565–15578 (2013).
47. Ghosh, T. S., Chaitanya, S. K. & Sankaramakrishnan, R. End-to-end and end-to-middle interhelical interactions: new classes of interacting helix pairs in protein structures. *Acta Crystallogr. D. Biol. Crystallogr.* **65**, 1032–1041 (2009).
48. Marqusee, S. & Baldwin, R. L. Helix stabilization by Glu...Lys+ salt bridges in short peptides of de novo design. *Proc. Natl Acad. Sci. USA* **84**, 8898–8902 (1987).
49. Lyu, P. C., Gans, P. J. & Kallenbach, N. R. Energetic contribution of solvent-exposed ion pairs to alpha-helix structure. *J. Mol. Biol.* **223**, 343–350 (1992).
50. Huyghues-Despointes, B. M., Scholtz, J. M. & Baldwin, R. L. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci.* **2**, 80–85 (1993).
51. Scholtz, J. M., Qian, H., Robbins, V. H. & Baldwin, R. L. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* **32**, 9668–9676 (1993).
52. Cid, H., Gazitua, F. & Bunster, M. Interactions that regulate the helical fold in proteins. *Biol. Res.* **29**, 213–225 (1996).
53. Andrew, C. D., Penel, S., Jones, G. R. & Doig, A. J. Stabilizing nonpolar/polar side-chain interactions in the alpha-helix. *Proteins* **45**, 449–455 (2001).
54. Andrew, C. D. et al. Stabilizing interactions between aromatic and basic side chains in alpha-helical peptides and proteins. Tyrosine effects on helix circular dichroism. *J. Am. Chem. Soc.* **124**, 12706–12714 (2002).
55. Shi, Z., Olson, C. A. & Kallenbach, N. R. Cation-pi interaction in model alpha-helical peptides. *J. Am. Chem. Soc.* **124**, 3284–3291 (2002).
56. Butterfield, S. M., Patel, P. R. & Waters, M. L. Contribution of aromatic interactions to alpha-helix stability. *J. Am. Chem. Soc.* **124**, 9751–9755 (2002).
57. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys salt bridges on  $\alpha$ -helical stability and folding kinetics. *Biophysical J.* **110**, 2328–2341 (2016).
58. Crick, F. The packing of [alpha]-helices: simple coiled-coils. *Acta Crystallogr.* **6**, 689–697 (1953).
59. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544 (1991).
60. Alber, T. Structure of the leucine zipper. *Curr. Opin. Genet. Dev.* **2**, 205–210 (1992).
61. Russ, W. P. & Engelman, D. M. The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911–919 (2000).
62. Dieckmann, G. R. & DeGrado, W. F. Modeling transmembrane helical oligomers. *Curr. Opin. Struct. Biol.* **7**, 486–494 (1997).
63. Walters, R. F. & DeGrado, W. F. Helix-packing motifs in membrane proteins. *Proc. Natl Acad. Sci. USA* **103**, 13658–13663 (2006).
64. Zhang, S. Q. et al. The membrane- and soluble-protein helix-helix interactome: similar geometry via different interactions. *Structure* **23**, 527–541 (2015).
65. Walther, D., Eisenhaber, F. & Argos, P. Principles of helix-helix packing in proteins: the helical lattice superposition model. *J. Mol. Biol.* **255**, 536–553 (1996).
66. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
67. Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
68. Egelman, E. H. The iterative helical real space reconstruction method: surmounting the problems posed by real polymers. *J. Struct. Biol.* **157**, 83–94 (2007).
69. Egelman, E. H. Reconstruction of helical filaments and tubes. *Methods Enzymol.* **482**, 167–183 (2010).
70. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
71. Zheng, W. et al. Cryoelectron-microscopic structure of the pKpQIL conjugative pili from carbapenem-resistant *Klebsiella pneumoniae*. *Structure* **28**, 1321–1328 (2020).
72. Xu, J., Dayan, N., Goldbourt, A. & Xiang, Y. Cryo-electron microscopy structure of the filamentous bacteriophage IKe. *Proc. Natl Acad. Sci. USA* **116**, 5493–5498 (2019).
73. Tarafder, A. K. et al. Phage liquid crystalline droplets form occlusive sheaths that encapsulate and protect infectious rod-shaped bacteria. *Proc. Natl Acad. Sci. USA* **117**, 4724–4731 (2020).
74. Zhou, J., Panaitiu, A. E. & Grigoryan, G. A general-purpose protein design framework based on mining sequence-structure relationships in known protein structures. *Proc. Natl Acad. Sci. USA* **117**, 1059–1068 (2020).
75. Frappier, V., Jenson, J. M., Zhou, J., Grigoryan, G. & Keating, A. E. Tertiary structural motif sequence statistics enable facile prediction and design of peptides that bind anti-apoptotic Bcl-1 and Mcl-1. *Structure* **27**, 606–617 (2019).
76. Kobe, B. & Kajava, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
77. Kajava, A. V. Tandem repeats in proteins: from sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).
78. Lupas, A. N. & Bassler, J. Coiled coils - a model system for the 21st Century. *Trends Biochem. Sci.* **42**, 130–140 (2017).
79. Woolfson, D. N. Coiled-coil design: updated and upgraded. *Subcell. Biochem.* **82**, 35–61 (2017).
80. Collie, G. W. et al. Shaping quaternary assemblies of water-soluble non-peptide helical foldamers by sequence manipulation. *Nat. Chem.* **7**, 871–878 (2015).
81. Gallardo, R., Ranson, N. A. & Radford, S. E. Amyloid structures: much more than just a cross- $\beta$  fold. *Curr. Opin. Struct. Biol.* **60**, 7–16 (2020).
82. Valery, C. et al. Biomimetic organization: octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proc. Natl Acad. Sci. USA* **100**, 10258–10262 (2003).
83. Hervas, R. et al. Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in *Drosophila*. *Science* **367**, 1230–1234 (2020).
84. Brodin, J. D., Smith, S. J., Carr, J. R. & Tezcan, F. A. Designed, helical protein nanotubes with variable diameters from a single building block. *J. Am. Chem. Soc.* **137**, 10468–10471 (2015).
85. Brodin, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382 (2012).
86. Görbitz, C. H. Nanotube formation by hydrophobic dipeptides. *Chem. – A Eur. J.* **7**, 5153–5159 (2001).
87. Reches, M. & Gazit, E. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **300**, 625–627 (2003).
88. Lu, K., Jacob, J., Thiyagarajan, P., Conticello, V. P. & Lynn, D. G. Exploiting amyloid fibril lamination for nanotube self-assembly. *J. Am. Chem. Soc.* **125**, 6391–6393 (2003).

89. Kurt, T. D. et al. Asparagine and glutamine ladders promote cross-species prion conversion. *J. Biol. Chem.* **292**, 19076–19086 (2017).
90. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
91. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
92. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
93. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
94. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Computational Chem.* **25**, 1605–1612 (2004).
95. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D.* **66**, 486–501 (2010).
96. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
97. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
98. Zhou, J. & Grigoryan, G. Rapid search for tertiary fragments reveals protein sequence–structure relationships. *Protein Sci.* **24**, 508–524 (2015).
99. Zhou, J. & Grigoryan, G. A C++ library for protein sub-structure search. Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.26.062612v1.full> (2020).
100. Mackenzie, C. O., Zhou, J. & Grigoryan, G. Tertiary alphabet for the observable protein structural universe. *Proc. Natl Acad. Sci. USA* **113**, E7438–E7447 (2016).
101. Mackenzie, C. O. & Grigoryan, G. Protein structural motifs in prediction and design. *Curr. Opin. Struct. Biol.* **44**, 161–167 (2017).
102. Zheng, F. & Grigoryan, G. Sequence statistics of tertiary structural motifs reflect protein stability. *PLoS ONE* **12**, e0178272 (2017).

## Acknowledgements

Cryo-EM imaging of the 36-31-3 filament was performed at the Cryo-EM facility (NCEF) at the National Cancer Institute's Frederick Laboratory for Cancer Research under contract HSSN261200800001E. Cryo-EM work on the Form II and Form IIA filaments was conducted at the Molecular Electron Microscopy Core facility at the University of Virginia, which is supported by the School of Medicine and built with NIH grant G20-RR31199. In addition, the Titan Krios (SIG S10-RR025067) and K3/GIF (U24-GM116790) were purchased in part or in full with the designated NIH grants. This study was supported by the Robert P. Apkarian Integrated Electron Microscopy Core (IEMC) at Emory University, which is subsidized by the School of Medicine and Emory College of Arts and Sciences. Additional support was provided by the Georgia Clinical & Translational Science Alliance of the National Institutes of Health under award number UL1TR000454. Negative stain TEM images gathered on a Hitachi HT7700 120 kV TEM at Emory University, which was supported by the Georgia Clinical and Translational Science Alliance under award number UL1TR002378. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health. This research was supported by grants from the NSF (DMR-1534317 (V.P.C.), DMR-1533958 (E.H.E.) and DMR-1534246 (G.G.)) and NIH

(GM132117 (G.G.)). The circular dichroism spectropolarimeter was acquired through funding from an NSF grant (DBI-1726544). Dartmouth cluster computer infrastructure was supported by NSF grant CNS-1205521.

## Author contributions

O.G., C.M. and C.X. prepared the samples and performed the initial analyses; F.W. and P.J. collected the cryo-EM data; F.W., L.B., Z.S. and E.H.E. performed image processing; F.W. did the structural modeling; G.G. performed the computational analysis of designability and computational peptide design; F.W., O.G., G.G. and V.P.C. prepared figures; F.W., G.G., E.H.E. and V.P.C. wrote the manuscript; F.W., G.G., E.H.E. and V.P.C. conceived the study.

## Competing interests

Dartmouth College has filed a provisional patent application protecting the technology behind dTERMen. G.G. is an employee and shareholder of an early-stage biotechnology venture, which may have an interest in using the dTERMen technology. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-20689-w>.

**Correspondence** and requests for materials should be addressed to V.P.C.

**Peer review information** *Nature Communications* thanks Anant Paravastu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021