A new building strategy in protein origami: design and self-assembly of a brick and staple artificial repeat protein pair leading to macroscopic tubular superhelices.

Laureen Moreaud,[a 1] Sébastien Viollet,[b 1] Agathe Urvoas,[b] Marie Valerio-Lepiniec,[b] Agnès Mesneau,[b] Inès Li de la Sierra-Gallay,[b] Jessalyn Miller,[b,f] Malika Ouldali,[b] Cécile Marcelot,[a] Stéphanie Balor,[c] Vanessa Soldan,[c] Cristelle Meriadec,[d] Franck Artzner,[d] Erik Dujardin[a, e 2] and Philippe Minard[b 2]

[a] Centre d'Elaboration des Matériaux et d'Etudes Structurales (CEMES), CNRS UPR 8011, 29 rue J. Marvig, B.P. 94347, F-31055 Toulouse, France
[b] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France**.**
[c] METi, Centre de Biologie Intégrative, Université de Toulouse, CNRS, UPS, 31062, Toulouse, France.
[d] Institut de Physique de Rennes (IPR), CNRS, UMR 6251, Université de Rennes 1, F-35042 Rennes, France
[e] Laboratoire Interdisciplinaire Carnot de Bourgogne, CNRS UMR 6303, Université de Bourgogne Franche-Comté, 21000 Dijon, France
[f] Department of Chemistry, Emory University, Atlanta, Georgia 30322, USA

[1] These authors contributed equally to the work.

[2] To whom correspondence may be addressed. Email: philippe.minard@i2bc.paris-saclay.fr or erik.dujardin@cnrs.fr

Author Contributions: SV, PM conceived the brick-staple concept. SV, AU, MVL, AM, JM synthesized the proteins and dye-tagged proteins, characterized the self-assembly by ITC, SDS-PAGE, DLS, Fluorimetry. LM, SV, JM, MO performed stained TEM imaging. ILdlSG solved the crystallographic structure of the staple-bait complex. LM, SB, VS prepared cryosamples and performed cryoEM and tomography imaging and processing. CMe, FA, LM performed SAXS experiments. LM, ED, FA, CMa analyzed all EM and SAXS data. PM, ED performed electrostatic and structural modelling. LM, ED, PM drafted the manuscript, which was amended by all co-authors. All coauthors have approved the final version of the text.

The authors declare no competing interest.

Classification: Major: Biological Sciences
Minor: Biochemistry

Keywords: Repeat proteins, protein origami, pair recognition, supramolecular assembly, protein crystal, directed evolution.

ORCID numbers:

S Viollet	0000-0001-7791-5928
L Moreaud	0000-0003-4092-5983
A Urvoas	0000-0002-6077-859X
M Valerio-Lepiniec	0000-0002-9940-0957
A Mesneau	N/A
I. Li de la Sierra-Gallay	0000-0003-2770-7439
J Miller	0000-0002-0396-8549
M Ouldali	0000-0002-7601-6397
C Marcelot	0000-0002-1562-6802
S Balor	0000-0001-6828-0377
V Soldan	0000-0002-6993-4202
C Meriadec	N/A
F Artzner	0000-0002-5613-576X
E Dujardin	0000-0001-7242-9250
P Minard	0000-0002-3256-528X

Abstract

A versatile strategy to create an inducible protein assembly with predefined geometry is demonstrated. The assembly is triggered by a binding protein that staples two identical protein bricks together in a predictable spatial conformation. The brick and staple proteins are designed for mutual directional affinity and engineered by directed evolution from a synthetic modular repeat protein library. As a first proof-of-concept, this article reports on the spontaneous, extremely fast and quantitative self-assembly of two designed alpha-repeat (αRep) brick and staple proteins into macroscopic tubular superhelices at room temperature. Small-angle X-ray scattering (SAXS) and transmission electron microscopy (TEM with staining agent and cryoTEM) elucidate the resulting superhelical arrangement that precisely matches the *a priori* intended 3D assembly. The highly ordered, macroscopic biomolecular construction sustains temperatures as high as 75°C thanks to the robust αRep building blocks. Since the α-helices of the brick and staple proteins are highly programmable, their design allows encoding the geometry and chemical surfaces of the final supramolecular protein architecture. This work opens new routes towards the design and fabrication of multiscale protein origami with arbitrarily programmed shapes and chemical functions.

Significance Statement

Spontaneous building of bio-inspired organization with both accurate morphologies and well-defined functions is still highly challenging. We illustrate a versatile approach to control assemblies of complementary "staple" and "template" proteins into supramolecular accurate architectures by characterizing *de novo* nanotube crystals. For this purpose, we exploit highly selective binding surfaces of repeat proteins to generate robust close-contacts. We design the brick protein with a semi-lock washer shape by splitting and appending the sequence of the partner protein to its terminal modules. Equimolar mixture results in sequential growth generating long tubular superhelices. This strategy paves a new way to chimeric proteins able to organize functions on designed structures by origami processes.

INTRODUCTION

Biological nanostructures originating from self-assembling proteins such as filaments, microtubules, cilia or flagella are ubiquitous in living cells. These sophisticated protein-based architectures have enabled complex and integrated functions such as intracellular transport, cell motility and cell division. Strategies to design and control ordered synthetic protein assembly may foster a design strategy whereby three-dimensional structure and chemical activity could be merged towards protein-templated functional nanomaterials synthesis,(1) design of self-assembling nanostructures,(2) self-healing or responsive materials,(3-5) spatially ordered multi-enzyme cascades,(6) and biomolecular display for atomically-resolved structure determination by electron microscopy.(7) In most reported studies, the proteins chosen as building blocks of these ordered assemblies are obtained by limited alteration of natural proteins rather than by *de novo* design.(8-11) Synthetic protein cages, tubes or filaments have been created by genetically linking two or more natural homo-oligomeric protein structures with rigid connections,(12-14) by embedding metal-binding sites(15) or small bi-specific molecules,(16), or by using covalent self-splicing intein.(17) Importantly, computational protein design methods have reached such an accurate level of prediction that complex self-assembled architectures such as protein filaments(18), arrays (19), rings (20) or cages (21) have been accurately produced from existing designed protein components.(22)

During natural evolution, structural innovation often proceeds by combining or repeating simple structural modules such as coiled coils, beta-alpha units or propeller blades, which leads to functional protein architectures.(23) This suggests that a higher degree of conceptual abstraction can be applied to artificial protein assembly design. Idealized sequence motifs characterizing self-compatible building blocks can be conceived from the phylogenetic analysis of natural repeated proteins. Such carefully designed peptides were shown to self-assemble into supramolecular structures including as helical nanotubes (24). This modular strategy, sometimes further optimized by combinatorial exploration, has also been successful in designing complex artificial proteins from simpler elementary building blocks (25). One archetypal example is the emergence of complex structures from the careful analysis of

engineered coiled coils. (26-28) When mutually orthogonal coiled coil pairs are inserted into a single polypeptide chain, each pair of cognate helices assembles and drives the macromolecule to fold into the targeted tetrahedron topology, (29) thereby coining the concept and creative research area of protein origami (30, 31).

In this Report, a new and highly versatile strategy is demonstrated to create inducible protein assembly with predefined geometry. The assembly is mediated by a binding protein that staples two identical protein bricks together in a predictable geometry. Bricks and staples are synthetic proteins engineered for mutual affinity by directed evolution from synthetic repeat protein libraries. Our choice of exploiting repeat proteins is motivated by their regular architecture and by an established design strategy.(32) Indeed, analysis of natural repeat sequence collections can be used to extract sequence profiles characteristic of each type of repeat and to design idealized sequences with excellent foldability. The sequence variability allowed in some surface positions is exploited to endow these idealized proteins with tailorable binding specificity as first demonstrated for DARPins (designed ankyrin repeat proteins) (25, 33, 34) and other types of repeat proteins such as αReps(32) or repebodies.(35) Furthermore, we have developed a phage display library of extremely stable artificial αReps, from which a wide range of specific binders against natural proteins (36, 37) as well as self-assembling αRep pairs (32, 38-40) have been generated.

By exploiting the highly regular structures of repeat proteins it is possible to embed a recognition surface initially generated in a target/binder pair into another protein of the same family chosen for its structural specificity. The host protein template is thereby entrusted with the recognition capability of the embedded sequence without impacting its structure. As a proof-of-concept, we herein describe the generation of brick and staple repeat proteins and report the direct observation of their spontaneous assembly into extremely long superhelices. The detailed structural organization of the assembled protein array is characterized using SAXS and cryo-electron microscopy. The resulting model points out the potential for chemical and electrostatic ordering on this multiscale bioinspired template.

RESULTS AND DISCUSSION

Proteins design. The αRep family has been designed as a repeated architecture of structurally similar modules.(32) Each repeated motif of 31 amino acids forms two α-helices. When stacked on the previous module, the successive repeats are bound together with an azimuthal angle of 22° that naturally leads to a curved toroid. Yet, an out-of-plane twist angle between two consecutive motifs distorts the overall donut shape in such a way that a 16-repeat αRep will adopt a lock-washer shape equivalent to one turn of a potential right-handed superhelix. In principle, an αRep fold could be extended by concatenating more repeats to form a right-handed superhelical macromolecule. However, synthesizing and purifying very long proteins, which tend to be poorly expressed and insoluble, is highly impractical. We demonstrate a general alternative strategy, which consists in connecting small and stable "brick" αRep proteins comprising a few motifs with a “staple” protein that forces their head-to-tail assembly into a predefined geometry that mimics the covalent αRep superhelix.

Our general concept is to create a heterotrimeric αRep junction in which a staple protein (Figs. 1a, b, yellow) mediates the assembly of two consecutive bricks (Figs 1a, b, blue) with the same geometry as an αRep with twice as many repeats. The propagation of this trimeric junction from an equimolar brick:staple mixture leads to infinite helical supramolecular architectures. In order to generate such junctions, we took advantage of the known crystal structure of a face-to-back complex between two αReps (PDB code 8AW4 and Supplementary Information Section 1). This complex involves two proteins herein referred to as the “bait” and “bBE3” (as “back-binder” E3). The bait αRep comprises a 3-repeat core (I_1-I_2-I_3, Fig. 1c) and two terminal repeats named N- and C-caps. The back binder bBE3 was selected from the αRep library(37) and is a specific binder of the bait protein (Figs. 1c-e and Supplementary Information Section 1). The crystal structure shows that bBE3 (Figs. 1d,e) binds to the convex back surface made by the (I_1-I_2-I_3) modules of the bait protein using its own hypervariable concave surface as is typical for all selected αRep binders. The side chains located on the convex - or “back” - surface of the I_1-I_2-I_3 internal repeats of the bait protein were modified to differ from the repeated sequence found in these positions in other αReps. Therefore, the back-binder of the bait protein is not able to bind its own back surface, which would presumably have prevented its selection from the

library. In the bBE3/bait complex, the variable side chains of the first two turns of the back-binder helices fit in the grooves located between the helices on the convex side of repeats I_1-I_2-I_3. A closer view of the interaction surface is shown in Figure 1e (See also Fig. S1 in Supplementary Information). The interface between the two proteins, as assessed by PISA, involves 24 residues from the bait protein and 29 residues from the back-binder. An extensive hydrogen bond network, salt bridges and hydrophobic interactions are established between the variable side chains (yellow) of the back-binder protein (orange) and the side chains (purple) of residues located on the convex surface of the bait protein repeats I_1-I_2 (blue) and I_3 (red). Isothermal titration calorimetry (ITC) shows that the bBE3/bait complex forms in a 1:1 stoichiometry and its dissociation constant, K_D = 68 nM, indicates a strong affinity in the range observed for other αRep-based protein pairs (Fig. 2a) (37). The crystal structure of a bBE3/bait complex was obtained with a variant of the bait protein possessing different side chains on its concave surface. This further demonstrates that the recognition of the “back” convex surface of the bait protein is independent of the specific sequence on the concave surface of the bait protein.

Next, we hypothesized that such an avid back-binder protein would template the recombination of its full cognate partner when the matching surface is split into two distinct molecules. Thereby, two proteins, each terminated by one half of the bait repeats (Fig. 1h) would be forced into a trimeric junction with the back-binder acting as a staple to recover its full recognition partner. By applying this principle to the two termini of a same αRep protein, with a circular permutation of the bait repeats, I_3 is appended to the N-terminal when the bait repeats I_1-I_2 are appended to the C-terminal of the cap-free αRep resulting in a generic "brick" protein shown in Figures 1e-g. The structural regularity of repeat proteins ensures that the appended I_1-I_2 and I_3 exhibit the same geometry as in the bait. The central modules of the brick do not contribute to the self-assembly and can be chosen with arbitrary external surfaces. When the staple protein is exposed to the brick protein, it associates with the C-terminal repeat of one brick and the N-terminal repeat of another identical brick (Figs. 1h-j), resulting in a potentially infinite head-to-tail polymerization of adjacent bricks held together laterally by the staples. A model of staples binding adjacent bricks was built based on the experimental structure of the bBE3/bait complex

and is shown in Figures 1a-b. Two bricks (in blue) joined by one staple (in orange) form the elementary motif that is repeated along a regular superhelix.

Protein synthesis. In this work, we demonstrate the self-assembly principle of bricks and staples by constructing an eight-repeat brick protein comprised of five central repeats plus the internal bait modules I_1-I_2 and I_3 fused at the C-end and N-end respectively, according to the circular permutation design schematized in Figure 1f. However, internal αRep repeats exhibit hydrophobic faces that tend to aggregate, making them difficult to mass-produce. To prevent this, αReps are terminated by first and last motifs with polar exofaces, named N- and C-cap. Yet, these caps would also block the desired reconstitution of the bait motif with the staple protein and therefore the self-assembly. To circumvent this, a TEV protease cleavage site is inserted between each cap and the brick sequence (Fig. 1e). The final whole sequence of the brick protein is presented in the Supplementary Information Section 1. Prior to mixing with the staple protein, the caps are cleaved by TEV protease. The cap-free brick is separated from the two cap peptides as well as from the TEV protease bearing poly-histidine tags (HT) using a Ni-NTA purification column. Although repeat proteins are commonly produced with the N and C cap repeats, it appears here that the folded brick protein remains stable and soluble in a cap-free form, once cleaved. The cleaved brick (Figs. 1f-g) is composed of 8 repeated motifs equivalent to one half-turn of the self-assembled superhelix (Figs. 1h-j) that will bear two staples per turn on diametrically opposed sites as shown in Figures 1a-b.

Characterization of the Brick and Staple assembly. When a 10 μM stoichiometric solution of brick and staple bBE3 is mixed, a white turbid precipitate rapidly forms (Fig. 2b). Time-resolved light scattering at 350 nm detects the onset of assembly within minutes after mixing and reaches a plateau after 15-20 min at 25°C suggesting a fast kinetics of the association (Fig. 2c). Lowering the temperature to 4°C virtually suppress macroscopic assembly as no turbidity at 350 nm is observed, even after several hours, indicating a significant contribution of the hydrophobic interaction to the assembly. The terminal surface between two adjacent bricks is identical to the one between two covalently bound repeats inside an αRep (See Supplementary Information Section 2). This repeat-repeat interaction comprises a large hydrophobic contribution from the side chain packing (Fig. S2a). The cap-free bricks expose

hydrophobic clusters at both termini resulting in significant hydrophobic interactions that contribute to the head-to-tail alignment of two consecutive bricks (Fig. S2b), locked by the staple binding. Each isolated brick or staple protein and equimolar mixtures is monitored by SDS-PAGE at 20°C as initial mixtures (M) and as supernatant (S) and pellet (P) after centrifugation, for final protein concentrations comprised between 1 and 8 μM (Figs. 2d-f). At 1-2 μM, both proteins remain soluble in pure or mixed solution and no pellet is formed. At 4 μM, the pure proteins remain soluble, but an insoluble pellet is observed for the mixture that contains both proteins in comparable amounts. This is even more pronounced at 8 μM (Fig. 2f). Unlike the pure solutions, the proteins appear in the precipitate above 4 μM and their concentration in the supernatant decreases. The brick concentration even vanishes at 8 μM suggesting a slight excess of staples in this particular experiment (Fig. 2f). These results indicate that macroscopic self-assembly is triggered by protein concentrations of 2 to 4 μM beyond which the amount of precipitate increases with the protein concentration. The onset of assembly occurs at a micromolar critical concentration that is significantly larger than the bBE3/bait dissociation constant K_D (68 nM). This can be attributed to the entropic cost of the discontinued bound surface of the bait protein when it is split into two distinct subparts as terminal sections of adjacent bricks. The macroscopic precipitate was characterized by fluorescence microscopy (See Supplementary Information S3), negative-stain transmission electron microscopy and small angle X-ray scattering as shown in Figures 3 and 4.

TEM and SAXS structural analysis. negative-stain transmission electron microscopy (TEM) has been used to identify the structural features composing the molecular architecture. In order to obtain individualized supramolecular objects, we have used 3 wt% (150 mM) ammonium molybdate $(NH_4)_2MoO_4$, a TEM contrast agent compatible with the near-neutral pH conditions initially used for the assembly. The proteins, which appear bright over a dark stained background (Fig. 3a), form filaments with a highly uniform apparent width of 7.6 ± 0.8 nm (Fig. 3b) in good agreement with the nominal 8.4 ± 0.8 nm outer diameter of the model of Figure 1. The filament length follows a standard lognormal distribution with a mean value of about 260 nm and the longest filaments reaching near 1000 nm (Fig. 3c). Increasing the concentration of the equimolar mixtures from 2.5 to 10 μM results in a higher concentration of filaments and also

augments both the mean and spreading of their length distribution (Fig. S4). This behavior suggests a growth mechanism where the number of nuclei is set by the protein concentration and the superhelix elongation proceeds by incremental addition of bricks and staples at the filament termini, as long as monomers are available. (41, 42) The low dissociation constant of the brick/staple complex ensures that unbinding events do not limit the lengthening nor induce the fragmentation of the supramolecular assembly. Magnified views of a filament as displayed in Figure 3f reveal a braid-like contrast with a pitch along the main axis of 55.6 ± 2.2 Å (Fig. 3d). As a negative stain, the molybdate is excluded from the core of the proteins and creates a white contrast where the staple and bricks are seen tangentially as depicted in Fig. 3g. The observed periodicity is consistent with the staple pitch of 60 Å in the molecular model. The braid-like pattern is also characterized by a uniform angle of 115° ± 2° between two opposite white ellipsoids (Fig. 3e), which matches the 115° angle measured on the lateral projection of the superhelix as shown in Figures 3g-h. These negative-stained TEM observations strongly suggest that the dried protein superstructures consist of the programmed self-assembled superhelices.

The massive supramolecular assembly obtained in the absence of added salt was monitored by Small Angle X-ray Scattering (SAXS). The well-defined set of diffraction peaks shown in Figure 4a demonstrates a highly ordered crystalline structure rather than isolated monodisperse objects.(43) Different batches of molecular assemblies result in similar patterns with very minor variations of the peak positions or intensities. Interestingly, the intense peak at q = 0.0108 $Å^{-1}$ corresponds to a distance of 58 Å that matches the pitch between two staples along the superhelical model presented in Figure 1.a-b. The global analysis of all the peaks found in this highly reproducible SAXS pattern could be matched to the orthorhombic space group with cell parameters a = 477.3 Å, b = 194.8 Å and c = 58.2 Å (Table S1). We observe the presence of the (2 1 1) and (3 3 1) peaks that indicate that these data correspond to a (6x3x1) superlattice of a generic sub-cell (a = 80 Å, b = 65 Å and c = 58 Å) that comprises two pairs of bricks and staples. The rectangular 80 × 65 Å cell originates from the diametrically opposed bBE3 decoration along the superhelix with a very strong staple interdigitation between adjacent superhelices. The 58 Å longitudinal periodicity matches the distance between staples

positioned on the same side of the helix. The orthorhombic space group stems from the crystal symmetries, which is, in turn, imposed by the protein chirality, leaving the two-fold or screw axes as the only compatible symmetry elements. The missing (*h00*) peaks with odd *h* (Table S1) are strongly in favor of a screw axis along ***a*** and an antipolar packing of the helical superstructures aligned head-to-tail along the ***c*** axis. The head-to-tail packing along ***a*** gives rise to a first x2 supercell. We interpret the additional x3 superlattice along ***a*** and ***b*** as a low amplitude distortion mode between neighboring helices along ***b***, that is identically propagated along ***a*** as suggested by the observation of the (3 3 1) peaks. The most likely distortion type is a tilt modulation that occurs in chiral fibers when the helical pitch does not exactly match the ***c*** parameter.(44) Such superlattice induction has been observed in the rare crystals of chiral fibers where the helices have to solve the frustration between their intrinsic symmetry and the crystal packing symmetry.(45, 46) Note that positional distortions, like the ones observed with DNA,(47) are improbable due to the tight packing. The occasional variations in the positions and intensities of some diffraction peaks are attributed to slight differences in the crystal packing from one assembly to another (Fig. S5). The thermal stability of the self-assembled superstructures was monitored up to 75°C (Fig. S6). No change in the SAXS pattern was observed up to 50°C, indicating a robust crystalline organization over a wide range of temperatures. Around 55°C, a 2-4 Å reduction of the (0 0 1) and (6 0 0) peak positions accompanied by an increase of the (12 0 0) intensity and a vanishing (3 3 1) intensity reveals the superlattice melting into the lowest symmetry crystal made of anti-parallel superhelices. This transition is fully reversible upon cooling albeit with a small (5°C) hysteresis. The higher temperature structure remains stable up to 75°C. Notwithstanding the minor supercoiling variations, the pairing of brick and staple bBE3 therefore appears to result in strongly interacting anti-parallel superhelices. We note a tighter packing of unstained helices in the absence of added salt since the (a,b) SAXS parameters are significantly smaller than the apparent outer diameter measured in negative-stained TEM (Fig. 3).

Cryo-EM imaging. The highly ordered unstained supramolecular assembly was investigated in a quasi-native and hydrated state by cryo-electron microscopy. Figures 4b-d show that the self-assembly results in extremely large bundles of parallel tubules reaching at least tens of

micrometers in length and several micrometers in diameter. In the terminal region of the bundles (Fig. 4c), the tubules split apart from each other, which strongly suggests that the bundles are made of individualized tubules rather than being a 3-dimensional co-crystal of both proteins. The extent of the bundled tubules confirms the massive self-assembly of the brick and staple proteins, which is a direct consequence of their high mutual affinity driven by the programmed trimeric junction recognition further stabilized by strong inter-superhelix interactions. Closer examination of cryo-fractured bundle segments aligned along different orientations with respect to the electron beam (Fig. 4d) reveals a few typical EM patterns that yield more precise information on the close packing of the superhelix inside the bundles. Four different patterns are analyzed in Figure 5. In Figure 5a, straight and parallel thick lines are clearly observed that show a highly regular, periodical and staggered decoration of darker dots separated by a brighter uniform space. A statistical analysis of the positions of the black dots relative to their three nearest neighbors is detailed in Section S8 of the Supplementary Information. It reveals that the distance between successive dark dots on the same side of the superhelix is 6.4 ± 1.2 nm, which is consistent with the superhelical pitch measured in stained TEM (5.56 ± 0.22 nm), SAXS (c = 5.8 nm) and the model shown in Fig. 1a (6.0 ± 0.5 nm). Further structural confirmation is obtained by the systematic analysis of fast Fourier transforms (FFT) of selected areas, as shown in Figure 5b. The large set of diffraction spots is fully indexed by taking into account the group symmetry and lattice parameters identified in SAXS and simulating a diffraction pattern using the Carine software. Without any other adjustment parameter, all spots in Figure 5b were indexed to specific (*hkl*) consistent with the zone axis {100}. Similarly, Figures 5d, 5g and 5j show cryoEM images along the zone axes {101}, {111} and {011} respectively. Their corresponding FFT patterns in Figures 5e, 5h and 5k were fully indexed by the same method using the exact same SAXS parameters. In particular, one can notice the clear extinctions of the spots (*h00*) in Figure 5k and (*0k0*) in Figures 5b,e with *h* and *k* odd in full agreement with the SAXS patterns. The packing model derived from SAXS and consistent with these four orientations is shown, in Figs. 5c,f,i,l, for 9 superhelices similar to Fig. 1a. The distances d_{hkl} between Miller planes in each zone axis are measured directly in Figures 5b,e,h,k and compared with a good agreement to the SAXS-derived values in Table 1.

A systematic increase of 6 ± 1% or 17 ± 2% is observed for the cryoEM d_{hkl} compared to the SAXS d_{hkl} that is attributed to the crystal isotropic expansion during the cryogenic vitrification of the cryoEM samples. This model establishes that the grey straight region with the black dots on either side in Figure 5a is the inner space of the superhelix with an apparent thickness of 2.8 ± 1.0 nm (See Fig S7). The black dots are the tangential segments of the brick proteins within the helix. The staples from neighboring tubules are aligned and strongly interdigitated along the ***a*** axis therefore accounting for the larger parameter a = 80 Å compared to b = 65 Å. Finally, several characteristic Moiré patterns were also observed that further confirm the crystalline organization of the antiparallel superhelices (See Supplementary Information S9).

Discussion

Taken together, the SAXS, electron microscopy and molecular model of the self-assembly yield a precise description of the supramolecular organization induced from the onset of the brick and staple recognition events. Figures 6a-c give a molecular model of the 2D crystalline ordering of anti-parallel αRep superhelices viewed along the three main crystal axes. The superhelices formed by the bricks are aligned parallel to each other. The staples sit at the junction between two successive bricks that are brought together by hydrophobic interactions of their cap-free ends (Fig. S2), therefore reforming the cognate interface of the staple / bait complex (Fig. 1e). One striking aspect visible in Figures 6a and 6b is the ridges-into-grooves packing of bBE3 staples from neighboring helices. This fortuitous assembly is made possible by the periodicity (2 staples per turn) and the steric compatibility of the protruding staple of one superhelix in register with the inter-staple space of the neighboring antiparallel superhelix (Fig. 6d). Due to high regularity and the length of the assembly, even a relatively weak local interaction between the staples of adjacent superhelix could become predominant in the highly regular interhelix final assembly. This prominent interdigitation is first favored by inter-staple hydrophobic interactions as detailed in section 10 of the Supplementary Information. Yet, the ionic strength dependence of the assembly suggests that the main contribution originates from the complementary surface charges between the outer surface of the staple and the inner groove of the helix made from the side surfaces of the bricks. This is clearly illustrated in the

models of surface charges shown in Figures 6e-g. At neutral pH and ionic strength of 0.2, the convex surface of the staple is positively charged and aligned in-register with negative charges lining the inner surface of the bricks thus accounting for an electrostatic zipping of one superhelix alongside its neighbor (Fig. 6f). Figure 6g presents the facing electrostatic charges at an interface normal to the ***a*** axis where the charge intercalation occurs. The overlapped representation shows in black regions where positive charges of one planar set of superhelices overlaps negative charges of the next planar set of superhelices. These regions create the attractive force responsible for the inter-helix cohesion that shows a characteristic +10° and -10° tilt resulting from the fact that the staples are immobilized on the helix at an 80° angle with respect to the main axis. The resulting electrostatic torques acting on the two sets of staples coming from two adjacent superhelices are counter-rotating (Fig. 6g, black arrows). This strain, which is sensitive to variations in ionic strength, local pH and temperature, could be responsible for the variable supercoiling and changes in packing detected in SAXS patterns from one batch to another. The temperature-dependent structural transition monitored by SAXS (Fig. S6) suggests that a moderate input of energy is sufficient to modify the staple-staple repulsive steric interaction allowing for further interpenetration. The (600) peak shifts from 82.5 Å to 78.5 Å, which, in turn, decreases the helical pitch period from 58.8 Å to 56.8 Å. Incidentally, in negative-stain TEM, the positively charged patches on the back surfaces of the staples and exposed bricks (Fig. 6d) are bound by the highly negative molybdate ions accounting for the separation of entirely negative individualized superhelices. The precise and programmable 3-dimensional spatial ordering of the staple and brick proteins is directly visualized by cryoelectron tomography as shown in Figures 6h-j and in the supplementary video (See Supplementary Information S11). CryoEM 120°-tilt series of superhelix crystals were recorded from which 3D tomograms were reconstructed as detailed in the Experimental Section. Figures 6h-j show the 3D model resulting from the segmentation of 25 consecutive sections from a 430×175×160 nm^3 sub-tomogram (pixel size: 1.12 nm). The on-axis (Fig. 6h) and basal (Fig. 6i) views of the model

confirm the layered close-packing of parallel superhelices with a hollow inner cavity and match the (001) and (100) zone axis as depicted in Figures 6a and 6b respectively.

Importantly, the massive hierarchical assembly shown here demonstrates that the design of the brick structure and the mutual recognition surfaces of the two αRep proteins based on molecular models does lead to the expected superstructure. When combined with the evolutionary optimization of the brick/back-binder interaction by phage display, this modular approach appears capable of producing not only supramolecular assemblies, e.g. superhelices, with programmed geometry, but also higher order architectures. The protein origami forms within minutes after mixing at room temperature. This is fast compared to DNA origami which requires elevated temperatures and extended time to reach the targeted folded structure. Despite the very mild formation conditions, the supra-structure is extremely robust even at 75°C, due to a combination of hydrophobic and strong electrostatic network plus an extended hydrogen bond network at the staple / brick cognate surface.

Future designs of the surface charges of targeted residues on the convex surface of the back-binder and the concave surface of the brick, as well as shifting the relative position of the staples along the superhelix, will allow tuning of the inter-superhelix interactions. We will then be able either to limit the self-assembly at the individual superhelix stage, even at low ionic strength or promote the organization of a 2D helix crystal sheet.

Known affinity-dependant protein assembly is limited to a few available high affinity binding pairs where a specific protein is associated with its small molecule cognate partner (e. g. streptavidin/biotin, DHFR/methotrexate,(16)). Yet, the recent concomitant development of highly diverse repeat protein libraries and selection methods allows the identification of specific binders with submicromolar K_D for most folded protein targets, thus expanding the pool of affinity protein pairs.(33, 35, 37) The main challenge resides in selecting binders for a specific area of the targeted bait structure. The careful design of the binder selection strategy, which could include a counter-selection step for unwanted binders, makes it possible to screen for such adequate binders.(48) With the principles demonstrated here, other types of affinity protein pairs can serve as a staple/split bait pair to be appended to a brick scaffold, thereby opening numerous possible routes to generalizing the concept of designable protein origami.

The choice of the bait protein to be split and appended to any arbitrary “spacer” unit could be derived from natural (49) or computationally designed (50) repeat proteins as their structural regularity would facilitate the brick and staple assembly mechanism. Further increase of structural and functional complexity of the brick itself, within the supramolecular complexes, could be designed with the recent advent *of AlphaFold2,(51) RoseTTAFold (52) and Protein MPNN* (53) computational platforms. Imagination is the sole limit of future expansion of our proof-of-principle that will certainly include controlling of the origami size, embedding polydentate branching structures towards on-demand geometry of the supramolecular assembly, and eventually enzymatic or nanomaterial functionalization strategies.

MATERIALS AND METHODS

Protein preparation. Bait, staple and brick proteins were expressed as recombinant protein *in E.Coli*, in native forms, purified by Ni-chelate chromatography and gel filtration using standards methods (see Supplementary Information for details). The brick protein is expressed and purified with N and C cap repeats. These caps repeat were cleaved by TEV protease, at the two specific TEV cleavage sites located in between the caps and the central core of the protein. The cleaved protein is no longer His tagged and can be separated from the cleaved caps and TEV protease using Ni-Chelate chromatography.

Superhelix origami assembly. The molecular assembly was induced by mixing the bBE3 and brick proteins typically at 10 μM at 20°C (see Supplementary Information for details). Only when both the brick and staple proteins are mixed, a white “precipitate” appears within min after mixing. The kinetics of assembly was monitored by the absorbance increase at 340nm.(48) The insoluble assembly of proteins can be separated from the soluble materials by centrifugation. The protein composition of the pellet and supernatant fractions were determined by SDS-PAGE electrophoresis.

Negative staining sample protocol. Specimens were prepared for electron microscopy using the conventional negative staining procedure. For the analysis, 10 µl of solutions was adsorbed on Formvar carbon-coated grids for 5 min, blotted, and negatively stained with ammonium molybdate (3 wt%) for 3 min. Grids were examined with a TEM (Jeol JEM- 1400, JEOL Inc, Peabody, MA, USA) at 80 kV. Images were acquired using a Gatan Orius digital camera (Gatan Inc, Pleasanton, CA, USA).

CryoTEM sample protocol. 3 µL of sample were deposited onto glow-discharged lacey carbon grids and placed in the thermostatic chamber of a Leica EM-GP automatic plunge freezer, set at 20°C and 95% humidity. Excess solution was removed by blotting with Whatman n°1 filter paper, and the grids were immediately flash frozen in liquid ethane at -185°C. The frozen specimens were placed in a Gatan 626 cryo-holder, and cryoEM was carried out on a Jeol 2100 microscope, equipped with a LaB6 cathode and operating at 200 kV, under low dose conditions. Images were acquired with SerialEM software, with defocus of 1–2 µm, on a Gatan US4000

CCD camera. This device was placed at the end of a GIF Quantum energy filter (Gatan, Inc.), operated in zero-energy-loss mode, with a slit width of 25 eV. Images were recorded at a nominal magnification of 4 000x corresponding to calibrated pixel sizes of 1.71 Å.

SAXS measurements. SAXS measurements were carried out on two distinct types of samples. Samples representative of standard assembly were prepared in NaP buffer as described in "Superhelix origami assembly". For samples prepared in the presence of salts (see Figure 3), the final salt concentration was either 150 mM, for ammonium molybdate, or 500 mM, for ammonium chloride, in order to reach ionic strength of 1.05 and 1.20 respectively. For all preparations**,** 10 µL of the self-assembled proteins mixture is inserted into a capillary of 1.3 to 1.6 mm in diameter. The capillaries are probed in a homemade SAXS-WAXS Guinier beam line with a bidimensional Pilatus detector placed at 273 mm of the sample.(54) The beam is generated by a 30 µm X-ray Copper source (Xenocs). The beam is focused on the detector and monochromatized (λ = 1.541 Å) by a toroidal multilayer mirror (Xenocs). 2D concentric scattering rings are radially integrated as a function of $\boldsymbol{q} = 4\pi.\sin\theta$, where 2θ is the scattering angle. Typical acquisition time is 1 hour per diffraction pattern.

Cryoelectron tomography (cryoET). Tilt-series were acquired on a Talos Arctica (Thermo Fisher Scientific) operated at 200 kV in parallel beam condition, a K2 Summit direct electron detector and a BioQuantum energy filter (Gatan Inc.) operated in zero-loss mode with a slit width of 20 eV. Data were collected using Tomo software (Thermo Fisher Scientific), at a nominal magnification of 63 000x with a calibrated pixel size of 2.8 Å and between -5 and -8 µm defocus. Tilt series were acquired following the dose symmetric scheme between +60° and −60° with a step size of 2°.(55) Each tilt image was acquired in electron counting mode with a total dose of 70 e^-/A^2.

Tomogram reconstruction and modelling. Frames were aligned using MotionCor2 to correct for beam-induced motion and reconstruction was performed in IMOD.(56, 57) Tilt-series were aligned using the gold beads deposited on the surface of the support film as fiducial markers. 3D reconstructions with final pixel size of 11.2 Å were obtained by Simultaneous Iterative Reconstruction Technique (SIRT). Sub-volumes of the experimental tomograms comprising a

single crystal of superhelices were extracted and reoriented. The 3D model of the protein assemblies in the sub-volumes was constructed and visualized with the 3dmod.(58) A descriptive video on one such model is provided as supplementary material.

ACKNOWLEDGEMENTS

The authors thank B. Gigand and V. Campanacci for technical assistance in kinetic absorbance measurements. This work was funded by the Agence Nationale de la Recherche (ANR, Contracts No. ANR-14-CE08-0004-ARTEMIS, ANR-16-CE09-0027-HYBNAP, ANR-18-CE44-0013-Scaffold-Art, ANR-21-CE09-0045-ProteOrigami). J. Miller acknowledges the support of the Office for Science & Technology of the French Embassy in the United States through a Chateaubriand Fellowship. This work has benefited from the Electron Microscopy, Photonic Microscopy and Protein Interactions Platforms of I2BC, supported by French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INBS-05. F. A. acknowledges the Région Bretagne and Rennes Métropole for the X-ray setup support.

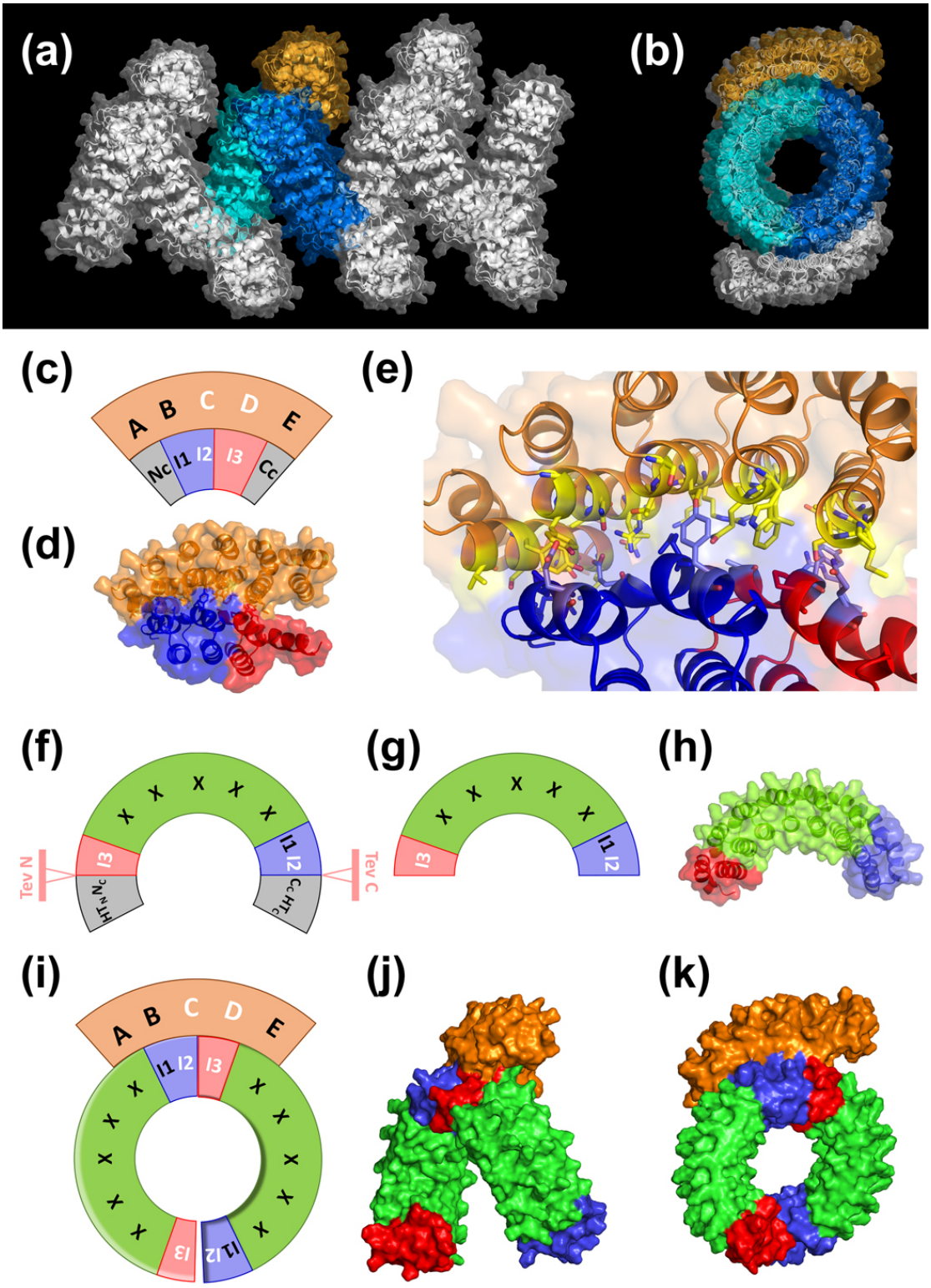

Figure 1 : Design principle of self-assembling helical protein origami. (a) Side and **(b)** axial view of a ribbon model of the supramolecular αRep helix based on the crystallographic structure reported in (PDB Code 8AW4). The elementary assembly motif comprises two brick proteins (blue and cyan) linked by a staple protein (orange). **(c)** Schematics and **(d)** crystallographic structure of the interaction pair formed by the 3-repeat "bait" αRep protein (N-

and C-capped $I_1I_2I_3$, blue-red) and the selected 5-repeat bBE3 back-binder (ABCDE, orange) used as origami staple. **(e)** Magnified view of the bBE3-bait interacting surface in (d) highlighting the variable side chains (yellow) of bBE3 that directly contact the side chains (purple) located on the convex surface of the bait. **(f-g)** Schematics and **(h)** computational ribbon model of the brick protein **(e)** before and **(f,g)** after cleaving the His-tagged N-cap (HT_NN_C) and C-cap (C_CHT_C) with TEV protease. X indicates internal repeats with arbitrary hypervariable sequences. **(i)** Schematic and **(j,k)** computational ribbon model of the heterotrimeric junction driving the origami assembly that is composed of two concatenated bricks stapled together by a bBE3 back-binder. The bait-like pairing partner of bBE3 is reconstituted from the C-terminal repeats (I_1I_2, blue) of one brick and the N-terminal repeat (I_3, red) of the next brick. **(j)** and **(k)** show the lateral and axial views of the assembled heterotrimeric motif that is propagated in the superhelical origami (see colored sections in **(a,b)**).

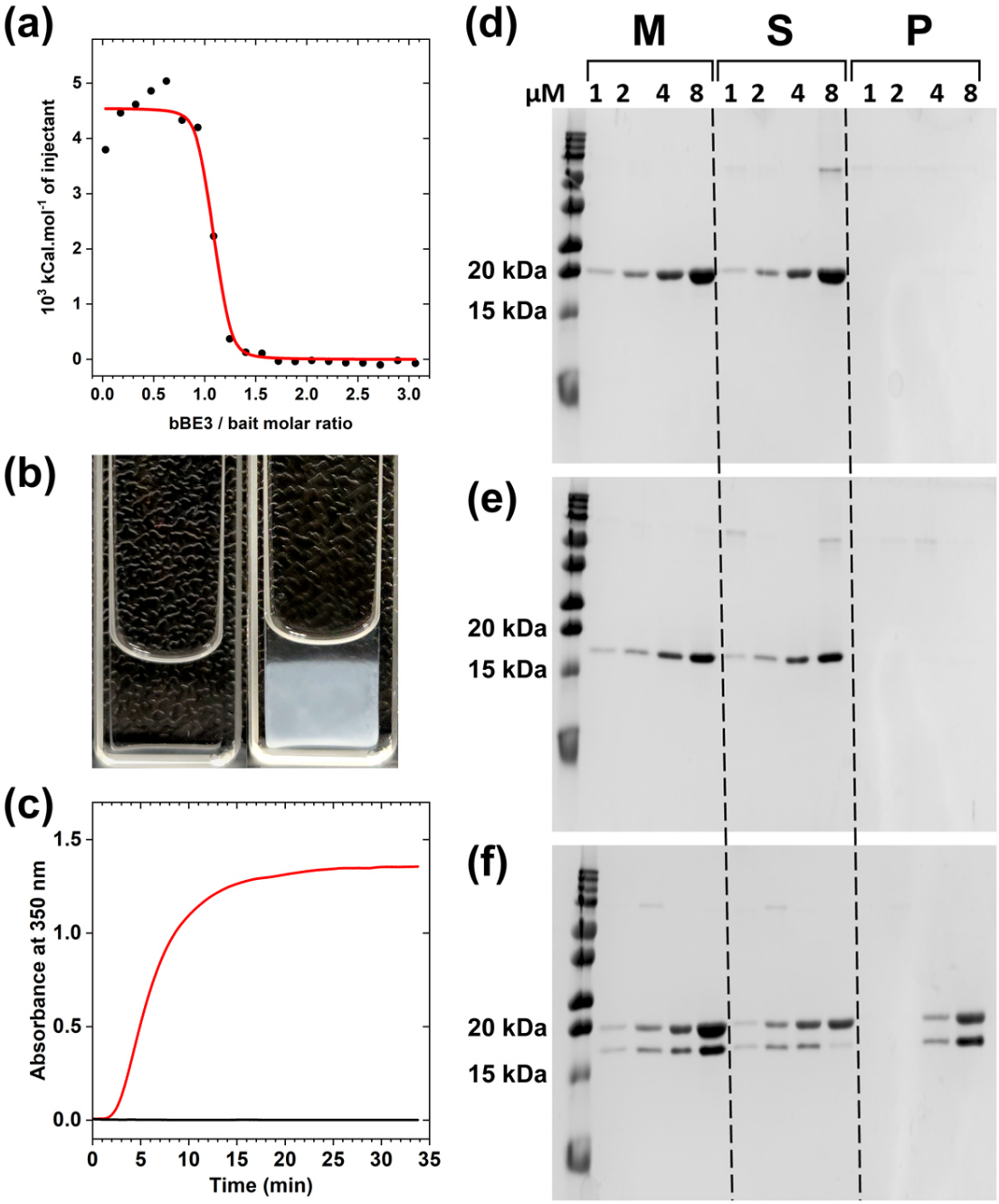

Figure 2: Fast and robust supramolecular assembly of brick / bBE3. (a) Calorimetric titrations of bait (25 µM) with the staple bBE3 (375 µM). The dissociation constant (K_D = 68 ± 23 nM) and stoichiometry (n = 1.05) are extracted from the saturation curve. **(b)** Photograph of the 20 µM bBE3 staple solution before mixing (left) and of the white and turbid suspension rapidly obtained when mixing identical volumes of 20 µM of each brick and bBE3 (right). **(c)** Turbidity measurement by light scattering at 350 nm as a function of time after mixing brick and staple bBE3 in equimolar ratio (total protein concentration 20 µM) at 4°C (black) and 25°C (red). **(d-f)** SDS-PAGE analysis of monomers and assembly in **(d)** a staple bBE3 fraction, **(e)** a brick fraction, **(f)** mixed staple and brick fractions before centrifugation (M), supernatants (S) and pellets (P) obtained after 30 min 14000g centrifugation and re-suspension in the same volume buffer (see material and methods). 10 µL of each sample were run after incubation of the mixtures at 1, 2, 4 and 8 µM concentrations.

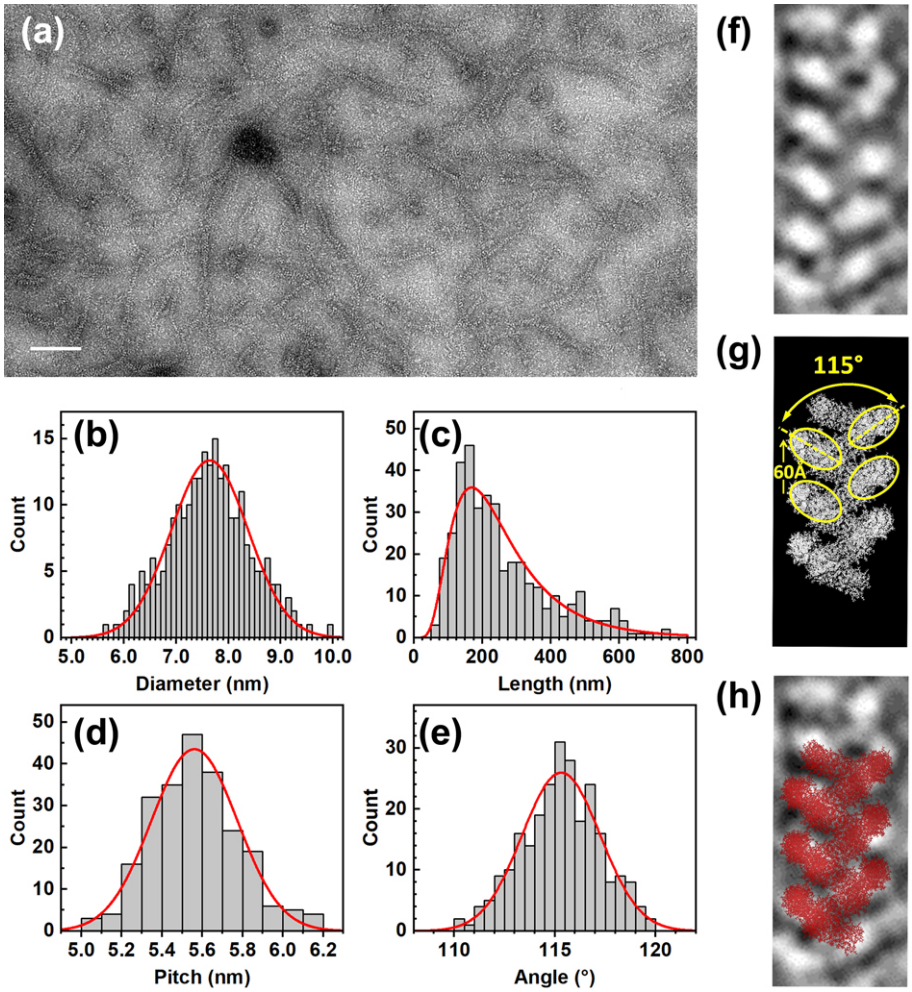

Figure 3: Electron microscopy of individualized protein origami superhelix. (a) Transmission electron micrograph of scattered self-assembled protein superhelices. TEM contrast is enhanced by negative staining with 150 mM ammonium molybdate. **(b)** Normal distribution of the diameter (7.6 ± 0.8 nm) and **(c)** lognormal distribution of the length (260 ± 150 nm) of the tubular superhelices. Normal distributions of **(d)** the pitch (55.6 ± 2.2 Å) and **(e)** relative directional angle (115° ± 2°) of the periodical white ellipsoids observed along the superhelices. **(f)** Magnified TEM image of one fibrillary structure showing periodical braid-like patterns. **(g)** Lateral projection of the structural model of the superhelix highlighting the dense areas (staple, tangential section of the bricks) from where the molybdate stain is excluded. They form pairs of ellipsoids with main directions at 115° from each other and with a pitch of ca. 60 Å. **(h)** Overlaid representation of **(f)** and **(g)**.

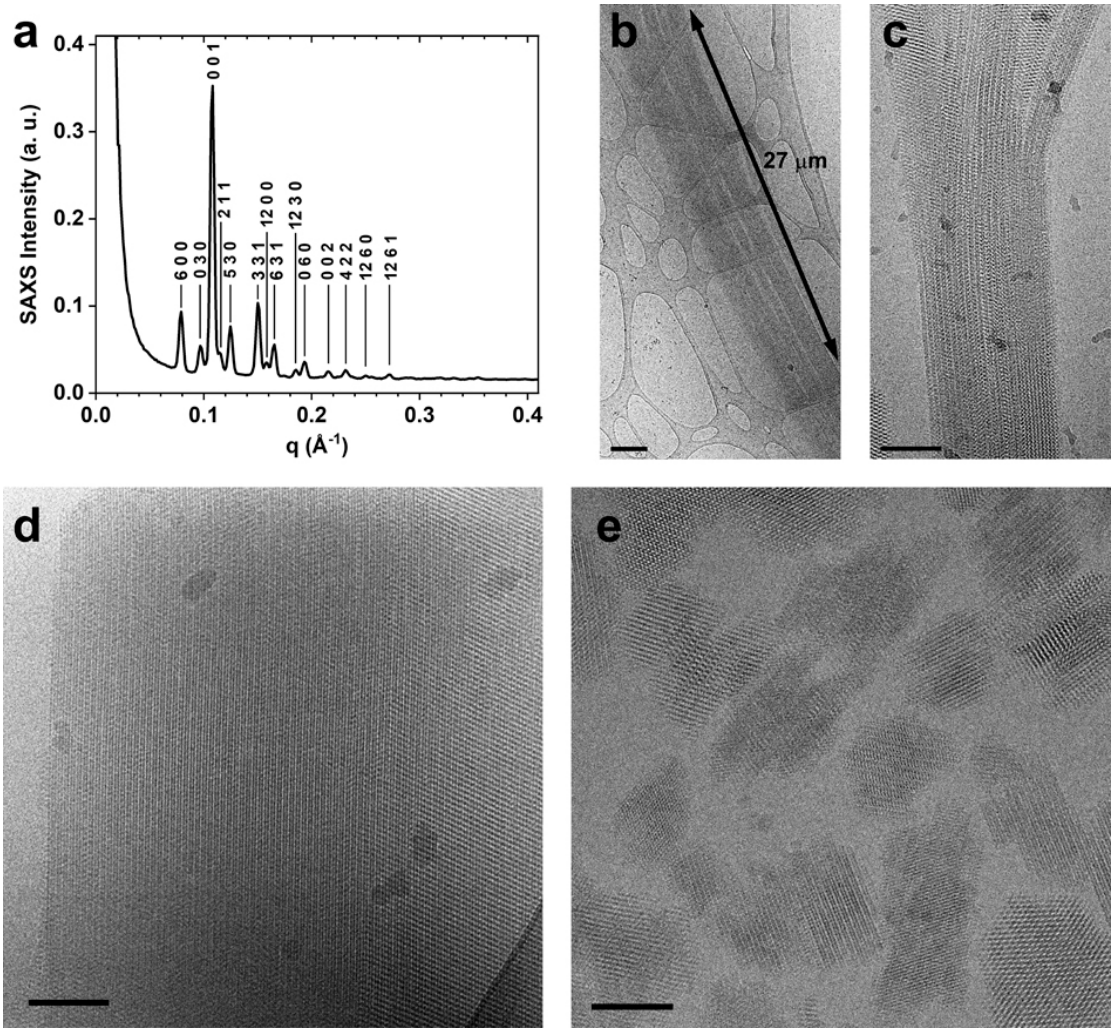

Figure 4: Massive ordering of superhelical protein origami into quasi-crystalline bundles. (a) SAXS diffraction pattern of a 10 μM supramolecular assembly suspended in water. The peaks are indexed to the $P2_122$ orthorhombic symmetry group, with elementary cell parameters a = 477 Å, b = 95 Å and c = 58.2 Å. This is a (x6, x3, x1) superlattice of a subcell (a = 80 Å, b = 65 Å and c = 58 Å). **(b, c)** Wide field and zoomed cryo-TEM images of brick-bBE3 superhelix aligned into large scale compact bundles. **(d)** Magnified view of a bundle tip showing highly ordered parallel tubules. **(e)** Multiple fragments of superhelix bundles with different orientations aligned along the electron beam. Scale bars are (b) 2 μm, (c,d,e) 100 nm.

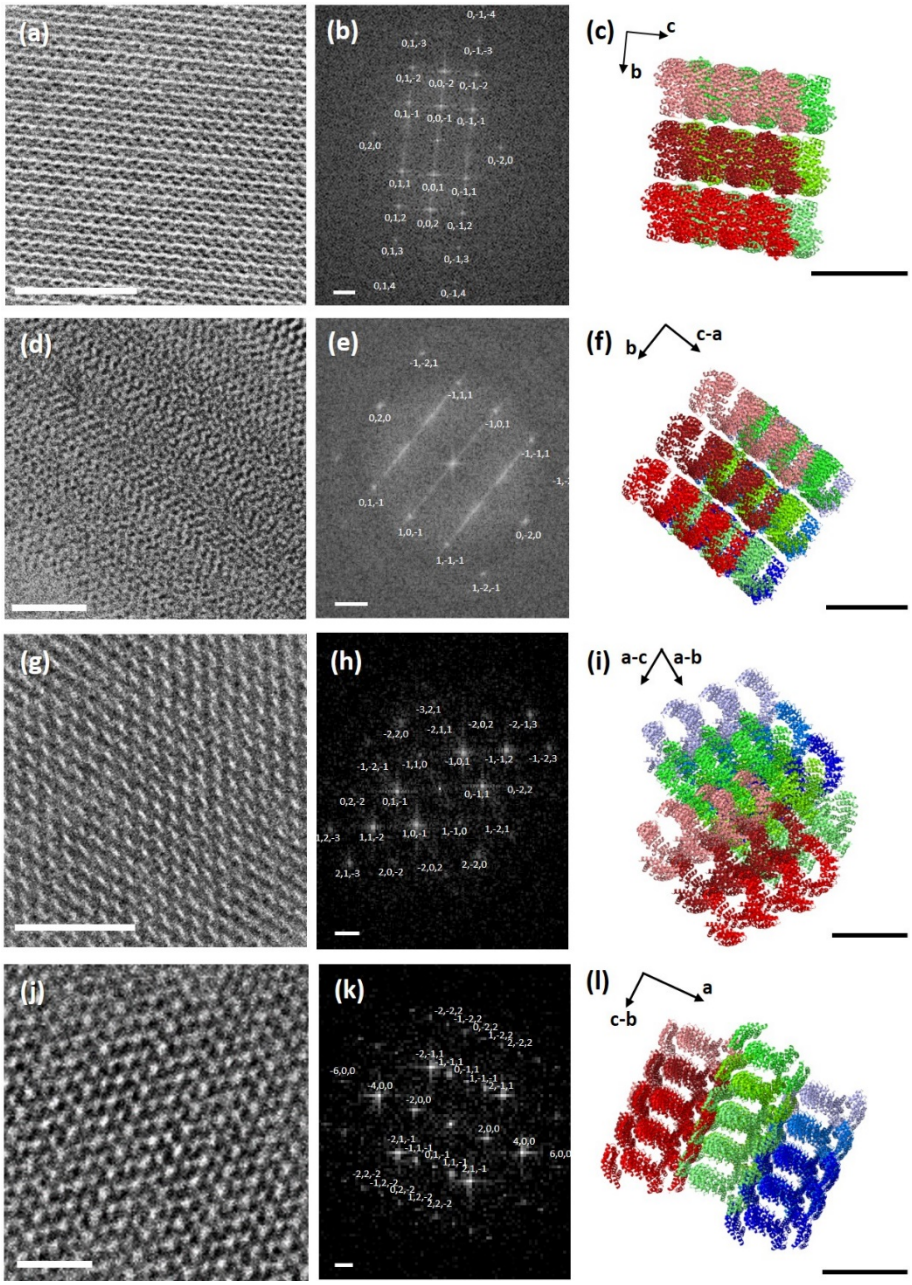

Figure 5: 2D crystals of superhelical protein origami. Zoomed areas of cryoEM images similar to Fig. 4d,e with different orientations with respect to the e-beam axis in (a,d,g,j) are Fast Fourier Transformed in (b,e,h,k) to emulate Selected Area Electron Diffraction patterns. (c,f,i,l) display a 3D model consisting of 3x3 superhelices shown in Fig. 1a, arranged in a $P2_122$ orthorhombic crystal with lattice parameters as determined by SAXS and oriented like the TEM images. **(a-c)** [100] zone axis viewed in **(a)** direct space cryoEM image, **(b)** fully indexed FFT image of (a) and **(c)** 3D model of 3x3 superhelices in the same orientation. Similar data are shown along the **(d-f)** [101], **(g-i)** [111] and **(j-l)** [011] zones axes. Note the extinctions of the (*h00*) spots in panel (k) and (*0k0*) spots in panels (b) and (e) with *h* and *k* odd. Scale bars are (a, d, g) 50 nm, (j) 20 nm, (b, e, h, k) 0.1 nm^{-1}, (c, f, i, l) 10 nm.

Zone axis	(hkl)	cryoEM d_{hkl} (nm)	SAXS d_{hkl} (nm)	Relative variation
{100}	(001)	6.23	5.82	+7%
	(011)	4.61	4.33	+7%
	(020)	3.39	3.24	+5%
{101}	(101)	4.93	4.70	+5%
	(111)	4.07	3.80	+7%
	(020)	3.46	3.24	+7%
{111}	(011)	5.12	4.33	+18%
	(101)	5.61	4.70	+19%
	(110)	5.97	5.03	+18%
{011}	(200)	4.59	3.98	+16%
	(011)	4.95	4.33	+14%
	(111)	4.41	3.81	+16%

Table 1: Miller interplanar distances d_{hkl} along zone axes {100}, {101}, {111} and {011} derived from FFT cryoEM images shown in Figure 5 and SAXS data. The relative variation with respect to the SAXS-derived values is shown in the rightmost column.

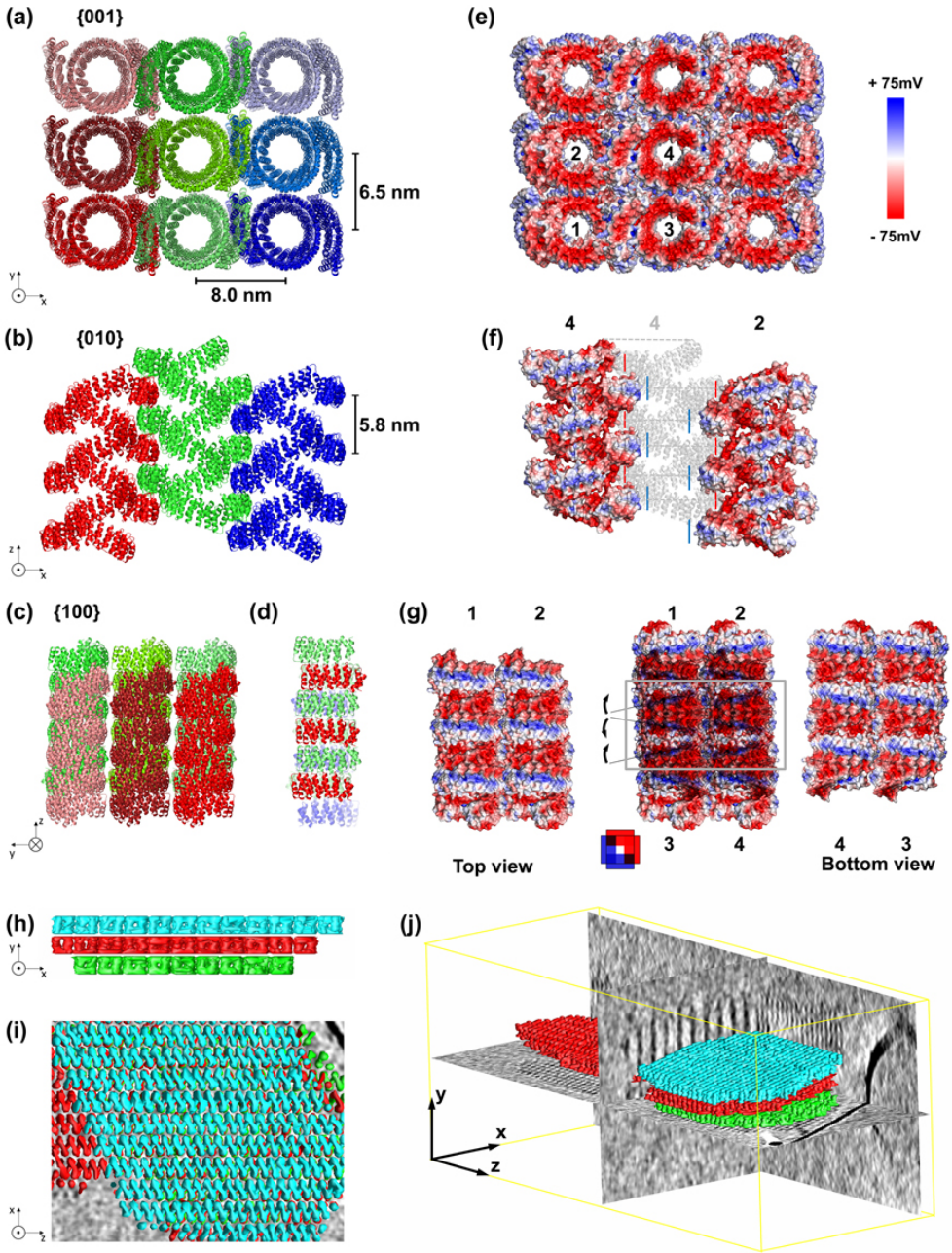

Figure 6: 3D structural and electrostatic modelling and cryotomography of the protein origami crystal. (a-d) Structural model of the superhelix 2D crystal derived from the SAXS and EM data and viewed along the (a) {001}, (b) {010} and (c) {001} zones axis. The strong interdigitation of the staples along the ***a*** axis is visible in (a) and (b) while the alternating black dots observed experimentally clearly visible in (c) as tangential segments of the superhelix. In (d) the brick proteins are hidden to show, along the ***a*** axis view, how the staples of the green superhelix interdigitate with the staples of the red superhelix in the front and the ones of the blue superhelix in the back. **(e-g)** Electrostatic surface of the superhelical crystal viewed along the (e) {001], (f) {010} and (g) {100} zone axis.(59, 60) In (f) The antiparallel superhelices are overlapped with a schematic of the positions and signs of the opposed charged surfaces: protruding positive surface from the back of the staple (blue segments) face the deep negative superhelix grooves (red segments). The shaded ribbon structure represents the position of the left superhelix 4 inside the crystal with respect to the right superhelix 2 as labelled in (e). (g) Top (left) and bottom (right) surfaces of the superhelices 1, 2 and 3,4 respectively viewed along the ***a*** axis showing the positively charged back side of the staple and negatively charged inner

groove of the superhelix. The central overlayed representation indicates, in black, regions where opposite charges coincide while the color is preserved when similar charges are aligned in this view along the ***a*** axis The black regions point where positively charged staples of one layer fit into the negatively charged grooves of the other layer. The relative angle between the two sets of staples is highlighted with the black lines. The resulting electrostatic torques acting on the upper and lower layers are counter-rotating (black arrows), accounting for the distortion detected in SAXS patterns. **(h-j)** Model views of a cryoelectron tomogram (cryoET) of a small crystal of superhelices. Out of a 430x175x160 nm^3 reconstructed tomogram (cubic pixel size 1.12 nm) delineated in yellow in (h), three layers of parallel superhelices could be modelled and are shown in (h) along the ***c*** axis, (i) along the ***b*** axis and (i) in 3D with three particular cryoET cross-sections. See also Supplementary Material.

References

1. A. R. Voet, J. R. Tame, Protein-templated synthesis of metal-based nanomaterials. *Curr Opin Biotechnol* **46**, 14-19 (2017).
2. I. W. Hamley, Protein Assemblies: Nature-Inspired and Designed Nanostructures. *Biomacromolecules* **20**, 1829-1848 (2019).
3. A. Beloqui, A. L. Cortajarena, Protein-based functional hybrid bionanomaterials by bottom-up approaches. *Curr Opin Struct Biol* **63**, 74-81 (2020).
4. T. Aida, E. W. Meijer, S. I. Stupp, Functional Supramolecular Polymers. *Science* **335**, 813-817 (2012).
5. M. A. Meyers, P. Y. Chen, A. Y. M. Lin, Y. Seki, Biological materials: Structure and mechanical properties. *Prog. Mater. Sci.* **53**, 1-206 (2008).
6. H. Lee, W. C. DeLoache, J. E. Dueber, Spatial organization of enzymes for metabolic engineering. *Metab Eng* **14**, 242-251 (2012).
7. T. O. Yeates, M. P. Agdanowski, Y. Liu, Development of imaging scaffolds for cryo-electron microscopy. *Curr Opin Struct Biol* **60**, 142-149 (2020).
8. N. H. Kim *et al.*, Supramolecular assembly of protein building blocks: from folding to function. *Nano Converg* **9**, 4 (2022).
9. N. Kobayashi, R. Arai, Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr Opin Biotechnol* **46**, 57-65 (2017).
10. Q. Luo, C. Hou, Y. Bai, R. Wang, J. Liu, Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chem Rev* **116**, 13571-13632 (2016).
11. J. Zhu *et al.*, Protein Assembly by Design. *Chem Rev* **121**, 13701-13796 (2021).
12. N. Kobayashi *et al.*, Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric de Novo Protein. *J Am Chem Soc* **137**, 11285-11293 (2015).
13. Y. T. Lai, K. L. Tsai, M. R. Sawaya, F. J. Asturias, T. O. Yeates, Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. *J Am Chem Soc* **135**, 7738-7743 (2013).
14. T. O. Yeates, Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu Rev Biophys* **46**, 23-42 (2017).
15. L. A. Churchfield, F. A. Tezcan, Design and Construction of Functional Supramolecular Metalloprotein Assemblies. *Acc Chem Res* **52**, 345-355 (2019).
16. J. C. Carlson *et al.*, Chemically controlled self-assembly of protein nanorings. *J Am Chem Soc* **128**, 7630-7638 (2006).
17. J. A. Harvey, L. S. Itzhaki, E. R. G. Main, Programmed Protein Self-Assembly Driven by Genetically Encoded Intein-Mediated Native Chemical Ligation. *ACS Synth Biol* **7**, 1067-1074 (2018).
18. H. Shen *et al.*, De novo design of self-assembling helical protein filaments. *Science* **362**, 705-709 (2018).
19. A. J. Ben-Sasson *et al.*, Design of biologically active binary protein 2D materials. *Nature* **589**, 468-473 (2021).
20. C. E. Correnti *et al.*, Engineering and functionalization of large circular tandem repeat protein nanoparticles. *Nat Struct Mol Biol* **27**, 342-350 (2020).
21. K. A. Cannon *et al.*, Design and structure of two new protein cages illustrate successes and ongoing challenges in protein engineering. *Protein Sci* **29**, 919-929 (2020).
22. F. Parmeggiani *et al.*, A general computational approach for repeat protein design. *J Mol Biol* **427**, 563-575 (2015).

23. V. Alva, J. Soding, A. N. Lupas, A vocabulary of ancient peptides at the origin of folded proteins. *Elife* **4**, e09410 (2015).
24. S. A. Hughes *et al.*, Ambidextrous helical nanotubes from self-assembly of designed helical hairpin motifs. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14456-14464 (2019).
25. Y. L. Boersma, A. Pluckthun, DARPins and other repeat protein scaffolds: advances in engineering and applications. *Curr Opin Biotechnol* **22**, 849-857 (2011).
26. C. L. Edgell, N. J. Savery, D. N. Woolfson, Robust De Novo-Designed Homotetrameric Coiled Coils. *Biochemistry* **59**, 1087-1092 (2020).
27. G. G. Rhys *et al.*, Navigating the Structural Landscape of De Novo alpha-Helical Bundles. *J Am Chem Soc* **141**, 8787-8797 (2019).
28. A. J. Scott *et al.*, Constructing ion channels from water-soluble alpha-helical barrels. *Nat Chem* **13**, 643-650 (2021).
29. H. Gradisar *et al.*, Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat Chem Biol* **9**, 362-366 (2013).
30. F. Lapenta, J. Aupic, Z. Strmsek, R. Jerala, Coiled coil protein origami: from modular design principles towards biotechnological applications. *Chem Soc Rev* **47**, 3530-3542 (2018).
31. F. Lapenta, R. Jerala, Design of novel protein building modules and modular architectures. *Curr Opin Struct Biol* **63**, 90-96 (2020).
32. A. Urvoas *et al.*, Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats. *J Mol Biol* **404**, 307-327 (2010).
33. A. Pluckthun, Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu Rev Pharmacol Toxicol* **55**, 489-511 (2015).
34. K. B. Uribe *et al.*, Engineered Repeat Protein Hybrids: The New Horizon for Biologic Medicines and Diagnostic Tools. *Accounts Chem. Res.* **54**, 4166-4177 (2021).
35. S. C. Lee *et al.*, Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3299-3304 (2012).
36. A. Chevrel *et al.*, Alpha repeat proteins (alphaRep) as expression and crystallization helpers. *J Struct Biol* **201**, 88-99 (2018).
37. A. Guellouz *et al.*, Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (alphaRep). *PLoS One* **8**, e71512 (2013).
38. K. L. Gurunatha *et al.*, Nanoparticles Self-Assembly Driven by High Affinity Repeat Protein Pairing. *ACS Nano* **10**, 3176-3185 (2016).
39. C. Leger *et al.*, Ligand-induced conformational switch in an artificial bidomain protein scaffold. *Sci Rep* **9**, 13 (2019).
40. M. Fernandez *et al.*, Hybrid gold nanoparticle-quantum dot self-assembled nanostructures driven by complementary artificial proteins. *Nanoscale* **12**, 4612-4621 (2020).
41. L. M. Blancas-Mejia, P. Misra, M. Ramirez-Alvarado, Differences in Protein Concentration Dependence for Nucleation and Elongation in Light Chain Amyloid Formation. *Biochemistry* **56**, 757-766 (2017).
42. E. T. Powers, D. L. Powers, The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the "supercritical concentration''. *Biophys. J.* **91**, 122-132 (2006).
43. C. Valery *et al.*, Biomimetic organization: Octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10258-10262 (2003).
44. J. Puiggali *et al.*, The frustrated structure of poly(L-lactide). *Polymer* **41**, 8921-8930 (2000).
45. A. A. Kornyshev, D. J. Lee, S. Leikin, A. Wynveen, Structure and interactions of biological helices. *Rev. Mod. Phys.* **79**, 943-996 (2007).

46. C. Salameh *et al.*, Origin of transparency in scattering biomimetic collagen materials. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 11947-11953 (2020).
47. D. Durand, J. Doucet, F. Livolant, A study of the structure of highly concentrated phases of DNA by X-ray diffraction. *J. Phys. II* **2**, 1769-1783 (1992).
48. V. Campanacci *et al.*, Selection and Characterization of Artificial Proteins Targeting the Tubulin alpha Subunit. *Structure* **27**, 497-+ (2019).
49. L. Paladin *et al.*, RepeatsDB in 2021: improved data and extended classification for protein tandem repeat structures. *Nucleic Acids Res.* **49**, D452-D457 (2021).
50. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580-+ (2015).
51. P. Cramer, AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* **28**, 704-705 (2021).
52. I. R. Humphreys *et al.*, Computed structures of core eukaryotic protein complexes. *Science* **374**, 1340-+ (2021).
53. J. Dauparas *et al.*, Robust deep learning-based protein sequence design using ProteinMPNN. *Science (New York, N.Y.)* **378**, 49-56 (2022).
54. T. Bizien, J. C. Ameline, K. G. Yager, V. Marchi, F. Artzner, Self-Organization of Quantum Rods Induced by Lipid Membrane Corrugations. *Langmuir* **31**, 12148-12154 (2015).
55. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191-198 (2017).
56. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331-332 (2017).
57. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71-76 (1996).
58. D. N. Mastronarde, Dual-axis tomography: An approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343-352 (1997).
59. M. Hebditch, J. Warwicker, Web-based display of protein surface and pH-dependent properties for assessing the developability of biotherapeutics. *Sci Rep* **9**, 9 (2019).
60. E. Jurrus *et al.*, Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112-128 (2018).

A new building strategy in protein origami: design and self-assembly of a brick and staple artificial repeat protein pair leading to macroscopic tubular superhelices.

Laureen Moreaud,[a 1] Sébastien Viollet,[b 1] Agathe Urvoas,[b] Marie Valerio-Lepiniec,[b] Agnès Mesneau,[b] Inès Li de la Sierra-Gallay,[b] Jessalyn Miller,[b,f] Malika Ouldali,[b] Cécile Marcelot,[a] Stéphanie Balor,[c] Vanessa Soldan,[c] Cristelle Meriadec,[d] Franck Artzner,[d] Erik Dujardin[a, e 2] and Philippe Minard[b 2]

[a] Centre d'Elaboration des Matériaux et d'Etudes Structurales (CEMES), CNRS UPR 8011, 29 rue J. Marvig, B.P. 94347, F-31055 Toulouse, France
[b] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, F-91198 Gif-sur-Yvette CEDEX, France
[c] METi, Centre de Biologie Intégrative, Université de Toulouse, CNRS, UPS, 31062, Toulouse, France.
[d] Institut de Physique de Rennes (IPR), CNRS, UMR 6251, Université de Rennes 1, F-35042 Rennes, France
[e] Laboratoire Interdisciplinaire Carnot de Bourgogne, CNRS UMR 6303, Université de Bourgogne Franche-Comté, 21000 Dijon, France
[f] Department of Chemistry, Emory University, Atlanta, Georgia 30322, USA

[1] These authors contributed equally to the work.

[2] To whom correspondence may be addressed. Email: philippe.minard@i2bc.paris-saclay.fr or erik.dujardin@cnrs.fr

Supplementary Information

Materials and Methods.

Protein production and purification. The *E. coli* strain M15 (pREP4) was transformed by the plasmid coding for the brick, staple bBE3 or bait proteins. Cells were grown at 37°C in 2YT medium containing 100 μg.mL^{-1} ampicillin until the optical density $OD_{600\ nm}$ = 0.6. Protein expression was induced by addition of IPTG (0.5 mM final concentration). The cells were incubated 4 h at 200 rpm and 37°C. Then they were harvested, suspended in phosphate buffered saline (PBS: 50 mM Na Phosphate, 150 mM NaCl, pH 8.0) supplemented with anti-protease (PIC Roche), treated with DNase 1 (Thermo Scientific) and sonicated. The His_6-tagged proteins were purified from crude supernatant using nickel-affinity chromatography (Ni-NTA agarose, Qiagen) followed by size-exclusion chromatography (Hiload 16/60 SuperdexTM 75) in HEPES buffer (HEPES 20 mM pH 8.0, 150 mM NaCl).

Digestion by TEV-Protease. N-terminal His_6-Cap and C-terminal Cap-His_6 were cleaved by mixing a ratio of 50:1 brick and TEV-protease, incubating overnight 4°C in TEV buffer (20mM HEPES pH 8.0, 5 mM EDTA,1 mM DTT). The cleaved protein was purified by loading the sample on Ni-NTA agarose and collected in the flow through fraction as the His_6-tag containing proteins (uncleaved, partially cleaved fragments and TEV protease) were retained on the Ni-NTA resin.

Protein assembly controlled by SDS-PAGE. The molecular assembly was triggered by mixing the bBE3 and brick proteins at various concentrations. 50 µL of a 20 µM brick protein buffered solution and 50 µL of a 20 µM staple bBE3 protein solution were mixed by pipetting up and down in an Eppendorf tube and briefly shaken. A white precipitate appears within 5-10 min after mixing. The mixture is left undisturbed for 30 min at room temperature (20°C). Each protein alone and the protein mix were incubated overnight at 20°C in HEPES buffer. For each protein or mix, 3 types of samples were prepared for further SDS-PAGE analyses, the mix (M), the supernatant (S) and the pellet (P). For the supernatant and the pellet samples preparations, 100 µL of sample were centrifuged at 14 000g, after saving the supernatant (S fraction), the pellet sample (P fraction) was obtained by re-suspension in 100 µL of HEPES buffer. The mix (M) samples corresponded to the equimolar mix of both proteins for each concentration. Samples were treated with Laemmli

denaturation buffer (1X final concentration) boiled for 5 min at 95°C, run on a 20% SDS-PAGE gel and stained with InstantBlueTM Coomassie protein stain (Expedeon). 10 μL of each sample were run for the mixtures at different concentrations (1, 2 4, 8 μM). Assembly kinetic experiments were monitored at 350 nm with a Cary 50 spectrophotometer (Agilent Technologies), using a 0.7-cm path length cuvette thermostated at 25° C or 4°C.

S1. Bait, brick and staple protein sequences. Structure of the bait/back binder complex.

The sequences of the bait, staple and brick proteins are shown below. Although αRep are composed of several repeats on a single polypeptide chain, the successive repeated modules of each protein are reported here on separated lines with the same colour code per repeat of the bait and brick proteins. The same colour code is used on the crystal structure of the complex Bait / back-Binder bBE3 shown in Figure S1. The residues of the back surface of the bait protein and the binding surface of the staple protein are underlined.

Bait (Mw 17.1 kDa)

N-Cap	TDPEKVDMYIENLRDEDPEVRARAAEALGKI
I1	GDERAVPALIEALKDEDSNVRKEAARALGEI
I2	GDPEAVEALIYALRDEDADVRKEAAAALGQI
I3	GDEAAVYPLIQALEDEDSDVRRAAAEALGRI
C-Cap	GDPRAEEALRRAREDEDPEVQKEAEKAEGEIR

Brick (Mw 37.3 kDa)

His tag	MRGSHHHHHH
N Cap	TDPEKVDMYIENLRDEDPEVRARAAEALGKI
TEV N cleavage site	GSGSGENLYFQ / GGSGSG
I3	GDEAAVYPLIQALEDEDSDVRRAAAEALGRI
Xa	GDERAVPALIEALKDEDPEVRKEAAKALGEI
Xb	GDERAVEPLIKALKDEDDVRKEAAAALGQI
Xc	GDERAVEPLIKALKDEDSDVRRAAAEALGRI
Xd	GDERAVEPLIKALKDEDSNVRKEAARALGEI
Xe	GDEAAVEPLIQALEDEDPEVRARAAEALGKI
I1	GDERAVPALIEALKDEDSNVRKEAARALGEI
I2	GDPEAVEALIYALRDEDADVRKEAAAALGQI
TEV C cleavage site	GSGSGENLYFQ/GGSGSG
C cap	GDPRAEEALRRAREDEDPEVQKEAEKA

Staple (Back Binder E3: bBE3) (Mw 22.7 kDa)

HisTag	MRGSHHHHHHTENLYFQG
	SPEKVEMYIKNLQDDSIVVRYSAASALGKI
	GDERAVEPLIKALKDEDGYVRQAAALALGQI
	GDERAVEPLIKALKDEDSTVRIRAARALGKI
	GDERAVEPLIKALKDEDWQVRLSAASALGKI
	GDERAVEPLIKALKDEDPSVRMAAANALGQI
	GGERVRAAMEKLAETGTGFARKVAVNYLETHKSLIS

The crystal structure of a bBE3/bait complex was obtained with a variant of the bait protein possessing different side chains on its concave surface. The sequence differences between the bait protein used for the selection (see sequence above) and the bait protein that was crystallized (PDB code 8AW4) are all located on the concave surface of the bait protein which does not interact with the “back binder” protein.

In Figure S1a, the variable surface of the “back binder” protein (bBE3 in orange) interacts with the outside surfaces of the first helices of repeat I_1 (green), I_2 (cyan) and I_3 (purple) of the bait protein. In the crystal structure, the N cap module and the second helix of the C cap of the bait protein were not visible in the electron density map.

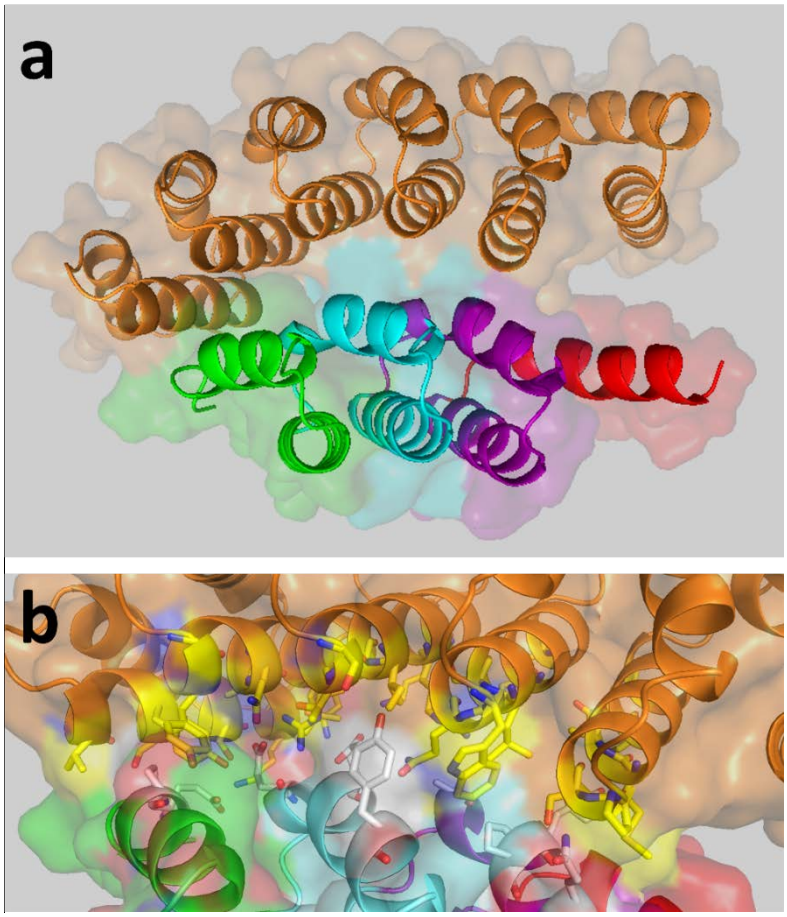

Figure S1: Structure of a bBE3/bait protein complex (PDB Code 8AW4). The structure of the bait protein is colored using the color code used for the sequence. (a) The variable surface of the back-binder protein (bBE3 in orange) interacts with the outside surfaces of the first helices of repeat I_1 (green), I_2 (cyan) and I_3 (purple) of the bait protein. In the crystal structure, the N cap module and the second helix of the C cap of the bait protein were not visible in the electron density map. (b) A closer view of the interaction surface. The side chains located in the interaction surface are shown in yellow and white sticks, for the back binder and the bait protein respectively.

A closer view of the interaction surface is shown in Figure S1b. The sequence of the bait protein is circularly permutated in the brick sequence in such a way that repeat I_1(green) and I_2 (Cyan) are located at the C extremity of the brick protein, while I_3 (Purple) is now located at the N terminal part of the sequence. The back-binder reconstitutes its cognate binding surface by binding the end repeats from two molecules of the brick protein in the same relative position that these repeats adopt in the bait protein.

S2. Hydrophobic interactions between bricks and staple

In order to understand the temperature effect monitored in ITC (Fig. 2), we have examined the hydrophobic patch distribution inside the back-binder/bait complex, at the junction between two consecutive bricks and near the staple-brick recognition site as shown in Figure S2.
We have used the Protein-Sol software (https://protein-sol.manchester.ac.uk/) and the ProteinTools site (https://proteintools.uni-bayreuth.de/clusters/structure).

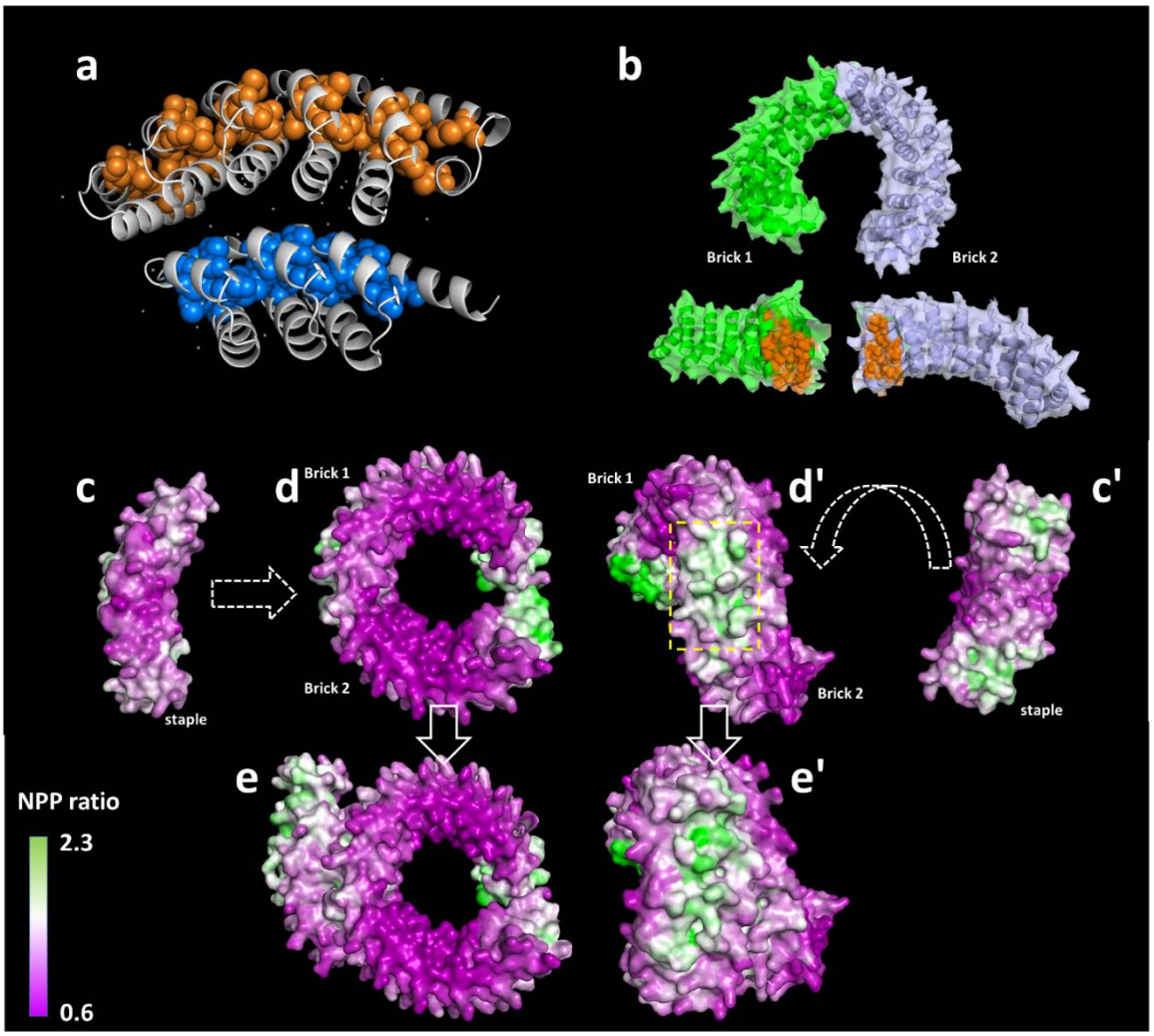

Figure S2. Hydrophobic interactions between brick and staple. (a) bBE3-bait complex structure with hydrophobic clusters highlighted in orange (bBE3) and blue (bait). **(b)** Axial and "open book" representation of two consecutive cap-free bricks from the superhelix. In the "open book" representation the hydrophobic clusters of both proteins are highlighted in orange. **(c-e, c'-e')** Axial (c-e) and transverse (c'-e') views of (c, c') the staple, (d, d') two consecutive bricks 1 and 2 and (e, e') the assembled complex colored coded in purple-white-green from low to high NPP (non-polar to polar) ratio. Green color indicates the most hydrophobic areas while purple color outlines the polar regions. Note that (c') displays the convex surface of the staple that is flipped by 180° onto the concave surfaces of the bricks shown in (d'). In (d') the yellow dashed box highlights the sole hydrophobic patch found on the concave surface of the bricks and located at the junction between bricks 1 and 2, precisely where the staple docks.

The highlight of the hydrophobic clusters in the back-binder/bait complex (Fig. S2a) confirms the strong hydrophobic interaction between repeats inside the proteins. When the αRep are used without caps, as in the case of the bricks, this implies significant hydrophobic interactions between two consecutive bricks in the superhelix. In Figure S2b, the last repeat of the brick 1 and the first repeat of the next one (brick 2) are shown in axial view (top) and in an "open book" representation (bottom) that reveals, in orange, the two hydrophobic clusters facing each other. Clearly, this head-to-tail interaction of the bricks themselves will therefore be less favored at 4°C than at 25°C.

This interaction needs to occur in tandem with the docking of the staple in order to recreate the recognition surface of the staple in order to assemble the superhelix. Figure S2a indicates that the backside of the bait also shows a moderately hydrophobic surface. This can be tracked when we observe that the backside of the bricks at junction between two consecutive bricks, where the staple is interacting as shown in Figs. S2c-e and Figs. S2c'-e', where the "Non-Polar to Polar" ratio (NPP) reveals the hydrophobic patches in green and polar ones in purple.

The yellow dashed box in Fig. S2d' shows the only hydrophobic patch existing on the outer rim of the superhelix, except the cap-free ends of the bricks that are buried inside the superhelix once the bricks are adequately positioned. The interacting surface of the staple shown on the top right exhibit a mixed polar-hydrophobic character. It is therefore possible that the hydrophobic contribution to the staple-brick recognition is also less efficient at low temperature.

Note that beyond the hydrophobic contribution to the interdigitation described in SI section S10, the crystal structure is also stabilized by electrostatic inter-helix interactions. These interactions are athermal, and, consequently, they should not modulate the nucleation-growth mechanism of the crystal.

So these are 3 reasons that could account for a slower or even prevent the superhelix assembly and crystallization at 4°C.

S3. Monitoring origami assembly by fluorescence microscopy

The precipitate obtained by mixing brick and staple proteins was monitored by confocal fluorescence microscopy. To this end, both bricks and staples were modified with a unique C-terminal cysteine and coupled to Alexa488 and Alexa594 dyes respectively.

Note that the amino-acids in hypervariable positions of the brick protein used in this experiment are different from the ones of the brick studied in the main text. However, these residues are exclusively located in the convex surface that is not interacting with the staple proteins nor with the adjacent bricks in the superhelix and therefore do not modify the self-assembling mechanism.

Fluorescence microscopy studies were performed using a Leica SP8 confocal microscope. Briefly, proteins with a free cysteine-tag at their C-terminal were conjugated to maleimide Alexa Fluor dyes 488 and 594 (Thermo Fischer Scientific) according to the manufacturer instructions. Labelled proteins were mixed as a final 10 µM equimolar mixture and incubated at room temperature for 1h. Origami assembly was dropped on a 1 mm 1% Agarose coated glass slide before fluorescence recording. Pictures were processed using ImageJ software.

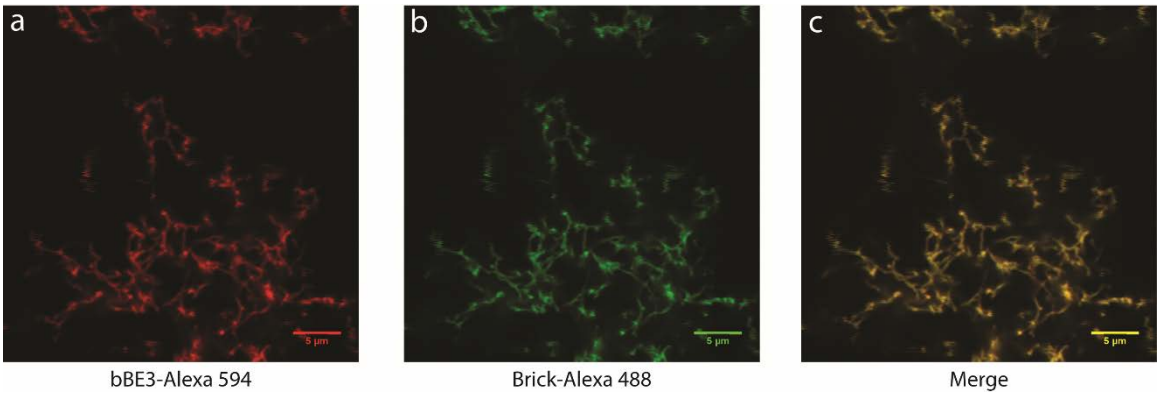

Figure S3: Confocal fluorescence microscopy images of an origami assembly. Final mixture of labelled proteins was dropped on an agarose coated glass slide before fluorescence recordings of the staple-Alexa 594 (a), the brick-Alexa 488 (b) and the composite picture (c). the scale bar corresponds to 5 µm.

The fluorescent images clearly show that the precipitate is a macroscopic porous network of fibrillary sub-micrometric structures and suggests a massive supra-molecular assembly. The strict overlap of the images of the tagged brick and staple (Fig. S3a, red) and (Fig. S3b, green) is illustrated by the merged image (Fig. S3c), which shows only yellow hues indicating the molecular proximity of both labelled proteins in the superstructure. This supplementary experiment was performed with a brick A presenting the same convex face as the brick used in the rest of the study.

S4. Superhelix length vs protein concentration.

While all experiments reported in the main text have been realized in 10 μM final concentration of proteins, the influence of the concentration on the morphology of the individual superhelices has been examined by stained TEM for lower final concentrations as shown in Figure S4.

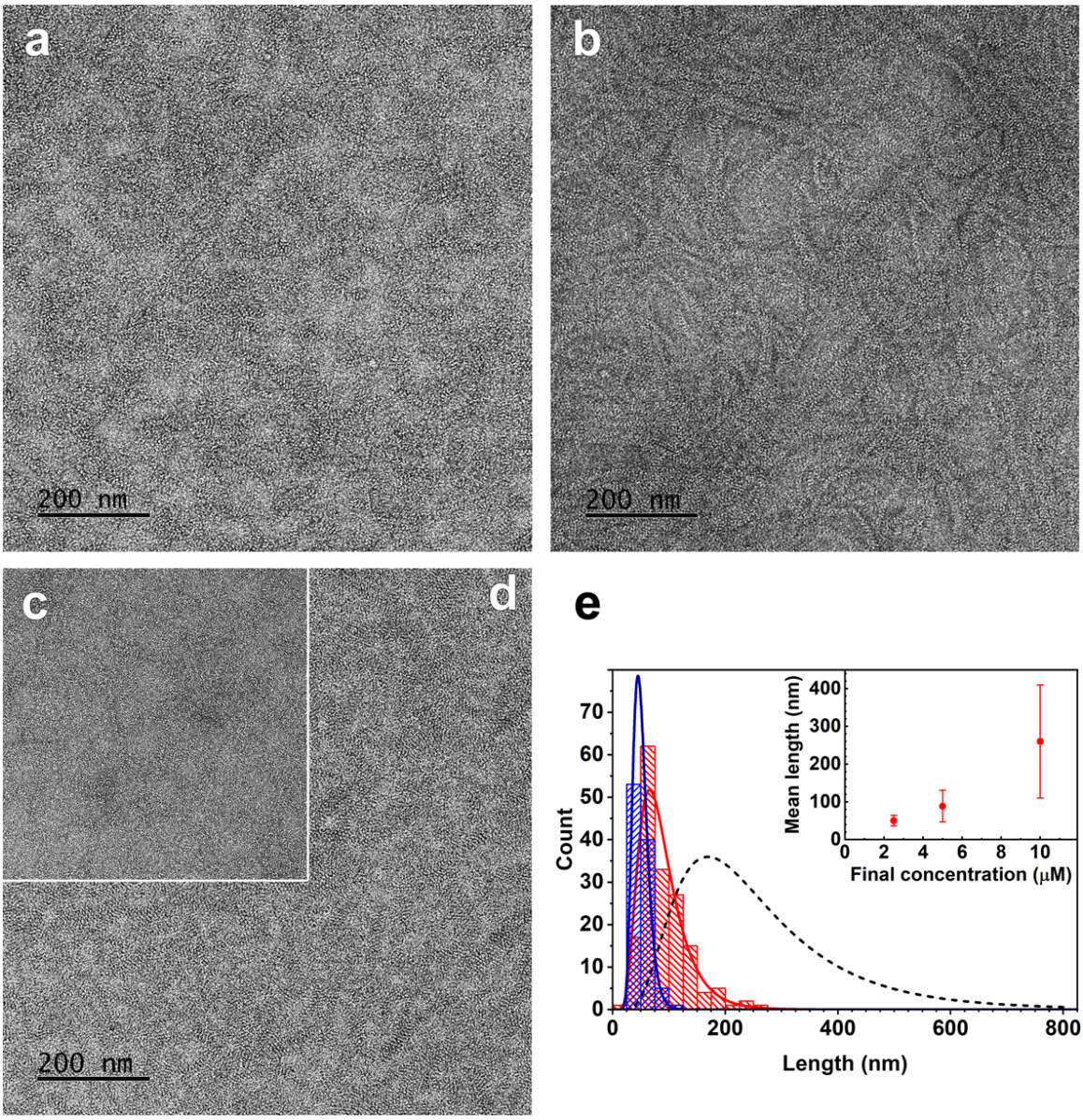

Figure S4: Stained TEM images of superhelices obtained for equimolar mixtures of the brick and staple proteins at final concentrations of (a) 2.5 μM, (b) 5 μM and (c,d) 10 μM. Samples (a,b,c) have been drop-casted from the suspension. Sample (d) is the same as in (c) but it has first been centrifuged and redispersed. (e) Histogram of superhelix length for a final protein concentration of 2.5 μM (blue) and 5 μM (red). The continuous lines are the corresponding lognormal distributions. The dashed line is the lognormal distribution from Figure 3c corresponding to a 10 μM final concentration. All micrographs share the same scale.

Structures resulting from the self-assembly of the brick and staple proteins are observed in the suspensions with 2.5, 5.0 and 10.0 μM final protein concentrations (Fig. S4a-c).
An increase in both the length and the concentration of superhelices is observed as the protein concentration increases.
In order to assess whether even longer objects are produced and precipitated, we centrifuge the 10 μM samples and redispersed them in the same volume. The corresponding image is shown in Figure S4d. Although a significant increase in the density of superhelices is patent, no significant size variation between (c) and (d) is observed.

Length measurements performed on micrographs pertaining to 2.5 and 5.0 μM final protein concentrations provide the superhelix length distributions shown in the histograms of Figure S4e. As in Figure 3c, a lognormal distribution is observed for both concentrations. The lognormal distribution of Figure 3c is recalled in Figure S4e as the black dashed line.
A clear shift and a broadening of the length distribution are observed as the protein concentration increases (Fig. S4e, inset). The mean length increases from 50.2 nm to 88.6 nm and 260 nm for 2.5, 5.0 and 10.0 μM respectively. Similarly, the spread of the length distribution increase with concentration as the standard deviation shifts from 14, to 42 and 150 nm.

S5. SAXS data and indexation

q_{obs}(Å^{-1})	q_{theo}(Å^{-1})	d_{hkl}(nm)		h	k	l
0.0789	0.0790	7.965		6	0	0
0.0968	0.0967	6.489		0	3	0
0.1079	0.1079	5.825		0	0	1
0.1164	0.1156	5.400		**2**	**1**	**1**
0.1247	0.1249	5.039		6	3	0
0.1502	0.1502	4.184		**3**	**3**	**1**
0.1581	0.1580	3.973		12	0	0
0.1652	0.1650	3.804		6	3	1
0.1850	0.1852	3.396		12	3	0
0.1935	0.1935	3.248		0	6	0
0.2156	0.2158	2.915		0	0	2
0.2314	0.2313	2.715		**4**	**2**	**2**
0.2500	0.2498	2.514		12	6	0
0.2719	0.2721	2.310		12	6	1

Table S1: Data corresponding to the SAXS diagram shown in Figure 4a and corresponding peak indexation in the $P2_122$ space group (***a***=477.3Å, ***b***=194.8Å, ***c***=58.2Å). Bold indices indicate the (x6, x3,x1) superlattice Bragg peaks. Note that no *h00* peak with *h*=2*n*+1 are observed indicating a two-fold screw axis along ***a***.

S6. SAXS vs pI and pH

The protein assembly proceeds rapidly in buffered solutions at room temperature leading to massive aggregates. In order to demonstrate the formation of individual tubular assemblies, electrostatic screening was shown to produce individualized filaments as discussed in Fig. 3 when using a polydentate anion such as molybdate in stained TEM sample preparation. The monitoring of the polydentate anion, ionic strength and pH effects was also monitored in solution by recording SAXS signal from protein mixtures in glass capillaries. Figure S5 show four typical diagrams obtained for two pH (6 and 8) and two ionic strength in the presence of 150mM ammonium molybdate and adjusted with ammonium chloride.

The presence of the 001 peak at 0.11 Å-1 in all conditions is consistent with the formation of the superhelical assembly with a pitch of 58Å. The main variations from one condition to the other concern the lateral packing into 2D crystal, which is observed for high ionic strength or alkaline pH but is absent when both pI and pH are low. These observations fully support the formation of individualized superhelices that pack sidewise very tightly under the electrostatic attractive force between the positively charged backside of the staple proteins and the strongly negatively charged superhelix inner cavity and side grooves (staple interdigitation being sterically allowed).

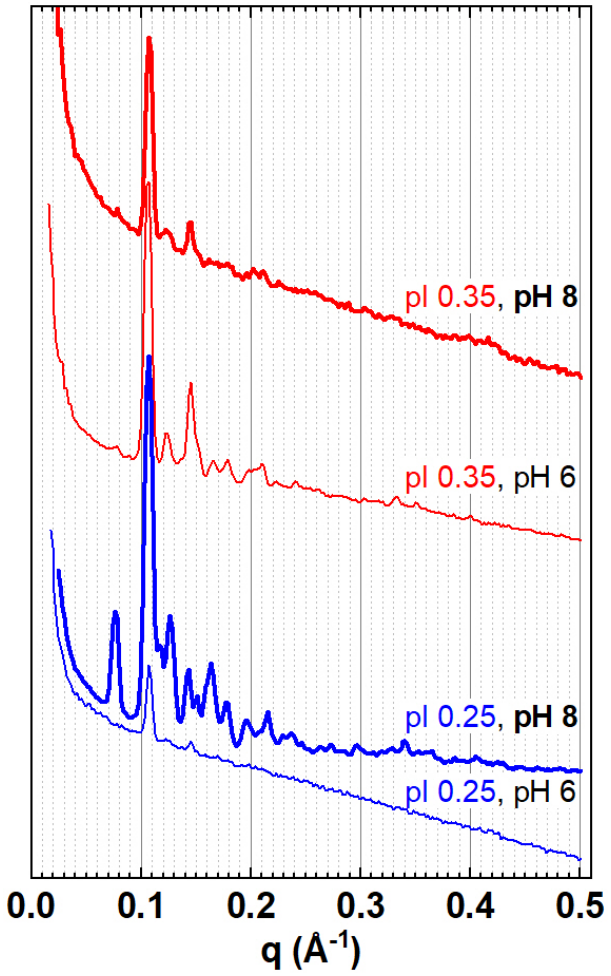

Figure S5: SAXS diagram of the precipitate produced by mixing brick and staple bBE3 at final concentration of 10 mM in buffer at pH 6 or 8 and in the presence of (150 mM) ammonium molybdate $(NH_4)_2MoO_4$ at adjusted ionic strength (pI) of 0.25 or 0.35.

S7. Temperature-dependant SAXS

Temperature-dependent SAXS monitoring was carried out from freshly prepared equimolar solutions of brick and staple bBE3 proteins (at 10 μM final concentration) placed in 1-mm glass capillaries. The temperature was ramped up to 70°C by steps of 5 to 10 degrees and then back down to 20°C. The SAXS diagrams obtained at the extremal temperatures are shown in Fig. S6a. A monotonous conversion from one diagram to another was observed for intermediate temperature with a rapid switch around 55°C.

Importantly, the protein crystal is shown to sustain high temperature (70°C) without major damage as it returns to an identical SAXS pattern upon cooling back to room temperature.

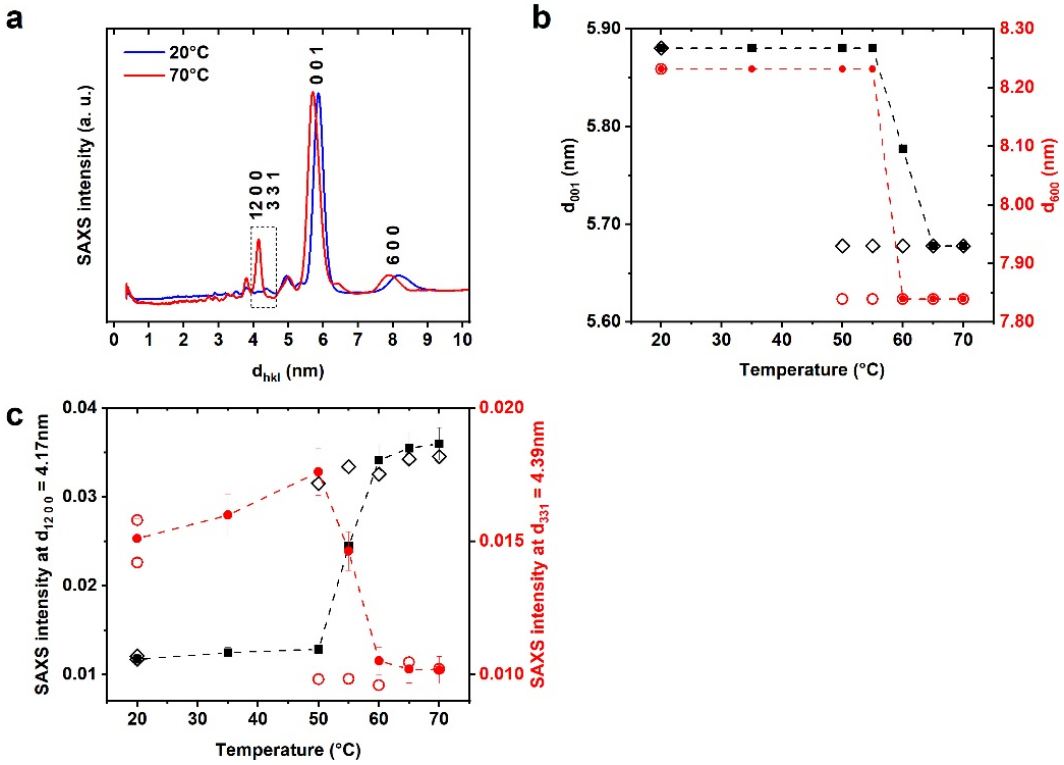

Figure S6: Temperature evolution of the structural parameter of the superhelix crystal monitored by SAXS. (a) Representative SAXS diagrams of the protein precipitate at minimal (20°C) and maximal (70°C) temperatures. (b) Evolution of the lattice parameters of the (0 0 1) and (6 0 0) plane families suggesting an abrupt contraction of the cell parameters at 55°C with limited hysteresis. (c) Intensity interconversion for the (12 0 0) and (6 3 1) peaks occurring almost reversibly at 55°C. Solid symbols are used for increasing temperature ramps while open symbols are related to SAXS data obtained when cooling the samples down.

S8. Statistical analysis of dot patterns in (100) zone axis TEM images

CryoEM images of the superhelix crystals along the [100] zone axis as shown in Figure 5a present a characteristic pattern of parallel dark and bright lines periodically decorated by oblong darker dots (See magnified view in Figure S7a). The highly regular dark dot pattern seems to reveal an alternating location of electron-absorbing areas that can be represented by considering one unitary motif composed of four neighbouring dots, three of which are taken on the grey line and one is taken as the nearest neighbour on the adjacent greyline to form a distorted diamond as schematized in orange in Figure S7a.

By pointing out the position of the three neighbours relatively to the position of the fourth dot, we aggregated the relative spatial position of the four apices as displayed in Figure 7c. Taking into account the direction of the parallel grey and bright lines that are also clearly seen in the FFT (Fig. 5b), we deduced a relative distance of 6.4 nm between dark dots located on the same side of the grey line. Moreover, using the model of the crystal based on pdb files of the constituent proteins as shown in Figure 5c, one can emulate the electron-beam absorption by plotting each atom as a black dot as displayed in Figure S7b. It appears that tangential regions of the superhelix produce oblong darker spots alternating on the side of the superhelix. Four neighbouring darker regions are marked with cyan ellipses in Fig. S7b that are overlaid with the scattered plot of the centroid of the dark cryoEM dots. A best agreement is obtained for the configuration shown in Figure S7 compared to other choices. This strongly suggests that the grey lines correspond to the core of the superhelix with an apparent width of 2.8 ± 1.0 nm when the inter-helix space (brighter lines) is about 3.1 ± 1.0 nm. Figure S7b shows how the model fits with the cryoEM and suggests that the water-filled interhelix space may have expanded during freezing compared to the highly symmetrical model based on SAXS cell parameters.

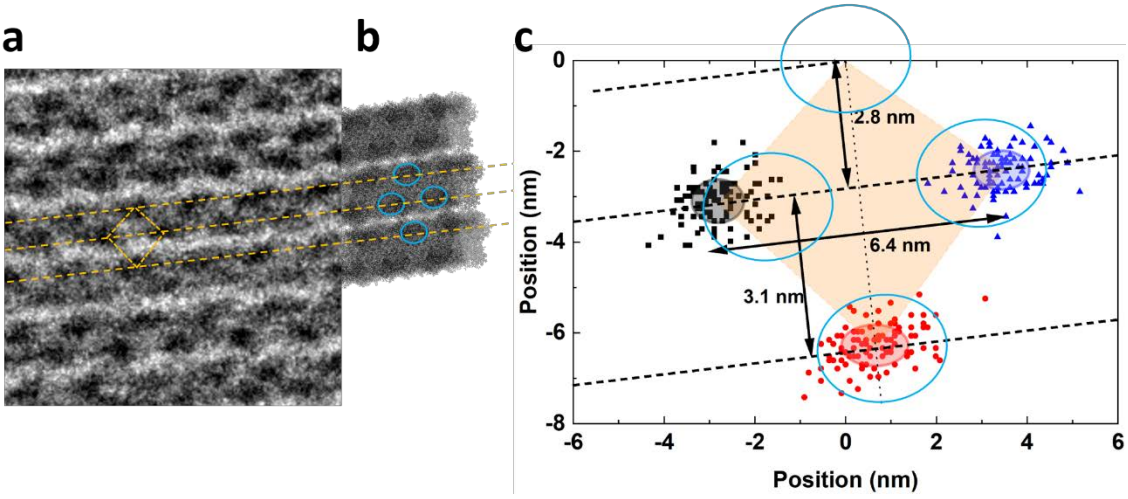

Figure S7: CryoEM measurements of the relative positions of four neighbouring dark dots observed periodically along the bright and dark lines of [100] zone axis views of the superhelix crystal. (a) Zoomed view of the cryoEM image presented in Figure 5a. The orange diamond highlights the elementary pattern of four neighbouring dark dots. The orange lines highlight the alignment of the dot series. (b) Model of the superhelix crystal view along the [100] direction with atoms marked with a single black dot marker. Regions denser in black markers are reproducing the alternating pattern of oblong dots observed in TEM and the diamond apices are highlighted with cyan ellipses. (c) Two-dimensional distribution of the dark dots centroids measured on cryoEM images taking the upper apex in (0,0) as a reference. The blue, black and red scattered distributions match the cyan ellipse and allow the extraction of geometrical features of the elementary diamond pattern.

S9. Moiré collection

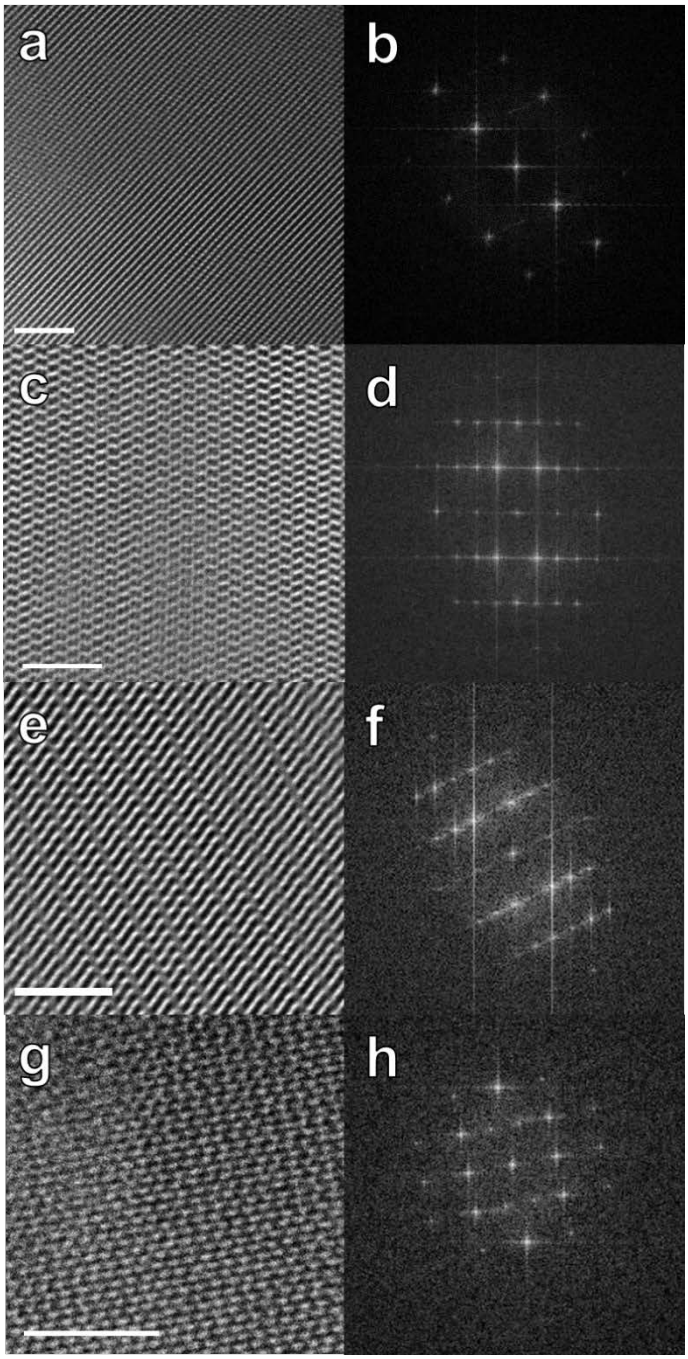

Figure S8: Direct space and Fast Fourier Transformed (FFT) images of superhelical protein origami crystals. In these particular areas, highly regular and periodical moiré patterns reveal the overlap of two or more crystals oriented each along different zone axis. The FFT shows the collection of spots equivalent to the overlapped electron diffraction patterns of all the zones axis.

S10. Model of staple interdigitation in superhelix crystals

We have examined the hydrophobic patches within the linear stacks of interdigitated staples as illustrated in Figure S9 in a fragment of the 3D model presented in Figure 6.

In Fig. S9a, the brick proteins have been omitted for clarity and replaced by a dashed cylinder. The staples of that superhelix are shown in green, as well as the interdigitated staples from the neighboring superhelices in red and blue. The interfaces between staples belonging to two different superhelices are highlighted by a black rectangular boxes and red-green or blue-green arrows.

We have used the Protein-Sol software (https://protein-sol.manchester.ac.uk/) to assess the hydrophobic patches present on the superhelices and, here, particularly along the stacks of interdigitated staples. This is shown in panel Fig. S9b, where the "Non-Polar to Polar" ratio (NPP) reveals that each staple possess one hydrophobic side (green) and one polar side (purple). The head-to-tail arrangement along the *a* axis allows the hydrophobic (resp. polar) sides of the two series of staples coming from neighboring superhelices to match as indicated by the green (resp. purple) arrows and rectangular boxes.

These hydrophobic interaction zones could contribute to the observed zipping effect alongside the strong electrostatic effect shown in Figure 6.

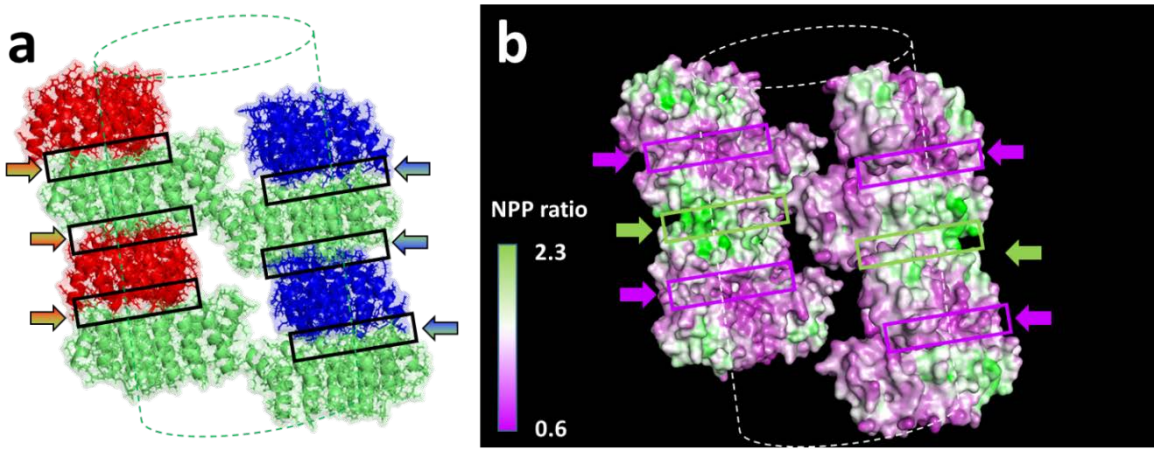

Figure S9. Staple interdigitation in superhelix crystals. (a) Structural model showing only the staples of one superhelix within the superhelix crystal in green and the staples of the direct neighboring superhelices in red and blue. The core of the central superhelix formed by the brick proteins is suggested bu the green dotted cylinder. The cores of the neighboring superhelices bearing the red and blue staples are omitted. The black boxes highlight the interfacial region between interdigitated staples and are marked by corresponding blue-green or red-green arrows. (b) Identical view of the interdigitated staples as in (a) but the surface is color-coded for the non-polar to polar (NPP) ratio. Green color indicates the most hydrophobic areas while purple color outlines the polar regions. Each green staple is engaged in a significant hydrophobic interaction surface with a neighboring (red or blue) staple where pointed out by green arrows.

S11. Cryotomography video

In complement to Figures 6h-j, a video of the corresponding cryoelectron tomogram and its corresponding 3D model is provided as supplementary material.

The video comprises five sequences :

1- *From 0 to 22 seconds*
Full scan of the 143 sections (430x175 nm^2) along the Z axis with a step size of 1.12 nm (total Z size 160 nm) and reverse scan revealing the 3D model that comprises three layers coded in blue, red and green.

2- *From 22 to 37 seconds*
Flat view of the model along the Z direction that coincides with the ***b*** axis of the superhelix crystal. Note that a dislocation is visible (for example at t = 29 s) that crosses the red layer approximatively in its middle: the superhelices of the left side are shifted by a/2 compared to the right domain. Zoom in the rightmost region (t = 32 s) and display of the three layers. In this configuration the model clearly reveals the right superhelices aligned parallel to each other. However, the model cannot differentiate the presence of the 2_1 screw axis (as identified in SAXS and cryoEM) rather than a direct translation.

3- *From 37 to 52 seconds*
The model is rotated towards views along the b axis (t = 46 s) and then along the ***a*** axis (t = 51 s). The 1-nm gap between layers is simply due to the fact that these specific tomogram sections could be equally modelled as belonging to the top or bottom objects. In the absence of an unambiguous assignment, no modelling was performed.

4- *From 52 sec. to 1min 15 sec.*
Views of the three modelled layers only, first along the ***c*** axis and then a full 360° rotation showing all orientations. The superhelices appear staked in registry in agreement with the packing order along b as shown in Figure 6. The zoomed view along the ***c*** axis (t = 56 s) clearly displays the inner cylindrical cavity inside the superhelices.

5- *From 1min 15 sec. to 1min 20 sec.*
The final sequence replaces the whole model within 3 particular X, Y, Z tomogram sections as shown in Figure 6j.

The aspect ratios of the cross sections of the superhelices and the helical pitch coincide with the pdb model and the experimental TEM and SAXS data and imply that the staples are interdigitated between neighboring superhelices of the same colored layer.