Title: De novo design of metal-oxide templating proteins

Authors: Amijai Saragovi*[1,2], Harley Pyles*[1,2], Timothy F. Huddy[1,2], Wenzhao Dai[1,4], Xinqi Li[5,6], Andrew J. Borst[1,2], Nikita Hanikel[1,2], Alexis Courbet[1,2,3,7], Paul Kwon[1,2], Fátima A. Dávila-Hernández[1,2], Ryan Kibler[1,2],Dionne K. Vafeados[1,2], Aza Allen[1,2], Kenneth D. Carr[1,2], Asim K. Bera[1,2], Alex Kang[1,2], Evans Brackenbrough[1,2], Sakshi Schmid[5,6], Yuna Bae[5,6], Lance Stewart[1,2], Shuai Zhang[5,6], James De Yoreo[5,6], David Baker[1,2,3]
* equal contributions
Corresponding author: amijais@uw.edu, dabaker@uw.edu

Affiliations:
1. Institute for Protein Design, University of Washington, Seattle, WA, 98105, USA
2. Department of Biochemistry, University of Washington, Seattle, WA, 98195, USA
3. Howard Hughes Medical Institute, University of Washington, Seattle, WA, 98105, USA
4. Department of Chemistry, University of Zürich, Winterthurerstrasse, 190, 8057, Zürich, Switzerland
5. Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
6. Department of Materials Science and Engineering, University of Washington, Seattle, WA, USA
7.Centre de Biologie Structurale, University of Montpellier, CNRS, INSERM. Montpellier, France

Abstract: Protein design now enables the precise arrangement of atoms on the nanometer length scales of inorganic crystal nuclei, opening up the possibility of templating the growth of metal oxides including semiconductors. We designed proteins presenting regularly repeating interfaces containing functional groups that organize ions and water molecules, and characterized their ability to bind to and template metal oxides. Two interfaces promoted the growth of hematite under conditions that otherwise resulted in the formation of magnetite. Three interfaces promoted ZnO nucleation under conditions where traditional ZnO-binding peptides and control proteins were ineffective. Designed cyclic assemblies with these ZnO nucleating interfaces lining interior cavities promoted ZnO growth within the cavity. CryoEM analysis of a designed octahedral nanocage revealed atomic density likely corresponding to the growing ZnO directly adjacent to the designed nucleation promoting interfaces. These findings demonstrate that designed proteins can direct the formation of metal oxides not observed in biological systems, opening the door to protein-semiconductor hybrid materials.

One Sentence Summary: We describe the design of structured protein interfaces that bind to, promote, and localize the growth of zinc oxide and hematite, inorganic materials which are not found in biological systems.

Main Text:

Introduction

Designing protein metal-oxide hybrid materials is challenging as detailed atomic structures of protein-inorganic interfaces are rare[1] and the vast majority of the universe of possible protein-inorganic hybrids are not explored by nature, including those containing inorganic semiconductors. Previous work has screened phage-display libraries of random amino acid sequences to identify peptides, typically 8 or 12 amino acids long, including ones which promote nucleation of semiconductors when densely displayed on the surface of phage[2–6].

Although they template growth by modifying the chemistry of existing surfaces, these peptides have not been shown to promote the heterogeneous nucleation of inorganic crystals when dispersed in bulk solution as self-contained molecular templates.

Larger structurally and chemically defined protein interfaces could in principle bind these minerals more strongly and specifically, and thereby exert greater control over nucleation and growth. However, the sequence space of the longer (> 50 amino acid) polypeptides needed to create such proteins is too large to adequately explore with random libraries, necessitating a more directed approach for the design of protein-inorganic interfaces. Protein design enables the construction of molecular scaffolds with a wide variety of surface geometries[7,8], and to pattern diverse chemistries across these surfaces. This enables specific arrangements of chemical moieties that promote the binding to an inorganic phase to be identified and interrogated. We reasoned that proteins which strongly bind an inorganic crystal could lower the interfacial free energy of nascent nuclei and template heterogenous growth of the material from supersaturated solution, and set out to design proteins presenting inorganic crystal -binding interfaces.

The optimal surface for interacting with inorganic lattices is unknown, and hence we explored three design hypotheses and sought to generate proteins presenting 1) periodic arrays of charges[9,10], 2) residues partially coordinating metal ions[1], and 3) ordered hydration layers promoting adsorption[11,12](**Figure 1A**). As the optimal geometry for protein-templated inorganic nucleation is unknown, we set out to explore this by presenting these interfaces on designed scaffolds with different topologies, periodicity, and curvature. These scaffolds included primarily alpha helical designed helical repeat (DHR) proteins[8,13], designed alpha-beta proteins[14–16](**Figure 1B, C**), and proteins containing flat beta-sheet surfaces including designed idealized beta-solenoid (DBS) repeat proteins (see Methods), a native antifreeze protein (RiAFP) and an idealized penta-peptide repeat (PPR) fold[17,18] (**Figure 1D**). We selected two target materials that can be synthesized in aqueous solutions but are not found in native bio-inorganic hybrid materials: zinc oxide (ZnO) and hematite (alpha-Fe2O3)[19]. Zinc oxide has diverse applications in emerging technologies[20], energy scavenging nanosystems[21], thin film transistors[22], field emitters[23] and can be detected using photoluminescence (PL) spectroscopy[24]. Hematite (alpha-Fe2O3) is a hydrogen generating photocatalyst[25] that can be distinguished from other iron oxides by color[26].

Results

We defined sets of interface amino-acid compositions based on the three above design hypotheses (**Figure 1A, Supplementary Figure 1**), histidines and cysteines were included to test metal coordination, threonines rich surfaces were used to mimic water organizing motifs in ice-binding proteins, and acidic and basic residues were used to achieve periodic arrays of charges. We placed these potential mineral-interacting interfaces on each scaffold (**Figure 1B-D**) with each composition in both periodic and aperiodic arrangements (**Figure 1E-Q**), resulting in 7×10^3 designed proteins (**Figure 1R**). Outside of the intended binding surface the sequence of the proteins remained fixed, and the original scaffold sequences were also included in the library.

We developed a yeast display flow cytometry (FC) screen that relies on the light scattering properties of NPs to identify and enrich yeast cells with increased binding propensity to ZnO or hematite nanoparticles (**Figure 2A, B**; **Supplementary Figures 2-4**). Next generation sequencing (NGS) was used to compare the abundance of designs in sixteen subpopulations of the library before and after sorting and identify enriched sequences (**Figure 2C; Supplementary Figure 5-6**; see Methods). Individual enriched clones were retested by FC on yeast (**Figure 2C; Supplementary Figure 7-8; Supplementary Table 1-6**). In cases where mutated, truncated, or chimeric sequences were abundant in the sorted subpools, we tested the observed sequences as well as the design with the highest identity, and we also tested 'fiber' versions of the repeat protein designs where the sequence capping features on the terminal repeat were removed to promote head-to-tail interactions **(Supplementary table 4;** see Methods). The identified ZnO binders included eight DBSs, two DHRs, two alpha-beta topologies, and a surface redesign of RiAFP. Five of these surfaces contained histidine and cysteine residues which can coordinate zinc ions[27], six contained large hydrophobic patches, five contained threonine motifs observed in ice-binding proteins[11], and eight interfaces contained periodic changes (**Figure 2D-H; Supplementary Figure 9; Supplementary Table 1, 3, 5**). We also identified hematite binders: two DHRs containing histidine-rich surfaces, two DHRs with charge-rich surfaces, and three beta-solenoids containing hydrophobic surface patches (**Figure 2I; Supplementary Figure 10; Supplementary Table 2, 4, 6**).

Biochemical characterization of metal-oxide binding proteins

We expressed these proteins in *E. coli* with 6xHis tags and purified them with immobilized metal affinity chromatography (IMAC; **Figure 2D-F, I, Supplementary Figures 11-12; Supplementary Table 7-8**). We selected five proteins (Z0, Z1, Z3, Z4, and Z5) which had substantial FC binding signals to ZnO NP and expressed well in *E. coli* as soluble proteins for further characterization. Z3 and Z5 were monodisperse by size exclusion chromatography (SEC), while Z0, Z1, and Z4 appeared to form large assemblies (**Supplementary Figure 11)**. Z0, Z3, Z4, and Z5 produced circular dichroism (CD) spectra consistent with their predicted secondary structure while Z1 appeared disordered (**Supplementary Figure 11)**. Z0 formed long fibers and bundles of fibers visible with negative-stain transmission electron microscopy (ns-TEM), presumably through strand pairing head-to-tail interactions (**Supplementary Figure 13A**). To regularize these interactions, we designed and produced a version without these capping features, Z0-fiber, and used it in further analysis of this design (**Supplementary Figure 13B**). In a pull-down assay followed by SDS-PAGE, all of the purified ZnO designs bound ZnO NPs and did not bind to anatase, rutile, or hematite NPs, while bovine serum albumin (BSA) and lysozyme controls showed no binding (**Supplementary Figure 14**). Atomic force microscopy (AFM) of the zinc-terminated (001) surface incubated with the proteins showed binders had similar (Z3, Z4) or greater (Z0-fiber) coverage in respect to controls (**Supplementary Figure 15**). The three designs which produced the highest yields of purified protein, Z0-fiber, Z3, and Z4, were selected for further study.

We also selected two hematite binders, H7 and H9, which expressed as soluble protein in *E. coli* and demonstrated robust hematite NP binding signals when displayed as single clones on yeast (**Figure 2I, Supplementary Figure 8, 12; Supplementary Table 2).** In order to increase the yield and improve the monodispersity of these proteins, we mutated non-interfacial residues to negative residues, resulting in H7-neg and H9-neg (**Supplementary Figure 16; Supplementary Table 6**). No protein associated with hematite NP was detected in a SDS-page pull-down assay (not shown), however hematite nanoparticles incubated with H7-neg showed decreased flocculation compared to solutions without protein or BSA (**Supplementary Figure 17**).

We next tested if the designs coordinate metal ions. To eliminate the potential effects of His-tags, the designs were produced with strep tags instead and purified with Strep-Tactin beads (**Supplementary Table 7-8;** see Methods). Using isothermal titration calorimetry (ITC) Z0-fiber was found to bind Zn^{2+} exothermically with a Kd of 4.1 ± 1.4 μM and Z3 was found to bind Zn^{2+} endothermically with 16 ± 5.0 (**Figure 2G**). A ratiometric dye (Mag-Fura-2) assay indicated that Z0-fiber binds Zn^{2+} ions with a best fit model indicating six binding sites, three with 3.8 ± 0.2 nM affinity and three with 60 ± 6 nM affinity (**Figure 2D**), and Z3 showed three binding sites with 200 ± 60 nM affinity (**Figure 2E**; the difference in measured affinities is likely due to the lower pH for the ITC measurements (see Methods)). Z0-fiber interfacial amino acid substitutions indicated that the two rows of repeating histidine residues are critical for Zn^{2+} ion binding (**Figure 2D**). In the Z3 interface, mutating glutamate, cysteine, and histidine residues reduced Zn^{2+} ion binding, while mutation of basic residues did not (**Figure 2E**). Mutating histidines in the interfaces of both Z0-fiber and Z3 also resulted in reduced ZnO NP binding when displayed on yeast (**Supplementary Figure 18**). No heat signal was detected upon titration of Zn^{2+} ions to Z4 (**Figure 2G**), and the Kd was estimated to lie above the quantification limit of the dye competition assay using Mag-Fura-2 (> 1 uM; **Supplementary Figure 19**). Z4 contains a threonine motif found in alpha-helical ice-binding proteins (**Figure 2F, H**), and four other enriched proteins (Z8, Z9, Z12, and Z14) contain another threonine motif found in beta-solenoid ice-binding proteins (**Figure 2H, Supplementary Figure 9**)[11]. A ratiometric dye assay using Mag-Fura-2 showed that both H7 and H9 coordinate Co^{2+} ions, a common analog for Fe^{2+} (Figure 2I). Co^{2+} is typically used to assess affinity since Fe^{2+} often oxidizes to Fe^{3+}, forming ferrihydrite. H7 exhibited strong nanomolar binding affinity, while H9 only weakly coordinated Co^{2+} with micromolar affinity (**Figure 2I)**.

<u>Assessment of metal oxide templating activity</u>

We next assayed the ability of the purified ZnO-binding proteins to promote the growth of ZnO NPs by incubating them in supersaturated solutions (3 mM $Zn(NO_3)_2$, 50 mM NaCl, 100 mM HEPES, pH 8.2) which allow heterogenous growth of ZnO, but not homogeneous growth (**Figure 3A**). We measured the photoluminescence (PL) of ZnO NPs to assess protein-induced growth in these conditions and used solutions seeded with 13 μg/mL of commercial ZnO NPs nanoparticles as a positive control, an ideal template. After 2 hours, neat solutions and solutions containing lysozyme showed no PL, while solutions seeded with 0.1 mg/mL of Z0-fiber, Z3, or Z4, had PL spectra consistent with ZnO NPs (**Figure 3B**)[28].

X-ray powder diffraction of the material grown in the presence of Z0-fiber, Z3, and Z4 produced diffraction patterns indicative of ZnO (**Figure 3C**). Tracking PL over time showed the increase induced by the ZnO NPs control began immediately, while the signal increased in solutions with the proteins began after a delay of 30 to 60 minutes. Additional control proteins bovine serum albumin (BSA), DNas I, and DHR10-mica6[9] had no observable effect on the level of the PL signal (**Supplementary Figure 20**). Z0-fiber exhibited inconsistent behavior among replicates, showing growth signals in only two out of six replicates, whereas all replicates containing Z3, Z4, and ZnO NP showed growth (**Supplementary Figure 21**).

To confirm the interfacial motifs we identified were responsible for the ZnO templating activity of the proteins, we tested the activity of the motifs when hosted by different scaffold proteins. Reasoning that the aggregation prone behavior of Z0-fiber may be impeding its function, we extracted the putative functional motifs and designed topologies with RFdiffusion[29] and MPNN[30] to host the Z0 motifs in alpha-beta proteins (**Figure 3E**). We expressed eleven of these 'zab' proteins in *E. coli* and selected four, zab3, zab5, zab10, and zab11, with CD spectra and SEC traces consistent with the models (**Supplementary Figure 22**). Two of the redesigns, zab3 and zab10, showed a strong signal in the ZnO NP growth assay (**Figure 3H**). Having observed threonines in the Z4 interface motifs (**Figure 3F**) which resemble a helical ice-binding motif (**Figure 2H**), we tested a designed ice-binding twistless-helical repeat protein (b-iTHR-201) which presents a larger array of this motif[31] and found it also has templating activity (**Figure 3I**; to account for potential variability, six replicates were run for each of these conditions (**Figure 3J**). Four previously reported ZnO binding peptides[3–6], fused to Small Ubiquitin-like Modifier (SUMO) to aid solubility and expression, showed no activity (**Figure 3J**; **Supplementary Figure 23, Supplementary Table 9**). For replicates that triggered ZnO growth, we compared the times of nucleation which we defined as the halfway point between the initial and maximum PL signal. Five design templates promoted an increase in photoluminescence at times greater or equal to the control NPs, while the increase promoted by zab3 occurred substantially earlier.(**Figure 3K**).

We examined if the H7 and H9 interfaces could direct the formation of hematite, the specific material they were selected for, over the variety of other iron oxide polymorphs Operating under anaerobic growth conditions in a magnetic field which typically yield magnetite (**Supplementary Figure 24**), we found that supersaturated solutions containing H7 were more red in color after 20 minutes incubation, consistent with the presence of hematite, compared to solutions containing H9, lysozyme, no additive, or hematite or magnetite NPs (**Figure 3L, M; Supplementary Figure 25; Supplementary Video 1**). The reaction was allowed to continue for 120 minutes, however the color was only compared within the first 20 minutes because subsequently the magnetite material began to be pulled out of solution by the external magnetic field. (**Figure 3L, Supplementary Video 1**). Raman spectroscopy of the reaction products indicated hematite was produced in solutions containing H7-neg and H9-neg while magnetite was produced in solutions without protein or with lysozyme (**Figure 3N; Supplementary Figure 25)**. We next examined the effect of a hematite binding protein on oxidation at the iron-water interface in aerobic conditions. After a 2-hour incubation on polished iron surfaces, solutions containing H7-neg accumulated less

rust than samples incubated without protein, an effect not seen for a lysozyme control (**Figure 3O**).

Confinement of zinc oxide mineralization

We examined the effect of confinement on the rate of nucleation by lining the interior cavities of oligomers with the ZnO nucleating Z4 interface. We used RFdiffusion[29] and MPNN[30] to design cyclic oligomers that host motifs from the Z4 interface within the interior of rings (**Figure 4A-D**, see Methods). We experimentally tested twenty-three oligomers and selected three designs forming monodisperse species of the expected size, Z4-C3i, Z4-C3ii, and Z4-C6, for further characterization (**Supplementary Figure 27**). The Z4-C3ii-XL trimer was derived from Z4-C3ii by adding another repeat of the Z4 interface to each monomer. The structures of Z4-C3i and Z4-C3ii were verified with X-ray crystallography (**Figure 4A, B**) and the structures of Z4-C6 and Z4-C3ii XL were verified with ns-TEM (**Figure 4C, D**). All four of these oligomer designs retained the templating activity of the Z4 monomer, while knockouts versions of Z4-C3i, Z4-C3ii, and Z4-C6 in which the threonine residues in the Z4 interfaces were mutated to glutamate (**Figure 4E)**, which disrupted the ZnO binding interface but not the assembly of the oligomer (**Supplementary Figure 28**), had no templating activity (**Figure 4F**). Cryo-EM of ZnO NPs grown with Z4-C6 showed both empty rings matching the designed hexamer and solid heterogeneous NP of the same diameter. (**Supplementary Figure 29**).

We next incorporated the ZnO templating interface into larger protein assemblies. To the Z4-C3i we incorporated a homodimeric protein-protein interface with an orientation that generates a 24 subunit octahedral nanocage. (**Figure 4G**). Z4-tube was designed by adding tetrameric domains to the N- and C-terminus of an extended version of Z4 such that the ZnO interacting interface surrounds a rod-like cavity with a high aspect ratio (**Figure 4H**). These assemblies were expressed in E. coli and purified with strep tags and SEC (**Supplementary Figure 30**), and their structures confirmed with ns-TEM (**Figure 4G, H**). In the ZnO NP growth PL assay, Z4-cage promoted a large increase in PL, while Z4-tube showed a substantially lower increase (**Figure 4I)**. There was no clear relationship between the size of the cavity or number of interfaces and the magnitude or rate of ZnO NP nucleation (**Figure 4I, J**).

CryoEM characterization of protein-ZnO interface

We used Cryo-EM to investigate the structure of ZnO NP formed within the Z4-cage. While the majority of Z4-cage particles appeared empty in raw micrographs **(Figure 5A)**, Cryo-EM micrographs indicated stochastic NP growth originating from a subset of designed protein nanocages, with three distinct morphologies observed **(Figure 5A,B)**: first, within the cage ("filled"), second, growth extending outward ("blebbed"), and third, connecting sets of cages ("bridged") **(Figure 5B)**.

2D class averages of Z4-cage particles revealed secondary structure elements consistent with the design **(Figure 5C; Supplementary Figure 31)**, and 3D refinements with relaxed

octahedral symmetry produced a ~3.8 Å cryo-EM structure closely aligning with the design model **(Figure 5D, E; Supplementary Figure 31; Supplementary Table 10)**. Despite achieving a moderately high-resolution map, the asymmetric and stochastic growth of ZnO, combined with the imposition of the octahedral symmetry operator, hindered clear visualization and reliable analysis of ZnO density at this resolution. Consequently, a global C1 3D refinement was performed using a near-spherical mask, allowing for detailed and unbiased analysis of the pores along both the C3 and C4 axes. This approach ensured an overall averaged representation of the Z4-cage **(Supplementary Figure 32)**. The resulting ~4 Å asymmetric map was low-pass filtered to 8 Å to highlight lower-resolution features corresponding to observed stochastic ZnO nucleation and growth patterns **(Figure 5A,B; Supplementary Figure 32)**. This approach revealed additional non-protein density, likely ZnO, localized along the center of each C3 axis and proximal to the designed ZnO binding interface **(Figure 5F; Supplementary Figure 32; Supplementary Table 10)**. In contrast, no extra density was observed along the C4 axes (distant from the designed ZnO binding interface) at any contour level **(Supplementary Figure 32)**. To evaluate whether ZnO nucleation occurred uniformly across all C3 trimers, we next performed octahedral symmetry expansion followed by 3D classification for all Z4-C3i trimers within the Z4-cage. As expected, a subset of C3 trimers displayed ZnO density along the C3 axis, with several 3D classes appearing empty, which is consistent with both the empty particle morphologies and asymmetry growth patterns of ZnO observed in raw micrographs **(Figure 5A,G)**.

3D local refinement of the threefold-symmetry axis of the mineralized Z4-cage yielded a near-atomic resolution cryo-EM map, ranging from 2.5 Å resolution in the center of the C3 axis to ~3.5 Å at the interface between adjacent Z4-C3i trimers in the nanocage **(Figure 5H; Supplemental Fig. 31; Supplementary Table 10)**. At this resolution, density consistent with solvent organization was visible along the ice-binding motifs in the Z4 designed interface[11] **(Figure 5H-J)**. Sixty-one waters were modeled into this density along the interface (see Methods), although the precise identity of these ordered species remains uncertain at this resolution. These results are consistent with stochastic ZnO NP nucleation within the trimers promoted by atomic ordering at the designed interface, followed by further ZnO growth into heterogeneous morphologies **(Figure 5)**.

Discussion

We present an approach to de novo design and screen de novo protein-mineral interfaces and apply it to identify proteins which bind ZnO and hematite NPs. We identified seven surfaces that promote hematite binding (**Supplementary Figures 10**) and demonstrate that two proteins, H7-neg and H9-neg, coordinate ions (**Figure 2I**), direct the formation of hematite in conditions which otherwise produce magnetite (**Figure 3L-N**), and appear to reduce the rate of oxidation at iron-water interfaces (**Figure 3O**). We identified thirteen de novo ZnO binding protein interfaces (**Supplementary Figure 9**) and demonstrate that three of them, Z0-fiber, Z3, and Z4, promote ZnO growth (**Figure 3B-D**; **Supplementary Figure 9)**. Our results suggest these interfaces function by two distinct mechanisms. Firstly there is evidence for direct ion mediated interactions; Z0-fiber and Z3 coordinate Zn^{2+} ions and mutating metal coordinating residues reduces their affinity to both ions and ZnO NPs (**Figure 2D, E, G**; **Supplementary Figure 18**). Secondly, ZnO binding interfaces including Z4 contain

water-structuring threonine motifs observed in ice-binding proteins (**Figure 2F, H; Supplementary Figure 9**), suggesting they may adsorb to ZnO via ordered water layers[11,12,32]. Hydrophobic interfaces were also selected in our assay but these proteins could not be purified from *E. coli* and the mechanism of their binding is unclear. While periodic charges were present in interfaces alongside other features (**Figure 2**, **Supplementary Table 1-2**), no interfaces composed entirely of arrays of charged groups, like those seen in previous mineral-templating and mineral-binding designed proteins[9,10], were observed. ZnO NP binding was more predictive of nucleation activity than binding to bulk zinc-terminated ZnO (**Fig. 3J, Supplementary Figures 14, 15**).
Our designed metal-oxide interacting proteins are larger, more ordered, and better structurally understood than the short peptides that were previously identified with inorganic-binding assays[2]. The protein-ZnO interfaces template ZnO growth in supersaturated solutions under conditions where ZnO-binding peptides have no effect and maintain this function when grafted into new structures (**Figure 2 D-F, 3 E-K**). When lining the interior of protein assemblies, the Z4 interface promoted ZnO formation within the assembly (**Figure 4, Figure 5 F,G, Supplementary Figure 32**). CryoEM characterization of a designed 18 nm octahedral nanocage with the Z4 interface lining a triangular cavity along the three fold axis showed that NP nucleation occurs adjacent to this interface (**Figure 3 F, Figure 5 H,I,J**).

Nature has evolved proteins which control the formation of hybrid organic-inorganic materials such as bone and nacre with greater complexity than is accessible with conventional top-down manufacturing[19]. Modern protein design tools have made protein structures and protein-protein interactions readily programmable, but inorganic-templating protein interfaces are inadequately understood. Our high-throughput design and evaluation approach provides a route to overcoming this limitation that can be readily extended to a wide range of inorganic crystals, opening the door to a new world of designed hybrid materials not explored by nature.

Acknowledgments We acknowledge support from the Alexandria Venture Investments Translational Investigator Fund (GF116545), the Nan Fung Life Sciences Translational Investigator Fund (GF126554), and the Washington Research Foundation Innovation Fellows Program (GF136997).

This work was funded in part by the Bill & Melinda Gates Foundation (INV-043758, GR019486), the Open Philanthropy Project through the Improving Protein Design Fund (GF129460) and the Universal Flu Vaccine Fund (GF129461), and The Audacious Project at the Institute for Protein Design (PG117879, PG117878, PG117866). We also acknowledge support from the Alfred P. Sloan Foundation (G-2021-16899, GR017147) and Spark Therapeutics (GR017125).

Additionally, we thank the Defense Threat Reduction Agency (HDTRA1-19-1-0003, GR008757; HDTRA1-21-1-0038, GR018007) and the Howard Hughes Medical Institute

(GR020267) for their contributions. This research was also supported by the National Institutes of Health, including grants from the National Institute of Allergy and Infectious Diseases (P01AI167966, GR014929; R01AI160052, GR010198) and the National Institute on Aging (R01AG063845, GR009173).

Crystallographic data were collected at National Synchrotron Light Source II (NSLS-II) Beamline 17-ID-1 (AMX). The Center for BioMolecular Structure is primarily supported by the NIH-NIGMS through a Center Core P30 Grant (P30GM133893) and by the Department of Energy Office of Biological and Environmental Research (KP1607011). NSLS-II is a U.S. Department of Energy Office of Science User Facility operated under contract no. DE-SC0012704. This publication is based on data collected through the NECAT BAG proposal no. 311950.

H.P. was supported by the Open Philanthropy Project Improving Protein Design Fund. T.F.H. was supported by the DARPA MIMA Seedling Grant (HR00112420369) and the Breakthrough Fund 100 Nanometer Assemblies Program from The Audacious Project.

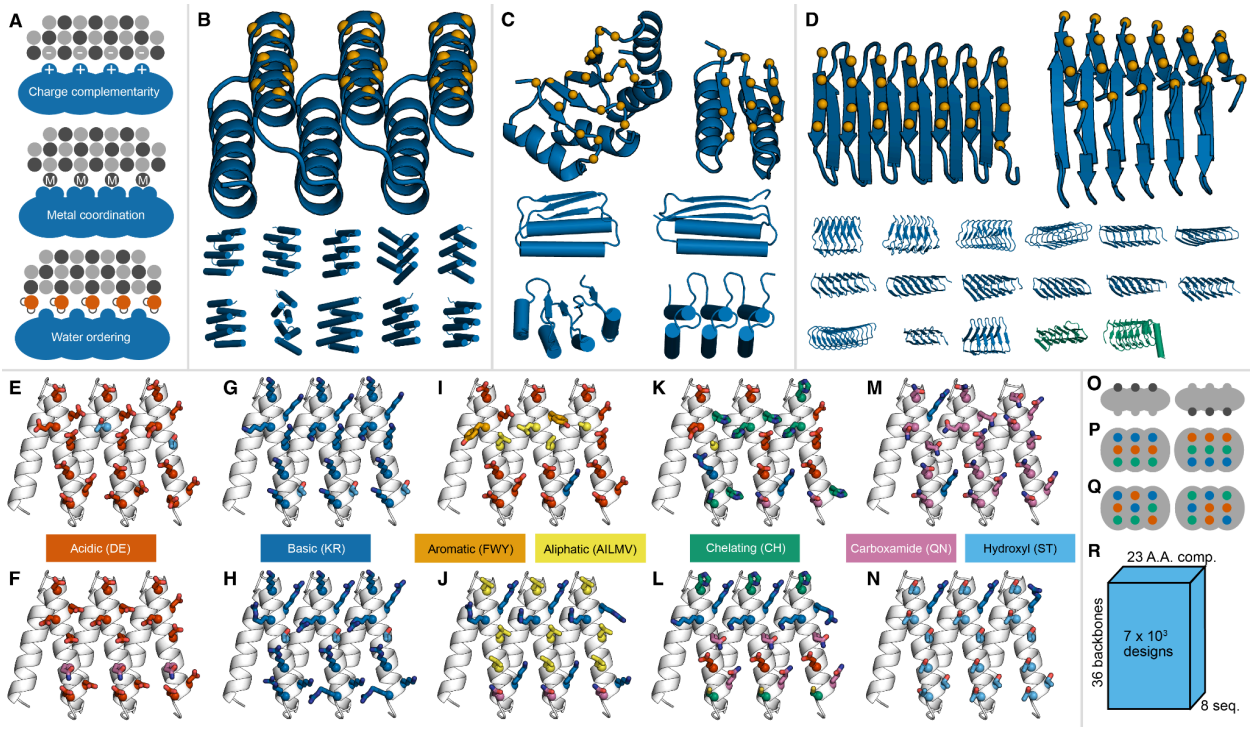

Fig. 1: Design principles for templating mineral growth. (**A**) Three hypothesized mechanisms of mineral binding: charge complementarity, metal coordination, and water ordering. (**B-D**) 36 protein backbones used in the library. Orange spheres show alpha carbons of residues in a designed interface. Scaffolds include 11 designed helical repeats (DHR) (**B**), 6 alpha-beta topologies (**C**), and 17 designed beta solenoids (DBS) and 2 native solenoids (in green) (**D**). (**E-N**) Example designed surfaces defined by 10 (out of 23) amino acid compositions applied to the DHR backbone DHR14. The colors of the amino acid side chains correspond to chemical categories identified in rectangles. Compositions include highly charged residues (**E-H**), charged and hydrophobic moieties (**I, J**), metal chelating residues (**K, L**), and non-charged hydrophobic residues (**M, N**). (**O-Q**) Illustration of the designs generated for each combination of scaffold and composition. (**O**) Two sets of interface residues are selected on opposing surfaces of protein. For a given surface and amino acid composition, two repeat surfaces (**P**), with the same residue in analogous positions in each repeat subunit, and two non-repeat surfaces (**Q**), wherein residues are scrambled among repeats. (**R**) 4 designed interfaces for each of 2 surfaces yields 8 sequences per combination of 36 protein backbones and 23 amino acid composition, resulting in a library of approximately 7×10^3 designs.

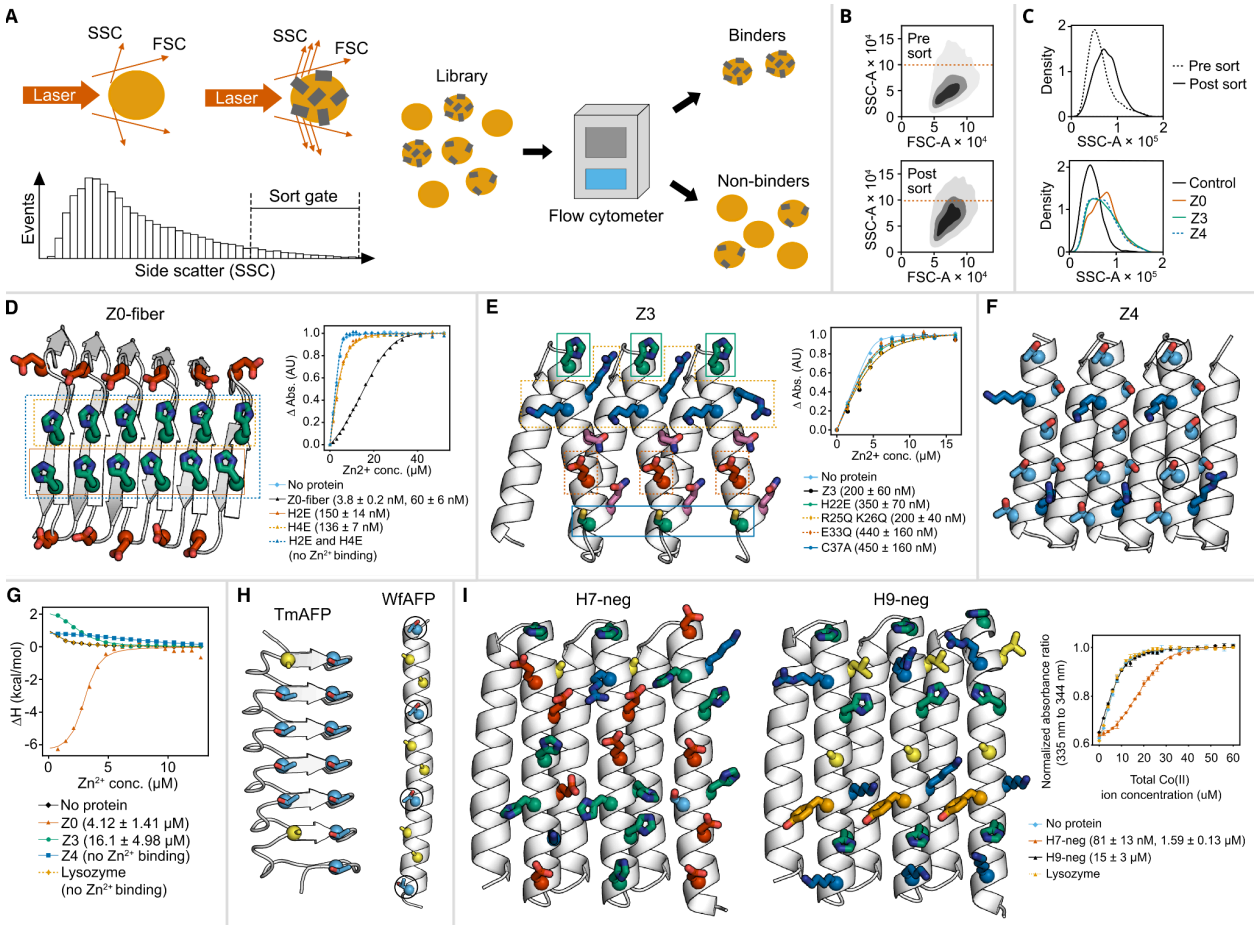

Fig. 2. Mineral binding screen. (**A**) Illustration of screening strategy. (**B**) SSC-A vs. FSC-A signals of yeast populations before and after ZnO enrichment sorting. Plots are kernel density plots estimates from 10,000 events. (**C**) SSC-A signals of a sample sublibrary before and after sorting, and yeast clones displaying three enriched designs vs. control yeast. (**D-E**) Z0-fiber and Z3 interfaces and mutational analysis of Zn^{+2} ion binding with ratiometric dye (Mag-Fura-2). Although mutants are named by the first residue that is mutated, all repeated instances of these residues, circumscribed by lines on the models, are mutated as well. (**F**) Model of the Z4 enriched design with circles highlighting threonines repeated in pattern resembling TXXXAXXXAXX motif in native antifreeze proteins (AFP). (**G**) Isothermal titration calorimetry analysis of Zn^{+2} against three designed proteins, lysozyme, and solutions without protein (protein concentration = 30 μM). (**H**) Example native ice-binding proteins: beta solenoid Tenebrio molitor AFP (TmAFP) containing threonine rich parallel beta-sheets (PDBid: 1ezg) and alpha-helical winter flounder AFP (WfAFP) with threonines in TXXXAXXXAXX motif circled (PDBid: 1wfa). (**I**) H7 and H9 interfaces and analysis of Co^{+2} ion binding with ratiometric dye (Mag-Fura-2). Colors of all amino-acid side chains correspond to the chemical categories defined in Figure 1.

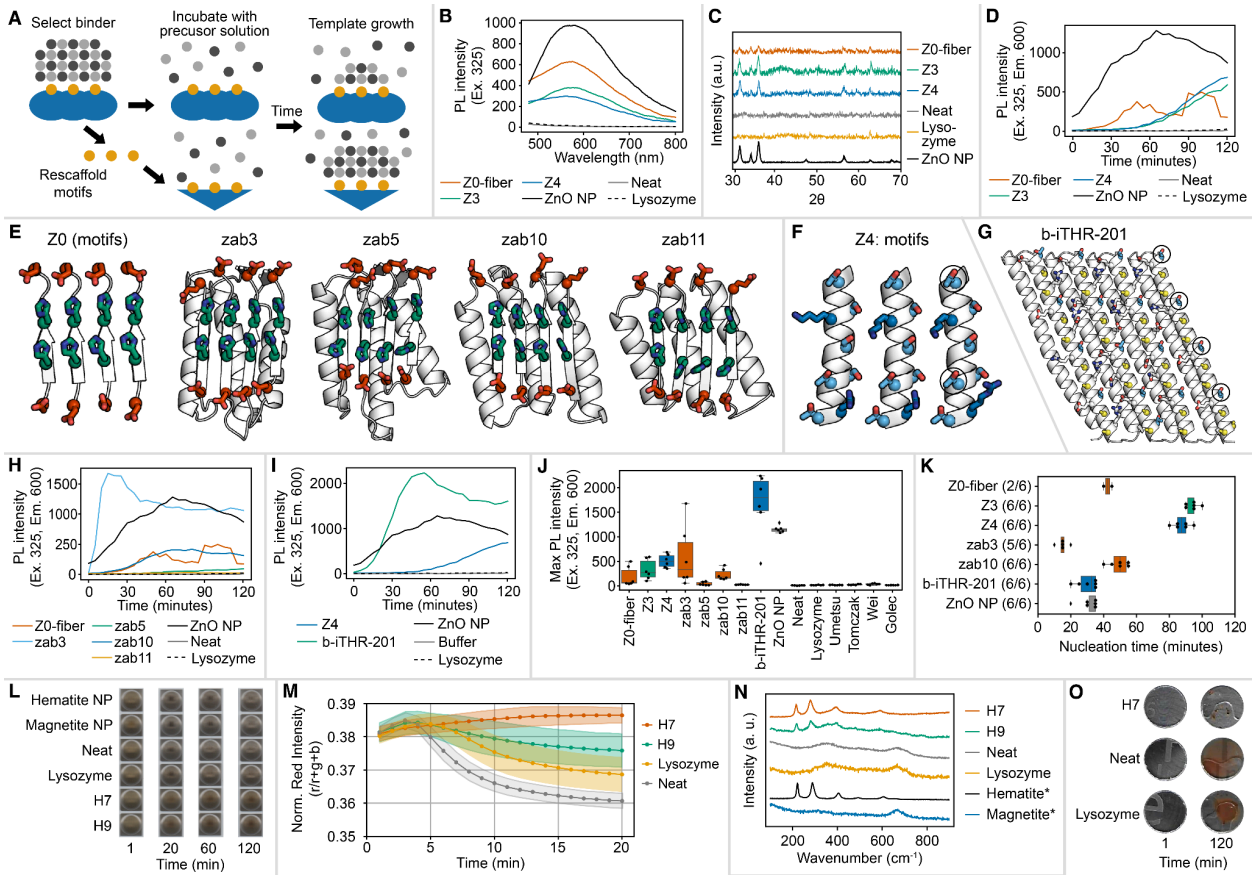

Fig. 3. Characterization of metal oxides nucleation by designed interfaces. (**A**) Schematic illustrating the nucleation assay and motif grafting methodology. (**B**) Emission spectra (Ex. = 325 nm) of solutions superstatured for ZnO seeded with selected ZnO-binding designs compared to ZnO NP, lysozyme and neat (no-additive) solution controls. (**C**) X-Ray diffraction analysis of synthesized products following two hour incubation with designed proteins (Z0-fiber, Z3, Z4), control protein (lysozyme), or solutions without additive (neat). A reference spectra taken of ZnO nanoparticles (ZnO NP) is also provided. (**D**) Plot of PL emission over time (Ex. 325, Em. 600) of supersaturated solutions containing selected designs, lysozyme, ZnO NP, or neat solutions. (**E**) Models of interface motifs taken from Z0 and four Z0-grafted alpha-beta (zab) designs. (**F**) Models of the Z4 interface motifs with threonine side chains resembling TXXXAXXXAXX motifs observed in native α-helical ice-binding proteins circled in one repeat. (**G**) Designed Ice-binding Twistless Helical Repeat (iTHR) with threonines in TXXXAXXXAXX motifs circled in one repeat. Colors of amino acid side chains in all panels correspond to the chemical categories defined in Figure 1. (**H-I**) PL emission over time (Ex. 325, Em. 600) of the zab redesigns vs. Z0-fiber and controls (**H**) and b-iTHR-201 vs. Z4 and controls (**I**). (**J-K**) Box plots showing 6 replicates of growth experiments using designed proteins, controls, and four previously reported ZnO-binding peptides, referred to by the last names of the first authors of the source publications (Umetsu, Tomczak, Wei, and Golec)[3–6]. (**J**) Average maximum PL intensity (n=6). (**K**) Nucleation times, defined as the halfway point between the initial and maximum PL signal, for the subset of samples that showed a PL signal of nucleation (maximum PL > 100), n values as indicated. All ZnO growth solutions contain 3 mM $ZnNO_3$, 50

mM NaCl, 100 mM HEPES pH 8.2, and contain 0.1 mg/mL protein or 13 μg/mL ZnO NP as indicated. (**L**) Digital scans taken of iron oxide growth reactions over time with protein (0.15 mg /mL) or NP ($1.6x10^{-5}$ g/ml Hematite NP; $1.5x10^{-5}$ g/ml Magnetite NP) additives as indicated. Solutions contain ammonium iron(II) sulfate hexahydrate 0.05 mM, Iron(III) nitrate nonahydrate 0.1 mM, KNO_3 50 mM, CHES 100 mM. Normalized red-intensity of the reactions shown in panel **L**. Values shown are the mean red intensity divided by the sum of the mean read, mean green, and mean blue intensity within the area of each well (n=6). (**N**) Raman spectroscopy of the products of the H7-neg, H9-neg, lysozyme, and neat reactions shown in panels **L** and **N,** with spectra taken of reference hematite and magnetite nanoparticles for comparison. (**O**) Images from iron substrate scans prior and following two hours incubation with 10ul samples containing H7-neg (0.6 mg/ml), Lysozyme (0.6 mg/ml) or buffer only.

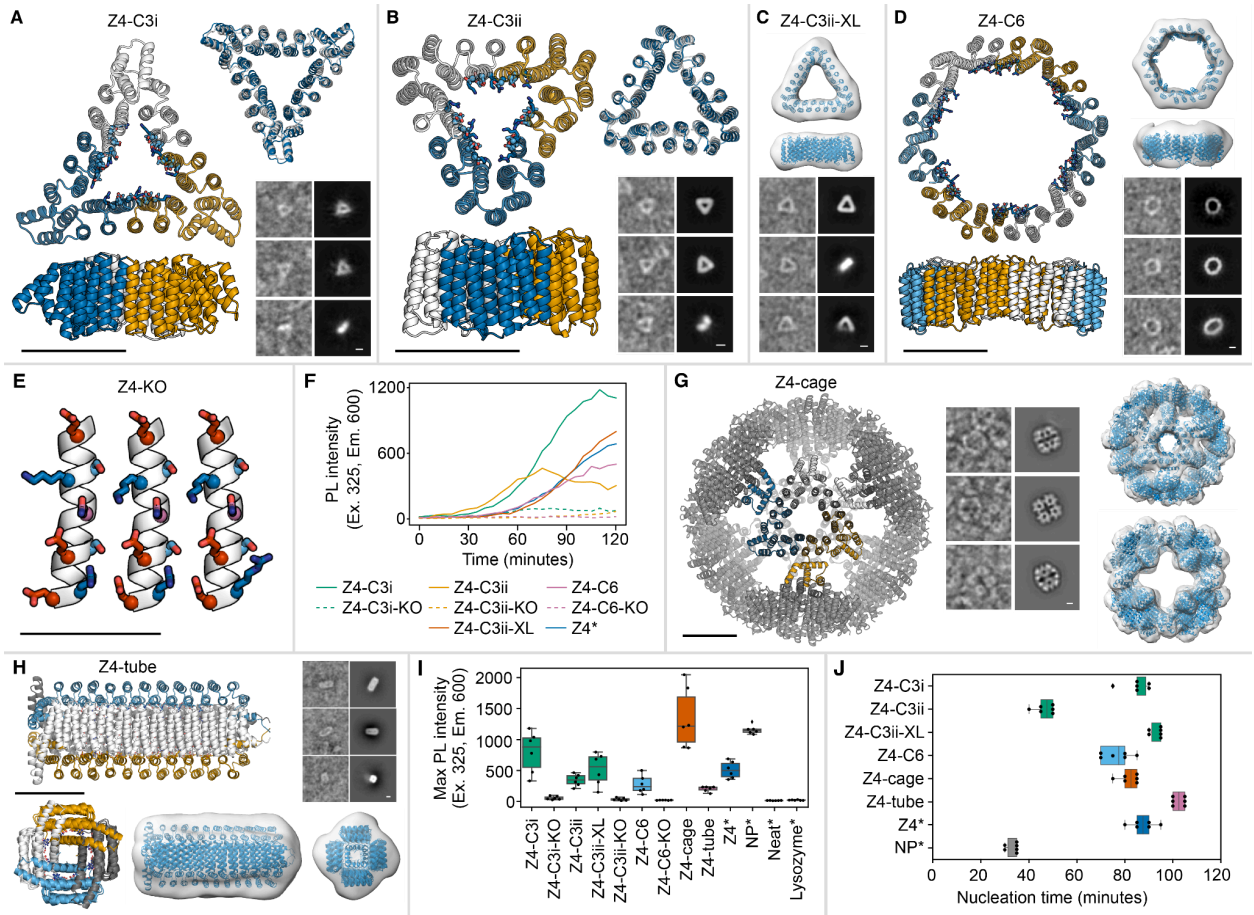

Fig. 4. ZnO nucleation by Z4 motif containing designed oligomers. (**A-D**) Cyclic oligomers with Z4 motifs lining interior cavities. (**A-B**) Two C3 designs with design models colored by subunit with Z4 motifs shown as sticks. Design model backbones shown in white superimposed with X-ray crystal structures in blue (PDB IDs: 9d92, 9cc4). Images at bottom right are representative ns-TEM picked particles (left) and 2D class averages (right). (**C**) Z4-C3i-XL model superimposed with ns-TEM 3D reconstruction, and examples of picked particles and 2D classes. (**D**) Z4-C6 colored by subunit with motifs shown as sticks, model superimposed with ns-TEM 3D reconstruction, and examples of picked particles and 2D classes. (**E**) Model of the mutated Z4 knock out (Z4-KO) interface with threonines replaced with glutamates. (**F**) Plot of PL emission over time (Ex. 325, Em. 600) of Z4-motif containing cyclic oligomers compared to oligomers with interface KO mutants and the Z4 monomer (*sample replicated from Figure 3). (**G**) Octahedral cage designed using Z4-C3i, ns-TEM picked particles and 2D class averages, and model superimposed with ns-TEM 3D reconstruction. (**H**) Z4-tube colored by subunit with Z4 motifs shown as sticks, ns-TEM picked particles, 2D-averages, and 3D reconstructions superimposed with models. (**I-J**) Effect of Z4-motif containing oligomers on ZnO nucleation compared to Z4 monomer, ZnO NP, and neat solutions (*samples are replicated from Figure 3). (**I**) Maximum PL intensity after nucleation assay (n=6). (**J**) Nucleation times, defined as the halfway point between the initial and maximum PL signal (n=6), for samples that showed a signal of nucleation (maximum PL > 100). Colors of amino acid side chains correspond to the chemical categories defined in Figure 1. All scale bars are 4 nm.

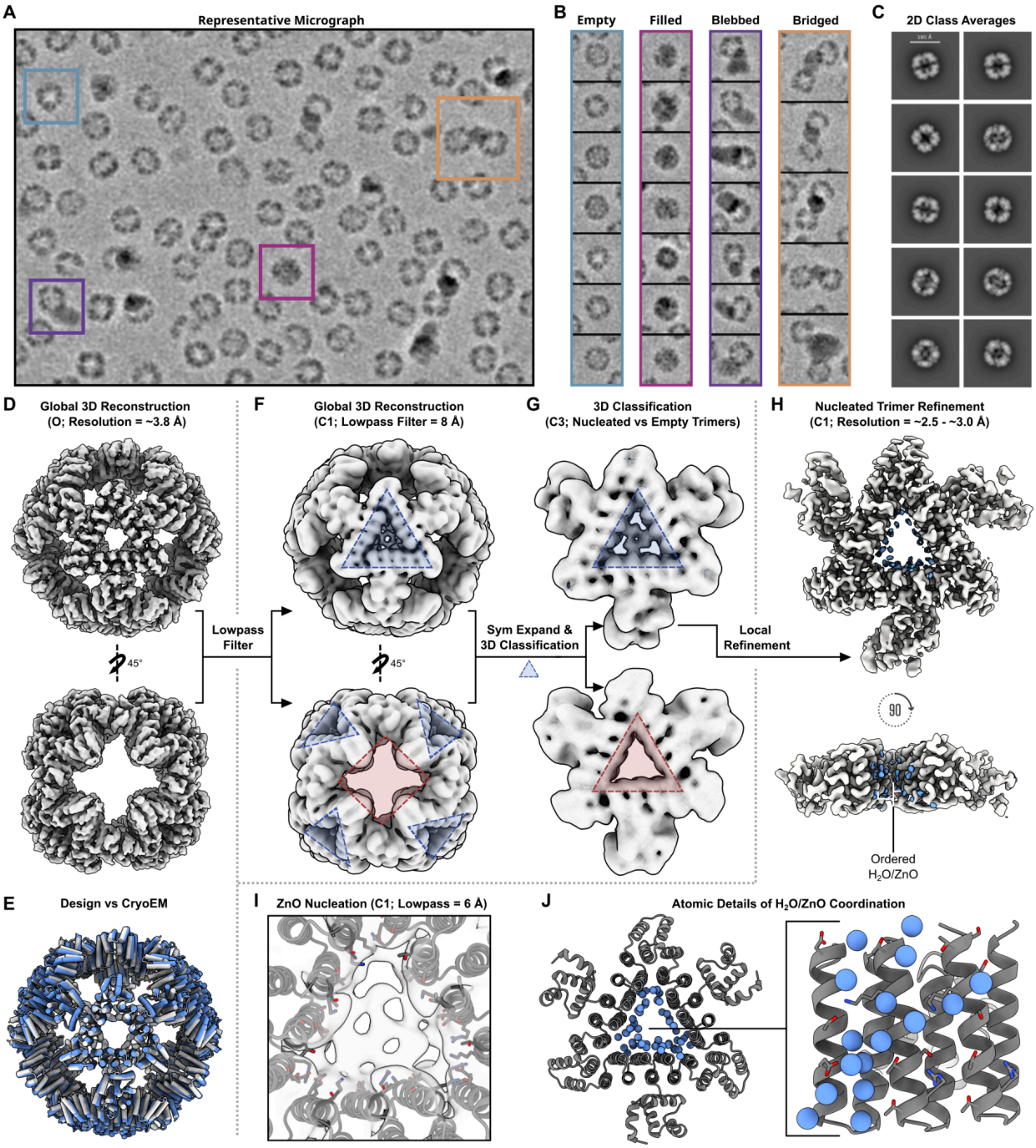

Fig. 5. Cryo-EM characterization of ZnO mineralized Z4-cage.
(A) Representative denoised micrograph of Z4-cage particles imaged in the presence of 3 mM $Zn(NO_3)_2$ for ~30 minutes prior to freezing. Highlighted are examples of "empty" (teal), "filled" (maroon), "blebbed" (purple), and "bridged" (orange) nanoparticle morphologies observed in a single micrograph. **(B)** Additional examples of each morphology observed in the dataset. **(C)** 2D class averages of Z4-cage. **(D)** 3.8 Å cryo-EM 3D reconstruction of Z4-cage refined with relaxed octahedral symmetry. **(E)** Comparison of the final built cryo-EM model of Z4-cage (grey) with

the computational design model (blue). **(F)** C1 reconstruction of Z4-cage, low-pass filtered to 8 Å, showing ZnO nucleation localized along the C3 symmetry axis (blue), and no ZnO density along the C4 axis (red). **(G)** Octahedral symmetry expansion and 3D classification of all Z4-C3i trimers from all Z4-cage nanoparticles show both nucleated (blue) and empty (red) trimers exist within the data. **(H)** Near-atomic resolution 3D reconstruction of nucleated Z4-C3i trimers in the Z4-cage nanoparticle, highlighting density from coordinated H_2O or ions (blue) near the designed ice-binding motif. **(I)** Low-pass filtered view of (H), zoomed in on the C3 axis, illustrating stochastic ZnO growth within the Z4-C3i trimer. **(J)** Placement of atoms into the density (see Methods) suggests that initial H_2O/Zn ion coordination and nucleation occurs along the designed Z4 ice-binding motif (sticks).

References

1. Ciambellotti, S., Pozzi, C., Mangani, S. & Turano, P. Iron biomineral growth from the initial nucleation seed in L-ferritin. *Chemistry* **26**, 5770–5773 (2020).
2. Seker, U. O. S. & Demir, H. V. Material binding peptides for nanotechnology. *Molecules* **16**, 1426–1451 (2011).
3. Umetsu, M. *et al.* Bioassisted room‐temperature immobilization and mineralization of zinc oxide—the structural ordering of ZnO nanoparticles into a flower‐type morphology. *Adv. Mater.* **17**, 2571–2575 (2005).
4. Tomczak, M. M., Gupta, M. K., Drummy, L. F., Rozenzhak, S. M. & Naik, R. R. Morphological control and assembly of zinc oxide using a biotemplate. *Acta Biomater.* **5**, 876–882 (2009).
5. Wei, Z., Maeda, Y. & Matsui, H. Discovery of catalytic peptides for inorganic nanocrystal synthesis by a combinatorial phage display approach. *Angew. Chem. Int. Ed Engl.* **50**, 10585–10588 (2011).
6. Golec, P., Karczewska-Golec, J., Loś, M. & Węgrzyn, G. Novel ZnO-binding peptides obtained by the screening of a phage display peptide library. *J. Nanopart. Res.* **14**, 1218 (2012).
7. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
8. Huddy, T. F. *et al.* Blueprinting extendable nanomaterials with standardized protein blocks. *Nature* **627**, 898–904 (2024).
9. 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).

10. Davila-Hernandez, F. A. *et al.* Directing polymorph specific calcium carbonate formation with de novo protein templates. *Nat. Commun.* **14**, 8191 (2023).

11. Davies, P. L. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem. Sci.* **39**, 548–555 (2014).

12. 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).

13. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).

14. Koepnick, B. *et al.* De novo protein design by citizen scientists. *Nature* **570**, 390–394 (2019).

15. Jiang, H. *et al.* De novo design of buttressed loops for sculpting protein functions. *bioRxiv* (2023) doi:10.1101/2023.08.22.554384.

16. Baker, D. & Zorine, D. Design of Repeat Alpha-Beta Proteins with Capping Helices. *bioRxiv* 2024.06.15.590358 (2024) doi:10.1101/2024.06.15.590358.

17. Hakim, A. *et al.* Crystal structure of an insect antifreeze protein and its implications for ice binding. *J. Biol. Chem.* **288**, 12295–12304 (2013).

18. Buchko, G. W. *et al.* Characterization of two potentially universal turn motifs that shape the repeated five-residues fold--crystal structure of a lumenal pentapeptide repeat protein from Cyanothece 51142. *Protein Sci.* **15**, 2579–2595 (2006).

19. Dove, P. M., De Yoreo, J. J. & Weiner, S. *Biomineralization*. (Walter de Gruyter GmbH & Co KG, 2018).

20. Oh, B.-Y. *et al.* Transparent conductive Al-doped ZnO films for liquid crystal displays. *J.*

Appl. Phys. **99**, 124505 (2006).

21. Qin, Y., Wang, X. & Wang, Z. L. Microfibre-nanowire hybrid structure for energy scavenging. *Nature* **451**, 809–813 (2008).
22. Nomura, K. *et al.* Thin-film transistor fabricated in single-crystalline transparent oxide semiconductor. *Science* **300**, 1269–1272 (2003).
23. Li, Y. B., Bando, Y. & Golberg, D. ZnO nanoneedles with tip surface perturbations: Excellent field emitters. *Appl. Phys. Lett.* **84**, 3603–3605 (2004).
24. Sharma, D. K., Shukla, S., Sharma, K. K. & Kumar, V. A review on ZnO: Fundamental properties and applications. *Materials Today: Proceedings* **49**, 3028–3035 (2022).
25. Avital, Y. Y. *et al.* Two-site H2O2 photo-oxidation on haematite photoanodes. *Nat. Commun.* **9**, 4060 (2018).
26. Schwertmann, U., Cornell, R. M. & Koch, C. B. *Iron Oxides in the Laboratory*. (Wiley-VCH Verlag, Weinheim, Germany, 2022).
27. Maret, W. & Li, Y. Coordination dynamics of zinc in proteins. *Chem. Rev.* **109**, 4682–4707 (2009).
28. Repp, S. & Erdem, E. Controlling the exciton energy of zinc oxide (ZnO) quantum dots by changing the confinement conditions. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **152**, 637–644 (2016).
29. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
30. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
31. de Haas, R. J. *et al.* Inhibition of ice recrystallization with designed twistless helical repeat

proteins. *bioRxiv* (2025) doi:10.1101/2025.03.09.642278.

32. Woll, C. The chemistry and physics of zinc oxide surfaces. *Prog. Surf. Sci.* **82**, 55–120 (2007).