Title: Controlling semiconductor growth with structured de novo protein interfaces

Authors: Amijai Saragovi*[1,2], Harley Pyles*[1,2], Paul Kwon[1,2], Nikita Hanikel[1,2], Fátima A. Dávila-Hernández[1,2], Asim K. Bera[1,2], Alex Kang[1,2], Evans Brackenbrough[1,2], Dionne K. Vafeados[1,2], Aza Allen[1,2], Lance Stewart[1,2], David Baker[1,2,3]
* equal contributions

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

Abstract: Protein design now enables the precise arrangement of atoms on the length scales (nanometers) of inorganic crystal nuclei, opening up the possibility of templating semiconductor growth. We designed proteins presenting regularly repeating interfaces presenting functional groups likely to organize ions and water molecules, and characterized their ability to bind to and promote nucleation of ZnO. Utilizing the scattering properties of ZnO nanoparticles, we developed a flow cytometry based sorting methodology and identified thirteen proteins with ZnO binding interfaces. Three designs promoted ZnO nucleation under conditions where traditional inorganic binding peptides and control proteins were ineffective. Incorporation of these interfaces into higher order assemblies further enhanced nucleation. These findings demonstrate the potential of using protein design to modulate semiconductor growth and generate protein-semiconductor hybrid materials.

One Sentence Summary: In this study we designed a library of potential inorganic binding de novo proteins and identified structured protein interfaces with the capacity to bind to and promote the growth of zinc oxide, a semiconductor material that is not observed in native 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[2–5]. Although they template growth by modifying the chemistry of existing surfaces, these peptides have not been shown to promote the heterogeneous nucleation of inorganics 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. 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. With recent advances in protein design[6,7], it is now possible to construct molecular scaffolds with a wide variety of surface geometries, and to pattern diverse chemistries across these surfaces, enabling 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.

We followed three guiding hypotheses while designing inorganic-binding interfaces in the library: first that proteins can bind via periodic charge complementarity[8,9], second that partial coordination of metal ions facilitates binding[1], and third that ordered hydration layers may promote adsorption[10,11](**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 (**Figure 1B, C**)[7,12–16]. To include flat beta-sheet surfaces we designed beta-solenoid (DBS) repeat proteins (see Methods), and we also included a native antifreeze protein (RiAFP) and an idealized native-like penta-peptide repeat (PPR) fold[17,18] (**Figure 1D**). We chose ZnO nanoparticles (NP) as a target material because it can be synthesized in aqueous solutions, can be detected using fluorescence spectroscopy[19], is not found in native bio-inorganic hybrid materials[20], and is a semiconductor with diverse applications in emerging technologies[21].

Results

We defined sets of interface amino-acid compositions based on the three above design hypotheses (**Figure 1A, Supplementary Figure 1**), and designed 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 ZnO binding surface the sequence of the proteins remained fixed, and the starting designed scaffolds were also included in the library.

We developed a yeast display flow cytometry (FC) screen that relies on the light scattering properties of ZnO NP to identify and enrich yeast cells with increased binding propensity (**Figure 2A, B**; **Supplementary Figures 2-3**). Next generation sequencing (NGS) was used to compare the abundance of designs displayed on yeast before and after sorting to identify enriched sequences (**Figure 2C; Supplementary Figure 4**; see Methods). Individual enriched clones were retested by FC on yeast (**Figure 2C; Supplementary Figure 5**). 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[22], six contained large hydrophobic patches, five contained threonine motifs observed in ice-binding proteins[10], and eight interfaces contained periodic changes (**Figure 2; Supplementary Figure 5**).

We selected five proteins (Z0, Z1, Z3, Z4, and Z5) which had substantial FC binding signals to ZnO NP and expressed them in *E. coli* as soluble proteins for further characterization (**Figure 2, Supplementary Table 1**). Z0, Z3, Z4, and Z5 produced circular dichroism (CD) spectra consistent with their predicted secondary structure while Z1 appeared disordered

(**Supplementary Figure 6)**. Z3 and Z5 were monodisperse by size exclusion chromatography (SEC), and Z0, Z1, and Z4 appeared to form large assemblies (**Supplementary Figure 6)**. Z0 formed long fibers and bundles of fibers visible with negative-stain transmission electron microscopy (TEM), presumably through strand pairing head-to-tail interactions (**Supplementary Figure 7A**). To regularize these interactions, we designed and produced a version –Z0-fiber– without these capping features and used it in further analysis of this design (**Supplementary Figure 7B**). In a pull-down assay followed by SDS-PAGE, all of the purified proteins bound ZnO NP and did not bind to anatase, rutile, or hematite NP. In contrast, bovine serum albumin (BSA) and lysozyme controls showed no binding (**Supplementary Figure 8**). Z0-fiber, Z3, and Z4 produced the highest yields of purified protein and were selected for further study.

We next tested if the binding of these proteins to ZnO was facilitated by the partial coordination of Zn^{2+} ions. Using isothermal titration calorimetry (ITC) and found Z0-fiber and Z3 bound Zn2+ with Kd's of 4.1 ± 1.4 μM and 16 ± 5.0 μM respectively (**Figure 2D**). A ratiometric dye binding assay indicated that Z0-fiber binds Zn^{2+} ions with a best fit model indicating six binding sites, three with 33 ± 2 nM affinity and three with 530 ± 50 nM affinity (**Figure 2E**), and Z3 showed three binding sites with 1.8 ± 0.7 μM affinity (**Figure 2F**). The difference in measured affinities is likely due to the ITC requiring a lower pH (see Methods). Z0-fiber interfacial amino acid substitutions indicated that the two rows of repeating histidine residues are critical for Zn^{2+} binding (**Figure 2E**). Mutating glutamate, cysteine, and histidine residues in the Z3 interface reduced Zn^{2+} binding, while mutation of basic residues did not (**Figure 2F**). Mutating histidines in the interfaces of both Z0-fiber and Z3 also resulted in reduced ZnO NP binding when displayed on yeast (**Supplementary Figure 9**). Z4 did not coordinate Zn^{2+} ions (**Figure 2D**), but contains a threonine motif seen in alpha-helical ice-binding proteins (**Figure 2G, H**), and four other enriched proteins (Z8, Z9, Z12, and Z14) contain another threonine motif seen in beta-solenoid ice-binding proteins (**Figure 2I, H**)[10].

We next assayed the ability of purified ZnO binding proteins to promote the growth of ZnO NP by incubating them in supersaturated solutions (3 mM $ZnNO_3$, 50 mM NaCl, 100 mM HEPES pH 8.2) which allow heterogenous growth of ZnO, but not homogeneous growth (**Figure 3A**). We measured the fluorescence of ZnO NP to screen for protein-induced growth in these conditions and used solutions seeded with 13 μg/mL of commercial ZnO NP nanoparticles as a positive control[19]. After 2 hours, neat solutions and solutions containing lysozyme showed no fluorescence, while solutions seeded with 0.1 mg/mL of Z0-fiber, Z3, or Z4, had spectra consistent with ZnO NP (**Figure 3B**). X-ray powder diffraction of the material grown in the presence of Z0-fiber, Z3, and Z4 produced spectra indicative of ZnO (**Figure 3C**). Tracking fluorescence over time showed the increase induced by the ZnO NP 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), DNasI, and Mica6[8] had no observable effect on the level of the fluorescence signal (**Supplementary Figure 10**). Z0-fiber showed 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 11**).

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[23] and MPNN[24] 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 12**), and found that two (zab3 and zab10) showed a strong signal in the ZnO NP growth assay (**Figure 3F**). Four previously reported ZnO binding peptides, fused to Small Ubiquitin-like Modifier (SUMO) to aid solubility and expression, showed no activity (**Figure 3G**). To account for potential variability, six replicates were run for each of these conditions (**Figure 3G**). For replicates that showed growth, we compared the times of nucleation which we defined as the halfway point between the initial and maximum fluorescent signal (**Figure 3H**).

We next explored the effect of confinement on the rate of nucleation by incorporating the Z4 interface to the interior cavities of oligomers. We used RFdiffusion[23] and MPNN[24] to design cyclic oligomers that host motifs from the Z4 interface within the interior of rings (**Methods; Figure 4 A, B)**. We experimentally tested twenty-three oligomers and selected two designs forming monodisperse species of the expected size, Z4 c3 and Z4 c6, for further characterization (**Supplementary Figure 13**). The Z4 c3 XL trimer was derived from Z4 c3 by adding an additional repeat of the Z4 interface to each monomer. A crystal structure of Z4 c3 was obtained (**Figure 4C**) and the structures of Z4 c6 and Z4 c3 XL were confirmed with negative-stain TEM (**Figure 4 D, E**). We compared the activity of these designs to the Z4 monomer from previous experiments and knockouts in which the threonine residues in the Z4 interfaces were mutated to glutamate (**Figure 4G, Supplementary Figure 14**). The knockouts eliminated the templating activity, and only the oligomer with the smallest internal volume, Z4 C3, accelerated the rate of growth (**Figure 4 H,I**).

Discussion

We present an approach to design and screen de novo protein-mineral interfaces and apply it to identify proteins which bind ZnO NP. We describe thirteen de novo ZnO binding protein interfaces and demonstrate that three of them, Z0-fiber, Z3, and Z4, promote ZnO growth. 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 NP (**Figure 2D-F**; **Supplementary Figure 9**). Secondly, interfaces including Z4 contain water-structuring threonine motifs observed in ice-binding proteins (**Figure 2G-I**), suggesting they may adsorb to ZnO via ordered water layers[10,11,25]. 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. No interfaces were composed entirely of arrays of charged groups, although periodic charges were present in interfaces alongside other features (**Supplementary Table 1**).

Our designed ZnO interacting proteins are larger, more ordered, and better structurally understood than previously reported ZnO binding peptides. Their size and sophistication likely allows them to serve as molecular templates for inorganic growth in supersaturated solutions under conditions where ZnO binding peptides have little effect. The motifs comprising the ZnO binding protein interfaces can be grafted into new structural contexts

and maintain their function. Transferring motifs from aggregation prone to monodisperse molecules enhanced their function (**Figure 3 D-F; Supplementary Figures 6, 7, 12**), presumably by increasing the number of accessible interfaces. While the time of nucleation was decreased by confining the Z4 motifs inside the 3 nm diameter internal cavity of Z4 c3, no change was observed within the 5 or 7 nm diameter cavities of Z4 c3 XL or Z4 c6, respectively (**Figure 4 C-E, I)**.

Nature has evolved proteins which control the formation of hybrid organic-inorganic materials such as bone and enamel with greater complexity than is accessible with conventional top-down manufacturing[20]. 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. Structural characterization of the hybrid protein-ZnO materials will be important for understanding the atomic details of the protein-mineral interfaces. Moving forward, our approach opens the door to a new world of designed hybrid materials not explored by nature.

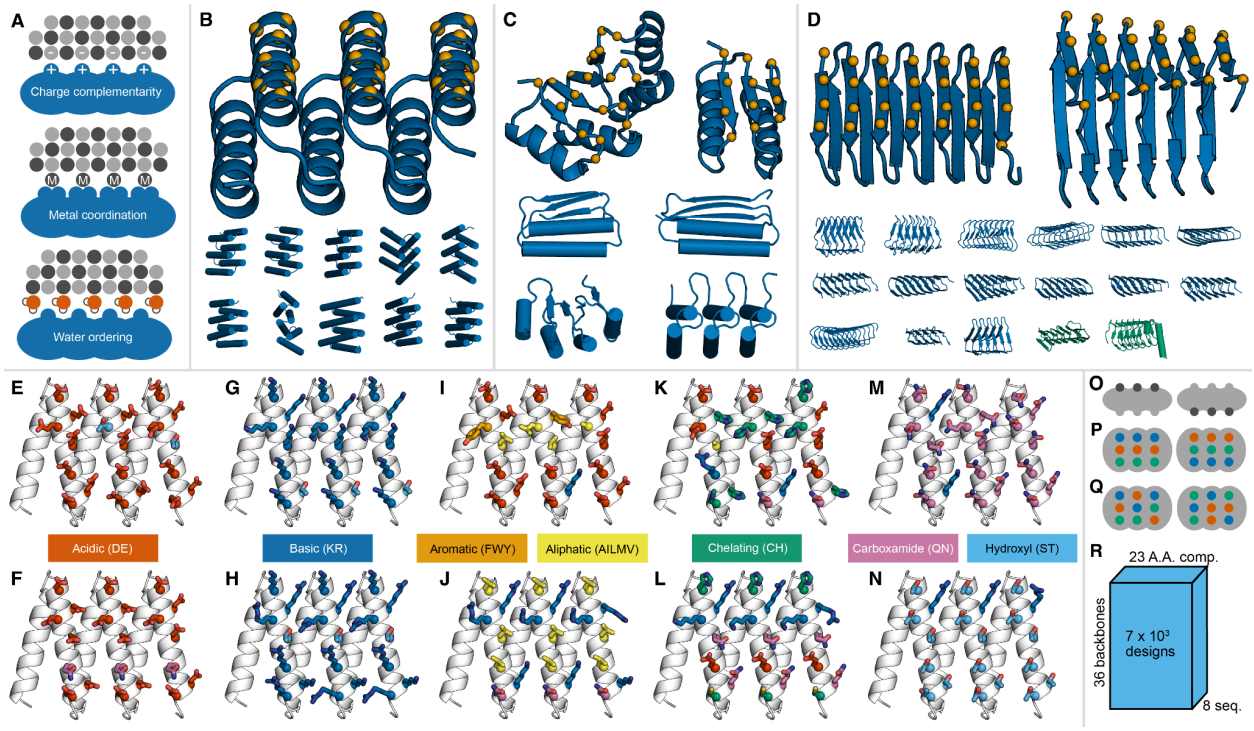

Fig. 1: Design principles for templating mineral growth. (**A**) Three hypothesized mechanisms of mineral binding: charge complementarity, metal coordination, and water ordering. (**B-D**) 36 protein backbones used in the library. Orange spheres show alpha carbons of residues in a designed interface. 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 two surfaces yields 8 sequences per combination of 36 protein backbones and 23 amino acid composition, resulting in a library of approximately 7×10^3.

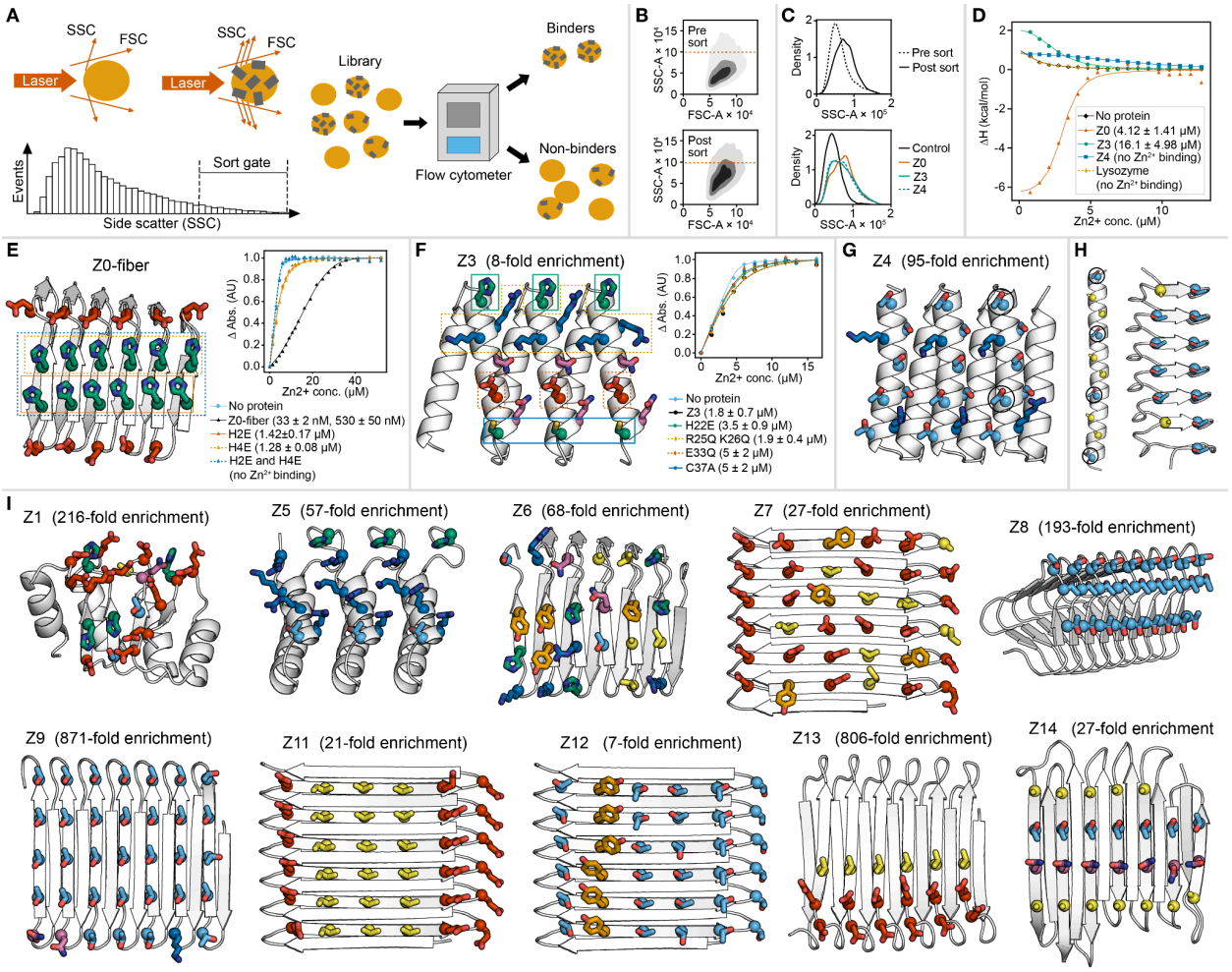

Fig. 2. Mineral binding screen. (**A**) Illustration of screening strategy. (**B**) SSC-A vs. FSC-A signals of yeast populations before and after ZnO enrichment sorting. Plots are kernel density plots estimates from 10,000 events. (**C**) SSC-A signals of a sample sublibrary before and after sorting, and yeast clones displaying three enriched designs vs. control yeast. (**D**) Isothermal titration calorimetry analysis of Zn^{+2} against three selected designs, lysozyme, and solutions without protein (protein concentration = 30 μM). (**E-F**) Z0-fiber and Z3 interfaces and evaluation of Zn^{+2} ion binding by ratiometric dye. (**E**) Mutational analysis of the Z0-fiber interface. Mutated residues are circumscribed by dotted lines on the model. (**F**) Mutational analysis of the Z3 interface. Mutated residues are circumscribed by dotted lines on the model. (**G**) Model of the Z4 enriched design with circles indicating repeated helical threonine motif. (**H**) Structure of native alpha-helical antifreeze peptide (left, PDB: 1wfa) containing the helical threonine motif, and beta solenoid AFP (right, PDB: 1ezg) containing threonine rich parallel beta-sheets. (**I**) Models of remaining enriched designs. Enrichment factors for individual designs are specified at the top of each model where available. Colors of amino acid side chains correspond to the chemical categories defined in Figure 1.

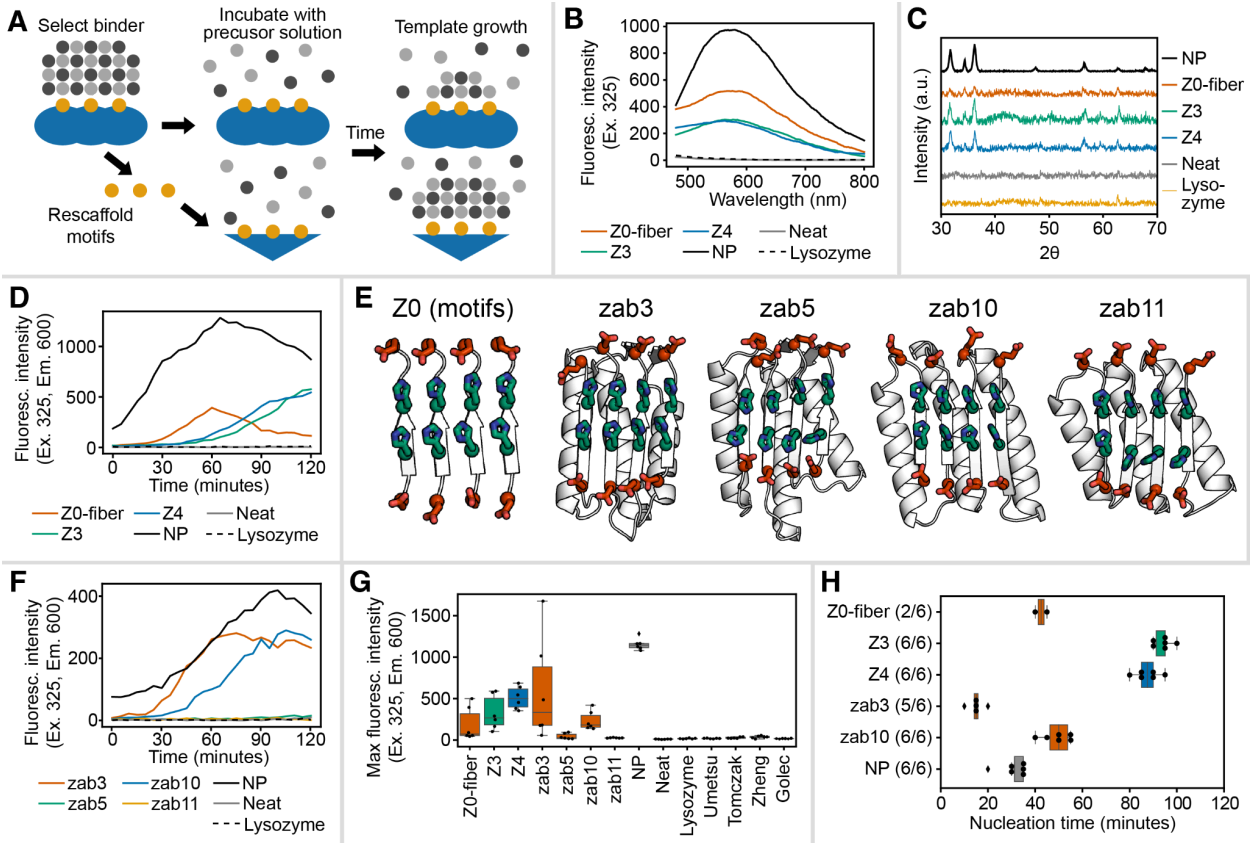

Fig. 3. Characterization of ZnO nucleation by designed interfaces. (**A**) Schematic illustrating the nucleation assay and motif grafting methodology. (**B**) Emission spectra (Ex. = 325 nm) of solutions superstatured for ZnO seeded with the Z0, Z3, and Z4 designs compared to ZnO NP, lysozyme and neat solution controls. (**C**) X-Ray diffraction analysis of synthesized product following two hour incubation. (**D**) Plot of fluorescence emission over time (Ex. 325, Em. 600) of supersaturated solutions containing selected designs, lysozyme, ZnO NP, or neat solutions. (**E**) Models of the Z0 interface motifs and four Z0-grafted alpha-beta (zab) designs. (**F**) Plot of fluorescence emission over time (Ex. 325, Em. 600) of the zab redesigns and controls. (**G, H**) Summary of experiment on designed proteins and experiment with four previously reported ZnO peptides. (**G**) Average maximum fluorescence intensity (n=6). (**H**) Nucleation times (defined as the halfway point between the initial and maximum fluorescent signal), for the subset of samples that showed a fluorescent signal of nucleation (maximum fluorescence > 100), n values as indicated. All superstatured 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. Colors of amino acid side chains correspond to the chemical categories defined in Figure 1.

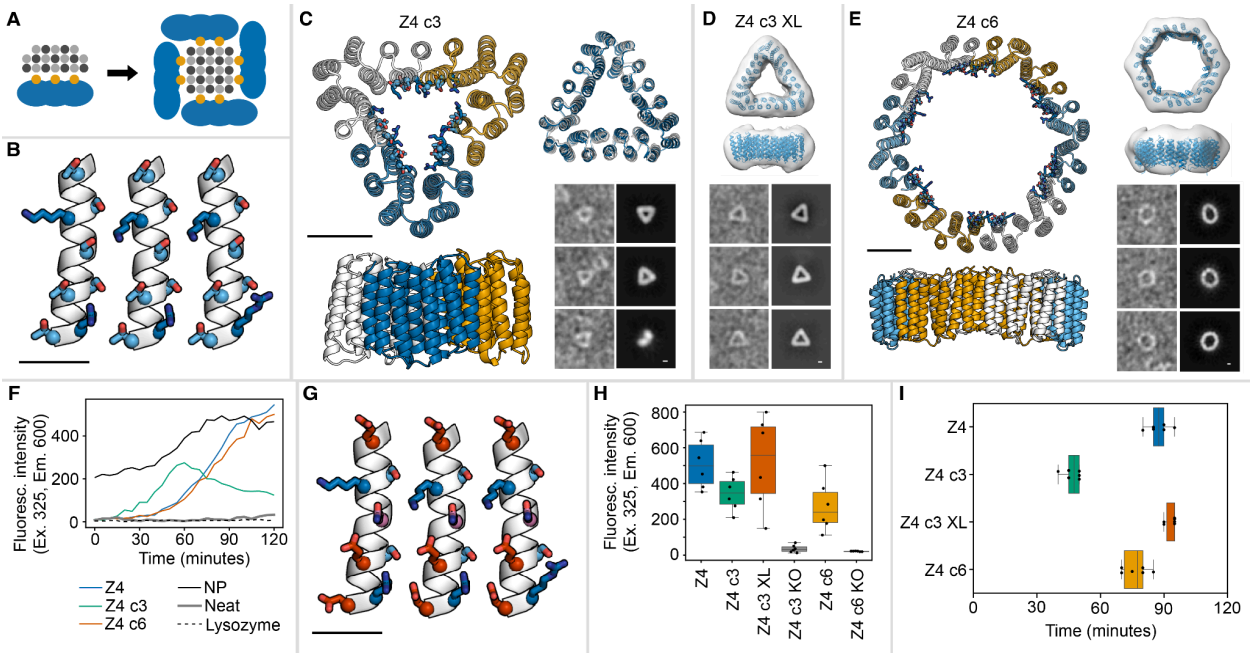

Fig. 4. ZnO nucleation by Z4 motif containing designed oligomers. (**A**) Illustration of strategy to test the effect of confinement on nucleation. (**B**) Motifs grafted into oligomer designs. (**C**) Z4 c3 oligomer model colored by chain in top and side view with motifs shown as sticks, model (grey) superimposed with crystal structure (blue; PDB: 9CC4), and examples of ns-TEM picked particles and 2D classes. (**D**) Z4 c3 XL model superimposed with ns-TEM 3D reconstruction, examples of picked particles and 2D classes. (**E**) Z4 c6 oligomer colored by subunit with motifs shown as sticks, model superimposed with ns-TEM 3D reconstruction, and examples of picked particles and 2D classes. (**F**) Plot of fluorescence emission over time (Ex. 325, Em. 600) of Z4 oligomers in respect to Z4 monomer (sample replicate from figure 3), lysozyme, ZnO NP, and neat solutions. (**G-I**) Analysis of interface mutation on ZnO nuclation. (**G**) Model of the mutated Z4 interface with threonines replaced with glutamates. (**H**) Maximum fluorescence intensity after nucleation assay n=6. (**I**) Nucleation time in solutions containing oligomers defined as the halfway point between the initial and maximum fluorescent signal (n=6). All superstatured 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. Colors of amino acid side chains correspond to the chemical categories defined in Figure 1.

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. 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).
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. 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).
6. Wicky, B. I. M. *et al.* Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
7. Huddy, T. F. *et al.* Blueprinting extendable nanomaterials with standardized protein blocks. *Nature* **627**, 898–904 (2024).
8. 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).
9. Davila-Hernandez, F. A. *et al.* Directing polymorph specific calcium carbonate formation with de novo protein templates. *Nat. Commun.* **14**, 8191 (2023).
10. Davies, P. L. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem. Sci.* **39**, 548–555 (2014).
11. 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).
12. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
13. Koepnick, B. *et al.* De novo protein design by citizen scientists. *Nature* **570**, 390–394 (2019).
14. Norn, C. *et al.* Protein sequence design by conformational landscape optimization. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
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. 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).

20. Dove, P. M., De Yoreo, J. J. & Weiner, S. *Biomineralization*. (Walter de Gruyter GmbH & Co KG, 2018).
21. 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).
22. Maret, W. & Li, Y. Coordination dynamics of zinc in proteins. *Chem. Rev.* **109**, 4682–4707 (2009).
23. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
24. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
25. Woll, C. The chemistry and physics of zinc oxide surfaces. *Prog. Surf. Sci.* **82**, 55–120 (2007).