LETTER

https://doi.org/10.1038/s41586-019-1361-6

Controlling protein assembly on inorganic crystals through designed protein interfaces

Harley Pyles[1,2,6], Shuai Zhang[3,4,6], James J. De Yoreo[3,4,7]* & David Baker[1,2,5,7]*

The ability of proteins and other macromolecules to interact with inorganic surfaces is essential to biological function. The proteins involved in these interactions are highly charged and often rich in carboxylic acid side chains[1–5], but the structures of most protein–inorganic interfaces are unknown. We explored the possibility of systematically designing structured protein–mineral interfaces, guided by the example of ice-binding proteins, which present arrays of threonine residues (matched to the ice lattice) that order clathrate waters into an ice-like structure[6]. Here we design proteins displaying arrays of up to 54 carboxylate residues geometrically matched to the potassium ion (K^+) sublattice on muscovite mica (001). At low K^+ concentration, individual molecules bind independently to mica in the designed orientations, whereas at high K^+ concentration, the designs form two-dimensional liquid-crystal phases, which accentuate the inherent structural bias in the muscovite lattice to produce protein arrays ordered over tens of millimetres. Incorporation of designed protein–protein interactions preserving the match between the proteins and the K^+ lattice led to extended self-assembled structures on mica: designed end-to-end interactions produced micrometre-long single-protein-diameter wires and a designed trimeric interface yielded extensive honeycomb arrays. The nearest-neighbour distances in these hexagonal arrays could be set digitally between 7.5 and 15.9 nanometres with 2.1-nanometre selectivity by changing the number of repeat units in the monomer. These results demonstrate that protein–inorganic lattice interactions can be systematically programmed and set the stage for designing protein–inorganic hybrid materials.

Insight into protein–inorganic interfaces has come from studies of designed peptides that modulate calcite growth[7,8], additive control of crystallization[9], designed helical peptides that assemble on carbon nanotubes[10], and designed biphasic β-sheet proteins[11] and β-sheet peptides[12] on graphite surfaces. Designing assemblies of large proteins on inorganic lattices presents a new challenge because extensive spatial matching must exist within individual subunits and be maintained in the protein assembly. Collagen forms ordered arrays on mica[13], but despite theoretical work[14] the physical basis of the observed alignment remains unclear. The explicit programming by design of proteins to bind to inorganic lattices in pre-defined orientations, and assemble into larger-scale arrays with different architectures, is a stringent test of our understanding of the principles of mineral-bound protein self-assembly.

Inspired by the lattice-matching of ice-binding proteins to ice crystals, and the carboxyl-rich nature of many proteins that interact with minerals, we explored the design of protein–inorganic interfaces based on the placement of carboxylate residues electrostatically and structurally matched to a crystalline inorganic surface. We chose muscovite mica as the model mineral system because it presents a well-defined K^+ sublattice on the (001) cleavage plane (Fig. 1a, b and Supplementary Fig. 1) onto which molecules and biomacromolecules have been shown to adsorb and, in some cases, assemble into ordered structures[15–19]. To achieve an extended geometrically matched binding interface, we sought a protein scaffold with a flat surface and a regularly repeating backbone with spacing equal to a multiple of the 5.2 Å nearest-neighbour distance between K^+ sites. We found that the de novo designed helical repeat (DHR) protein DHR10 satisfied these criteria[20]. To introduce a mica binding surface, one side of the protein was redesigned so that all residues were either glutamates lattice-matched to mica K^+ ions or alanines. We call this protein DHR10-mica*X* where *X* is the number of repeat subunits. The version with $X = 18$, DHR10-mica18, was designed to have a large lattice-matched interface (1.8 nm by 18.7 nm) (Fig. 1a) and an exaggerated aspect ratio (3.6 nm wide, 20 nm long, and 2.5 nm high) to make the binding orientation of the protein readily detectable by atomic force microscopy (AFM). In Rosetta docking calculations, DHR10-mica18 exhibits a very strong preference for the designed binding orientation (Supplementary Fig. 2). DHR10-mica18 expressed in *Escherichia coli* was α-helical (Supplementary Fig. 3), monomeric (Supplementary Figs. 4, 5a), had a small-angle X-ray scattering (SAXS) profile consistent with the design model[21,22] (Supplementary Fig. 6), and showed directional adsorption in liquid AFM experiments on freshly cleaved mica in the presence of K^+ consistent with the design scheme (Fig. 1c–e, Supplementary Figs. 7, 8).

The coverage and order of DHR10-mica18 on the surface increased with increasing salt concentration (Fig. 2a–c, Supplementary Figs. 9, 10a). In 10 mM KCl, DHR10-mica18 molecules adsorbed as stationary monomers (Supplementary Fig. 11a) with their repeat axes aligned to the three close-packed K^+ directions (Figs. 1d, 2a and Supplementary Figs. 7, 8). In 100 mM KCl, the proteins became highly mobile (Supplementary Fig. 11b) and assembled into a two-dimensional liquid-crystal-like phase with discrete co-aligned protein domains also aligned along the three close-packed K^+ directions (Fig. 2b, Supplementary Fig. 12). In 3 M KCl, the protein mobility increased further (Supplementary Fig. 11c), and the arrays completely covered the surface with one lattice direction predominating over multi-millimetre distances (Fig. 2c, Supplementary Figs. 10b, 13 and Supplementary Table 1), corresponding to billions of co-aligned monomers. The long-range order probably accentuates a well-known structural anisotropy in the underlying muscovite crystal lattice[23], adding to the natural entropy-driven tendency of nanorods to align[24]; on a truly three-fold-symmetric form of mica that lacks this anisotropy, domains of coaligned nanorods form with similar probabilities along the three K^+ sublattice directions, even at 3 M KCl (Fig. 2g). A redesigned protein (DHR10-mica14-checker) that has the designed mica-binding interface but considerably altered electrostatic patterning on its back side (Supplementary Fig. 14) forms ordered arrays on mica (Fig. 2f, Supplementary Fig. 15a–c) very similar to those of DHR10-mica18 (Fig. 2c, Supplementary Fig. 9a–c), suggesting that the designed interface, not the back side, directs mica binding.

To determine the length dependence of assembly, we took advantage of the modular nature of the repeat protein monomer and varied the number of repeat units. As expected for liquid-crystal phases, the ability

[1]Department of Biochemistry, University of Washington, Seattle, WA, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA, USA. [3]Physical Sciences Division, Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. [4]Department of Materials Science and Engineering, University of Washington, Seattle, WA, USA. [5]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [6]These authors contributed equally: Harley Pyles, Shuai Zhang. [7]These authors jointly supervised this work: James J. De Yoreo, David Baker. *e-mail: james.deyoreo@pnnl.gov; dabaker@uw.edu

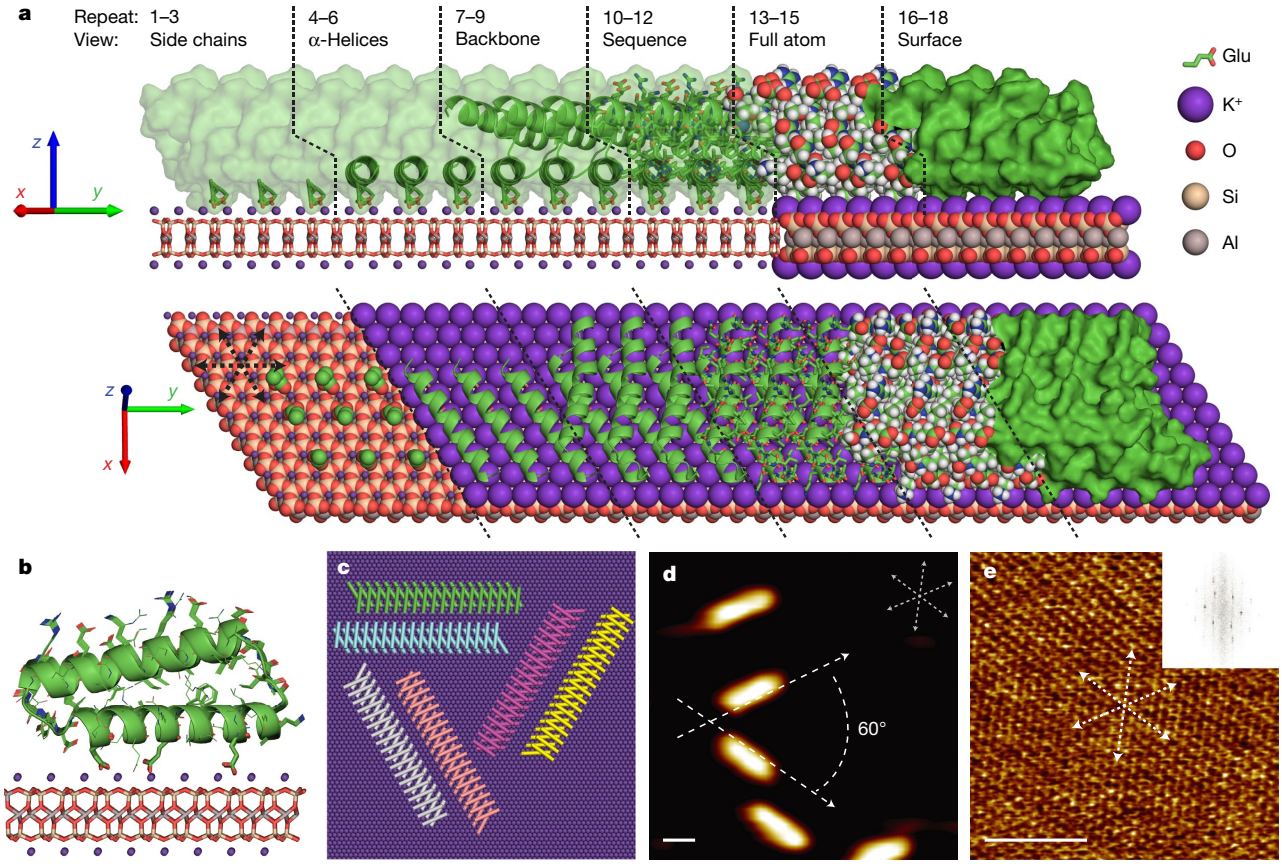

Fig. 1 | Design and characterization of lattice-matched protein monomers. a, Model of DHR10-mica18 protein lattice-matched to mica (001) surface through the K^+ sublattice. One 18-repeat DHR10-mica18 molecule is shown from the side (top panel) and from the top (bottom panel). Repeats 1–3 illustrate the lattice-matched glutamate (Glu) side chains. Repeats 4–6 show the α-helical secondary structures that contain the matched glutamates. Repeats 7–9 show the full DHR protein backbone. Repeats 10–12 show the full backbone and all amino acid side chains. Repeats 13–15 show all atoms as spheres. Repeats 16–18 show the external surface of the protein. The arrows indicate the orientations of the mica lattice. **b**, Projection view looking along the *y* direction. Lattice-matched glutamates are shown as sticks and other side chains as thinner lines. **c**, DHR10-mica18 bound to K^+ sublattice in the six symmetry-equivalent orientations predicted by the protein–mica interface design model. **d**, AFM image of DHR10-mica18 adsorbed on mica, showing two of the three predominant orientations that are distinguishable by AFM (anti-parallel orientations look similar). The protein concentration is 0.1 μM and the buffer contains 10 mM KCl and 20 mM Tris. The arrows in the top right corner indicate the orientations of the K^+ sublattice and mica lattice. The scale bar is 5 nm. **e**, AFM of mica (001) beneath the proteins in **d**, showing the mica lattice directions (dashed arrows). The fast Fourier transform is shown in the inset. The scale bar is 5 nm.

to order was strongly dependent on aspect ratio[24]. With 14 to 18 repeat units, high coverage and long-range order at high KCl concentration (high [KCl]) were maintained (Fig. 2f, Supplementary Fig. 15e, f). With only 6 repeats, adsorbed protein was still observed at 10 mM but not at 100 mM KCl or 3 M KCl (Fig. 2d, e and Supplementary Fig. 15d); in the latter cases the proteins probably transition to a two-dimensional liquid phase, as expected for rod-shaped particles with low aspect ratios[24], and move too rapidly to be observed as stationary objects by conventional AFM.

We measured the mica coverage at 10 mM KCl for DHR10-mica10 versus protein concentration and DHR10-mica*X* for $X = 4$, 6, 10 and 18 (Fig. 2h). The amount of bound protein scales exponentially with *X*, implying that the free energy of binding scales linearly with the number of repeats; the average binding energy per repeat is 3.0 kJ mol^{-1} ($1.2k_BT$ at room temperature) at 10 mM KCl.

We next sought to direct the pattern of the self-assembling monomers on mica by designing interfaces that can stabilize particular monomer–monomer arrangements. To maintain the overall alignment on mica, we restricted the designed interactions to those preserving, in the resulting assembly, the geometric match between the structure of each individual monomer and the mica lattice (Fig. 3a, b). We first explored increasing the effective nanorod length by designing end-to-end hydrophobic interactions between 'non-capped' monomers (Fig. 3a–c). Because the interactions are relatively weak, in solution the capless protein does not assemble into fibres (see Methods, Supplementary Fig. 4) and was monomeric, as assayed by native mass spectrometry[25,26] (Supplementary Fig. 16).

On mica, the non-capped DHR10-mica18-NC proteins form extensive single-protein-diameter wires (Fig. 3d, g). This assembly was maintained for much shorter 6 and 2 repeat constructs (Fig. 3e, f, h, i). DHR10-mica6-NC formed micrometre-length nanowires (Supplementary Fig. 17), whereas the capped version of DHR10-mica6 did not assemble (Fig. 2e), demonstrating the importance of the designed protein–protein interactions. The long-range order of the nanowire arrays increased with [KCl]: the wires were straighter and better aligned to the lattice, and the effect was more pronounced for shorter monomers (Supplementary Fig. 18), probably because of the increased mobility and more favourable lateral interactions between the proteins caused by electrostatic screening. Even at 3 M KCl, some domains remain aligned to one of the other two lattice directions on muscovite mica (Fig. 3g, h). All three orientations were observed for three-fold-symmetric mica (Supplementary Fig. 19). The misalignment at low [KCl] and alternative domain orientations at high [KCl] probably reflect kinetic trapping[27]; the net affinity of the long fibres to the mica surface is greater than for individual DHR10-mica18 monomers.

Next, we explored the design of protein–protein interfaces that direct formation of symmetric assemblies spatially matched with the hexagonal K^+ sublattice. Interfaces with C2, C3 or C6 point symmetry are

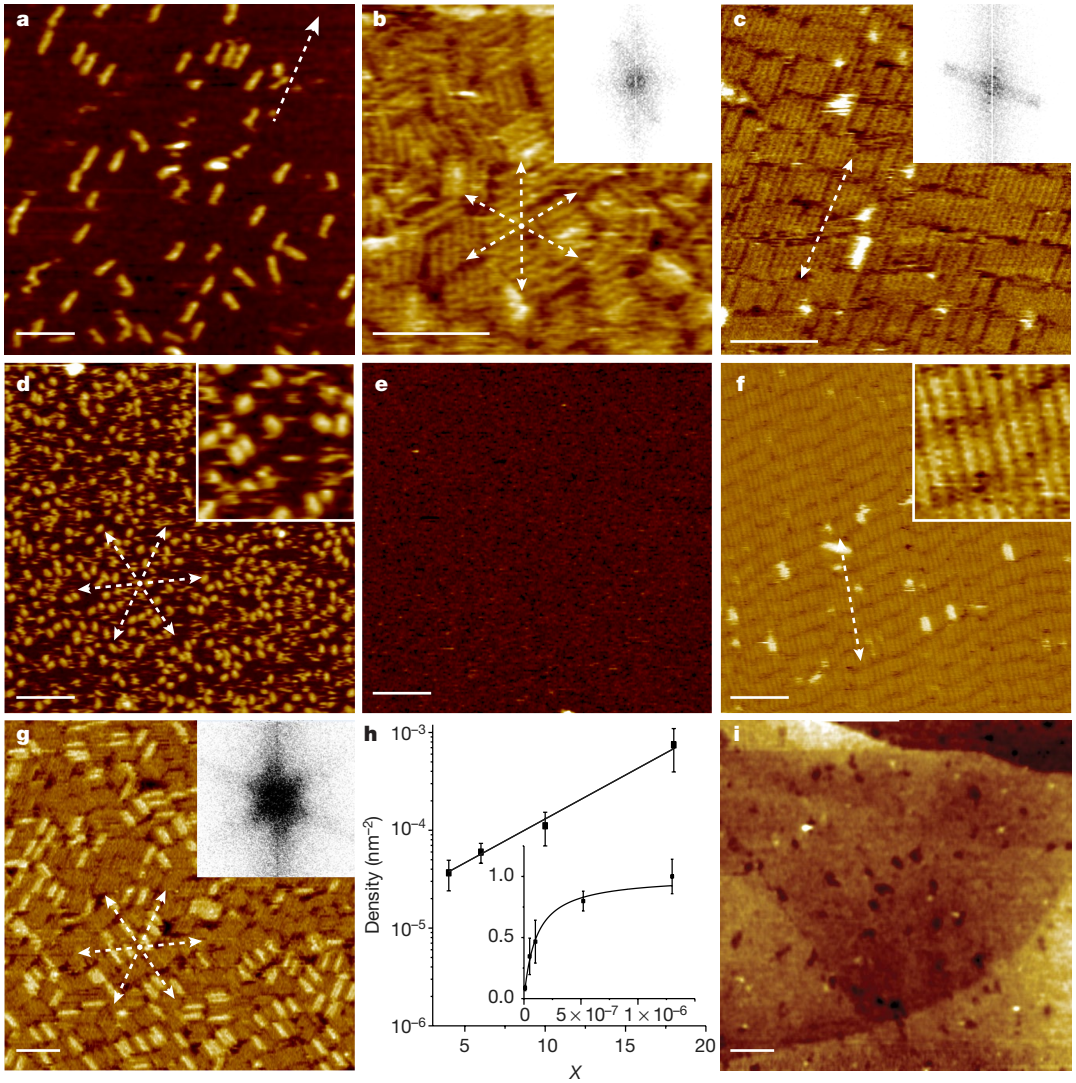

Fig. 2 | Assembly of DHR10-mica*X* molecules on mica (001). **a**–**c**, DHR10-mica18 on mica at 10 mM, 100 mM and 3 M KCl, respectively. Fast Fourier transforms are shown in the insets of **b** and **c**. Increasing [KCl] shifts the protein adsorption from individual lattice-aligned monomers to a liquid-crystal-like phase that extends beyond the imaging field. Reducing the length to 6 repeat units (DHR10-mica6) leaves the monomeric adsorbed state at 10 mM KCl largely unchanged (**d**; the inset is a 60 nm × 60 nm magnified image), but at 3 M KCl (**e**) no assembly is observed and the surface is probably covered in a mobile liquid phase. **f**, DHR10-mica14-checker forms a two-dimensional liquid-crystal phase at 3 M KCl. The inset shows the smectic alignment (40 nm × 40 nm). **g**, DHR10-mica18 on fluorophlogopite mica. A fast Fourier transform is shown in the inset. **h**, Number density (nm^{-2}) of protein molecules versus the number of repeat units (*X*) of DHR10-mica*X* at 10 mM KCl with 0.1 μM DHR10-mica*X* (mean of measurements at 7, 6, 8 and 10 imaging locations for $X = 4$, 6, 10 and 18, respectively; data are mean ± s.d.). The inset shows fractional surface coverage of DHR10-mica10 versus concentration (M) (mean of measurements at 6, 8, 6, 5 and 5 imaging locations for low to high concentrations, respectively; data are mean ± s.d.). The solid line is a fit based on a Langmuir dependence. **i**, DHR10-mica18 on plasma-pretreated HOPG at 10 mM KCl. The concentrations of DHR10-mica18, DHR10-mica6 and DHR10-mica14-checker are 0.1 μM, 0.3 μM and 0.1 μM, respectively. All solutions were buffered with 20 mM Tris-HCl. Scale bars are all 50 nm. Dashed arrows indicate the orientations of DHR10-mica*X* molecules.

compatible with the surface symmetry. Of these, packing constraints preclude C6 arrangements of the monomers and C2 arrangements would be difficult to distinguish from the head-to-tail arrangements of the monomers (as in Figs. 2 and 3), so we chose to design a C3 interface. In symmetric protein oligomer design, continuous sampling along angular and translational degrees of freedom is generally carried out to find a subunit arrangement that can accommodate a designed interface[28], but in this case the arrangements of lattice-matched proteins are constrained to translations of 5.2 Å × *n* (where *n* is any integer) along the mica lattice and 120° and 240° rotations around lattice C3 symmetry axes. C3 lattice-matched docks were stabilized by a combination of backbone remodelling, sequence design (see Methods, Fig. 4c), and in some cases fusion to previous de novo designed trimeric helical bundles[29]. Six trimer designs were tested and found to form trimers to different extents on mica (Supplementary Fig. 20). The largest fraction of aligned trimers was observed for one of two designs that lack the helical bundle, DHR10-mica18-trimer-V1 (version 1); however, the majority of proteins still adsorbed as monomers.

To reduce the affinity between monomers and the mica lattice, we reduced the number of repeat units from 18 to 5, and to direct formation of a hexagonal lattice, we incorporated a C2 interface at the exposed end of each trimer arm (Supplementary Fig. 21). At 3 M KCl this designed protein, hexagonal DHR10-mica5-H, formed an extensive and regular hexagonal lattice on mica (Fig. 4k, p) that was more open than the original design model (Supplementary Fig. 21). Rosetta calculations (see Methods, Supplementary Fig. 22) identified a low-energy lattice-matched C2 arrangement of the DHR10-mica5-H trimers (Fig. 4a, b) that generates a honeycomb lattice very similar to that observed (Fig. 4f). At lower [KCl], the protein adsorbed but did not form the hexagonal network (Supplementary Fig. 23).

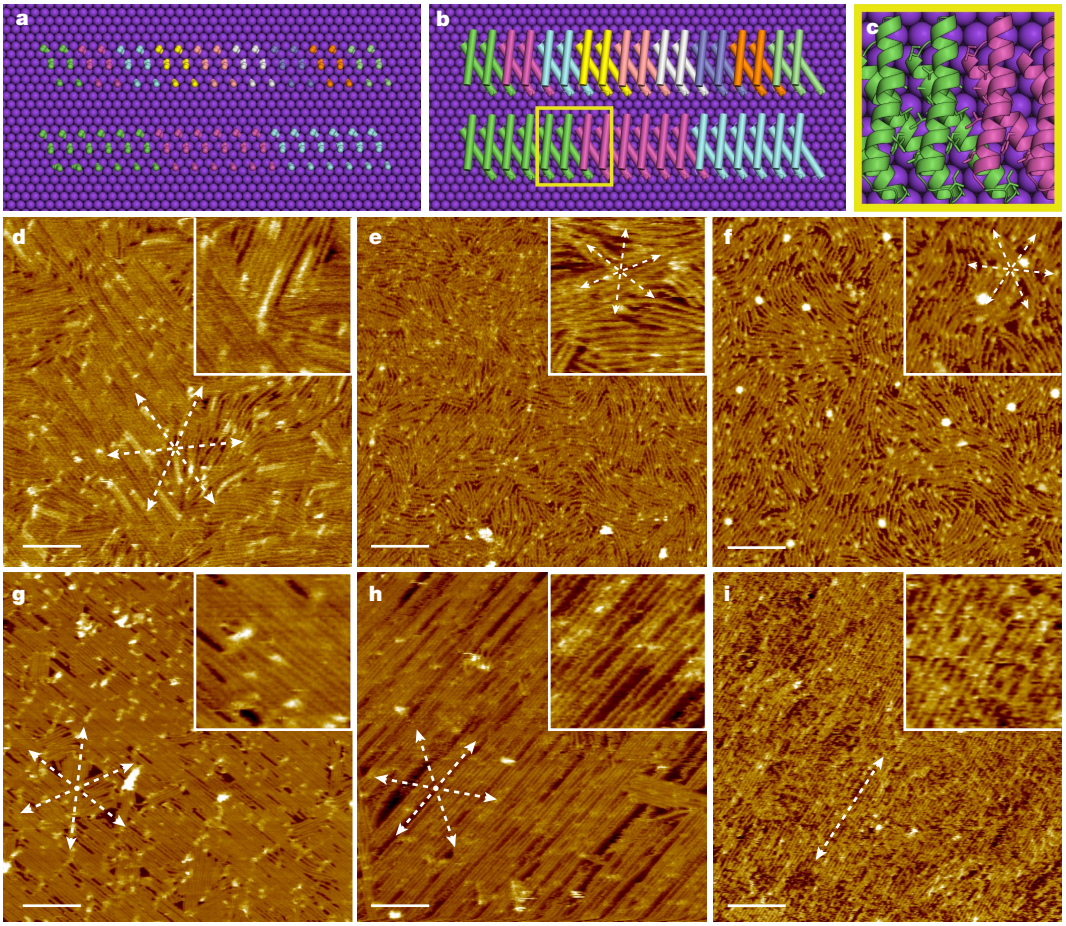

Fig. 3 | Liquid crystals formed by single-protein-diameter DHR-mica*X*-NC nanowires. a–c, Nanowire design concept. **a**, Lattice-matching side chains from nine molecules of DHR10-mica2-NC (top) and three molecules of DHR10-mica6-NC (bottom) in epitaxially matched fibres. **b**, DHR10-mica*X*-NC fibre model backbones. **c**, Magnification of the yellow square in **b** showing analogous hydrophobic interactions at the intramolecular repeat–repeat interface and the intermolecular fibre interface. **d–i**, Self-assembly of DHR-mica18-NC (**d**, **g**), DHR-mica6-NC (**e**, **h**), and DHR-mica2-NC (**f**, **i**) in 100 mM KCl (**d–f**) and 3 M KCl (**g–i**). At low [KCl], the degree of co-alignment increases with length, and at high [KCl], the shorter proteins are better aligned. The insets are magnified images (100 nm × 100 nm, except for the inset to **f**, which is 200 nm × 200 nm). Scale bars are 100 nm. The arrows indicate the orientations of the nanowires. The concentrations of DHR10-mica18-NC, DHR10-mica6-NC and DHR10-mica2-NC are 0.1 μM, 0.3 μM and 0.9 μM, respectively. All solutions were buffered with 20 mM Tris-HCl.

We sought to tune the spacing of the honeycomb lattice by taking advantage of the modular repeat architecture of DHR10. The individual repeat units each contribute 1.04 nm to the length of the monomer, and hence adding (or subtracting) a single repeat unit changes the hexagon side length by 2.08 nm. We generated a series of DHR10-mica*X*-H constructs with the amino (N)-terminal trimer and carboxy (C)-terminal dimer interfaces kept constant and the number of repeat units varied from 3 to 7. These proteins formed extensive and regular hexagonal lattices on mica up to several square micrometres with geometries consistent with the Rosetta models (compare Fig. 4d–h to Fig. 4i–m and Fig. 4n–r). All exhibited sharp six-fold-symmetric fast Fourier transform patterns and the lattice parameters exhibited ratios of 3.0 to 4.0 to 4.8 to 6.1 to 6.9 for the 3, 4, 5, 6 and 7 repeat constructs, respectively. The measured nearest-neighbour distances of the hexagonal arrays were within 5% to 15% of the corresponding Rosetta model values of 7.5 nm, 9.6 nm, 11.7 nm, 13.8 nm and 15.9 nm, respectively (Supplementary Fig. 22). In the versions with 4 to 7 repeats, individual domains are orientationally aligned, even when not in contact, suggesting a shared lattice-match (Fig. 4j–m, Supplementary Fig. 24). The three-repeat version forms domains in two different orientations (see Methods, Fig. 4i, Supplementary Fig. 25).

To test the importance of specific protein–mineral interactions beyond overall electrostatic attraction to the observed assembly geometries, we varied both the properties of the surface and those of the protein. Oriented binding of DHR10-mica18 in 10 mM K^+ was not observed on highly ordered pyrolytic graphite (HOPG) and molybdenum disulfide, six-fold-symmetric hydrophobic substrates previously used to adsorb β-sheet proteins[11,30], or to negatively charged plasma-treated HOPG (Fig. 2i, Supplementary Fig. 26). To test the importance of the designed side chain interactions, we generated four DHR10-mica4-H variants with the same overall charge on the mica-interacting surface, but with different residue placements (Supplementary Fig. 27). Although all variants retain some residual complementarity to the mica lattice because the backbone spacing is fixed, and the protein–protein interfaces were left unchanged, three out of four such constant surface charge variants failed to form the honeycomb lattice (Supplementary Fig. 28), demonstrating that the specific positioning of the charged protein side chains is critical to patterned assembly.

Our results highlight the subtle balance of forces governing the structure of protein–mineral composite systems. From the perspective of colloidal physics, our designed proteins may be viewed as highly charged, high-aspect nanorods packing in two dimensions on the charged mica surface. The forces determining structure are primarily entropic packing effects, modulated by electrostatics, that drive ionic-strength-dependent liquid-crystal formation. From the atomic-scale, biochemical perspective, the nanorods are patterned to be lattice-matched to the surface and the interactions between the rods are tuned by modulating the configuration of atoms on their surfaces

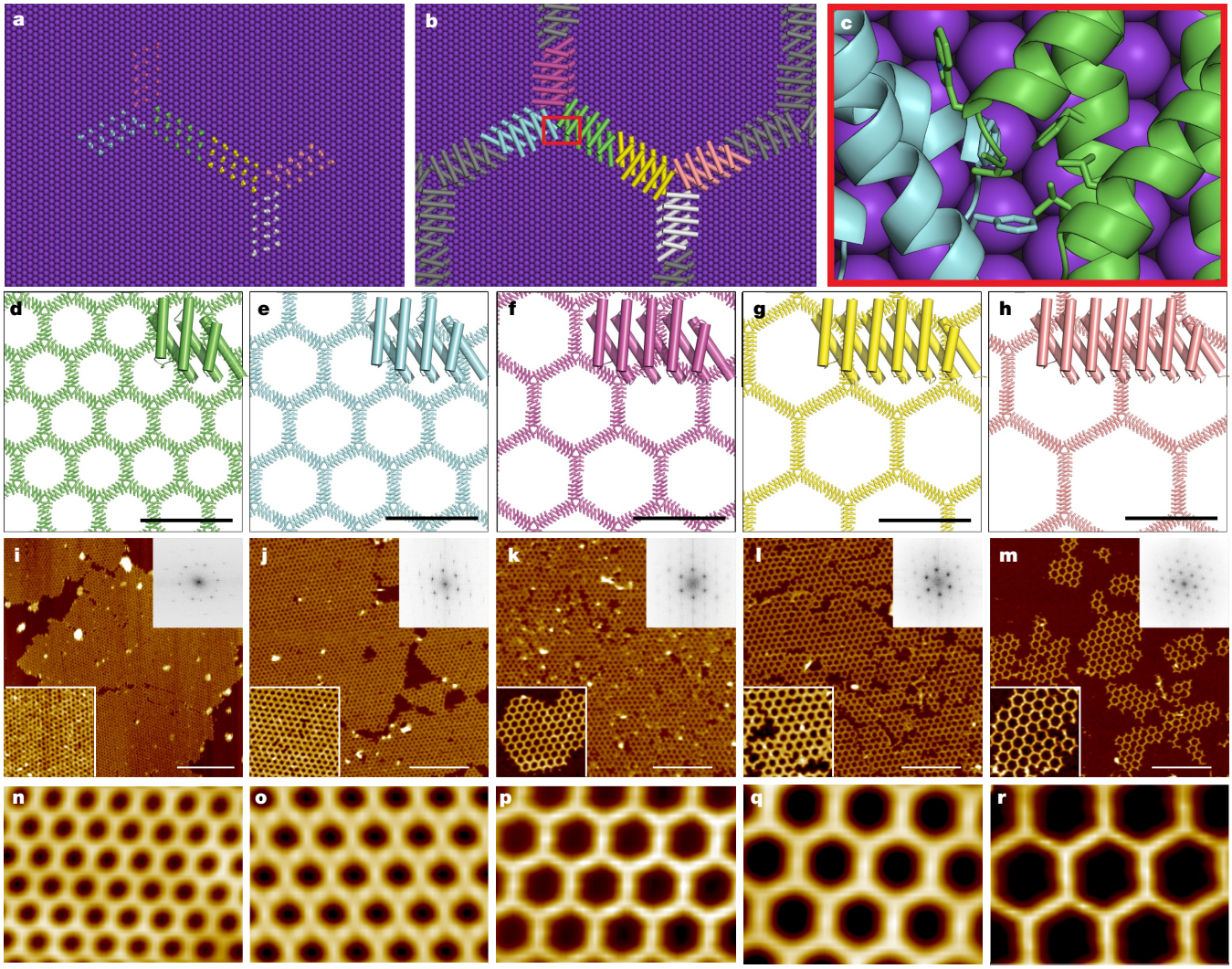

Fig. 4 | Protein–protein interactions drive formation of the tunable hexagonal lattice. a–**c**, Hexagonal lattice design concept. **a**, Lattice-matching side chains from six molecules of DHR10-mica5-H in the honeycomb lattice. **b**, DHR10-mica5-H proteins containing the side chains shown in **a**. **c**, Magnification of the red square in **b** showing packing of hydrophobic side chains at the designed trimer interface. **d**–**h**, Computational models of DHR10-mica*X*-H hexagonal arrays for 3, 4, 5, 6 and 7 repeat units. Insets at top right corners show one monomer with different numbers of repeat units. The scale bars are 20 nm. **i**–**m**, AFM images of DHR10-mica*X*-H lattices with $X = 3$, 4, 5, 6 and 7 repeat units. The scale bars are 200 nm. Insets are higher-magnification views (200 nm × 200 nm; lower left corner) and Fourier transforms (upper right corner). **n**–**r**, Averaged images of **i**–**m**. The field of view is 80 nm by 60 nm. The concentrations of DHR10-mica*X*-H with 3, 4, 5, 6 and 7 repeat units are 0.76 μM, 0.62 μM, 0.225 μM, 0.044 μM and 0. 017 μM, respectively. The buffer is 20 mM Tris-HCl and 3 M KCl.

by protein design. We are able to control some aspects of the assemblies by balancing the first set of contributions. Monomers that are too short, in the DHR10-mica2-NC and DHR10-mica6-NC nanowires (Fig. 3e, f) and the DHR10-mica3-H hexagonal lattice (Fig. 4i), form assemblies with lower orientational order than longer monomers because the penalty for misalignment is too small. When the interactions are too strong, as in the end-to-end assembling nanorods in Fig. 3, order is reduced because of kinetic trapping. Increasing the KCl concentration weakens protein–mica interactions, increasing mobility and lowering kinetic barriers to annealing into the lowest-energy state.

Because of the control afforded by computational protein design and the use of rigid, designed building blocks, the honeycomb lattices with tunable unit cell parameters are exceptionally well ordered protein–inorganic lattice hybrids that are amenable to detailed structural characterization (Fig. 4). These lattices enable the patterning on mica of nanoscale wells with precisely controlled sizes that may be useful for diagnostics, high-throughput biochemistry and other applications. More generally, our results should inform the design of new protein–mineral hybrid materials.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1361-6.

Received: 24 October 2018; Accepted: 10 May 2019;
Published online 10 July 2019.

1. Sodek, J., Ganss, B. & McKee, M. D. Osteopontin. *Crit. Rev. Oral Biol. Med.* **11**, 279–303 (2000).
2. Addadi, L., Joester, D., Nudelman, F. & Weiner, S. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry* **12**, 980–987 (2006).
3. Shaw, W. J. Solid-state NMR studies of proteins immobilized on inorganic surfaces. *Solid State Nucl. Magn. Reson.* **70**, 1–14 (2015).
4. Staniland, S. S. & Rawlings, A. E. Crystallizing the function of the magnetosome membrane mineralization protein Mms6. *Biochem. Soc. Trans.* **44**, 883–890 (2016).
5. Fukushima, T. et al. The molecular basis for binding of an electron transfer protein to a metal oxide surface. *J. Am. Chem. Soc.* **139**, 12647–12654 (2017).
6. Davies, P. L. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem. Sci.* **39**, 548–555 (2014).

7. DeOliveira, D. B. & Laursen, R. A. Control of calcite crystal morphology by a peptide designed to bind to a specific surface. *J. Am. Chem. Soc.* **119**, 10627–10631 (1997).
8. Masica, D. L., Schrier, S. B., Specht, E. A. & Gray, J. J. De novo design of peptide-calcite biomineralization systems. *J. Am. Chem. Soc.* **132**, 12252–12262 (2010).
9. Song, R. & Cölfen, H. Additive controlled crystallization. *CrystEngComm* **13**, 1249–1276 (2011).
10. Grigoryan, G. et al. Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076 (2011).
11. Brown, C. L., Aksay, I. A., Saville, D. A. & Hecht, M. H. Template-directed assembly of a de novo designed protein. *J. Am. Chem. Soc.* **124**, 6846–6848 (2002).
12. Mustata, G.-M. et al. Graphene symmetry amplified by designed peptide self-assembly. *Biophys. J.* **110**, 2507–2516 (2016).
13. Leow, W. W. & Hwang, W. Epitaxially guided assembly of collagen layers on mica surfaces. *Langmuir* **27**, 10907–10913 (2011).
14. Tao, J. et al. Energetic basis for the molecular-scale organization of bone. *Proc. Natl Acad. Sci. USA* **112**, 326–331 (2015).
15. Akutagawa, T. et al. Formation of oriented molecular nanowires on mica surface. *Proc. Natl Acad. Sci. USA* **99**, 5028–5033 (2002).
16. Loo, R. W. & Goh, M. C. Potassium ion mediated collagen microfibril assembly on mica. *Langmuir* **24**, 13276–13278 (2008).
17. Shin, S.-H. et al. Direct observation of kinetic traps associated with structural transformations leading to multiple pathways of S-layer assembly. *Proc. Natl Acad. Sci. USA* **109**, 12968–12973 (2012).
18. Aghebat Rafat, A., Pirzer, T., Scheible, M. B., Kostina, A. & Simmel, F. C. Surface-assisted large-scale ordering of DNA origami tiles. *Angew. Chem. Int. Edn Engl.* **53**, 7665–7668 (2014).
19. Ma, X. et al. Tuning crystallization pathways through sequence engineering of biomimetic polymers. *Nat. Mater.* **16**, 767–774 (2017).
20. Brunette, T. J. et al. Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
21. Dyer, K. N. et al. High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol.* **1091**, 245–258 (2014).
22. Schneidman-Duhovny, D., Hammel, M., Tainer, J. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
23. Kuwahara, Y. Comparison of the surface structure of the tetrahedral sheets of muscovite and phlogopite by AFM. *Phys. Chem. Miner.* **28**, 1–8 (2001).
24. Boles, M. A., Engel, M. & Talapin, D. V. Self-assembly of colloidal nanocrystals: from intricate structures to functional materials. *Chem. Rev.* **116**, 11220–11289 (2016).
25. Ruotolo, B. T. & Robinson, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **10**, 402–408 (2006).
26. Sahasrabuddhe, A. et al. Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl Acad. Sci. USA* **115**, 1268–1273 (2018).
27. Whitelam, S. et al. Common physical framework explains phase behavior and dynamics of atomic, molecular, and polymeric network formers. *Phys. Rev. X* **4**, 011044 (2014).
28. Fallas, J. A. et al. Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
29. Boyken, S. E. et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
30. Wang, J. et al. Differential modulating effect of MoS_2 on amyloid peptide assemblies. *Chemistry* **24**, 3397–3402 (2018).

Acknowledgements We thank T. Brunette, P. Huang, F. Parmeggiani, Y. Hsia, W. Sheffler, T. Craven, S. Boyken and Z. Chen for suggestions; D. Alonso, L. Goldschmidt and P. Vecchiato for supporting computational resources; B. Legg and J. Tao for AFM support; B. Legg for providing the model of mica (001); L. Carter for size-exclusion chromatography with multi-angle light scattering (SEC-MALS) support. F. Busch and V. H. Wysocki provided native mass spectrometry support (NIH National Institute of General Medical Sciences award number P41GM128577). Development of the imaging protocols was supported by the Laboratory Directed Research and Development Office through the Materials Synthesis and Simulations Across Scales Initiative at Pacific Northwest National Laboratory (PNNL). AFM experiments on the DHR10-mica*X* and DHR10-mica*X*-NC were supported by the US Department of Energy (DOE), the Office of Basic Energy Sciences (BES), Biomolecular Materials Program (BMP) at PNNL. AFM experiments on DHR10-mica*X*-H formation of hexagonal lattices and analysis of DHR10-mica*X* binding kinetics were performed at PNNL and supported by the US DOE BES Energy Frontier Research Center CSSAS (The Center for the Science of Synthesis Across Scales) located at the University of Washington (award number DE-SC0019288). Design and synthesis of DHR10-mica*X*-H and its variants were performed at the University of Washington and supported by the US DOE BES BMP (award number DE-SC0018940). H.P. was supported by the Institute for Protein Design Materials Science Research Gift Fund and Michelson Medical Research Foundation, Protein Design Initiative Fund, and DOE Biomolecular Materials Program (award number DE-SC0018940). D.B. is funded by the Howard Hughes Medical Institute and Bruce and Jeannie Nordstrom / Patty and Jimmy Barrier Gift for the Institute for Protein Design Directors Fund. We thank the staff at the Advanced Light Source SIBYLS beamline at Lawrence Berkeley National Laboratory, including K. Burnett, G. Hura and J. Tainer for the services provided through the mail-in SAXS programme, which is supported by the DOE Office of Biological and Environmental Research Integrated Diffraction Analysis program DOE BER IDAT grant (award number DE-AC02-05CH11231) and NIGMS supported ALS-ENABLE (award number GM124169-01). PNNL is a multi-programme national laboratory operated for the Department of Energy by Battelle under contract number DE-AC05-76RL01830.

Reviewer information *Nature* thanks Roberto A. Chica, Rajesh R. Naik and Sarah S. Staniland for their contribution to the peer review of this work.

Author contributions H.P. designed, expressed, and purified the proteins. S.Z. developed the protocol to assemble protein on mica and performed the AFM experiments. S.Z. and J.J.D.Y. analysed the AFM data. D.B. and J.J.D.Y. supervised the project. All authors wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1361-6.

Correspondence and requests for materials should be addressed to J.J.D. or D.B.

METHODS

Designing and modelling a mineral-matched repeat protein. Our strategy for the design of protein–mineral interfaces is to match functional groups on protein side chains to corresponding groups on the mineral lattice. To do this, we attempt to satisfy three conditions. First, the protein should have a repeating structure with a distance between the repeating units that is an integer multiple of a lattice parameter of the mineral lattice. Second, multiple side chains in one repeat unit should make favourable contacts with the mineral lattice. Third, the protein should be nearly flat, so that it can interact with a crystal face over an extended area.

Our method for fulfilling these conditions for a particular mineral lattice is as follows. First, we identify a mineral surface and a designed repeat protein with compatible repeat spacings. Second, we design or identify side chains with functional groups capable of interacting with the lattice—at least 2 to 3 within each repeat unit—that interact with sites on the mineral surface. Third, we relax the protein with constraints to perfectly match the target lattice, and predict the lowest-energy protein–mineral interface.

Mica (001) potassium sublattice. The (001) cleavage plane of the mica surface organizes a well-defined hexagonal K^+ sublattice with a 5.2 Å lattice parameter (see Supplementary Fig. 1). The positive K^+ layer was targeted instead of the negatively charged mica (001) surface so [K^+] could be used to tune binding affinity and so that charge-complementary designs would avoid the solubility problems associated with proteins having excessively positive surfaces[31].

Selecting the DHR10 starting scaffold. Design models of DHR proteins[20] were examined to determine their capacity to geometrically match the mica lattice. The DHR protein DHR10 was selected because it has a distance between repeats of approximately 10.4 Å (twice the mica lattice parameter of 5.2 Å) and is very flat (lacking both curvature and twist). We hypothesized that these features would let DHR10 geometrically match the mica (001) surface just as antifreeze repeat proteins match ice[6].

Representation of mica (001) K^+ sublattice. The mica (001) K^+ sublattice was represented as 2,269 K^+ ions in a hexagonal grid with a 5.2 Å nearest-neighbour distance using default Rosetta parameters for K^+ ions, provided in the Rosetta source code as database/chemical/residue_type_sets/fa_standard/residue_types/metal_ions/K.params (see 'Code availability'). This representation of K^+ has a +1 electrostatic charge and Lennard–Jones properties from CHARMM27[32]. The coordinates of the K^+ ions were fixed in place during modelling by preventing sampling across the jumps between them in Rosetta's fold-tree representation of connectivity.

Rosetta modelling protocols. In the following description specific protocols are referred to by number and letter combinations which correspond to the directories in the DHR10micaX github repository that contain them (see 'Code availability').

Designing the DHR10-mica interface. DHR10 was observed to contain 12 glutamates (residues 28, 35, 42, 78, 85, 92, 128, 135, 142, 178, 185 and 192) whose C_α coordinates form a triangular grid with spacing of about 10.4 Å. To allow these glutamate residues to reach the mica (001) K^+ sublattice, the surrounding residues were mutated to alanine with Rosetta Script protocol 1. We call versions of DHR10 with these mutations DHR10-mica4, where 4 refers to the four 50-residue repeat units it contains. The amino-acid sequence of DHR10-mica4 consists of a unique N-terminal repeat, two identical internal repeats, and a unique C-terminal repeat. To cap the structure and improve solubility, the first and last repeats have additional hydrophilic residues at positions that are only solvent-accessible because they are on the ends of the protein. Protein versions with different numbers of repeats are called DHR10-mica*X*, where *X* is the number of repeat units, and can be generated by changing the number of identical internal repeats.

Modelling the lattice-matched repeat protein DHR10-mica18. To model a lattice-matched interface, we decided to use the 18-repeat DHR10-mica18 protein for three reasons. First, we predicted that a more extensive interface would be more likely to bind along the lattice. Second, we wanted a protein with an exaggerated aspect ratio so that its binding orientation could be observed with AFM. Third, cloning a DNA construct encoding an 18-repeat version was relatively convenient (see 'Plasmid design and synthesis').

The DHR10-mica18 sequence was generated by expanding the number of internal repeats from 2 to 16. The backbone torsion angles (phi, psi and omega) from an internal repeat of the DHR10-mica4 model were copied and duplicated 18 times to create an 18-repeat backbone model with Pyrosetta protocol 2a. The DHR10 model had minor structural differences between repeats and the backbone extrapolated from one repeat was not as flat as that of the starting model. Rosetta Scripts protocol 2b used Monte-Carlo sampling to find a set of 50-backbone torsion angles that, when repeated in tandem 18 times, produced a DHR10-mica18 model that is perfectly flat and has a repeat spacing of 10.4 Å (double the 5.2 Å spacing on mica). Three sets of constraints were used during sampling. First, C_α coordinate constraints based on residues 2 to 74 in the DHR10 crystal structure (Protein Data Bank (PDB) ID: 5CWG). Second, 10.4-Å-spaced C_α coordinate constraints that were extrapolated from residues 29, 36 and 43 in the crystal structure for lattice-matching residues in DHR10-mica18 (28, 35, 42, 78, 85, 92, 128, 135, 142, 178, 185, 192, 228, 235, 242, 278, 285, 292, 328, 335, 342, 378, 385, 392, 428, 435, 442, 478, 485, 492, 528, 535, 542, 578, 585, 592, 628, 635, 642, 678, 685, 692, 728, 735, 742, 778, 785, 792, 828, 835, 842, 878, 885 and 892). Third, 10.4 Å distance constraints between C_α atoms in adjacent repeats.

Modelling the lattice-matching binding mode in DHR10-mica18. DHR10-mica18 was docked onto a model of the K^+ sublattice consisting of 2,269 ion coordinates arranged in a hexagonal grid with 5.2 Å spacing. Rosetta Scripts docking protocol 3a was used. In each docking trajectory, the protein was randomly spun about the *z* axis in a 60° window, translated a random distance (between 0 Å and 5.2 Å), and then minimized with a relax protocol. During the relax step, side chains of residues in and around the surface-matching interface were allowed to repack and the rigid-body orientation of the protein on the sublattice was minimized. Throughout the docking protocol the protein backbone conformation and the structure of the K^+ sublattice were fixed. The docked models with the lowest Rosetta Energy Units scores were lattice-matched, meaning that the protein's repeat direction was aligned to the 5.2 Å lattice direction and identical interactions were seen between repeated sites in the protein and the sublattice (Fig. 1a–c, Supplementary Fig. 2).

Designing non-capped DHR10-mica*X* fibres. For solubility, the original DHR monomers have polar residues at the exposed ends. To enable the same set of repeat–repeat interactions between as within the monomers, we replaced the polar, 'capping' residues with the non-polar residues at corresponding positions in the internal repeat units, producing a protein–protein interface that resembles the hydrophobic packing between repeat units (Fig. 3c). Because the repeat spacing is the same within and between monomers, the geometric match to the mica lattice is preserved across the monomer–monomer interface. To make the DHR10-mica18-NC model, the sequences of the first and last repeats in DHR10-mica18 were changed to the sequence of the internal repeat with Rosetta Scripts protocol 4. Models of DHR10-mica2-NC and DHR10-mica6-NC fibres were made by splitting four 100-residue and three 300-residue segments (respectively) of the DHR10-mica18-NC model into separate chains, repacking all side chains, relaxing the backbone of three residues flanking the cut-point, and minimizing the rigid-body orientation of each protein chain on the K^+ sublattice (Rosetta Scripts protocols 5a and 5b).

Designing symmetric interfaces between lattice-matched proteins. When designing symmetric oligomers with proteins lattice-matched to a symmetric substrate, the symmetry between the proteins must be compatible with the symmetry of the surface. This severely restricts oligomer sampling in two ways. First, rotations are limited to spins around symmetry axes on the substrate. Second, translations are restricted to lattice-matched slides along the substrate. These sampling restrictions present a challenge because protein oligomer design typically depends on fine sampling of rigid-body degrees of freedom.

To overcome this, the monomer structure can be redesigned to fit into a specific oligomeric context. Reductive modifications (trimming a component to fit) are simpler than additive ones (for example, designing a new domain). Removing residues from the protein termini is easy, while removing residues from the middle of the polypeptide chain is harder because a new backbone connection between residues must be formed.

Designing the DHR10-mica*X* trimer interface. Designing the DHR10-mica*X* trimer involved generating trimer conformations that preserved each component's register to the mica (001) K^+ sublattice, removing residues from the N terminus, and making amino-acid substitutions to form a protein–protein interface. The prominent N-terminal helix was trimmed of 1–12 residues to increase the number of non-clashing configurations (Pyrosetta protocol 6a). Three distinguishable C3 axes that are compatible with the surface symmetry were identified (Supplementary Fig. 29). The trimmed N termini of DHR10-mica*X* were translated across a hexagonal grid with 5.2 Å intervals along two lattice directions and then symmetrized around the three C3 axes.

The trimer conformations were generated with Rosetta Script protocol 6b, which also replaces bulky residues with alanine and scores the energy and DDG (delta-delta G; the energy difference between the bound and unbound states) of potential docks. Positive DDGs reflect clashes and DDG = 0 indicates that the subunits are not in proximity, so only trimer configurations with negative DDGs were considered further. There were 104 potential trimers meeting these criteria, and one of these was selected for interface design based on visual examination in PyMOL (https://pymol.org/2/). The selected dock was chosen based on the lack of buried polar backbone atoms, minimal perturbation of the termini (only two residues removed), and the presence of multiple α-helices close enough to form hydrophobic packing interactions.

The trimer interface was designed by combining amino-acid substitutions from Rosetta Scripts design protocols with human-defined substitutions designed with PyMOL and Foldit. After deciding on a sequence, the final model of the interface was made with Rosetta Scripts protocol 6c. Design rationales for the 17-trimer interface substitutions included in the final sequence are as follows. The core of the

interface consists of Phe18 and Phe24 from one chain packing into Val2, Phe35, Val39 and Leu53 of a symmetric copy (K2V, K18F, T24F and R39V). Trp3 was included to sterically block alternative close-packing interfaces (that is, implicit negative design) and to form an interfacial cation-pi interaction with Arg11 (E3W and K11R). Five alanine substitutions were made to sterically accommodate other interface substitutions or other symmetric units in the trimer (E7A, V14A, E25A, E29A and E36A). Three residues surrounding a triangular cavity in the trimer interface were mutated from arginine to other charged residues to form intermolecular salt-bridges (R6K, R28E and R32K). Two substitutions were made so the sequence of the first repeat better matches the internal repeats (I12E and E43K). Ser1 is the end of a Gly–Ser linker (GSGGS) connecting the protein to a N-terminal thrombin-cleavable His6-tag (E1S).

After repacking the designed trimer interface symmetrically, a five-repeat version of the trimer was placed on the K^+ sublattice and was relaxed with Rosetta Script protocol 6d, which repacks rotamers and minimizes the rigid-body dock of each protein subunit. This is to confirm that the subunits remain in the designed conformation and in register with the surface when the subunits are allowed to move freely on the sublattice.

Designing the DHR10-mica5-H tiles (inaccurate model). After observing the behaviour of DHR10-mica18-trimer V1 on mica (Supplementary Fig. 20), we noted that among the adsorbed proteins some trimers were observed, but monomers predominated. We hypothesized that this was because the monomers could pack closer together and cover more of the surface, so we set out to design a symmetric layer with better surface coverage. A version with fewer repeats, DHR10-mica5-trimer, was used because it tessellates better than the longer-armed versions. A DHR10-mica5-trimer model was symmetrized around four sublayer-compatible C2 symmetry axes (Supplementary Fig. 29) and translated along the lattice to sample dimer configurations that, when combined with the existing trimer interface, form a two-dimensional layer with P6 symmetry that is compatible with the substrate lattice.

When choosing this C2 interface, we prioritized full coverage of the surface, not a designable protein–protein interface as we did for the C3 interface. Dimer-of-trimers configurations were screened with Rosetta Scripts protocol 7a. One of these was very closely packed, but residues 216–222 in the last repeat clashed with the first and second repeat. To excise this region and allow the dense arrangement the helices in the C-terminal repeat were shortened and a new loop connecting them was designed.

The loop was prototyped in Foldit[33] before a final model was built with Remodel and relaxed with hydrogen-bond constraints in Rosetta Scripts. In Foldit the pre-existing loop was cut and the end points were trimmed and reconnected using the delete and wiggle tools with cut-point distance constraints. Loop residues were mutated to flexible residues (Gly214, Gly215 and Ser216) and a helix-breaking proline was added (Pro217). A blueprint file describing the loop's connectivity, secondary structure and sequence was written for Rosetta Remodel. Remodel followed this blueprint to replace a 24-residue segment with a 17-residue segment via fragment insertion (Remodel protocol 7b). Other amino-acid substitutions were made at this time to prevent clashes in the context of the close-packed arrangement. Subsequently Rosetta Script protocol 7c was used to relax the region with side-chain backbone hydrogen-bond constraints. We call this protein with the N-terminal trimer interface and modified C-terminal repeat DHR10-mica5-H (H for hexagonal symmetry).

Two DHR10-mica5-H trimers docked in the close-packed C2 arrangement were used to generate a P6 symmetry definition file (Rosetta Scripts protocol 7d and PyMOL protocol 7e). Rosetta Script protocol 7f symmetrically repacked rotamers and minimized the rigid-body orientation of subunits with constraints to prevent large movements to make a model of this 'tiling' layer.

Modelling DHR10-mica5-H honeycomb based on observed layer. After AFM imagining of DHR10-mica5-H, we realized that the tiling model was not accurate; instead a honeycomb-like layer had formed. It seemed that the modified C termini formed a C2 interface, pushing the C3 axes further apart than they were in the tiling model.

To model the C2 interface in the observed layers, we resampled lattice-compatible dimer configurations between DHR10-mica5-H, this time with the C-terminal helices packed end-to-end (Rosetta Scripts protocol 8a). These conformations were named by the translation applied before applying C2 symmetry (from a starting conformation designated 0.0X_0.0Y). Full-atom and polyalanine backbone scans together found 19 lattice-matching dimers with a non-zero DDG of below 1,000 Rosetta Energy Units. These candidate dimer interfaces were evaluated by repacking the full sequence (which had already been designed in the context of the tiling layer) while allowing small (<0.5 Å) movements across the C2 interface (Rosetta Scripts protocol 8b).

The candidate C2 interface with the lowest full-atom DDG (9.10X_6.76Y; DDG = −21.2 Rosetta Energy Units) resembled the observed layers (Fig. 4b, f, k, p) and has a nearest-neighbour distance of 11.7 nm between C3 axes (Supplementary Fig. 22a). These interfaces form a P6 symmetric layer that we call the honeycomb model. Rosetta scripts protocols 8c and 8d were used to repack and rigid-body minimize this layer with P6 symmetry. Six chains from the P6 layer were then relaxed on the K^+ sublattice with Rosetta Scripts protocol 8e to ensure each subunit preserved its lattice match. Two other candidate interfaces had similar DDGs and also resembled the layers. Honeycomb models were made with these C2 interfaces as well: 10.40X_4.50Y (DDG = −18.5 Rosetta Energy Units, nearest-neighbour distance = 11.4 nm; Supplementary Fig.. 22b) and 11.70X_2.25Y (DDG = −19.3 Rosetta Energy Units, nearest-neighbour distance = 11.2 nm; Supplementary Fig.. 22c).

Modulating the DHR10-mica*X*-H honeycomb pore size. The DHR10-mica5-H monomer model was modified into three-, four-, six- and seven-repeat versions by deleting and inserting repeats in a list of its backbone torsion angles and its amino-acid sequence. As each repeat is 1.04 nm and two copies of the protein separate the trimer nodes the nearest-neighbour distance is increased by 2.08 nm per repeat addition and reduced by 2.08 nm per repeat subtraction. Pyrosetta protocol 9a generated backbone models from the torsion angles and sequence extrapolated from DHR10-mica5-H. The new DHR10-mica*X*-H monomers were superimposed with structurally analogous regions in the DHR10-mica5-H-honeycomb model and slid with 10.4 Å translations to accommodate the varying number of repeats (PyMOL protocol 9b). The resulting trimer-of-dimer structures were used to generate symmetry definition files for a Rosetta Scripts protocol to repack and minimize the dock in P6 symmetry with constraints to prevent large movements (PyMOL protocol 9c and Rosetta Scripts protocol 9d). Finally, hexamers extracted from the low-energy P6 layers were relaxed with the K^+ sublattice without constraints with Rosetta Scripts protocol 9e.

Expressed protein sequences. The designed protein sequences were expressed with a N-terminal addition containing a His6 tag and a thrombin protease cut site. The full amino-acid sequences of the proteins expressed are in Supplementary Table 2.

Plasmid design and synthesis. A gene encoding DHR10-mica4 was purchased from Genscript in a pet21b plasmid and genes encoding longer DHR10-mica*X* proteins were constructed from it with recursive directional ligation by plasmid reconstruction[34]. DNA encoding the 2 internal repeats was recursively doubled to encode the 4, 8 and 16 internal repeats of DHR10-mica6, DHR10-mica10 and DHR10-mica18, respectively. With restriction digests and DNA ligation the expanded genes were combined with genes encoding modified N and C terminal repeats to construct genes encoding the DHR10-mica*X*-NC and DHR10-mica18-trimer designs. The construct encoding DHR10-mica14-checker was the longest version obtained during repeated attempts to clone an 18-repeat chequered version; it may be the product of homologous recombination in *Escherichia coli*. Genes encoding DHR10-mica*X*-H proteins were designed using the Codon-Scrambler web server version 1.0[35] and purchased from IDT in pet29b plasmids.

Protein expression and purification. Proteins were expressed in BLR(DE3) *E. coli* cells from Novagen using Studiers autoinduction media (M2) with 0.5-litre cultures in 2-litre flasks at 37 °C for 24 h. Subsequently, 50 μg ml^{-1} carbenicillin or 30 μg ml^{-1} kanamycin was included in the media for cells containing pet21b or pet29b vectors, respectively. Cells were pelleted at 4,000*g* for 20 min, resuspended in 30-ml lysis buffer (20 mM Tris HCl pH 8, 150 mM NaCl, 30 mM imidazole, 10 mM lysozyme, 1 mM DNase, 1 mM phenylmethylsulfonyl fluoride) and lysed with a microfluidizer (Microfluidics M110P) at 18K psi (pounds force per square inch). Lysate was clarified at 17,000*g* and the soluble fraction batch bound to 1-ml Ni-NTA resin (Qiagen) for an hour. Lysate and resin were transferred to a gravity column and washed with 20-ml wash buffer three times (20 mM Tris HCl pH 8, 150 mM NaCl, 30 mM imidazole) before eluting the target protein with 12-ml elution buffer (20 mM Tris HCl pH 8, 150 mM NaCl, 250 mM imidazole). The eluate was dialysed in 3,500 molecular weight cut-off (MWCO) dialysis cassettes (Thermo) into 4 litres of TBS (20 mM Tris HCl pH 8, 150 mM NaCl). Thrombin (Novagen) was added to cleave the His-tag and the sample was incubated at room temperature. After 3 h phenylmethylsulfonyl fluoride was added to a final concentration of 1 mM to inactive the thrombin. Sample was again incubated with Ni-NTA resin to bind the cleaved His-tags and uncleaved His-tagged proteins. The flow-through fraction containing the cleaved proteins was concentrated in a 3,000 MWCO centrifugal filter (Amicon Ultra-15).

Measuring protein concentration. A Nanodrop 8000 spectrometer (Thermo Scientific) was used to measure the absorbance of 280-nm-wavelength light in 2 μl of protein samples. The concentration was determined from the measured absorbance at 280 nm and the calculated extinction coefficient following the Beer–Lambert law.

Size-exclusion chromatography. The concentrated, His-tag fused proteins were fractionated by size with an AKTA pure chromatography system on a Superdex 200 Increase 10/300 GL column. Most samples were run twice with an intervening storage period of 24–36 h. A TBS running buffer (150 mM NaCl and 20 mM Tris pH 8) buffer was used. The elution profiles of designs with protein–protein interfaces

(DHR10-mica*X*-NC and DHR10-mica*X*-H) were compared to DHR10-mica*X* versions to check their oligomeric state. DHR10-mica*X*-NC proteins eluted as a higher-order species of limited size, which was collected for further characterization (and assembly on mica). For DHR10-mica*X*-H proteins a species that elutes like DHR10-mica*X* proteins was isolated and used in all subsequent experiments. Size-exclusion chromatography (SEC) profiles of DHR10-mica*X* proteins of various *X* (number of repeats), with and without the non-capped (NC) and honeycomb (H) protein–protein interfaces, are shown in Supplementary Fig. 4. In SEC experiments run in KCl buffer, the concentration of the protein in the peak was determined by measuring the A280 in the 0.5-ml fraction that contains the peak.

Multi-angle light scattering. The molecular weights of DHR10-mica*X* and DHR10-mica*X*-NC proteins were determined by multi-angle light scattering as described previously[28]. All measurements were taken in TBS buffer (150 mM NaCl, 20 mM Tris, pH 8).

Circular dichroism. Circular dichroism spectra were measured with a AVIV Model 420 circular dichroism spectrometer. Samples were in 20 mM NaPi pH 7 or 20 mM Tris pH 8 buffers in a 1-mm cuvette. Units were converted to mean residue ellipticity by dividing the raw spectra by $N \times C \times L \times 10$, where N is the number of residues, C is the concentration of protein, and L is the path length (0.1 cm).

SAXS. SAXS structural data from DHR10-mica18 was collected at the SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California[21]. Beam exposures of 0.3 s for 10.2 s resulted in 33 frames per sample. Data was collected at low (0.32 mg ml^{-1}, 3.4 μM) and high (1.8 mg ml^{-1}, 19.3 μM) protein concentrations in TBS buffer (150 mM NaCl and 20 mM Tris pH 8). To check for concentration-dependent effects, the averaged profiles from each concentration condition were used to calculate the radius-of-gyration R_g from a linear Guiner region with unbiased residuals with the ScÅtter Java application[36]. The averaged scattering profile from the high concentration was used in subsequent analyses. ScÅtter was used to determine the data resolution (q_{max}) as indicated by the linear region in the SIBYLS plot[37]. The FoXS webserver[22] was used to compare the experimental scattering profile to a profile computed from the design model and calculate χ^2 and the predicted R_g. The theoretical minimum R_g of unfolded DHR10-mica18 was calculated following Flory's equation with values from chemically denatured proteins ($R_o = 1.9$, $\nu = 0.570$)[38].

Native mass spectrometry. Sample purity and integrity was analysed by on-line buffer exchange mass spectrometry using an UltiMate 3000 RSLC (Thermo Fisher Scientific) coupled to an Exactive Plus EMR Orbitrap instrument (Thermo Fisher Scientific) modified to incorporate a quadrupole mass filter and to allow for surface-induced dissociation[39]. In brief, 40-pM protein (5 μl of 8 μM protein in TBS) was injected and on-line buffer exchanged to 200 mM ammonium acetate, pH 6.8 by a self-packed buffer exchange column[40] (P6 polyacrylamide gel, BioRad) at a flow rate of 100 μl per minute. Mass spectra were recorded for 1,000–12,000 *m/z* at a resolution of 8,750 as defined at 200 *m/z*. The injection time was set to 200 ms. Voltages applied to the transfer optics were optimized to allow ion transmission while minimizing unintentional ion activation. Mass spectra were deconvoluted with UniDec version 2.6.5[41] using the following processing parameters: sample mass every 1 Da; peak full-width at half-maximum 1 Thompson, Gaussian peak shape function. Organic source-corrected average masses calculated with the NIST Mass and Fragment Calculator v1.32[42] from the His-tag-cleaved sequences were listed as the expected masses.

AFM. The protein stock solution was diluted to the desired concentration with an incubation buffer. The incubation buffer contained 20 mM Tris-HCl (pH 7), 10 mM–3 M KCl depending on the experimental requirements. Then, 100 μl of diluted protein solution was dropped onto freshly cleaved substrates, muscovite mica (Ted Pella), HOPG (Ted Pella) both with and without pre-60-s plasma treatment, molybdenum disulfide (Manchester Nanomaterials), fluorophlogopite (SPI Supplies), to incubate for 0.5–2 h in a sealed Petri dish at room temperature. The muscovite and fluorophlogopite surfaces were then rinsed with fresh incubation buffer, before imaging, to remove un-adsorbed protein molecules. Tris-HCl buffer (pH 7, 1 M) and KCl were bought from Sigma-Aldrich. Nuclease-free water was bought from Ambion.

The images were captured in aqueous solution by a Cypher ES atomic force microscope (Asylum Research). The imaging mode was amplitude modulation. The probes OTR4-B and ORC-8-C (Olympus, Japan) were used. The imaging force was adjusted to minimize any interruption in self-assembly. The offline AFM data processing was all done with software SPIP (Image Metrology, https://www.imagemet.com/). The imaging buffer was 20 mM Tris-HCl (pH = 7) with 10 mM–3 M KCl depending on the requirements of the experiment.

Definition of parameter *S*. The definition of the two-dimensional nematic-order parameter S is:

$$S = \frac{1}{N}\left(\sum_{i=1}^{N} \cos(2\theta_i)\right)$$

Where θ_i is the angle of the ith molecules with the nematic director[43].

Secondary structure characterization. Circular dichroism spectra (Supplementary Fig. 3) of DHR10-mica*X*, DHR10-mica*X*-NC, and DHR10-mica*X*-H proteins show a characteristic α-helical signal with minima at 208 nm and 222 nm and positive values below 202 nm.

Tertiary structure characterization. SAXS profiles of DHR10-mica18 were measured at SIBYLS[21] and averaged within each concentration condition. ScÅtter[36] was used to determine the values $R_g = 53.5$ Å and $R_g = 49$ Å from the low and high concentrations, respectively. The FoXS[22]-computed R_g of the design model was 56 Å. The averaged high-concentration profiles have data resolution limited by $q_{max} = 0.35$ (ref. [37]). The q_{max} trimmed dataset matched a profile FoXS computed from the design model with a $\chi^2 = 2.88$ (Supplementary Fig. 6).

Quaternary structure characterization. DHR10-mica4, DHR10-mica10 and DHR10-mica18 have solution-state molecular weights as determined by MALS that are 0.97, 0.93 and 0.84 times their design value (Supplementary Fig. 4). DHR10-mica*X* proteins were used as monomeric references when evaluating the size-fractionation profiles of proteins with designed protein–protein interfaces (DHR10-mica*X*-NC and DHR10-mica*X*-H), an approach used by those designing solution-state oligomers from other DHR proteins[28]. DHR10-mica18 remains soluble in Tris buffer containing 3 M KCl, and its SEC profile with 3 M KCl (Supplementary Fig. 5a) resembles its profile with 150 mM NaCl (Supplementary Fig. 4).

When fractionated by size during SEC purification, DHR10-mica*X*-NC proteins elute as a larger species than DHR10-mica*X* proteins of the same *X* (Supplementary Fig. 4). The molecular weights of DHR10-mica2-NC, DHR10-mica6-NC, and DHR10-mica18-NC determined by MALS were 1.9, 2.1 and 1.6 times the molecular weight of a monomer (Supplementary Fig. 4). These species were predominantly monomeric, as assayed by native mass spectrometry with some dimers of DHR10-mica6-NC also detected (Supplementary Fig. 16). DHR10-mica6-NC had a similar SEC profile when in solution with 100 mM KCl (Supplementary Fig. 5b) and 150 mM NaCl (Supplementary Fig. 4). When exchanged into buffer containing 3 M KCl (and 20 mM Tris pH 8) samples of 16 μM DHR10-mica6NC and 5 μM DHR10-mica18NC precipitated out of solution, forming a film on the side of the centrifugal filter.

After lysis DHR10-mica*X*-H proteins tended to form soluble aggregates but we could isolate species with size-fractionation profiles like DHR10-mica*X* proteins of corresponding *X* (Supplementary Fig. 4); these species showed no additional signs of aggregation during purification and were used in all subsequent experiments.

Discussion of solution behaviour. DHR10-mica*X*, DHR10-mica*X*-NC and DHR10-mica*X*-H proteins have α-helical circular dichroism spectra that resemble the spectra of DHR10, the DHR protein that they are based on[20]. They were purified with SEC in TBS buffer (150 mM NaCl, 20 mM Tris pH 8) and other standard purification protocols for soluble proteins.

We analysed the solution-state SAXS scattering profile of DHR10-mica18 in two ways. First, we determined that its R_g value, 53.5 Å or 49 Å (at about 3 μM and 20 μM, respectively) was slightly smaller than the design model ($R_g = 56$ Å) and substantially smaller than the value predicted by Flory's equation[38] for an unfolded 906-residue protein ($R_g > 92$ Å). Second, we found the measured SAXS profile (Supplementary Fig. 6) matched a profile computed from the DHR10-mica18 design model (Fig. 1a) with $\chi^2 = 2.88$.

DHR10-mica*X* proteins are monomeric during purification in TBS as assayed by SEC and MALS (Supplementary Fig. 4). Their SEC elution volume shifts predictably based on their number of repeats, making them useful standards when looking at profiles of designs with protein–protein interfaces. Comparison of SEC profiles in Supplementary Figs. 4 and 5 indicates that DHR10-mica18 remains monomeric in solution with 3 M KCl at greater protein concentration (3 μM) than we used to incubate monolayers on mica (0.1 μM) (Fig. 2c).

In DHR10-mica*X*-NC (non-capped) designs, the charged residues that 'cap' the ends of the repeat protein and prevent end-to-end association were removed. Theoretically the interactions could extend indefinitely but the proteins elute from SEC as a higher-order species of limited size (Supplementary Fig. 4). Their peaks skew towards greater elution volumes (smaller species). Multi-angle light scattering (MALS) indicates they are dimers (Supplementary Fig. 4), but they are monomeric upon dilution as assayed by native-mass spectroscopy (Supplementary Fig. 16). Together, these data suggest that DHR10-mica*X*-NC proteins are purified as concentration-dependent oligomers in TBS. DHR10-mica6-NC has a similar size-fractionation profile during SEC with 100 mM KCl (Supplementary Fig. 5b) at 19 μM protein, a greater concentration than used to incubate the adsorbed layers shown in Fig. 3e (0.3 μM) in 100-mM KCl buffer. DHR10-mica6-NC and DHR10-mica18-NC samples (at 16 μM and 5 μM protein concentrations, respectively) precipitated out of solution containing 3 M KCl.

Although the behaviour of DHR10-mica*X*-H (honeycomb) proteins in solution was also affected by their protein–protein interfaces we were able to isolate species with SEC profiles that look monomeric (Supplementary Fig. 4). This is consistent

with our observations of DHR10-mica5-H growing hexagonal domains on the substrate (Supplementary Fig. 24) and absorbing without forming the array at KCl concentrations with lower mobility on the surface (Supplementary Figs. 11, 23).

After observing the rod-like DHR10-mica18 proteins assembly into arrays on muscovite mica in solutions with 3 M KCl (Fig. 2c), we wanted to try to modulate the structure of adsorbed monolayers with designed protein–protein interactions. Although designs with these interfaces (DHR10-mica*X*-NC and DHR10-mica*X*-H proteins) behaved differently in solution, we were able to isolate oligomeric or monomeric species with standard techniques for soluble proteins. When exchanged into solutions containing 3 M KCl they precipitate out of solution, but the decreasing solubility or 'salting out' of proteins with increasing electrolyte concentration is well known. We still wanted to incubate the monolayers on mica in 3 M KCl, because we had found earlier that the mobility and order of our proteins when adsorbed on mica increases with [KCl] (Fig. 2a–c and Supplementary Fig. 11). To do this we purified the proteins in TBS, dialysed them into 20 mM Tris buffer with no salt, and finally diluted them into a solution containing 3 M KCl and $<1\,\mu$M protein immediately before incubating on mica. With this procedure, we again saw that greater [KCl] increased alignment to the mica lattice for both the mica*X*-NC (Fig. 3d–i and Supplementary Fig. 18) and mica*X*-H proteins (Fig. 4k and Supplementary Fig. 23).

Further discussion of DHR10-mica*X* assembly with [K^+]. The marked increase in mobility with increasing KCl concentrations may arise from competition for the protein carboxylates between solution K^+ ions and K^+ ions on the mica surface, as is seen with other cations[44], along with development of a strongly bound hydration layer[45]. That the mica surface is devoid of DHR10-mica6 proteins in 3 M KCl (Fig. 2e) is unlikely, given the high coverage at 10 mM (Fig. 2d), the increasing coverage seen for DHR10-mica14 (Fig. 2f and Supplementary Fig. 15e, f) and DHR10-mica18 at 3 M KCl (Fig. 2a–c and Supplementary Fig. 9), and the fact that the mica lattice would be visible if it were protein-free. More likely, the proteins transitioned to a two-dimensional liquid phase, as expected for rod-shaped particles with low aspect ratios[24], and move too rapidly to be observed as stationary objects by conventional AFM.

Further discussion of DHR10-mica*X*-H honeycomb assembly. At 100 mM KCl, DHR10-mica5-H adsorbed on mica but did not assemble into the lattice (Supplementary Fig. 23a) and even at 1.5 M, the assembled network had many errors (Supplementary Fig. 23b). In situ AFM of DHR10-mica5-H assembly in 3 M KCl (Supplementary Fig. 24) showed the hexagonal domains growing on mica are orientationally aligned, even when not in contact, suggesting that their co-alignment is caused by a shared lattice-match to the substrate.

Unlike DHR10-mica*X*-H with $X = 4$, 5, 6 and 7 repeat units, which forms orientationally aligned domains on mica, DHR10-mica3-H domains aligns both to the K^+ sublattice and at 30° to the sublattice (Supplementary Fig. 25); for $X = 3$, the difference in the lattice match between the nine carboxylates and the underlying K^+ sites along the 30° orientation may be too small to inhibit hexagon formation.

Data availability

Design models in PDB format are available on Github (https://github.com/pylesharley/DHR10micaX/tree/master/rosetta_models). Source data for Supplementary Figs. 2–6 are provided with the paper. All other data not included in the manuscript are available upon reasonable request from the corresponding authors.

Code availability

The Rosetta Macromolecular Modelling suite is available for non-commercial use at (https://www.rosettacommons.org). The specific Rosetta applications used were Rosetta Scripts[46], Remodel[47], and Pyrosetta[48]. Foldit[33], a graphic user interface to Rosetta, was used as well. PyMOL[49] was used to view the design models and prepare input files for Rosetta protocols.

A Github repository (https://github.com/pylesharley/DHR10micaX) contains the Rosetta protocols, input files and Python scripts used to model the protein assembles on mica, and corresponding README.txt files with instructions.

31. Chan, P., Curtis, R. A. & Warwicker, J. Soluble expression of proteins correlates with a lack of positively-charged surface. *Sci. Rep.* **3**, 3333 (2013).
32. MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
33. Kleffner, R. et al. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics* **33**, 2765–2767 (2017).
34. McDaniel, J. R., Mackay, J. A., Quiroz, F. G. & Chilkoti, A. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **11**, 944–952 (2010).
35. Tang, N. C. & Chilkoti, A. Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nat. Mater.* **15**, 419–424 (2016).
36. Rambo, R. P. ScÅtter: Software for SAXS Analysis. Version 3.0 http://www.bioisis.net/tutorial/9 (Diamond Light Source and SIBYLS beamline (12.3.1) of the Advanced Light Source, 2016).
37. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrisically unstructured biological macromoleucles by SAS using the Porod–Debye law. *Biopolymers* **95**, 559–571 (2011).
38. Bernadó, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8**, 151–167 (2012).
39. VanAernum, Z. L. et al. Surface-induced dissociation of noncovalent protein complexes in an extended mass range orbitrap mass spectrometer. *Anal. Chem.* **91**, 3611–3618 (2019).
40. Waitt, G. M., Xu, R., Wisely, G. B. & Williams, J. D. Automated in-line gel filtration for native state mass spectrometry. *J. Am. Soc. Spectrom.* **19**, 239–245 (2008).
41. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).
42. Kilpatrick, E. L., Liao, W. L., Camara, J. E., Turko, I. V. & Bunk, D. M. Expression and characterization of ^{15}N-labeled human C-reactive protein in *Escherichia coli* and *Pichia pastoris* for use in isotope-dilution mass spectrometry. *Protein Expr. Purif.* **85**, 94–99 (2012).
43. Frenkel, D. & Eppenga, R. Evidence for algebraic orientational order in a two-dimensional hard-core nematic. *Phys. Rev. A* **31**, 1776–1787 (1985).
44. Newcomb, C. J., Qafoku, N. P., Grate, J. W., Bailey, V. L. & De Yoreo, J. J. Developing a molecular picture of soil organic matter–mineral interactions by quantifying organo–mineral binding. *Nat. Commun.* **8**, 396 (2017).
45. Martin-Jimenez, D., Chacon, E., Tarazona, P. & Garcia, R. Atomically resolved three-dimensional structures of electrolyte aqueous solutions near a solid surface. *Nat. Commun.* **7**, 12164 (2016).
46. Fleishman, S. J. et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).
47. Huang, P. S. et al. RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6**, e24109 (2011).
48. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
49. The PyMOL Molecular Graphics System. Version 2.1.1 https://pymol.org/2/ (Schrödinger, 2018).