





# Assembly of a patchy protein into variable 2D lattices via tunable multiscale interactions

Shuai Zhang <sup>1,2,5</sup>, Robert G. Alberstein <sup>3,5</sup>, James J. De Yoreo <sup>1,2</sup>✉ & F. Akif Tezcan <sup>3,4</sup>✉

Self-assembly of molecular building blocks into higher-order structures is exploited in living systems to create functional complexity and represents a powerful strategy for constructing new materials. As nanoscale building blocks, proteins offer unique advantages, including monodispersity and atomically tunable interactions. Yet, control of protein self-assembly has been limited compared to inorganic or polymeric nanoparticles, which lack such attributes. Here, we report modular self-assembly of an engineered protein into four physicochemically distinct, precisely patterned 2D crystals via control of four classes of interactions spanning Ångström to several-nanometer length scales. We relate the resulting structures to the underlying free-energy landscape by combining in-situ atomic force microscopy observations of assembly with thermodynamic analyses of protein-protein and -surface interactions. Our results demonstrate rich phase behavior obtainable from a single, highly patchy protein when interactions acting over multiple length scales are exploited and predict unusual bulk-scale properties for protein-based materials that ensue from such control.

<sup>1</sup>Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. <sup>3</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA. <sup>4</sup>Materials Science and Engineering, University of California, San Diego, La Jolla, CA 92093, USA. <sup>5</sup>These authors contributed equally: Shuai Zhang, Robert G. Alberstein. ✉email: [james.deyoreo@pnnl.gov](mailto:james.deyoreo@pnnl.gov); [tezcan@ucsd.edu](mailto:tezcan@ucsd.edu)

Self-assembly provides a powerful means to organize matter at all length scales and is responsible for the emergence of living systems from inanimate nanoscale objects<sup>1</sup>. For example, although discrete proteins generally adopt well-defined structures, assembly into higher-order architectures considerably broadens the scope and complexity of their capabilities to fulfill the stringent requirements of life. Synthetically, even hard, non-interacting nanoparticles can spontaneously arrange into close-packed structures to maximize system entropy<sup>2,3</sup>. Yet, as component complexity increases through the introduction of specific interactions and anisotropy (i.e., patchiness)—so does the resulting structural and functional diversity<sup>3–5</sup>.

As patchy nanoparticles, proteins present several advantages: they are chemically monodisperse and can be tailored with atomic precision at different levels to control assembly: locally (e.g., through discrete chemical bonds)<sup>6,7</sup>, regionally (e.g., through associative noncovalent patches or charged surfaces)<sup>8,9</sup>, or globally (e.g., through shape complementarity and overall charge)<sup>10,11</sup>. Yet, the structural and chemical complexity of proteins also renders energy landscapes controlling their self-assembly highly intricate, with many minima of similar energies separated by small barriers<sup>12,13</sup>. This characteristic is universally exploited by living systems to create dynamic and reconfigurable architectures whose structures can be altered by environmental conditions<sup>14</sup> or physical templates (e.g., chaperones<sup>15</sup> and 2D membranes<sup>16</sup>), but translates to poor predictive control in synthetic systems. Accordingly, design methodologies for artificial protein assemblies have primarily focused on introducing a single type of enthalpy-driven interaction<sup>6,8,17,18</sup> to circumvent this complexity. Consequently, self-assembled products have generally been singular in that one building block leads to one assembled structure, residing in a deep energy well and exhibiting little dynamic or responsive behaviour<sup>19,20</sup> unless the designed interactions themselves are inherently flexible or tunable<sup>6,7,21,22</sup>. In other words, the patchiness of proteins has not been fully exploited to date, despite offering the opportunity to create multipurpose synthetic building blocks whose assembly can be directed toward a diverse set of outcomes, provided the relative depths of the free-energy minima (and corresponding barrier heights) can be manipulated.

2D materials are particularly attractive as targets for directing diverse assembly outcomes because they enable use of surfaces as templates (whose physical and chemical properties can be systematically varied) to modulate the free-energy landscape across which assembly takes place. Here we report the experimental realization of complete synthetic control over protein self-assembly pathways, whereby a singular patchy protein is arranged into alternate 2D crystalline structures through modulation of its surface templated self-assembly via environmental conditions. Our results demonstrate that the concept of patchiness can be adapted to protein design by considering the magnitudes and length scales of different interactions.

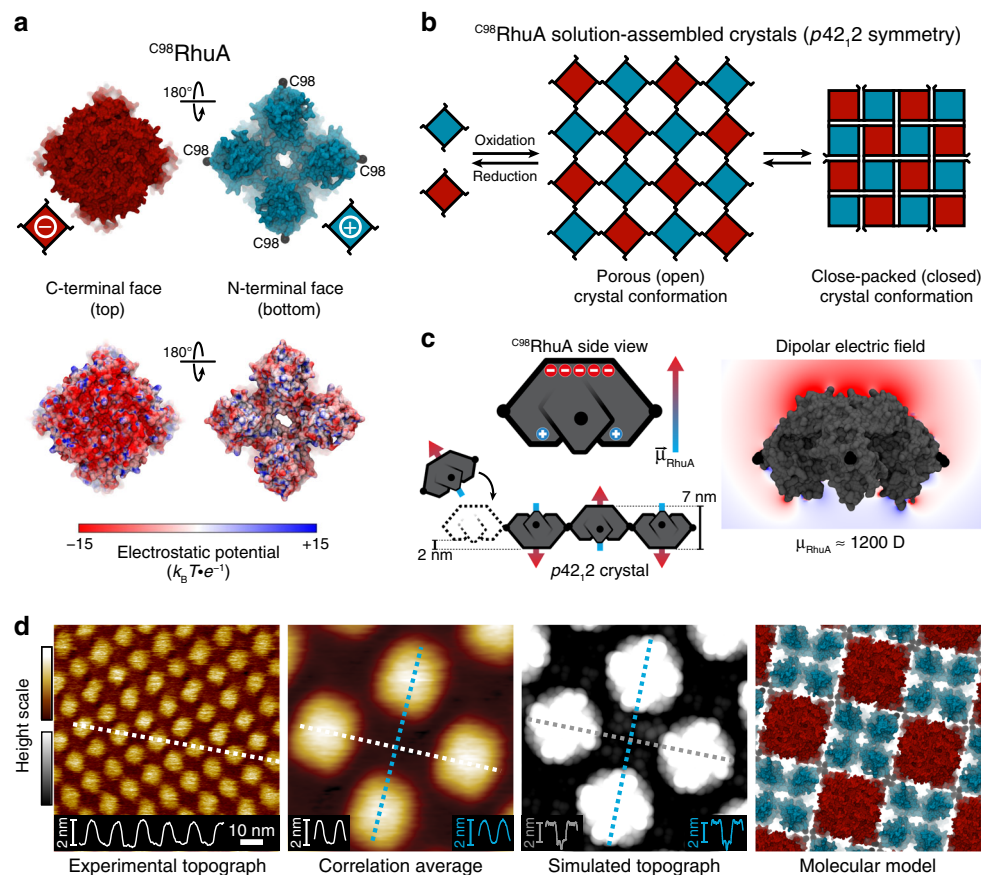
## Results

**AFM characterization of solution-assembled C<sup>98</sup>RhuA crystals.** We utilized the highly patchy protein building block L-rhamnulose-1-phosphate aldolase, a C<sub>4</sub>-symmetric protein modified at its corners with Cys residues (C<sup>98</sup>RhuA) (Fig. 1a), which we previously showed self-assembles in solution into μm-sized, 2D crystals via intermolecular disulfide bonds and 3D crystals through stacking of the layers<sup>7</sup>. While disulfide bond reversibility enabled defect-free self-assembly under thermodynamic control, bond flexibility endowed the 2D lattices with coherent dynamic behaviour, whereby the lattices opened and closed via correlated in-plane rotations of C<sup>98</sup>RhuA molecules (Fig. 1b)<sup>7,23</sup>. Interestingly,

transmission electron microscopy and electron diffraction analyses suggested that the lattices exclusively possessed *p*42<sub>1</sub>2 symmetry<sup>7</sup>, whereby neighbouring C<sup>98</sup>RhuA molecules adopted an out-of-plane antiparallel (up-down) arrangement. Atomic force microscopy (AFM) imaging of solution-assembled C<sup>98</sup>RhuA crystals directly confirms this periodic, alternating arrangement of C<sup>98</sup>RhuA molecules throughout individual lattices (Fig. 1d). The lower protein layer is most clearly resolved at step edges, where tip-protein convolution artefacts on the terrace are minimized (Supplementary Fig. 1).

**Dipolar interactions mediate molecular patterning of C<sup>98</sup>RhuA.** Given that the only interprotein connections in the C<sup>98</sup>RhuA lattices are disulfide bonds<sup>23</sup> (all protein-protein contacts are >5 Å, which is too distant even for weak hydrogen bonds, salt-bridging, or hydrophobic contacts), we originally attributed this distinct up-down arrangement to an energetic bias imposed by differences in disulfide bond conformational strain resulting from different protein orientations<sup>7</sup>. However, intuition would instead suggest that the flexibility of the disulfide hinges should prevent any such substantial bias. Therefore, we carried out all-atom molecular dynamics (MD) simulations, which indeed showed that the conformational energies of disulfide bonds are averaged out within the 2D C<sup>98</sup>RhuA lattice regardless of symmetry (Supplementary Fig. 2). This finding implies there must be indirect inter-protein interactions acting over several nm's (>5 nm's) to yield the alternating C<sup>98</sup>RhuA registry.

We next considered the possible impact of dipole-dipole interactions, which have been shown to promote antiparallel packing in covalent organic frameworks<sup>24</sup>, small-molecule crystals<sup>25</sup>, and nanoparticle lattices<sup>26–28</sup>, and have been invoked in stabilization of biomolecular complexes<sup>10</sup>, but never implicated as a determinant of protein self-assembly. Examination of the C<sup>98</sup>RhuA structure reveals a highly anisotropic charge distribution, with the flat C-terminal top and four-legged N-terminal bottom being negatively and positively charged, respectively (Fig. 1a, c), yielding a sizeable macrodipole moment (*ca.* 1200 D), which is greater than those of 90–95% of proteins<sup>10</sup>. This polarized structure allows C<sup>98</sup>RhuA to be effectively modelled as a physical dipole (Fig. 1c, Supplementary Discussion), as done previously for electrostatically anisotropic inorganic nanoparticles<sup>26,28,29</sup>. Calculation of the pairwise C<sup>98</sup>RhuA dipole-dipole interaction over 7–10 nm (dimensions accessible by lattice conformations) reveals a smooth, funnel-shaped energy landscape invariably favoring the antiparallel arrangement of *p*42<sub>1</sub>2 crystals—relative to parallel—by 0.13–1.11 kcal·mol<sup>−1</sup> (0.22–1.87 *k<sub>B</sub>T*), even at such long distances (Supplementary Fig. 3). Despite this energy being on the order of weak noncovalent bonds, its magnitude is rapidly amplified fourfold (up to 4.2 kcal·mol<sup>−1</sup> or 7 *k<sub>B</sub>T*) by nearest-neighbour interactions (Supplementary Fig. 4), propagating *p*42<sub>1</sub>2 symmetry to newly incorporated monomers during solution-phase crystallization. This strongly suggests that the collective effect of a multitude of weak, long-range, orientation-dependent interactions may play an important role in controlling the precise 2D, nanoscale patterning of C<sup>98</sup>RhuA assemblies. In addition, it was previously postulated based on a statistical analysis of protein crystals that space- or plane-group symmetries that offer more degrees of configurational freedom would be preferentially populated over those that offer less (e.g., *p*42<sub>1</sub>2 over *p*4)<sup>30,31</sup>. Thus, in the case of C<sup>98</sup>RhuA self-assembly in solution, *p*42<sub>1</sub>2 symmetry is both statistically and thermodynamically (through dipolar interactions) preferred over *p*4 symmetry. Determining the relative energetic contributions of these effects and predictably engineering them to modulate protein self-assembly over different length scales represent an exciting avenue for future studies.



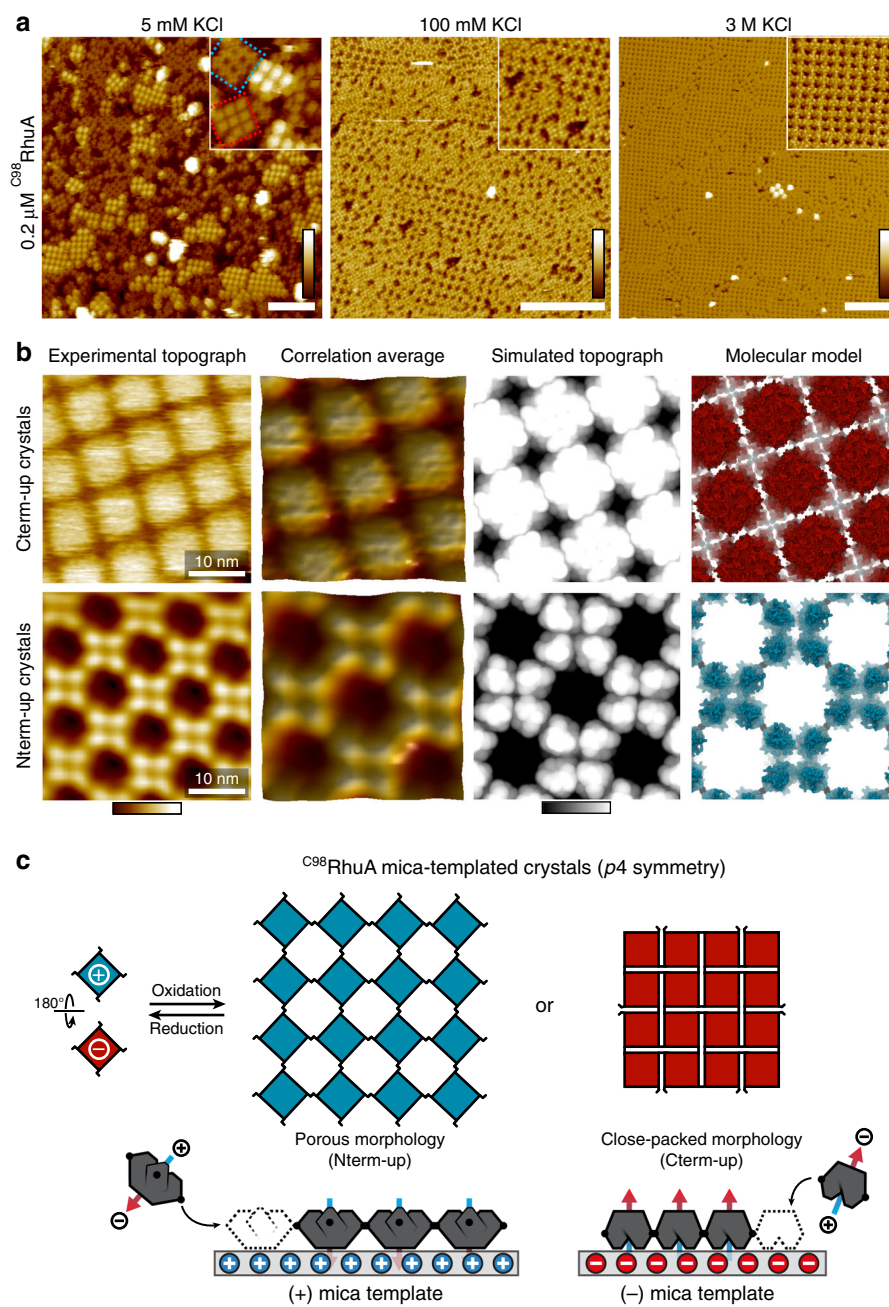
**Fig. 1** Solution self-assembly of  $C^{98}$ RhuA. **a** Solvent-excluded surfaces and electrostatic potential maps of  $C^{98}$ RhuA termini faces, colored according to charge. Depth cueing is used to highlight the two distinct topographies of each terminus. C98 residues are shown as black spheres. **b** Schematic depicting  $C^{98}$ RhuA solution self-assembly into 2D  $p4_2,2$  crystals and their emergent auxetic behavior. **c** Origin of the  $C^{98}$ RhuA macrodipole moment and its global-scale electric field (red to blue color range:  $-7.5$  to  $+2.5 k_B T \cdot e^{-1}$ ). Cartoon arrows represent the dipole orientation of each protein. **d** High-resolution AFM topographs and corresponding structural model of solution-assembled close-packed crystals. The height profiles correspond to the dashed lines. Height ranges (gradient): 5 nm (experimental), 2 nm (simulated).

**Surface-templated self-assembly of  $C^{98}$ RhuA crystals.** The above findings suggest the charge anisotropy of  $C^{98}$ RhuA building blocks and their propensity to form 2D lattices might be exploited to control their assembly by using surfaces that can modulate the balance of forces. As a template, we used muscovite mica (*m*-mica) (001): a pseudo-hexagonal tessellation of negatively charged cavities occupied by native  $K^+$  ions or other cations at a coverage that can be varied to tune the net surface charge<sup>32</sup>. We found that  $C^{98}$ RhuA readily formed large domains of 2D square lattices on *m*-mica at  $\geq 750$ -fold lower concentrations than required for solution assembly ( $\leq 0.2 \mu M$  vs.  $\geq 150 \mu M$ ). These lattices exhibited both open and closed pore conformations resembling those of solution-grown crystals, indicating the designed intermolecular disulfide bonding was preserved and that  $C^{98}$ RhuA-*m*-mica interactions were mediated by the charged surfaces of  $C^{98}$ RhuA (Fig. 2). Indeed, a variant in which Cys98 residues were replaced with Ser ( $S^{98}$ RhuA) adsorbed to *m*-mica but failed to form 2D lattices even at concentrations of  $50 \mu M$  (Supplementary Fig. 5).  $C^{98}$ RhuA assembly on *m*-mica could also be prevented (reversed) in the presence of 1 mM reductant, which prevents (reverses) disulfide bonding (Supplementary Fig. 6).

In contrast to the antiparallel registry of solution-assembled lattices, crystals formed on *m*-mica possessed  $p4$  symmetry, whereby neighbouring proteins adopted a parallel arrangement, despite the repulsive dipole-dipole interactions, demonstrating that protein-*m*-mica interactions could selectively template alternative

crystal structures (Fig. 2). In 5 mM KCl, two classes of  $p4$ -symmetric  $C^{98}$ RhuA crystals formed concurrently: close-packed crystals with C-terminus up (Cterm-up crystals) and porous crystals with N-terminus up (Nterm-up crystals) (Fig. 2a, red and blue dashed boxes). As the resolution of the AFM images were sufficient to resolve the central dip between the N-terminal lobes of individual  $C^{98}$ RhuA molecules (Supplementary Fig. 7), the relative orientations of the adsorbed proteins were readily determined from their surface topologies (Figs. 1a, and 2b). Interpretations of the images were corroborated by simulated surface profiles calculated from structural models of  $p4$ -symmetric  $C^{98}$ RhuA crystals (Fig. 2b). The distribution, size, and degree of order of these two crystal classes could be tuned via the KCl concentration: At 5 mM  $K^+$ , both Nterm- and Cterm-up crystals formed (Fig. 2a, left). At 100 mM  $K^+$ , poorly ordered Nterm-up crystals dominated (Fig. 2a, middle) and at 3 M  $K^+$ , self-assembly exclusively yielded highly ordered 2D Nterm-up lattices containing hundreds to thousands of monomers (Fig. 2a, right).

This variability in  $C^{98}$ RhuA assembly can be understood from the charge state of *m*-mica (Fig. 2c), which follows a well-documented trend towards positive values with increasing KCl<sup>32</sup>. Freshly cleaved *m*-mica surfaces are negatively charged in neutral aqueous solution. The 5 mM  $K^+$  is insufficient to occupy all negatively charged sites on the surface and the cations are distributed heterogeneously, preferentially forming extensive domains of negative and positive charge<sup>33</sup>. At  $[K^+] > 100$  mM,

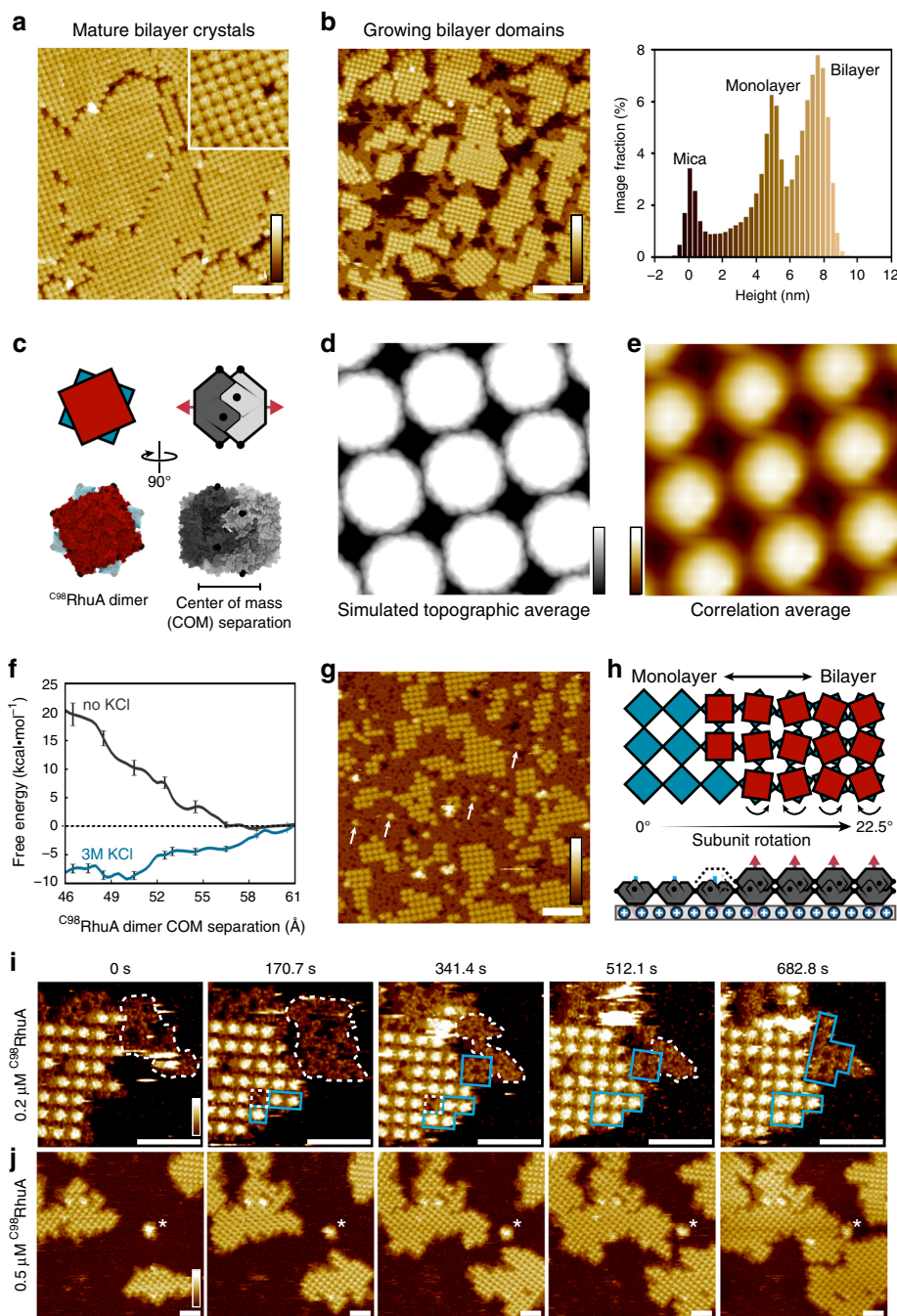


**Fig. 2** Surface-templated self-assembly of  $p4$   $C^{98}$ RhuA crystals. **a** Growth of monolayer  $C^{98}$ RhuA crystals on *m*-mica under increasing  $[K^+]$ . Scale bars (white) are 100 nm. Height ranges (gradient): 20 nm, 10 nm, 10 nm (left to right). **b** High-resolution AFM topographs and structural models of mica-templated crystals. Height ranges (gradients): 7 nm (experimental), 3 nm (simulated). **c** Cartoon overview of mica-templated self-assembly. Blue and red colors represent positive and negative charges, respectively.

the surface sites are more than 50% occupied, fully compensating the surface charge and resulting in positive zeta potentials<sup>32,34</sup>. Accordingly, while 0.2  $\mu$ M solutions of  $C^{98}$ RhuA yielded a mixture of Cterm-up and Nterm-up crystals at 5 mM  $K^+$ , only Nterm-up crystals were observed at  $>1$  M  $K^+$  (Fig. 2a). At much lower  $[C^{98}$ RhuA], Cterm-up crystals formed exclusively at  $[K^+] \leq 5$  mM, indicating preferential attachment to the negatively charged bare mica surface; this morphology was never observed at  $[K^+] > 1$  M (Supplementary Fig. 8). Moreover, selectivity for monomer/crystal binding orientation followed well-known trends in specific ion adsorption to *m*-mica:<sup>32,35</sup> low-affinity counterions (e.g.,  $Zn^{2+}$ ) only produced Cterm-up crystals, while  $Rb^+$ —a known substitute for  $K^+$ <sup>36</sup>—and  $Mg^{2+}$  (widely used to induce surface-charge

reversal on *m*-mica for DNA adsorption<sup>37</sup>) recapitulated the behaviour with  $K^+$  (Supplementary Fig. 9).

**Hierarchical self-assembly of bilayer  $C^{98}$ RhuA crystals.** When  $[C^{98}$ RhuA] was increased to  $>0.2$   $\mu$ M and  $[K^+]$  was held at 3 M, a third 2D lattice arrangement with a height of 9 nm emerged (Fig. 3). The surface topography resembled an array of Cterm-up monomers (Fig. 3a), but this array formed exclusively atop domains of 2D Nterm-up crystals (Fig. 3b, left) and was shorter than twice the height (*ca.* 10 nm) of the underlying Nterm-up monolayer (Fig. 3b, right), implying the  $C^{98}$ RhuA monomers in each layer must interact via interpenetration of the corrugated N-term faces (Fig. 3c). We hypothesized that this tail-to-tail packing is isostructural with the



**Fig. 3 Hierarchical self-assembly of bilayer  $C^{98}$ RhuA crystals.** High-resolution AFM images of **a**, mature and **b**, growing bilayer crystals ( $0.4\text{--}0.5\ \mu\text{M}$   $C^{98}$ RhuA), with image height value histogram shown on right. **c**  $C^{98}$ RhuA dimer structure. **d** Simulated and **e**, experimental average topographs for the bilayer structure. **f** Potential of mean force for  $C^{98}$ RhuA dimerization in the presence/absence of 3 M KCl. Error bars: s.e. of the free-energy gradient, as calculated using the block-averaging method (“Methods”). **g** Self-assembly of  $5\ \mu\text{M}$   $C^{98}$ RhuA, resulting in disordered monolayers which cannot template bilayer domains. Isolated 2nd layer monomers (arrows) are rare outside of crystalline domains. **h** Proposed self-templated growth mechanism for bilayer crystals. **i** N-term up  $C^{98}$ RhuA monolayers grow via nonclassical amorphous (white contour)-to-crystalline (blue regions) transitions at edges, which subsequently template second layer growth through classical (monomer-by-monomer) addition. **j** Bilayer growth kinetics are limited by the growth rate of the first layer. Scale bars (white): 100 nm (**a/b/g**), 50 nm (**i/j**). Height ranges (gradient): 10 nm (**a/b/j**), 2 nm (**d**), 4 nm (**e**), 12 nm (**g**), 8 nm (**i**).

obligate, 9.5-nm tall dimers of the  $F^{88}/C^{98}$ RhuA variant, whose formation during growth of crystals in bulk solution is stabilized by hydrophobic, Phe88-mediated interactions between the N-term faces<sup>7</sup>. Indeed, the simulated topography for a bilayer based on the  $F^{88}/C^{98}$ RhuA lattice structure (Fig. 3d, Supplementary Fig. 10) closely resembles the high-resolution AFM images (Fig. 3e).

Importantly, the unique circular shape of the protein units in bilayer crystals seen by AFM is reproduced upon averaging of both substructures of the  $F^{88}/C^{98}$ RhuA lattice (Supplementary Fig. 10). Furthermore, free-energy (i.e., potential of mean force) calculations show that  $C^{98}$ RhuA Nterm-Nterm intercalation is weakly favoured in the presence of 3 M KCl due to preferential hydration of ions

(i.e., by salting out) and protein shape complementarity, but highly disfavoured at low ionic strength (Fig. 3f). N-term intercalated  $C^{98}$ RhuA dimers can therefore be populated only when incorporated into the matrix of the bilayer, wherein they self-associate via lateral disulfide bonds to form the second layer. As expected from the weak interlayer interactions combined with the strong in-plane disulfide bonds, lowering the ionic strength ( $[KCl] = 100$  mM) after the formation of bilayers at high salt yielded free  $p4$ -symmetric monolayer lattices (likely through exfoliation) (Supplementary Fig. 11) that cannot form without the template.

The importance of disulfide bonding is reflected by the rarity of isolated 2nd layer monomers (i.e., N-term intercalated dimers) (Fig. 3g). If adsorbed randomly with no energetic bias for adsorption next to neighbouring proteins, the expected coverage of isolated monomers for the total 2nd layer coverage seen here (23%) would be 7%<sup>38</sup>, which is 20 times the observed 0.35% coverage of isolated monomers (Supplementary Fig. 12). Thus, the bilayer morphology is realized via a self-templating effect by the underlying  $p4$ -symmetric monolayer, whose subunits necessarily undergo a  $22.5^\circ$  rotation to bring the monomers in the second layer into the proximity required for disulfide bonding (Fig. 3h, Supplementary Fig. 10). All-atom MD simulations carried out in 3 M KCl show that the C-termini of  $m$ -mica-adsorbed  $C^{98}$ RhuA molecules preferentially reside 8–9 Å away from the mica surface (above the ionic slipping plane) (Supplementary Fig. 13), suggesting a negligible energetic barrier to in-plane rotation.

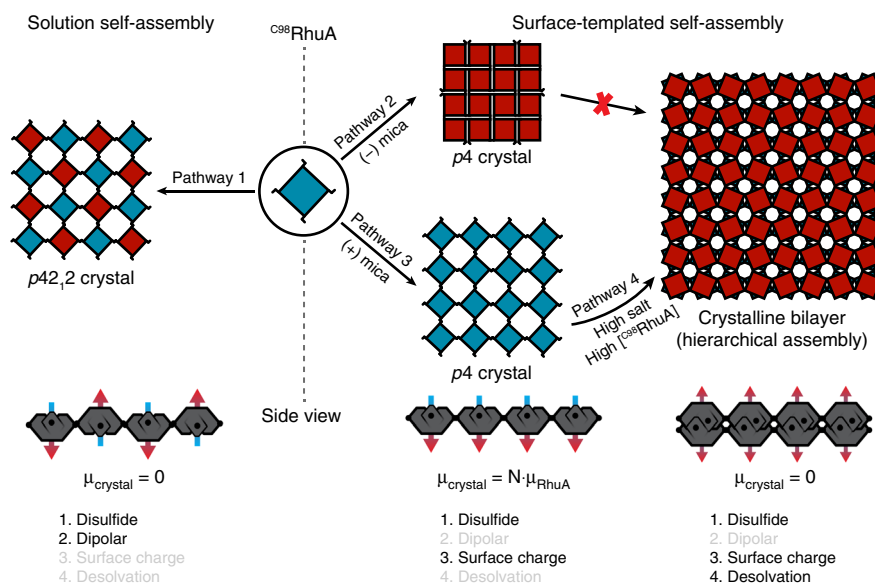
Seeking deeper insight into the bilayer formation mechanism, we used *in-situ* AFM to follow assembly and found two different growth mechanisms acting concurrently. At  $0.2 \mu\text{M}$  [ $C^{98}$ RhuA], the first (N-term-up) layer crystallized via a nonclassical two-step nucleation process<sup>39,40</sup>, whereby the growing edges (Fig. 3i) advanced by initial formation of amorphous regions (white dashed contour), which then spontaneously crystallized (blue dashed square). (Interaction with the adjacent nucleus that has already crystallized may catalyse the transition, a behaviour that has been observed previously during surface crystallization of S-layer proteins<sup>13</sup>.) Additionally, the  $m$ -mica lattice likely assists in the transition, as evidenced by the alignment of crystal domains with specific  $m$ -mica (001) lattice directions (Supplementary Fig. 14), which must be imposed at the time of the transition. This

capability for reorganization and ordering implies relatively weak  $C^{98}$ RhuA- $K^+$  surface interactions and is likely facilitated by the reversibility of disulfide bonds.

Unlike the first layer, the second layer grew via a classical mechanism with single monomers binding to existing crystalline monolayers (Fig. 3i, j). At much higher [ $C^{98}$ RhuA] of  $5 \mu\text{M}$ , the  $m$ -mica surface became fully covered by the amorphous layer before crystallization could occur, leaving the film kinetically trapped in a disordered state (Fig. 3g). Consequently, only small bilayer domains formed and were surrounded by an amorphous monolayer (Supplementary Fig. 12).

The ability of the surface-templated crystalline monolayer to further enable epitaxial growth of a second layer and the emergence of a stepwise self-assembly mechanism demonstrates how the balance of relative interaction energies can modulate assembly pathways to obtain a specific thermodynamic product, while modulating the protein flux via solution concentration can, instead, select for kinetic products<sup>41</sup>. Such tunability underscores the principal advantage of using proteins as patchy building blocks for self-assembled materials.

**Diverse morphologies from a singular patchy protein.** In contrast to our previous work<sup>7</sup>, which described the solution self-assembly of different RhuA variants into distinct 2D crystals (i.e., three building blocks—three assemblies), our findings here (Fig. 4) demonstrate that  $C^{98}$ RhuA can be exploited as a singular patchy particle capable of forming multiple morphologies owing to a multiplicity of tunable interactions that govern self-assembly. Key to all four observed morphologies and their attendant pathways are strong but reversible disulfide bonds (operating at a local level) that ensure self-assembly into 2D crystals. Untemplated solution self-assembly through Pathway 1 requires high protein concentrations and leads to the antiparallel out-of-plane registry of  $C^{98}$ RhuA. This arrangement likely is a consequence of weak interactions between neighbouring macrodipoles (operating at a global level) in addition to the anticipated statistical preference of  $p4_2,2$  symmetry over  $p4$ <sup>30</sup>. In Pathways 2 and 3, inclusion of the charged mica template drastically lowers the protein concentration needed for self-assembly due to favourable electrostatic interactions between  $m$ -mica and anisotropically



**Fig. 4 Self-assembly pathways afforded by  $C^{98}$ RhuA patchiness and template effects.** Overview of all  $C^{98}$ RhuA growth pathways and the self-assembly interactions responsible for their formation. The configuration-dependent dipole moment of each crystal morphology is also indicated.

charged  $C^{98}$ RhuA. These interactions (at a regional level) are strong enough to overcome the dipole-dipole interactions between  $C^{98}$ RhuA molecules, forcing a parallel in-plane registry, while alternating the mica surface charge affords control (also at a global level) over the absolute crystal orientation. Upon increase of ionic strength and protein concentration in Pathway 4, favourable desolvation (regional) interactions become operative between the shape-complementary N-term faces of  $C^{98}$ RhuA, leading to formation of bilayer crystals. Such rich phase behaviour from a single building block is rare in nanoparticle self-assembly<sup>4</sup> and unique among self-assembled protein lattices—natural or designed—because intermonomer interaction surfaces in such lattices are too extensive to allow reconfigurability.

### Emergent functional properties of template-grown $p4$ crystals.

Varying the arrangement of molecules or nanoparticles within a material is known to give rise to different emergent bulk-scale properties<sup>42,43</sup>. Consequently, we considered the effect of the dipolar interactions within the surface-templated 2D phases on their properties.  $C^{98}$ RhuA crystals with  $p4$ -symmetry (one of only 10 polar crystal classes) contain an extended array of coaligned dipoles (Fig. 4) and are formally electrets: polarized dielectric materials possessing a permanent electric field<sup>44</sup>. Although electrets are metastable configurations typically manufactured via thermal annealing in the presence of a strong field, they were achieved in this case through surface-template effects and the strength/topology of the covalent disulfide bonds that preserve structure after assembly. The resulting parallel configuration of the dipoles is indeed predicted to confer unique properties: all-atom MD simulations of infinitely-periodic open-state  $p4$ -symmetric  $C^{98}$ RhuA crystals immersed in a 200-mM NaCl solution predict that  $Na^+$  and  $Cl^-$  ions will rapidly segregate across the lattice, moving against their electrochemical gradients until achieving steady-state within 2–3 ns. Importantly, such behaviour is not observed for  $p4_2,2$ -symmetric crystals (Fig. 5a), consistent with their nonpolar crystal classification (Fig. 5b). At equilibrium, a charge differential of  $0.08\text{--}0.1 e \cdot nm^{-2}$  was attained across the membrane, corresponding to a reversal potential of 50–60 mV (Fig. 5a). Both the charge differential ( $40 e$ ) and membrane potential are in excellent agreement with analytical calculations (Supplementary Discussion) based exclusively on the density of permanent electric dipoles within the crystal (polarization density), indicating that the effect can be treated as purely dipolar in nature.

Given that the polarization density ( $0.008\text{--}0.016 C \cdot m^{-2}$ ) depending on crystal conformation, Fig. 5c) and resulting electric field strength is intrinsically coupled to in-plane compression and expansion,  $p4$ -symmetric  $C^{98}$ RhuA crystals are thus expected to exhibit piezoelectric properties<sup>44</sup>, while  $p4_2,2$ -symmetric crystals are not. Subsequent all-atom simulations directly confirmed this prediction, achieving density-dependent reversal potentials of  $\geq 100$  mV at sufficiently compressed conformations (Fig. 5c, Supplementary Fig. 15). This example illustrates the prospect of obtaining specific bulk-scale physical properties through precise control over the nanoscale patterning of protein building blocks, which in turn can be designed at the molecular level. Polarized membrane materials such as  $p4$ -symmetric  $C^{98}$ RhuA crystals should also be promising candidates for applications in biological nanoelectromechanical systems, including as nanogenerators and passive charge-storage devices, with tunable electrical properties coupled to mechanical in-plane lattice compression.

### Discussion

2D materials have garnered much interest due to their unique physical and chemical properties and applications as membranes, molecular displays and templates for inorganic or biological

patterning<sup>45–47</sup>. In these contexts, reconfigurable systems like  $C^{98}$ RhuA lattices offer distinct advantages in that the densities and patterns can be facilely altered without the need to engineer multiple building blocks. Unlike previous work, such as our own that reported the synthesis of singular structures ( $p4_2,2$  crystals) from a single protein<sup>7</sup>, the findings presented above demonstrate that  $C^{98}$ RhuA represents a patchy particle that is pleomorphic—i.e., it is capable of forming multiple structures from a single building block—a largely unexplored phenomenon for engineered proteins.

Our findings reinforce general design principles for patchy particle self-assembly derived from Monte Carlo simulations<sup>41,48</sup> and underscore the importance of the relative strengths and length scales of interactions over their specific nature in determining the outcome of self-assembly. This substantiates previous analogies between proteins and anisotropic nanoparticles<sup>49,50</sup> by demonstrating that this conceptual framework is not limited to the latter, but instead represents a generalizable phenomenon for obtaining a diversity of self-assembly outcomes from nanoscale objects<sup>4</sup>. The results also highlight the potential ability of macrodipole moments to influence protein self-assembly at extended length scales ( $>5$  nm), provided they are not overpowered by short-range protein-protein contacts and the system in hand has access to multiple conformational states through flexible interactions. Thus, dipolar forces may represent a powerful addition to the toolkit for designing extended protein-based materials, although systematic studies will be required to firmly establish their magnitude and their general applicability. Properly exploited, the unique sensitivity and angular dependence of polarized building blocks (such as  $C^{98}$ RhuA) should permit the use of external fields to direct their molecular-scale organization in solution, potentially enabling the template-free self-assembly of electrically anisotropic biomaterials.

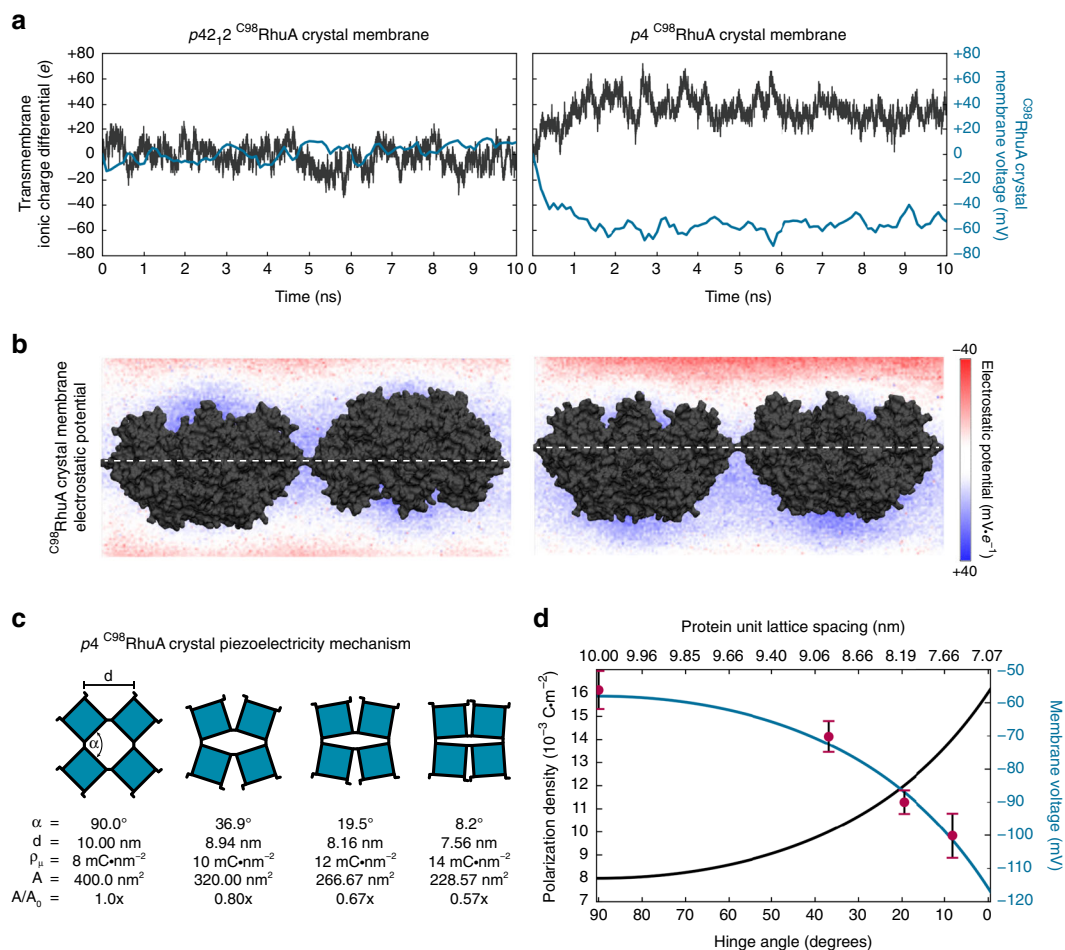
Abstracting from the details of the interactions specific to  $C^{98}$ RhuA, we propose three general guidelines for the rational design of pleomorphic building blocks: (1) selectivity for particular structures can be achieved by incorporating interactions across local, regional, and global scales, (2) intrinsic structural/molecular properties (e.g., dipole moment and symmetry) may impart sufficient energetic bias to influence self-assembly over nm distances without direct contact, but ultimately (3) interaction flexibility between building blocks (e.g., disulfide bonds) is essential to enable access to alternative configurations.

### Methods

**Protein purification and mutagenesis.**  $C^{98}$ RhuA and  $F^{88}/C^{98}$ RhuA were purified following established protocols (see refs. 7,23 for alternative descriptions). Briefly, for each protein, a pJ414 plasmid containing the RhuA coding sequence was transformed into BL21 (DE3) *E. coli* cells (New England Biolabs; Catalog #C25271) via heat-shock, grown to high density in LB + 100 mg·mL<sup>-1</sup> ampicillin, overexpressed by overnight 1 mM IPTG induction, pelleted, resuspended in 20 mM Tris-HCl (pH 7.5) + 10 mM  $\beta$ -mercaptoethanol ( $\beta$ ME), and lysed by sonication. The resulting solution was clarified by centrifugation (5000 rpm, 15 min), treated with 1.5% Polymyxin-P, reclarified, and purified via NaCl step gradient on a DEAE gravity column at 4 °C. Peak fractions were pooled and RhuA was precipitated using 1.7 M  $(NH_4)_2SO_4$ , gently stirred for 30 min, then separated by centrifugation. The precipitate was dialyzed into 20 mM sodium acetate (pH 5) + 10 mM  $\beta$ ME, exchanging 3–4 times over 3 days. The dialysate was sterile filtered, loaded onto S columns via FPLC and purified via NaCl gradient. RhuA elutes at  $\sim 200$  mM NaCl for both columns. Peak fractions ( $>90\%$  purity) were pooled prior to concentration and storage.

Overexpression and purification of  $S^{98}$ RhuA was carried out analogously to  $C^{98}$ RhuA except for the omission of  $\beta$ ME in the purification buffers. All purified proteins were dialyzed into 20 mM Tris-HCl (pH 7.5) and 10 mM reduced L-glutathione (GSH), concentrated to 100–150  $\mu$ M, flash-frozen in liquid nitrogen, and stored at  $\leq -60$  °C. The plasmid for  $S^{98}$ RhuA was generated from the  $C^{98}$ RhuA parent plasmid via site-directed mutagenesis using the following primers:

RhuA S98 Forward:  
GTTAAGGTGGATAGCAGCGGTGCAGGTTACCACATCC.



**Fig. 5** *p4*<sup>C98</sup>RhuA crystals are electrets and piezoelectric materials. **a** Net charge imbalance due to ion conduction across the disulfide bonds (half of the simulation box) and resulting electric potential difference across open-state crystals. Dipole-less *p4212* crystals induce negligible charge separation, while *p4* crystals generate a steady-state 50–60 mV potential difference across the lattice (membrane voltage). **b** Volumetric slices of the electrostatic potential across the central plane of each system directly illustrate the polarization. **c** Piezoelectric properties of *p4*-symmetry *C98*RhuA crystals, arising from mechanical coupling between the polarization density ( $\rho_{\mu}$ ) and lattice conformation (left). The magnitude of this effect was predicted analytically (right) from the  $\rho_{\mu}$  over all lattice conformations (as defined by the rotating squares model<sup>7</sup>), while the points reflect numerically determined voltages from all-atom simulations at discrete conformations (Supplementary Fig. 15). The associated hinge angle ( $\alpha$ ), lattice spacing (**d**),  $\rho_{\mu}$ , and absolute (A) and relative (A/A<sub>0</sub>) unit cell areas for each point are indicated at left. Error bars represent the s.d. of the membrane voltage over the last 5 ns of simulation.

RhuA S98 Reverse:  
GGATGTGGTAACCTGCACCGCTGCTATCCACCTTAAC.

**Solution self-assembly of *C98*RhuA.** Crystallization of *C98*RhuA was induced via hand-thawing of frozen RhuA aliquots, which were then placed on a shaking platform at 4 °C and allowed to mature. Nucleation typically occurred within 3–7 days, and crystals fully matured over 2–3 weeks, consistent with previous reports<sup>7,23</sup>. Crystal suspensions were clarified 2–3x by low-speed (*ca.* 3000 rpm) centrifugation in a benchtop centrifuge followed by replacement of the supernatant with fresh buffer to remove unincorporated proteins. This procedure facilitated the binding of *p4212* crystals onto the mica substrate, but also increased the population of stacked 2D crystals. The suspension of *p4212* crystals was diluted to 25  $\mu$ M using 10 mM Tris-HCl (Sigma-Aldrich) pH 7.0 prior to imaging by AFM. 10  $\mu$ l of diluted crystal suspension was incubated on polylysine-treated mica for 10 min, then rinsed with 10 mM Tris-HCl buffer prior to imaging. Polylysine-treated mica was obtained by incubating 10  $\mu$ l polylysine solution (0.01%; Sigma-Aldrich) on freshly cleaved mica for 1 min, rinsing with water, and drying under nitrogen prior to use.

**Mica-templated self-assembly of RhuA.** Frozen stock solutions of *C98*RhuA were thawed at room temperature and diluted to the final concentration with incubation buffer (10 mM Tris-HCl pH 7.0, 1–10  $\mu$ M GSH, and desired concentration of KCl, RbCl, MgCl<sub>2</sub>, or ZnCl<sub>2</sub>). 100  $\mu$ l of diluted protein solution was deposited onto freshly cleaved mica and incubated for 24 h in a sealed petri dish at room temperature. The mica surface was rinsed with fresh incubation buffer prior to imaging by AFM. GSH, KCl, RbCl, MgCl<sub>2</sub>, and ZnCl<sub>2</sub> were purchased from Sigma-Aldrich, nuclease-free water from Ambion, and muscovite mica from Ted Pella.

**Atomic force microscopy.** All AFM images of mica-templated and solution-grown *C98*RhuA crystals were captured in PeakForce Tapping<sup>TM</sup> mode on a MultiMode<sup>TM</sup> VIII AFM (Bruker, CA) using HYDRA4V-100NG or HYDRA6V-100NG (AppNANO, CA) probes in the incubation buffer used for each specific sample. The peak-force set-point was continuously adjusted to minimize any possible manipulation or damage from probes. The effective imaging force ranged from 100 to 200 pN, within the typical force range for AFM imaging of biomolecules<sup>51,52</sup>. All offline data processing was done using the SPIP<sup>TM</sup> software package (Image Metrology, Denmark).

**Simulated AFM topographs.** AFM topographs were calculated from atomic models with different monomer configurations (extracted from equilibrated 2 × 2 RhuA crystals; see below) using a custom Tcl script executed within VMD<sup>53</sup>, taking the mica surface to be the plane  $z = 0$ . Accordingly, crystal models were recentered at  $(x = 0, y = 0)$  and moved along the  $+z$  axis until the minimum  $z$  position of the protein Ca atoms was at 0. The probe tip was modeled as a sphere with radius 10.0 Å and center  $z_{\text{tip}}$  at the end of a cone with half-angle 20° and truncated at  $z_{\text{cutoff}}$  (see Supplementary Fig. 16). Parametric equations which define the volume enclosed by this tip shape were determined and used to check for overlaps with any protein heavy atoms within the conical or spherical volumes for a given  $(x, y, z_{\text{tip}})$ . This was critical to capture tip convolution artifacts arising from the finite size of the tip, not simply the tip radius. Tapping was simulated by moving the probe to  $(x, y, \text{radius})$  and increasing  $z_{\text{tip}}$  until the number of overlaps was 0 (or below a given threshold to account for the flexibility of disordered protein loops). The height for that  $xy$  position was then recorded as  $(z_{\text{min}} = z_{\text{tip}} - \text{radius})$ . This calculation was repeated row-by-row, with the resolution determined by  $\Delta x = \Delta y =$



0.50 Å, to generate a 2D array of height values analogous to a true AFM topograph. Output was directly visualized and processed using Gwyddion. This code (and associated utilities/example files) can be accessed at <https://doi.org/10.5281/zenodo.3924893>.

**Evaluation of the C<sup>98</sup>RhuA dimerization free-energy landscape.** The C<sup>98</sup>RhuA dimer structure was prepared from the A<sup>88F</sup>RhuA crystal structure (PDB ID: 2UYU) by performing the corresponding D98C and F88A mutations using PyMOL<sup>54</sup>. PSFGEN<sup>53</sup> was used to add missing hydrogens and assign atom types. The final dimer structure was centered at the origin with the axis of symmetry aligned along the z-axis, then each monomer was translated ±10 Å along the z-axis to yield a starting distance between each protein's center of mass (COM separation) of 65.5 Å. The separated dimer was then solvated with 45,082 water molecules (CHARMM TIP3P) and either neutralized with 32 K<sup>+</sup> ions (no KCl starting structure) or 2302 K<sup>+</sup> ions and 2270 Cl<sup>-</sup> ions (3 M KCl starting structure). The CHARMM36 force field<sup>55</sup> was used for protein atoms and Joung-Cheatham parameters<sup>56</sup> were used for monovalent ions. Both systems were minimized for 5000 steps with all protein atoms held fixed, followed by a 5000 step minimization without restraints. Protein Ca atoms were constrained to their starting positions by a 10 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> restraint for equilibration. The systems were equilibrated for 1 ns in the isobaric-isothermal (NPT) ensemble (1 atm, 300 K) with the cross-sectional xy proportions held constant, yielding final box dimensions of 104.7 × 104.7 × 145.6 Å. Monomers were then linearly pulled towards each other to a final COM-COM distance of 45.5 Å over 5 ns using a 100 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> moving restraint, with the xy coordinates of the Ca remaining constrained to prevent rotation of the monomers. Initial coordinates for umbrella sampling windows were extracted from this pulling simulation and maintained with weaker force constants (see Supplementary Table 1 for details). All windows were equilibrated for 25 ns, of which the last 10 ns were used for calculation of the PMF using the WHAM algorithm<sup>57</sup>. The 100 kcal·mol<sup>-1</sup>·deg<sup>-2</sup> harmonic restraints were employed during sampling to prevent rotation of each monomer about their axis of symmetry in order to preserve their relative orientations from the F<sup>88</sup>RhuA crystal structure, which simplifies the dimerization coordinate to 1D (COM separation along the z-axis). Errors in the free energy estimates were calculated using the block averaging method<sup>58</sup>. PMF simulations were carried out using NAMD 2.12<sup>59</sup>.

**Modelling and simulation of protein binding to m-mica.** Parameters and initial coordinates for muscovite mica were taken from the INTERFACE force field package<sup>60</sup>. A 5 × 3 m-mica supercell with Al substitutions in agreement with <sup>29</sup>Si NMR data was then tiled to form a 5 × 5 array of dimensions 129.795 × 135.230 × 19.9452 Å. Periodic bonds were reorganized using TopoTools<sup>61</sup> within VMD to yield a “full-size” single layer of m-mica corresponding to a 25 × 15 × 1 supercell of the crystallographic m-mica unit cell. Copies of this full-size layer were then translated along multiples of the vector (2.005, 0.000, 19.9452 Å) to yield a “5stack” bulk-like m-mica structure of final dimensions 129.795 × 135.230 × 99.726 Å, which was then merged into a single structure using TopoTools. This system was equilibrated at constant pressure without minimization in a fully-flexible orthorhombic periodic cell for 100 ps using the particle mesh Ewald method for full-system electrostatics. The two outermost layers were then removed from the equilibrated structure to yield the final starting m-mica coordinates (appx. 60 Å in height), with all surface vacancies on both sides of the mica occupied by K<sup>+</sup> ions. These ions were free to exchange with the solution, and at equilibrium 80–85% of sites were occupied by K<sup>+</sup> ions. This state was also attained if the system was initiated with a bare (K<sup>+</sup> occupancy = 0) m-mica surface and the same numbers of ions in solution, confirming that it reflects thermodynamic equilibrium.

A pre-equilibrated C<sup>98</sup>RhuA structure was recentered at the origin and solvated in a pre-equilibrated 128.8 × 134.0 × 129.9 Å box (68,667 waters) prior to neutralization with 16 K<sup>+</sup> ions. An additional 3463 K<sup>+</sup> ions and 3463 Cl<sup>-</sup> ions were added, bringing the [KCl] to ca. 2.67 M (in 61,725 waters). This system was minimized for 1000 steps with all protein atoms fixed and subsequently equilibrated in the NPT ensemble (1 atm, 300 K) for 1 ns keeping the xy area constant. The final box dimensions were 128.8 × 134.0 × 125.1 Å.

The equilibrated protein system was merged with the equilibrated mica layers into a single structure using TopoTools such that the lowest position of protein Ca atoms on the C-terminal face was 25 Å above the average z coordinate of the bridging oxygen atoms on the mica surface (henceforth treated as position z = 0). Overlapping waters and ions were translated along the +z direction by the periodic c vector from the last frame of the protein simulation to minimize bad contacts. Finally, 750 Cl<sup>-</sup> ions were added to the solution to neutralize the system, bringing the solution ion concentrations to ~2.67 M K<sup>+</sup> and 3.33 M Cl<sup>-</sup>, close to experimental conditions. Solution-exposed K<sup>+</sup> ions on both faces of the mica slab were converted to the Joung-Cheatham atom type to reflect their occupancy by solution-phase ions, while K<sup>+</sup> ions on the interior retained their typing as prescribed by INTERFACE. The resulting system was minimized for 100 steps to remove bad contacts then equilibrated for 1 ns in the NPT ensemble, keeping the periodic a and b vectors constant at the equilibrated mica dimensions of 129.25 Å and 134.45 Å, respectively. All mica atoms were lightly constrained (1 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>) to their initial positions to prevent drift of the slab. The protein was restrained (by its Ca atoms) to its initial COM height (z = 50.0 Å relative to the mica surface oxygens), xy position, and angle of rotation using 100 kcal·mol<sup>-1</sup>

restraints during initial equilibration. The average c vector over the last 50 ps of equilibration (185.5 Å) was then used to define the periodic cell dimensions for all subsequent runs. Mica hydrogen atoms were then released from constraints and the system was equilibrated for a further 2 ns at constant volume. In order to qualitatively determine whether C<sup>98</sup>RhuA adsorption occurs directly at the surface or above the ionic slipping plane, the former configuration was obtained by pulling the protein 20 Å towards the mica surface over 5 ns using a 100 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> moving restraint, such that the minimum Ca position was at z ≈ 5 Å.

The protein was allowed to adopt a preferred binding configuration (height, rotation) by releasing the COM restraints on the protein z position and rotation angle. Instead, a flat protein orientation was maintained in an unbiased fashion by restraining the difference in Ca COM z positions of diagonal protein chains (A-C and B-D) at 0 Å using two 100 kcal·mol<sup>-1</sup> restraints; this allowed the full protein to move freely while preserving the relative positions of its individual subunits. Existing constraints on the protein xy position and mica heavy-atom coordinates were preserved to simplify the adsorption/desorption pathway. Sampling was carried out over 10 ns of equilibration, with the protein initially relaxing away from the surface by ca. 3 Å, indicating that water and ions interact more strongly with the surface and prevent direct adsorption of the protein.

**APBS calculations.** Continuum electrostatics calculations of the C<sup>98</sup>RhuA protein (appropriately protonated using PROPKA 3.1<sup>62</sup>) were carried out using APBS 1.5<sup>63</sup> to reflect experimental solution self-assembly conditions (pH 7.5, [ions] = 20 mM, ε<sub>solvent</sub> = 78, ε<sub>protein</sub> = 2).

**Equilibration of 3D periodic crystals.** Open-state 2 × 2 C<sup>98</sup>RhuA crystals were constructed as previously described<sup>23</sup> for both p4 and p4<sub>2</sub> symmetries. These crystals were then crosslinked to their periodic images via disulfide bonds, solvated in a 200.0 × 200.0 × 114.0 Å box (120,609 and 120,531 waters), and neutralized with Na<sup>+</sup> ions (Joung-Cheatham). Initial coordinates were first minimized for 2000 steps with all protein atoms held fixed, followed by another 1000 step unrestrained minimization. The systems were then equilibrated for 5 ns (followed by 2 ns of production sampling) at constant pressure, keeping the xy-dimensions of the box held constant (final equilibrated periodic c dimension ≈ 104.5 Å). 100 kcal·mol<sup>-1</sup>·protein<sup>-1</sup> restraints were employed to maintain the difference in Ca COM xy and z positions of diagonally opposed proteins at 0 Å, preserving the structure of the crystal in a minimally biased fashion. The coordinates for the open-state 2 × 2 p4 crystal (extracted after 1 ns of equilibration) were manipulated to generate all other p4 lattice conformations for the piezoelectricity simulations via rigid-body rotations/translations of the individual protein units to their idealized dimensions (while preserving the periodic bonding topology). Each conformation was solvated in its own periodic water box of dimensions a × b × 114.0 Å (where a = b = the xy unit cell dimensions of that conformation) to maintain the periodic disulfide bonding within the crystal plane and equilibrated at constant pressure as described above. Analogous simulations were carried out for F<sup>88</sup>/C<sup>98</sup>RhuA (final periodic box size: 182.5 × 182.5 × 124.8 Å). The average protein coordinates over the last 1 ns of simulation served as the model structures for all simulated tapping calculations.

**Calculation of the C<sup>98</sup>RhuA macrodipole moment.** The dipole moment of C<sup>98</sup>RhuA was calculated from equilibrated protein structures using the VMD “measure dipole” command and visualized using the dipole watcher plugin, both of which yielded values of ca. 1200 D. Hydrogens and Zn ions were excluded. We verified the magnitude of the vector using the Protein Dipole Moments Server<sup>10</sup> using PDB IDs: 1OJR, which yielded a value of 310 D per protein subunit (1240 D total) and 1GT7 (1134 D, full tetramer). We used 1200 D for all dipole potential energy calculations as a representative (and reflective of the equilibrated protein structure) estimate.

**Electret/piezoelectric simulations.** The last frame of the equilibrated p4 and p4<sub>2</sub> 3D periodic crystal simulations were used as starting coordinates for slab-geometry electret simulations. The protein crystals were recentered such that their disulfides lay at z = 0, and solvent/ions were translated across periodic c vector until there were equal numbers of water molecules on each side (within tolerance of <0.1% difference). Ions were added to each side of the disulfides to concentrations of 200 mM NaCl such that there were equal numbers of each type of ion on both sides. Systems with neutral net charge and without neutralizing ions (−64 e) were constructed to test the effect of system neutrality on membrane potential (none observed). Protein Ca atoms were constrained to their starting z position using a 100 kcal·mol<sup>-1</sup> restraint to maintain an equal number of waters on each side over the course of the simulations. Equilibration was carried out for 10 ns using constant-area 2D periodic boundaries (a = b = 200.0 Å for p4<sub>2</sub> lattices, and spanning 151.19–200.00 Å for different p4 lattice conformations; see Supplementary Table 2 for details), with full-system electrostatics calculated using the Multiscale Summation Method<sup>64</sup> in NAMD 2.13, which is compatible with both 2D periodic boundaries as well as non-neutral systems. This permitted the calculation of membrane properties using a single crystal by avoiding the problem of ion diffusion back across the periodic boundary, which otherwise would require a double membrane system consisting of >900,000 atoms. Gentle boundary

conditions (1 kcal·mol<sup>-1</sup> potentials) were enforced above and below the liquid phase to prevent the evaporation of water molecules using the TclBC module of NAMD. Simulations were carried out for 10 ns to ensure full equilibration.

**Disulfide dihedral energy calculations.** Disulfide dihedral angles (as depicted in Supplementary Fig. 2) for the nonperiodic disulfide bonds of *p4* and *p42<sub>12</sub>* crystals were extracted from the 2 ns (2000 coordinates) of production sampling from the 3D periodic crystal simulations. The potential energy associated with each disulfide bond conformation was calculated according to the empirical formula described in ref. 65.

**Numerical C<sup>98</sup>RhuA electrochemical potential calculations.** To directly calculate the steady-state voltage drop arising from ion segregation across RhuA crystals, volumetric electrostatic potential maps (exclusively due to ions) were computed for all 2D electret simulations using the PMEpot<sup>66</sup> plugin in VMD for all 10 ns of sampling (10,000 configurations), split into 100-frame blocks. The maps were calculated over  $a \times b \times 200$  Å volumes ( $a = b = 151.19$ – $200.00$  Å depending on lattice conformation) at 1 Å intervals, incorporating the influence of periodic images in the *xy* dimensions, while remaining pseudo-non-periodic in the *z* dimension by providing an approx. 100 Å air gap between periodic images. An Ewald factor of  $0.25$  Å<sup>-1</sup> was used to smooth the charge potentials. The linear voltage drop across the crystal was quantified by integrating over the *xy* voxels within each 3D volume and projecting the average potential onto the *z* dimension using a Python script. The difference in electrostatic potential due to the ion distribution across the membrane was calculated, conservatively scaled by the dielectric of pure water ( $\epsilon = 78$ ), and multiplied by the factor  $25.87$  mV · ( $k_B T$ )<sup>-1</sup> for a 300 K simulation. The 5–10 ns time-averaged membrane potentials for *p4* and *p42<sub>12</sub>* crystals were  $-56.09 \pm 5.71$  mV and  $-3.09 \pm 5.85$  mV, respectively. This potential for *p4* crystals is in very close agreement with the potential calculated analytically ( $-57.96$  mV) for open-state *p4* crystals based purely on the density of dipoles within the lattice (see Supplementary Discussion), strongly suggesting that the macrodipole moment of RhuA crystals can indeed account for their predicted bulk properties.

**Pairwise C<sup>98</sup>RhuA nanoparticle electrostatic potentials.** The electrostatic potential between individual C<sup>98</sup>RhuA proteins was computed using the pairwise potential between two charged nanoparticles ( $U_{ij}$ ) proposed by Phillies<sup>29</sup> for dilute solutions of proteins and charged colloids (Eq. 1), as employed previously for modelling nanoparticle and protein interactions<sup>26,28,67</sup>.

$$U_{ij}(r_{ij}, \theta_{i,j}, \varphi_{i,j}) = k_e \frac{q_i q_j}{r_{ij}} e^{-\kappa r_{ij}} C_0^2 + k_e \frac{q_i \mu_j \cos \theta_j + q_j \mu_i \cos \theta_i}{r_{ij}^2} e^{-\kappa r_{ij}} C_0 C_1 + k_e \frac{\mu_i \mu_j}{r_{ij}^3} \left\{ \cos \theta_i \cos \theta_j [2 + \kappa r_{ij} + (\kappa r_{ij})^2] + \sin \theta_i \sin \theta_j \cos(\varphi_i - \varphi_j) [1 + \kappa r_{ij}] \right\} e^{-\kappa r_{ij}} C_1^2, \quad (1)$$

where  $C_0$  and  $C_1$  are given by:

$$C_0 = \frac{e^{\kappa a}}{1 + \kappa a}, \quad (2)$$

$$C_1 = \frac{3e^{\kappa a}}{2 + 2\kappa a + (\kappa a)^2 + \frac{(1+\kappa a)}{\epsilon}}. \quad (3)$$

Here,  $q_i$  and  $q_j$  are the net charge of each C<sup>98</sup>RhuA protein ( $-16e$ ),  $\mu_i$  and  $\mu_j$  are the dipole moments of each C<sup>98</sup>RhuA protein (1200 D),  $a$  the particle radius (chosen to be 4.0 nm, the radius for a circle of equivalent area), and  $r_{ij}$  is the center-center distance between particles. The angles  $\theta_{i,j}$  are the dipole- $r_{ij}$  angle for each protein, while  $(\varphi_i - \varphi_j)$  is the dihedral angle between the two dipoles about the  $r_{ij}$  vector ( $\Delta\varphi_{ij}$ ).  $k_e$  is Coulomb's constant and given as  $(4\pi\epsilon_0\epsilon)^{-1}$ ;  $\epsilon_0$  the permittivity of vacuum;  $\epsilon$  the relative solvent permittivity (78); and  $\kappa^{-1}$  is the Debye screening length (2.1508 nm, corresponding to a solution of 20 mM ions).

The requirement of lateral disulfide bonding within the 2D plane of C<sup>98</sup>RhuA crystals enables Eq. 1 to be simplified considerably. It geometrically restricts the dipole moment of each particle pair to be oriented normal to the lattice (so  $\theta = 90^\circ$ ), and  $\Delta\varphi_{ij} = 0$  (*p4* symmetry) or  $180$  (*p42<sub>12</sub>* symmetry). We can then reduce all  $\sin(\theta)$  terms to unity, and nullify all  $\cos(\theta)$  terms (comprising the entire charge-dipole term and first half of the dipole-dipole term). This gives us the reduced form of  $U_{ij}$  (Eq. 4):

$$U_{ij}(r_{ij}, \Delta\varphi_{ij}) = k_e \frac{q_i q_j}{r_{ij}} e^{-\kappa r_{ij}} C_0^2 + k_e \frac{\mu_i \mu_j}{r_{ij}^3} \cos(\Delta\varphi_{ij}) [1 + \kappa r_{ij}] e^{-\kappa r_{ij}} C_1^2. \quad (4)$$

We can see that the first term is purely repulsive due to the identical net charge on both C<sup>98</sup>RhuA proteins, while the dipolar interaction is dependent on the pair dihedral angle, making it repulsive for *p4* crystals ( $\cos(0) = 1$ ) but attractive for *p42<sub>12</sub>* crystals ( $\cos(180) = -1$ ). The dihedral angle can also be allowed to vary continuously (*i.e.*, accessible to a newly-attached monomer to the growing lattice edge), and results in an energy funnel towards the minimum energy configuration of antiparallel packing (Supplementary Fig. 3). This illustrates the origin of the thermodynamic selectivity for *p42<sub>12</sub>* lattices in solution (Supplementary Fig. 4).

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

## Data availability

Data supporting the findings of this manuscript are available from the corresponding authors upon reasonable request. A reporting summary for this article is available as a supplementary information file.

## Code availability

The VMD script (v1.0.1) capable of performing the AFM tapping simulations (and associated 10.5281/zenodo.3924893 helper utilities and example files) is available online from R.G.A. via GitHub repository (<https://doi.org/10.5281/zenodo.3924893>) under a GNU General Public License v3.0.

Received: 1 April 2020; Accepted: 7 July 2020;

Published online: 28 July 2020

## References

- Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84**, 551–575 (2015).
- Frenkel, D. & Ladd, A. J. C. New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.* **81**, 3188–3193 (1984).
- Thorkelsson, K., Bai, P. & Xu, T. Self-assembly and applications of anisotropic nanomaterials: A review. *Nano Today* **10**, 48–66 (2015).
- Glotzer, S. C. & Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature* **6**, 557–562 (2007).
- Gröschel, A. H. et al. Guided hierarchical co-assembly of soft patchy nanoparticles. *Nature* **503**, 247–251 (2013).
- Brodin, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382 (2012).
- Suzuki, Y. et al. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373 (2016).
- Yeates, T. O. Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu. Rev. Biophys.* **46**, 23–42 (2017).
- Kostiainen, M. A. et al. Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nat. Nanotech.* **8**, 52–56 (2013).
- Felder, C. E., Prilusky, J., Silman, I. & Sussman, J. L. A server and database for dipole moments of proteins. *Nucleic Acids Res.* **35**, W512–W521 (2007).
- Simon, A. J. et al. Supercharging enables organized assembly of synthetic biomolecules. *Nat. Chem.* **11**, 204–212 (2019).
- Yuan, Y., Tam, M. F., Simplaceanu, V. & Ho, C. New look at hemoglobin allostery. *Chem. Rev.* **115**, 1702–1724 (2015).
- 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. U.S.A.* **109**, 12968 (2012).
- Engel, A. & Müller, D. J. Observing single biomolecules at work with the atomic force microscope. *Nat. Struct. Mol. Biol.* **7**, 715–718 (2000).
- Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. Energetic landscape of  $\alpha$ -lytic protease optimizes longevity through kinetic stability. *Nature* **415**, 343–346 (2002).
- Sleytr, U. B., Schuster, B., Egelseer, E. M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **38**, 823–864 (2014).
- Koga, N. et al. Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
- Churchfield, L. A. & Tezcan, F. A. Design and construction of functional supramolecular metalloprotein assemblies. *Acc. Chem. Res.* **52**, 345–355 (2019).
- Ringler, P. & Schulz, G. E. Self-assembly of proteins into designed networks. *Science* **302**, 106–109 (2003).
- Bale, J. B. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
- Hoersch, D., Roh, S.-H., Chiu, W. & Kortemme, T. Reprogramming an ATP-driven protein machine into a light-gated nanocage. *Nat. Nanotechnol.* **8**, 928–932 (2013).
- Pyles, H., Zhang, S., De Yoreo, J. J. & Baker, D. Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **571**, 251–256 (2019).
- Alberstein, R., Suzuki, Y., Paesani, F. & Tezcan, F. A. Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly. *Nat. Chem.* **10**, 732–739 (2018).

24. Joshi, T. et al. Local electronic structure of molecular heterojunctions in a single-layer 2D covalent organic framework. *Adv. Mater.* **31**, 1805941 (2019).
25. Steiger, D., Ahlbrandt, C. & Glaser, R. Crystal potential formula for the calculation of crystal lattice sums. *J. Phys. Chem. B* **102**, 4257–4260 (1998).
26. Tang, Z., Zhang, Z., Wang, Y., Glotzer, S. C. & Kotov, N. A. Self-assembly of CdTe nanocrystals into free-floating sheets. *Science* **314**, 274–278 (2006).
27. Talapin, D. V., Shevchenko, E. V., Murray, C. B., Titov, A. V. & Král, P. Dipole–dipole interactions in nanoparticle superlattices. *Nano Lett.* **7**, 1213–1219 (2007).
28. Sinyagin, A. Y., Belov, A., Tang, Z. & Kotov, N. A. Monte Carlo computer simulation of chain formation from nanoparticles. *J. Phys. Chem. B* **110**, 7500–7507 (2006).
29. Phillies, G. D. J. Excess chemical potential of dilute solutions of spherical polyelectrolytes. *J. Chem. Phys.* **60**, 2721–2731 (1974).
30. Wukovitz, S. W. & Yeates, T. O. Why protein crystals favour some space-groups over others. *Nat. Struct. Mol. Biol.* **2**, 1062–1067 (1995).
31. Chruszcz, M. et al. Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? *Protein Sci.* **17**, 623–632 (2008).
32. Bourq, I. C., Lee, S. S., Fenter, P. & Tournassat, C. Stern layer structure and energetics at mica–water interfaces. *J. Phys. Chem. C* **121**, 9402–9412 (2017).
33. Ricci, M., Spijker, P. & Voitchovsky, K. Water-induced correlation between single ions imaged at the solid–liquid interface. *Nat. Commun.* **5**, 4400 (2014).
34. Martin-Jimenez, D. & Garcia, R. Identification of single adsorbed cations on mica–liquid interfaces by 3D force microscopy. *J. Phys. Chem. Lett.* **8**, 5707–5711 (2017).
35. Ido, S. et al. Immunoactive two-dimensional self-assembly of monoclonal antibodies in aqueous solution revealed by atomic force microscopy. *Nat. Mater.* **13**, 264–270 (2014).
36. Pinteá, S. et al. Solid–liquid interface structure of muscovite mica in CsCl and RbBr solutions. *Langmuir* **32**, 12955–12965 (2016).
37. Woo, S. & Rothemund, P. W. K. Self-assembly of two-dimensional DNA origami lattices using cation-controlled surface diffusion. *Nat. Commun.* **5**, 4889 (2014).
38. Tao, J. et al. Control of calcium phosphate nucleation and transformation through interactions of amelatin and amelogenin exhibits the “goldilocks effect”. *Cryst. Growth Des.* **18**, 7391–7400 (2018).
39. De Yoreo, J. J. et al. Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* **349**, aaa6760 (2015).
40. Vekilov, P. G. Two-step mechanism for the nucleation of crystals from solution. *J. Cryst. Growth* **275**, 65–76 (2005).
41. Whitelam, S. Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions. *Phys. Rev. Lett.* **105**, 088102 (2010).
42. Dandekar, P., Kuvadia, Z. B. & Doherty, M. F. Engineering crystal morphology. *Annu. Rev. Mater. Res.* **43**, 359–386 (2013).
43. Chung, H. & Diao, Y. Polymorphism as an emerging design strategy for high performance organic electronics. *J. Mater. Chem. C* **4**, 3915–3933 (2016).
44. Hillenbrand, J. & Sessler, G. M. Piezoelectricity in cellular electret films. *IEEE Trans. Dielectr. Electr. Insul.* **7**, 537–542 (2000).
45. Moll, D. et al. S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl Acad. Sci. U.S.A.* **99**, 14646–14651 (2002).
46. Mann, S. et al. Crystallization at Inorganic–organic Interfaces: Biominerals and Biomimetic Synthesis. *Science* **261**, 1286–1292 (1993).
47. Hinchet, R., Khan, U., Falconi, C. & Kim, S.-W. Piezoelectric properties in two-dimensional materials: Simulations and experiments. *Mater. Today* **21**, 611–630 (2018).
48. Haxton, T. K. & Whitelam, S. Do hierarchical structures assemble best via hierarchical pathways? *Soft Matter* **9**, 6851–6861 (2013).
49. Kotov, N. A. Inorganic nanoparticles as protein mimics. *Science* **330**, 188–189 (2010).
50. Glotzer, S. C., Solomon, M. J. & Kotov, N. A. Self-assembly: from nanoscale to microscale colloids. *AIChE J.* **50**, 2978–2985 (2004).
51. Dong, M., Husale, S. & Sahin, O. Determination of protein structural flexibility by microsecond force spectroscopy. *Nat. Nanotech.* **4**, 514–517 (2009).
52. Pfreundschuh, M., Harder, D., Ucurum, Z., Fotiadis, D. & Müller, D. J. Detecting ligand-binding events and free energy landscape while imaging membrane receptors at subnanometer resolution. *Nano. Lett.* **17**, 3261–3269 (2017).
53. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
54. Schrödinger, L. L. C. *The PyMOL Molecular Graphics System, Version 1.3*. <https://pymol.org/2/support.html> (2010).
55. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
56. Joung, I. S. & Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
57. Kumar, S., Rosenber, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
58. Zhu, F. & Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **33**, 453–465 (2012).
59. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
60. Heinz, H., Lin, T.-J., Kishore Mishra, R. & Emami, F. S. Thermodynamically consistent force fields for the assembly of inorganic, organic, and biological nanostructures: The INTERFACE force field. *Langmuir* **29**, 1754–1765 (2013).
61. Kohlmeier, A. & Vermaas, J. TopoTools. <https://doi.org/10.5281/zenodo.545655> (2017).
62. Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
63. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
64. Hardy, D. J. et al. Multilevel summation method for electrostatic force evaluation. *J. Chem. Theory Comput.* **11**, 766–779 (2015).
65. Katz, B. A. & Kossiakoff, A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J. Biol. Chem.* **261**, 15480–15485 (1986).
66. Aksimentiev, A. & Schulten, K. Imaging  $\alpha$ -hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.* **88**, 3745–3761 (2005).
67. Bratko, D., Striolo, A., Wu, J. Z., Blanch, H. W. & Prausnitz, J. M. Orientation-averaged pair potentials between dipolar proteins or colloids. *J. Phys. Chem. B* **106**, 2714–2720 (2002).

## Acknowledgements

We thank Rohit Subramanian, Chris Mundy, Jiarun Zhou, and Mike Gilson for helpful discussions. This material is based upon work supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (BES), as part of the Energy Frontier Research Centers program: CSSAS–The Center for the Science of Synthesis Across Scales—under Award Number DE-SC0019288. Development of imaging protocols was supported by the Laboratory Directed Research and Development Office through the Materials Synthesis and Simulations Across Scales Initiative. Protein synthesis and characterization at UCSD was supported by the National Science Foundation (Division of Materials Research; DMR-1602537 to F.A.T.). Simulations were supported by the DOE-BES (Division of Materials Sciences, Biomolecular Materials, Award DE-SC0003844 to F.A.T.). R.G.A. was partially supported by a UCSD Distinguished Graduate Student Fellowship. AFM experiments were performed at the Department of Energy’s Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for Department of Energy by Battelle under Contract No. DE-AC05-76RL01830.

## Author contributions

S.Z. developed the mica-templated self-assembly protocol and performed AFM experiments. S.Z. and J.J.D.Y. analysed the AFM data. R.G.A. prepared the protein constructs and the solution-grown crystals, provided experimental guidance, and performed the simulations and theoretical analyses. F.A.T. and J.J.D.Y. conceived and oversaw the project. R.G.A., F.A.T., S.Z., and J.J.D.Y. 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/s41467-020-17562-1>.

Correspondence and requests for materials should be addressed to J.J.D.Y. or F.A.T.

Peer review information *Nature Communications* thanks Todd Yeates and other, anonymous, reviewers for their contributions to the peer review of this work.

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

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



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

© The Author(s) 2020