COMPLEXITY THROUGH SIMPLICITY IN THE CHEMICAL DESIGN OF PROTEIN ASSEMBLIES

F. AKIF TEZCAN[†]

[†]*Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA*

Proteins as synthons for biological complexity

The molecular complexity of life is driven by proteins, at least in the eyes of this chemist. In fact, no other type of molecule epitomizes the term chemical complexity as much as a protein, which is perhaps best manifested by the fact that proteins may represent entirely different things to different scientists. To a polymer chemist, proteins are sequence-defined linear polymers that somehow manage to fold into discrete 3D structures in an atomically precise fashion. To an inorganic chemist, proteins are complex scaffolds that control the reactivity of metal ions beyond what is possible with small, synthetic ligands. To a nanoscientist or a supramolecular chemist, proteins may represent infinitely patchy colloidal particles or nanoscale synthons whose surface chemistry can be tuned with atomic precision. To a biochemist, proteins are the sophisticated end-products of millions of years of evolution, dissected and scrutinized through detailed top-down studies to uncover their structure-function relationships.

Regardless of which view one holds of proteins, they are distinguished from all other types of molecules by their spatially precise yet dynamic organization of 20 different chemical functionalities, which in turns allows them to interact with or act upon any other type of organic or inorganic entity of any arbitrary shape and composition with high specificity. Importantly, it also follows that individual proteins can organize into larger supramolecular assemblies that carry out some of the most challenging biochemical tasks (e.g., photosynthetic water oxidation or nitrogen fixation) or extended materials that fulfill essential mechanical functions (e.g., intracellular transport, cell motility and division) in living cells. Despite such functional diversity and sophistication, natural proteins and protein assemblies constitute an infinitesimally small fraction of 3D structures that can be obtained using the available amino acid sequence space. Furthermore, these natural systems have only evolved for functional adequacy in the very restricted chemical environment of a cell. These limitations logically lead to the following questions: Can we use proteins as building blocks for new biological machines and materials with structural

and functional properties that surpass those that evolution has produced? Can we incorporate new strategies unrestrained by the limitations of cellular chemistry in this design process to broaden the structural and dynamic complexity of proteins?

Chemical Design of Dynamic and Functional Protein Assemblies

From the outset, our group took an alternative approach to purely symmetry-based [1] or computational design strategies [2, 3] by viewing protein self-assembly through the perspective of a supramolecular or an inorganic chemist (rather than a biochemist or a protein designer). We surmised that properly engineered metal coordination motifs on a protein surface can enable control of protein self-assembly on small design footprints (owing to the higher strength of metal coordination bonds relative to non-covalent interactions), while also imposing symmetry and providing chemical control and reactivity by the metal ion [4]. Through this metal-directed strategy, sometimes complemented by computational design or the use of synthetic ligands, various natural protein building blocks (in particular, the monomeric hemeprotein cytochrome cb_{562}) were engineered to assemble into discrete oligomers, 0-D cages, 1-D filaments and nanotubes, as well as 2-D and 3-D lattices [5]. This chemical-bonding-focused approach was further expanded through the implementation of disulfide bonds and host-guest interactions to yield defect-free 1- and 2-D nanotubes and crystalline arrays [6, 7]. A distinguishing characteristic of all of these protein architectures is the reversibility and minimal surface footprints of the chemical interactions. These features not only render the assembly-disassembly process inherently responsive to external stimuli (e.g., redox, pH, temperature, metal ions, ligands, light, etc.), but also endow the resulting protein architectures with dynamic properties. A compelling case in point are disulfide-mediated 2D crystals of the protein RhuA, which undergo coherent opening-closing motions due to the flexibility of the disulfide linkages, yielding an auxetic material with the thermodynamically smallest possible Poisson's ratio of –1 (Fig. 1) [6]. Another example is the protein-metal-organic frameworks (protein MOFs) composed of ferritin building blocks assembled into porous, 3D lattices via surface metal coordination and ditopic organic linkers [8, 9]. Owing to their sparsely connected, soft-crystalline lattices, some ferritin-MOFs undergo highly cooperative first-order transitions with large volumetric changes near room temperature and wide thermal hysteresis windows – properties that may be exploited in sensing and memory devices [10].

The ability to chemically control protein-protein interactions has also allowed us to finely tune and explore the energy landscape of protein self-assembly. This, in turn, revealed previously unconsidered properties of proteins and unusual assembly behaviors that informed subsequent design efforts. For example, detailed structural analyses of disulfide-linked 2D RhuA crystals revealed that the individual RhuA molecules are oriented in an alternating, up-down pattern with respect to the 2D plane due entirely to the favorable dipolar interactions between the RhuA molecules (1200 debye per RhuA; Fig.

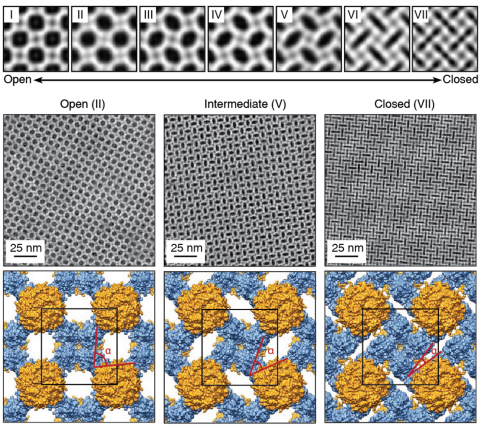

Fig. 1. Coherent dynamics of ^{C98}RhuA crystals. Top row: reconstructed 2D images of seven distinct conformational states of the 2D crystals (I-VII); middle row: magnified views of states II, V and VII; bottom row: structural models of conformations II, V and VII with unit cells and angles (α) between RhuA molecules. Adapted from Ref. [6].

2a) which can act over distances of >5 nm [11]. These dipolar interactions can be overcome by electrostatic interactions if the RhuA self-assembly takes place on a charged mica surface (Figs. 2b and 2c), which results in uniform alignment of the RhuA dipoles and yields a 2D material with permanent electric polarization and piezoelectricity. When the RhuA molecules are interconnected by host-guest interactions (e.g., β-cyclodextrin-azobenzene) instead of disulfide bonds, the original antiparallel arrangement of RhuA molecules is preserved. Yet, now, the longer host-guest linkages allow the curving of the 2D assemblies into helical 1D nanotubes [7]. Interestingly, a kinetic analysis showed that these RhuA nanotubes self-assemble without a nucleation barrier. This isodesmic polymerization behavior, which is highly unusual for helical structures, was found to stem from the ability of host-guest pairs in the lattice to readily exchange with one another during self-assembly. These examples demonstrate how a single protein building block can be engineered with different “chemical interaction motifs” to engender a wide variety of structural outcomes, emergent physical properties and new dynamic behaviors.

The dynamicity and the functional complexity of artificial protein assemblies can be further bolstered through their hybridization with abiological materials. For example, the mesoporosity of metal-directed 3D ferritin crystals allows them to be fully infiltrated by synthetic polymer precursors (e.g., acrylate, acrylamide), which can subsequently be polymerized *in crystallo* [12].The resulting polymer-integrated protein crystals (PIX) (Fig. 3) possess several outstanding properties including fully reversible, >500% volumetric

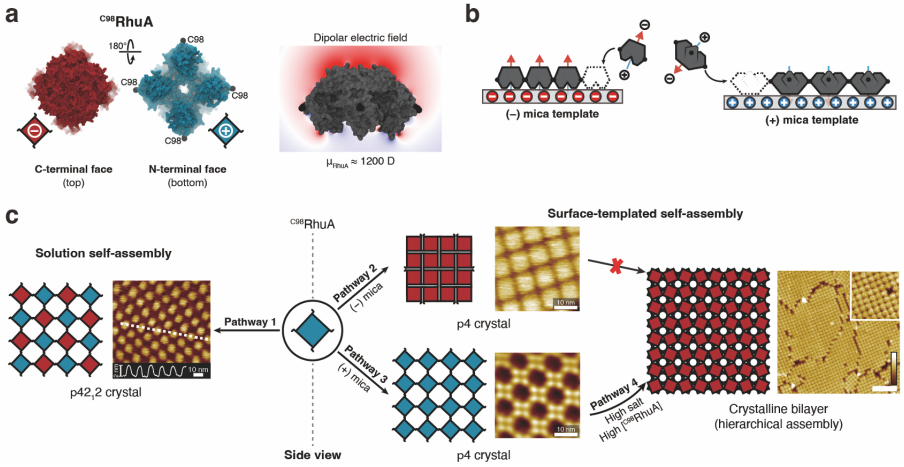

Fig. 2. (a) The negatively charged C-terminal face and partially positively charged N-terminal face of ^{C98}RhuA (left) give rise to a very large macrodipole moment (right). (b) Preferential adsorption of the N-terminal or the C-terminal face of RhuA can be controlled by the charge state of the mica template. (c) ^{C98}RhuA assembly pathways in solution and on the mica surface. Adapted from Ref. [11].

expansion while retaining crystallinity, efficient self-healing behavior, and the ability to selectively uptake and release macromolecules that are significantly larger than the original lattice pores [12-14]. Each of these is an emergent property that arises from the integration of protein, crystal and polymer components, and cannot be achieved with any one (or two) of the three components alone.

Outlook

As shown by the examples above, the field of protein design or "protein nanotechnology" has transitioned from a purely structure-building stage to a property- or function-building phase. One outstanding challenge is to increase the structural and compositional complexity of the protein assemblies beyond one- or two-component systems. Most natural biomachines (e.g., Photosystem II or nitrogenase) are composed of multiple types of protein units, each of which fulfill an individual role and cooperate together to enable the complex functional outcome. As mentioned before, AI/ML-based computational tools have certainly facilitated the design of sophisticated multi-component protein architectures, however, attaining cooperative functions (apart from binding target molecules) and requisite structural dynamics for such functions remain a distant goal. In parallel, directed evolution methods coupled with high-throughput screening have enabled

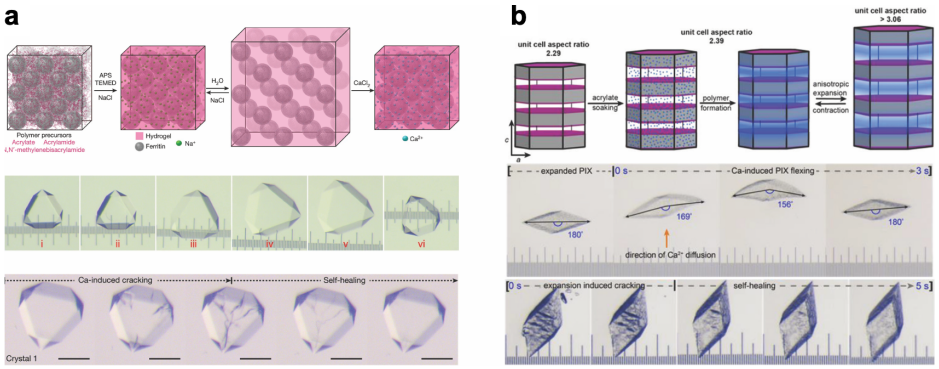

Fig. 3. (a) Schematic representation (top) and light micrographs (middle) showing the formation, expansion and contraction of ferritin-PIX. Light micrographs of ferritin-PIX (bottom) showing the self-healing behavior of cracks that appear during Ca-induced contraction. Adapted from Ref. [12]. (b) Schematic representation of the reversible anisotropic expansion/contraction of rhombohedral ^{raft}-ferritin-PIX (top). Light micrographs of the rhombohedral ^{raft}-ferritin-PIX crystal showing cation-induced bending motion (middle) and self-healing behavior (bottom). Adapted from Ref. [14].

the optimization enzyme activities and discovery of new-to-nature catalytic reactions [15]. However, these studies have almost exclusively been geared toward single-component protein systems with relatively easy-to-screen binding or catalytic functions. Evolving complex multi-step reactions or dynamic/mechanical properties in artificial protein assemblies will require considerable ingenuity and almost certainly involve a combination of rational chemical and computational design, exploitation of the ever-expanding AI/ML toolkit [16, 17], and the incorporation of the assemblies into the life cycles of natural or artificial cells (wherein they can be subjected to functional screening or natural selection).

Another important goal is to translate the cellular functions and materials properties of protein assemblies into real-world applications in non-cellular environments. There are prominent examples of industrially important enzymes that have been engineered for operation under non-physiological conditions and proteins that have been incorporated into functional devices and used in practical applications outside their native biological contexts (e.g., blood sugar monitors, antigen detection kits, and biosensors/DNA sequencing devices. However, again, all of these cases have relied on the inherent structure and function of a single protein as an individual unit and not in the context of a supramolecular architecture. The fundamental challenge here is how to stabilize protein assemblies to withstand harsh conditions (e.g., high temperatures, extreme pH's, organic solvents, low humidity, mechanical stress), while also maintaining their dynamics, inherent functions and chemical specificities. Our experience thus far indicates that the metal- or disulfide-directed assembly of proteins into extended crystalline arrays can also increase their

thermal and chemical stabilities to the extent that they maintain their structures and functions in boiling water and polar organic solvents [18]. More recently, our group reported that the dynamic 2D crystals of RhuA can be engineered to selectively open and close in response to hydrogen cyanide (HCN) and these crystals can be integrated with porous Si photonic chips as selective molecular gatekeepers for remote HCN sensing [19]. This study illustrates that properly engineered protein assemblies can operate as dynamic components of a solid-state device under conditions normally considered to be incompatible with proteins, opening up new avenues in the design of functional protein-based materials.

Acknowledgments

The author acknowledges DOE, NSF, NASA and ARO for funding the research described herein.

References and citations

1. T. O. Yeates, Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **46**, 23-42 (2017).
2. S. Gonen, F. DiMaio, T. Gonen, D. Baker, Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* **348**, 1365-1368 (2015).
3. N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida *et al.*, Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **336**, 1171-1174 (2012).
4. E. N. Salgado, R. J. Radford, F. A. Tezcan, Metal-Directed Protein Self-Assembly. *Acc. Chem. Res.* **43**, 661-672 (2010).
5. J. Zhu, N. Avakyan, A. Kakkis, A. M. Hoffnagle, K. Han *et al.*, Protein Assembly by Design. *Chem. Rev.* **121**, 13701-13796 (2021).
6. Y. Suzuki, G. Cardone, D. Restrepo, P. D. Zavattieri, T. S. Baker *et al.*, Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369 (2016).
7. Z. Zhang, H. T. Chiang, Y. Xia, N. Avakyan, R. R. Sonani *et al.*, Design of light- and chemically responsive protein assemblies through host-guest interactions. *Chem* **11**, 102407 (2025).
8. J. B. Bailey, L. Zhang, J. A. Chiong, S. Ahn, F. A. Tezcan, Synthetic Modularity of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **139**, 8160-8166 (2017).
9. P. A. Sontz, J. B. Bailey, S. Ahn, F. A. Tezcan, A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. *J. Am. Chem. Soc.* **137**, 11598-11601 (2015).
10. J. B. Bailey, F. A. Tezcan, Tunable and Cooperative Thermomechanical Properties of Protein–Metal–Organic Frameworks. *J. Am. Chem. Soc.* **142**, 17265-17270 (2020).

11. S. Zhang, R. G. Alberstein, J. J. De Yoreo, F. A. Tezcan, Assembly of a patchy protein into variable 2D lattices via tunable multiscale interactions. *Nat. Commun.* **11**, 1-12 (2020).
12. L. Zhang, J. B. Bailey, R. H. Subramanian, F. A. Tezcan, Hyperexpandable, self-healing macromolecular crystals with integrated polymer networks. *Nature* **557**, 86-91 (2018).
13. K. Han, Y. Na, L. Zhang, F. A. Tezcan, Dynamic, Polymer-Integrated Crystals for Efficient, Reversible Protein Encapsulation. **144**, 10139-10144 (2022).
14. K. Han, J. B. Bailey, L. Zhang, F. A. Tezcan, Anisotropic Dynamics and Mechanics of Macromolecular Crystals Containing Lattice-Patterned Polymer Networks. *J. Am. Chem. Soc.* **142**, 19402-19410 (2020).
15. Y. Wang, P. Xue, M. Cao, T. Yu, S. T. Lane *et al.*, Directed Evolution: Methodologies and Applications. *Chem. Rev.* **121**, 12384-12444 (2021).
16. J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim *et al.*, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023).
17. J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte *et al.*, Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49-56 (2022).
18. J. D. Brodin, J. R. Carr, P. A. Sontz, F. A. Tezcan, Exceptionally stable, redox-active supramolecular protein assemblies with emergent properties. *Proc. Natl. Acad. Sci. USA* **111**, 2897-2902 (2014).
19. S. Vijayakumar, R. G. Alberstein, Z. Zhang, Y.-S. Lu, A. Chan *et al.*, Designed 2D protein crystals as dynamic molecular gatekeepers for a solid-state device. *Nat. Commun.* **15**, 6326 (2024).